ACLA WORKING PAPER SERIES

THE EFFECT OF STUDENTS WITH SPECIAL NEEDS ON THEIR PEERS' COGNITIVE OUTCOMES: RESULTS FROM STUDENT-LEVEL POPULATION DATA IN THE NETHERLANDS

LAURENS DE CROES
NIENKE RUIJS
ILJA CORNELISZ
CHRIS VAN KLAVEREN

Working Paper 20221

http://www.acla.amsterdam/workingpapers-wp20221

AMSTERDAM CENTER FOR LEARNING ANALYTICS
van der Boechorststraat 7
1081BT, Amsterdam

The effect of students with special needs on their peers' cognitive outcomes: results from student-level population data in the Netherlands

Laurens de Croes[ab*], Nienke Ruijs[ab] , Ilja Cornelisz[a] & Chris van Klaveren[a]

a Vrije Universiteit Amsterdam, ACLA Amsterdam

b Inspectorate of education

*Corresponding author. *E-mail address:* l.x.de.croes@vu.nl (L. de Croes)

This paper presents evidence that students with special needs (SN) have no impact on the academic achievement of their peers in primary and secondary schools in the Netherlands. We use nation-wide administrative panel data on all Dutch students leaving primary and secondary education for the years 2015-2018 and estimate student fixed effects and school fixed-effect models. We consistently find no effects of inclusion on students without SN: the point estimates are precise and the coefficients are generally insignificant and inconsistent in sign. We apply a conceptual framework for learning from null results to establish if these results represent a true null finding and conclude that the null results are robust. These results nuance recent empirical findings, and provide further evidence that inclusive education can be implemented without harming the academic achievement of students without special needs.

Keywords: inclusive education; special needs; peers; achievement; learning from null results.

Acknowledgements

We gratefully acknowledge valuable comments from Inge de Wolf, Kaat Iterbeke, Rafiq Friperson and Gloria Moroni. We would like to thank session participants at the Eleventh International Workshop Applies Economics of Education (IWAEE) and at the XXIX Meeting of the economics of Education Association (AEDE). Further, we thank Statistic Netherlands (CBS) and the Education Executive Agency of the Dutch Ministry of Education, Culture and Science (DUO) for supplying the data for this study.

Sources of financial support

Not applicable

1 Introduction

The inclusion of students with special needs in regular schools has not only been a major topic in education policy for the past 30 years, but also a controversial area of education research. Following the UNESCO Salamanca Statement in 1994, countries worldwide have adopted more inclusive education practices for students with special needs (SN) (UNESCO 2020). Despite this policy trend, there is currently no conclusive evidence of either the potential benefits or drawbacks of inclusive education for both students with and without special needs (Aizer 2008; Carrell and Hoekstra 2010; Figlio 2007; Fletcher 2010; Hanushek, Kain and Rivkin 2002; Neidell and Waldfogel 2010). Yet, recent years have shown increasing evidence for negative effects of the exposure to students with behavioral or emotional problems on the cognitive outcomes of their peers (Ahn and Trogan 2017; Balestra, Eugster and Liebert 2020; Carell, Hoekstra and Kuka 2018; Contreras et al. 2020; Kristoffersen et al. 2015; Rangvid 2016). Potential mechanisms for these negative effects are that the (behaviour of) SN students cause the teacher to allocate relatively more time to SN students (Lazear 2001) and that SN students negatively affect the classroom climate (Huang, Lu and Zhu 2021). If there are indeed such negative externalities, policies towards inclusive education should take account of ways to compensate for these negative effects.

However, contrary to the aforementioned results, Friesen, Hickey, and Krauth (2010) and Ruijs (2017) find no significant effects of the inclusion of students with special needs. Contreras et al. (2020) conclude the negative effect decreases or even vanishes when sufficient resources are provided. Hanushek, Kain, and Rivkin (2002) conclude that disabled peers may even produce achievement gains among students without special educational needs, though the precise underlying causes for such a positive relationship are not readily identified. Whereas these findings were based on specific groups of SN students, this paper contributes to the existing literature by identifying a more comprehensive group of students with SN, and by using a conceptual framework for learning from null results to establish if we can interpret the result as true null findings.

First, we explore a broad group of students with SN. Previous null findings in the Dutch context were based only on students with diagnosed severe special educational needs with schools receiving substantial additional funding for their inclusion in regular classes (Ruijs 2017). However, the increasing

empirical evidence found for negative effects is largely based on proxies identifying (potentially) disruptive peers. We, therefore, select a broader group of students with SN by identifying four different subgroups of students with SN: students with prescribed medication for attention deficit disorder (ADHD), students with prescribed anti-depressants or tranquillizers, students receiving care in an ambulant or outpatient setting, and students receiving child custody or family supervision orders. Whereas there is some overlap between these students and students with diagnosed severe special educational needs, schools do not receive additional funding for these groups of students with SN.

The presence of students with SN may be related to school, teacher, or student characteristics. For example, high-quality schools may be more able to educate students with special needs, and may therefore be less likely to refer students to special education. The relationship between inclusion and student outcomes is then biased by school quality. We circumvent these endogeneity issues by using two complementary empirical strategies. First, we exploit that students in secondary schools take different courses. This creates within-student-between-course variation in classroom composition, and we use student fixed effects to compare student achievement in courses with varying numbers of SN students. The second strategy follows a widely used approach in the literature on peer effects (Hoxby 2000; Lavy, Paserman and Schlosser 2021) by using school fixed effects that exploit within school cohort-to-cohort variation in the numbers of SN students.

The second contribution of our paper is that we apply criteria from a conceptual framework for learning from null results to interpret the results. We first use the criteria formulated by Jacob et al. (2019) to establish if we can interpret the result as a true null finding: is the estimated impact small and estimated with enough precision? Was the effect size realistic considering the intervention intensity? Further, we perform robustness checks to test for non-linear effects and the proportion instead of the number of SN students and another type of SN: untreated disruptive peers. Finally, we discuss the proximity of the outcome as a possible concern. Achievement, as measured in our study, can be seen as a somewhat distal outcome. It is affected by many factors, such as student ability, teacher quality, and school quality. We provide evidence that our design matches the conditions to find impacts comparable to other studies with similar outcomes and discuss potential mechanisms that may explain our null result. We conclude that our results meet the criteria for a true null finding.

Next, we use the framework to explore potential reasons for not finding a negative effect of SN students on their peers' cognitive outcomes. To establish whether the findings are related to unobserved aspects of within-school cohort quality, we obtain additional evidence for primary education by utilizing neighborhood variation in the number of SN students. To address concerns about teacher quality we perform robustness checks on samples restricted to courses and cohorts with 30 or fewer students. After the methodological concerns, we explore other potential explanations of our findings related to the underlying mechanisms of peer effects. Potential peer effects might only be present for certain populations. For instance, low-achieving students might be more easily affected by classroom disruptions. We, therefore, check for this type of differential effects, which were also found in recent studies (Balestra, Eugster and Liebert 2020; Feld and Zölitz 2017). The results indicate that schools are not selectively referring to special education and that our results are robust to several ways of modeling the relation between inclusive education and the academic achievement of students without SN.

This paper is structured as follows. Section 2 summarizes the current literature on the effects of exposure to students with SN. Section 3 describes the Dutch educational context. Section 4 describes the data and section 5 the empirical strategies. The main results are presented in section 6 and the details and application of the conceptual framework for learning from null results are presented in section 7. Section 8 concludes.

2 Literature review

The studies on the effects of exposure to students with SN vary in methodology and type of needs of SN students and, perhaps because of this, yield different results.[1] Some researchers use proxies to identify students with externalizing problems. Carrell and Hoekstra (2010) use family violence as a proxy for disruptive behaviour, and find negative effects of the presence of children from troubled families on the reading and math scores of their peers in the US. Using a similar identification strategy with school-by-grade and school-by-year fixed effects, Carell, Hoekstra and Kuka (2018) find that being

[1] A related strand of literature is on peer effects of low-achieving students. Huang and Zhu (2020) show that the share of students in a class who are low achievers has a significant negative impact on the academic achievement of regular students, Lavy, Paserman and Schlosser (2012) find significant negative effects from low achieving peers. Contrary, Bietenbeck (2020) documents that students who are exposed to low-achievers show long-term educational attainment. In this paper, we specifically focus on students with special needs.

exposed to children from families linked to domestic violence has long-run effects on test scores, college attendance, and completion and earnings as well. Figlio (2007) uses an IV strategy that exploits the misbehavior of boys having girls' names to identify disruptive peers and finds that these students harm their peers' mathematics achievement in the first year of middle schools in the US.

Another often-used proxy is diagnostic information. Kristoffersen et al. (2015) identify three groups of potentially disruptive children from Danish register data: children with divorced parents, children with parents convicted of a crime, and children with a psychiatric diagnosis. They find that the presence of these groups lowers the reading test and math scores of their peers by about -0.02 SD in elementary schools. Rangvid (2016) exploits variation in the presence of SN students induced by a policy reform and finds that exposure to SN students returning to regular classrooms harms test score gains (-0.036 SD). Aizer (2008) uses the timing of ADD diagnoses to identify students with SN in elementary schools in the US. She concludes that the presence of students with ADD harms peers' reading and math test scores before the diagnosis, but the achievement of peers improves after the diagnosis. Balestra, Eugster, and Liebert (2020) identify SN students in secondary schools by using the information on students counselled by the School Psychological Service in Switzerland. They find a negative effect on students' outcomes of 0.025 standard deviation when adding one SN student to a class of 20. Fletcher (2010) uses a national survey of kindergartners to identify children with diagnosed emotional problems in the US. He exploits school-level and student-level fixed effects and finds evidence for negative peer effects in reading and mathematics scores. Friesen, Hickey, and Krauth (2010) identify disabled peers based on diagnostic categories for funding and use data on students in grades 4 – 7 in the Canadian province of British Columbia to investigate the peer effect on the province-wide numeracy and reading exam scores of their non-disabled peers. They find statistically insignificant effects.

Other studies proxy for disruptiveness in a different way. Ahn and Trogdon (2017) study the relationship between student achievement and the exposure to (severe) peer delinquency in North Carolina middle schools, using offenses recorded in the administrative data. They find a negative effect: a 10% increase in the exposure to "major" incidents is associated with 0.06 standard deviation in math score. Neidell and Waldfogel (2010) use a particular form of education (preschool) instead of

characteristics to examine peer effects in early education in Washington. They estimate value-added models and find significant and robust spillover effects from preschool on math and reading outcomes. Ohinata and Van Ours (2013) analyze how the share of immigrant children in the classroom affects the educational attainment of native Dutch children and find no strong negative spillover effects.

Some studies focus on students with diagnosed severe special educational needs. Contreras et al. (2020) use administrative data from Chile to identify students who are diagnosed with special educational needs. Using fixed-effects models at school-, student-, and time level, they find that having a classmate with SN is associated with a small negative effect on standardized test scores. During the policy change that increased resources for special educational needs, this effect decreased or became slightly positive. Ruijs (2017) uses diagnosed severe special educational needs to identify special educational needs students in primary and secondary education in the Netherlands and finds no effect on academic achievement. Horoi and Ost (2015) use the diagnosis of an emotional disability to proxy for disruption. They find that in public schools in North Carolina some teachers' value-added to math and reading scores of non-disruptive students is penalized because of disruptive students.

To summarize, the picture that emerges from the literature is that the presence of students with special educational needs, especially students with behavioural problems, negatively affects the academic achievement of their peers without special needs. This may be caused by distraction because of problematic behaviour or the need for teachers to divert their attention to students with special needs (Gottfried 2014).

However, several studies do not find negative effects caused by special needs students (Contreras et al. 2020; Friesen, Hickey and Krauth 2010; Ruijs 2017). Remarkably, these three studies all took place in education systems that provide specific and substantial additional funding to students with special needs. Although we cannot directly investigate whether it is the additional funding that mitigates potential negative effects, we can increase our understanding of peer effects of special needs students by focussing again on special needs students in the Netherlands. Where Ruijs (2017) focuses on students with severe and diagnosed special educational needs, for whom schools receive additional funding, we shift the focus to broader groups of students with special needs. Whereas there is some overlap between these students and students with diagnosed severe special educational needs, schools do not necessarily

receive additional funding for these students. It is interesting to investigate whether these students affect the academic achievement of their peers differently.

3 The Dutch context

3.1 Primary and secondary education in the Netherlands

The Dutch school system gives parents wide freedom of choice and all schools are completely publicly funded.[2] The government funding of schools is to a large extent dependent on student numbers, in which money follows the student. In 2020, the nationally determined average was €7.400 per student in primary education and €9.100 per student in secondary education (Ministerie van Onderwijs, Cultuur en Wetenschap 2020). Schools can be awarded additional funding for children from disadvantaged backgrounds. The extent of the additional funding in primary education is based on five criteria, including the educational level of both parents and the mother's country of origin. In secondary education, the additional funding is based on neighborhood characteristics.

[Insert figure 1 here]

Dutch primary education starts at age 4 and lasts until age 12. Students are generally educated by one or two teachers for an entire school year. At the end of primary education, students have to take a final test to determine their mathematics, Dutch, and comprehension skills. The result of this standardized final test and the recommendation of the primary school teacher determines which track a student takes in secondary education. We consider the final test score as a measure of student performance in primary education. Some pupils in the participating schools are exempted from the test, like pupils with special needs or students who are likely to attend special education. As schools have some discretion in deciding which students are exempted within the general regulations, they can choose

[2] Private education is virtually non-existent: less than 1% of the students attend a private school (Inspectorate of Education, 2021).

to exclude bad-performing students, which is why student-level final test participation is considered as an outcome besides the final test score.

The Dutch secondary school system is highly tracked. It distinguishes three academic ability tracks, illustrated in figure 1. Schools can offer one or more tracks. The basic track (pre-vocational secondary education [in Dutch abbreviated as vmbo]) prepares students for vocational training and lasts four years. Within the pre-vocational track are three different levels, each giving access to different levels of vocational education programs. We denote them by the numbers I to III, with I being the lowest level. The middle track (senior general education [in Dutch abbreviated as havo]) takes five years and gives access to higher professional education. The advanced track (vwo, pre-university education [in Dutch abbreviated as vwo]) takes six years and gives access to university education. Depending on student achievement and school policies, students can switch tracks during secondary education or they can decide to take a higher track after first graduating from a lower track. In the second half of secondary education, students choose their exam subjects. Several exam subjects are clustered in a specialization, such as “Economics and society”, other subjects are obliged for all students and some subjects can be freely chosen. The number of the freely chosen subjects is limited but always more than one. This implies students follow different subjects with different classroom peers, whichallows us to exploit this feature for within-student-between-course variation.

3.2 Inclusive education

In the Netherlands, special schools coexist next to inclusive education. In 2018, 2.1% of the students in primary education and 3.7% of the students in secondary education attended special education schools (Inspectorate of education 2020). A reform in 2014 abolished the pupil-bound budget, which students with severe special educational needs could receive to be mainstreamed in regular education schools. Instead of this national centralized system, regional alliances became responsible for organizing a comprehensive system of facilities for pupils who need extra support. There are 77 regional alliances for additional support in primary education and 75 for additional support secondary education. One of the objectives of this policy is to allow as many pupils as possible to go to a mainstream school.

Special education schools should remain available for those who need it: students with severe special educational needs. Despite these objectives, the share of students attending special education seems stable. After an initial decrease, the percentage of pupils in special (secondary) education is slowly growing since 2016 (Inspectorate of education 2020) and reached in 2022 the same level as it was when the reform was implemented (Inspectorate of education 2022).

4 Data

4.1 Students with special needs

For primary and secondary education we exploit student-level data from Statistic Netherlands (CBS) and the Education Executive Agency of the Dutch Ministry of Education, Culture and Science (DUO). To identify students with special needs, we use proxies identifying externalizing or internalizing behavioural problems, emotional problems, or a dysfunctional family situation as observed in the data from Statistics Netherlands.

The data include information on the entire Dutch population about the type of medication that is prescribed by doctors, the type of indicated psychological youth care, and whether a child is supervised under child protective services. The data also include school identifiers and socio-economic background information about pupils and their parents.

We identify four different subgroups of students with SN. First, we study students with externalizing behavioural problems by using information on which students use psychostimulant medication for treating externalizing behaviour problems, such as ADHD, in the current school year. Second, we study students who use prescribed medication for internalizing problems, like anti-depressants or tranquillizers.[3] This identifies a group of students with emotional and potential behavioural problems, such as social withdrawal. In the third group, we use receiving psychological youth care in an ambulant or outpatient setting as a proxy for special needs. This is a broader group of

[3] The ATC4 classification codes are psychostimulants, agents used for ADHD and nootropics (N06B) for externalizing behavioural problems, and antidepressants (N06A) and anxiolytics (N05B) for internalizing behavioural problems.

potentially disruptive and emotionally unstable peers as the legal reasons to be indicated for this care vary from mild psychosocial problems to drug abuse (CBS 2018). Fourth, we use child custody and family supervision orders, which are far-reaching measures ordered by a judge. On the grounds of the Dutch Civil code[4] a minor can be placed under supervision by the juvenile court "if he is growing up in such a way that his moral or spiritual interests or his health are at serious risk, and other means to avert this risk have failed or are expected to fail". Hence students under supervision' are likely to have suffered from a severe situation that was a developmental danger for them.

Table 1 and 2 report descriptive statistics on the four groups of SN and non-SN students in primary and secondary education, combined with information from the government organization that finances schools and administrates educational data (DUO), described in the next paragraphs. It shows that in primary and secondary education disruptive peers in regular classrooms are more often male (72 and 66%) compared to non-SN students (48 and 49%). This is expected, as ADHD is relatively more common among boys (Aizer, 2008). Also, disruptive peers are more often Dutch and from higher educated parents. There are more peers under supervision order with (one of their) parents with a non-western migration background[5] and with additional funding based on expected educational disadvantage. This is as expected, as the funding is based on parental education and migration background.

Further, the tables illustrate that students with SN in primary education are somewhat older than students without SN (for example 11.15 for disruptive students and 11.02 for non-disruptive students). Whereas there are some differences in background characteristics between the different groups of SN students, the differences in age, test participation, and test scores in primary education illustrate that students with SN have more difficulties in their school careers. Before 2014, inclusive education was facilitated by a pupil-bound budget to fund additional support to educate children with severe special educational needs in regular classrooms. The table shows there is a higher share of students in the

[4] Article 1:254, a minor can be placed under supervision by the juvenile court "if he is growing up in such a way that his moral or spiritual interests or his health are at serious risk, and other means to avert this risk have failed or are expected to fail"

[5] Turkey or a country in Africa (save Indonesia and Japan) or Latin America

different groups of SN in primary and secondary education who received this 'backpack funding' compared to the non-SN groups.

4.2 Data primary education

For primary education, data from the nation wide test is available on all Dutch students in 2015, 2016, 2017, and 2018. The purpose of this nation wide test is to track children in to secondary education levels[6]. The data are provided by the government organization that finances schools and administrates educational data (DUO). The data include information on final test scores and student background characteristics, such as gender, ethnicity, and student zip code area. The final test scores are standardized for ease of interpretation. As described earlier, not every student of a school participates in the final test. Table A.1 shows that non-SN students who don't take the final test are less often from a Dutch background and slightly older. This school dependent selection of participating students illustrates the relevance of including final test participation as an outcome measure in our main analyses.

Table 1 reports descriptive statistics on the number of (SN) students and the number of (SN) students per school cohort in primary education. By including 5240 schools the data covers around 70 percent of all the Dutch schools for primary education. The four SN groups differ in size, 55207 students with youth care and 18128 disruptive peers being the most prevalent and 611 peers with emotional problems and 4011 peers under supervision being less prevalent.

The final rows of table 1 show the means and standard deviations of the number of primary school students with and without SN per school leaving primary education in the same year (cohort). The average cohort size is 39.6 and the average number of SN students per cohort ranges between 0.3 and 4.5, dependent on the type of special needs. In our analyses, we focus on the number of SN students per school cohort, for both practical and substantive reasons. Practically, our data do not allow us to identify the exact classes within schools but do allow us to identify cohorts. Substantively, students and

[6] Since 1940, most schools used the 'cito toets' as final test (85%). When it became mandatory to have a final test in primary education in 2015, more providers entered the marked and five tests were approved by the minister. Even though there are quality criteria, the results of the new tests are not comparable over time and between tests as they differ in standardization and were not validated in a similar way (CPB 2019). In this study, we therefore focus on schools that take the 'cito toets' (around 70 percent of all the Dutch schools for primary education).

teachers may not be randomly allocated to classes within cohorts, which may create bias. When students with special needs are educated by better teachers, for example, the estimated coefficients may be smaller than the actual effects. By focussing on the cohort level, the results can be interpreted unambiguously based on the within-school-between-year variation in SN students. Yet, it is conditional on the notion that schools may mitigate potential negative effects by selective assignment of teachers to SN students. Therefore, we also perform a robustness check on smaller school cohorts in Section 7, in which we are more certain that the student with SN is present same class.

[Insert Table 1 here]

4.3 Data secondary education

For secondary education, administrative data on all Dutch students taking secondary school exams in 2015, 2016, 2017, and 2018 are used. The data from the Education Executive Agency of the Dutch Ministry of Education, Culture and Science contain exam grades for each subject for each student and the background characteristics gender, age, and ethnicity. We merge the data with data on parental education from Statistics Netherlands. To facilitate interpretation, exam grades are standardized within school tracks. If a student is in the sample twice because of failing the exams the first time or attaining a higher school track (7% of the observations), each observation is treated as a separate entry. Since students get educated within their school track, we define a school cohort as students in the same track, in the same school, in the same year. We summarize the data from the different tracks within schools to "all tracks".

In contrast to table 1 with the descriptive statistics for primary school, table 2 shows that in secondary education compared to students without SN, there are more girls with SN with medication for emotional problems (73% versus 50%) an youth care (58% versus 49%). There are fewer students with a non-western background who receive youth care. Again, disruptive peers in regular classrooms are compared to non-SN more often male, Dutch, and from (highly) educated parents.

In level I of the pre-vocational track of secondary education, students with an average intelligence quotient but learning disabilities can (temporarily) receive extra educational care, for example, attend smaller classes or receive homework assistance. The table shows there is a higher share of students who receive this extra educational care in the different groups of SN students. Together with the higher share of backpack funding, we assume the students with SN have more difficulties in their school careers. It might seem surprising that three types of SN students have slightly higher exam grades than students without that specific SN. This may be explained by secondary schools assigning SN students to a school track that is lower than their ability. The Dutch Inspecorate of Education (2022) demonstrates that students with special needs are more often in a lower track in year 3 of secondary education than indicated by their primary school advice than students without special needs. Additionally, students with special needs more often receive a school track advise that is lower than their ability as indicated on the final test in primary education compared to students without special needs. In primary education, where students are not tracked, SN students score lower than non-SN students.

The final rows of table 2 show the means and standard deviations of the number of students with and without SN for all tracks per school cohort. The average cohort size is 99.7 and the average number of SN students per cohort ranges between 0.4 and 7.7, dependent on the type of special needs.

[Insert Table 2 here]

5 Empirical strategy

5.1 Student fixed effects

There are several endogeneity concerns in estimating the peer effects of SN students on the academic achievement of students without SN. The prevalence of SN students in a school is potentially endogenous and may be related to the quality of the school. Weaker schools may have more problems educating SN students and could be more likely to decide to reject a SN student Or vice versa, strong schools could have more applications and have fewer incentives to keep students in regular education. To address these endogeneity concerns, we use two main identification strategies.

The first main strategy is exclusively for secondary education and is based on the fact that secondary school students are free to select different courses within the secondary school program. Each course has its exam. As such, the class composition can vary per course. We exploit this within-student-between-course variation in the number of SN students using student fixed effects. The first strategy evaluates student achievement in courses with varying numbers of SN students.

$$y_{iclst} = \delta_1 SN_{clst} + W'_{clst}\gamma_1 + Crs_{1,cl} + \zeta_{1,i} + \varepsilon_{1,icst} \quad (1)$$

Here, y_{iclst} denotes the standardized exam grade of student i, for course c, in track l in school s in year t. SN_{clst} indicates the number of students with special educational needs in course c in track l in school s in year t, which makes δ_1 the parameter of interest. W'_{clst} is a vector of peer characteristics such as the percentage of boys in course c in track l in school s in year t. $Crs_{1,cl}$ are the course- and track fixed effects, which include general differences in performances between course and tracks. $\zeta_{1,i}$ are the student fixed effects, picking up general differences between students, for example, ability. Then, $\varepsilon_{1,iclst}$ is an individual course-specific error term. As we estimate this equation for each subgroup of SN students, the non-SN students in the sample are defined by not having the specific SN of the subgroup we estimate. As the number of students without that specific SN differs over the subgroups, the number of students in the regressions differ as well.

This strategy solves several potential selection problems. The student fixed effects capture all student constant characteristics and the characteristics of the school and course. What remains is variation in exam grades within students between different courses. This strategy will uncover whether these differences are systematically related to the number SN students in a course. What is important here, is that the number of SN peers is not constant within individuals. Figure A.1 illustrates that the number of SN students indeed varies within students. The figure displays the frequencies differences between the maximum and the minimum number of students with special needs in a course. When the difference is 0, there are either no students with SN in any of a students' courses, or the number of SN students is constant over all the courses.The figure shows that the two most prevalent subgroups (disruptive peers and peers with youth care) have the most variation, where as in the two less prevalent

subgroups there are more students with zero variation. The identifying assumption in this strategy is that the prevalence presence of SN students in a course is random within individuals: students should not select into courses because of (avoiding) certain peers. Given the importance of course choice for future education and employment possibilities, this is very likely.

5.2 School fixed effects

While the student fixed effects strategy works for secondary education, there is no within-student-between-course variation in primary education, as class composition is constant across subjects. To identify the effect the within-school-between-year variation in the numbers of SN students is exploited. We use the following specification to estimate the school fixed effects:

$$y_{ist} = X_{ist}^{'}\beta_2 + P_{st}^{'}\gamma_2 + \delta_2 SN_{st} + \mu_{2,s} + \nu_{2,t} + \varepsilon_{2,ist} \qquad (2)$$

y_{ist} indicates achievement of student i in school s in year t. $X_{ist}^{'}$ is a vector of individual students characteristics, such as ethnicity and parental education. $P_{st}^{'}$ is a vector or peer characteristics in school s in year t such as the percentage of boys. SN_{st} denotes the number of SN students in school s in year t, which makes δ_2 the parameter of interest. $\mu_{2,s}$ are the school fixed effects which will pick up time-invariant differences between schools, $\nu_{2,t}$ indicates the year fixed effects and $\varepsilon_{2,ist}$ is an individual-specific error term. We assume $\varepsilon_{2,ist}$ to be exogenous apart from school and year-fixed effects. Again, as we are interested in the effect of the different groups of SN students, the specific SN group is dropped from the sample. Since the school fixed-effect strategy exploits year-to-year variation in the number of SN students, the number of SN students must differ between cohorts in the same school. Figure A.2 shows the difference between the maximal and minimal number of SN students per cohort within schools and the number of schools for primary and secondary education with this difference.

The school fixed effects strategy only picks up time-invariant differences between schools. The outcomes may be explained by unobserved school factors related to the changes in the number of SN students over time. Equation 3, therefore, includes a school-specific linear time trend, containing the

predicted values of the number of SN students in a school(track) in a certain year. We estimate this by using a linear trend through the actual number of SN students in the different years.

$$y_{ist} = X'_{ist}\beta_3 + P'_{st3}\gamma_3 + \delta_3 SN_{st} + \theta_3 PREDSN_{st} + \mu_{3,s} + \nu_{3,t} + \varepsilon_{3,ist} \quad (3)$$

The potential for non-random assignment of SN students to classes within schools introduces endogeneity concerns (e.g. if SN students are more likely to be assigned to high-quality peers and/or teachers). The school fixed strategy addresses this by exploiting within-school between-year variation. The identifying assumption then is that the presence of SN students in cohorts is unrelated to changes in overall peer and/or teacher quality. The balancing tests in table A.2 yield support for this. Out of the 140 estimated coefficients, 6 turn out significant at the 5% level, which concurs with what can be expected by chance. Further, the magnitude of these differences is very small. For example, when looking at primary education, boys have, on average 0.019 fewer peers with youth care in their cohort. When testing for joint significance by regressing the proportion of SN peers on gender, age, migrant background, and disadvantaged neighbourhood, we cannot reject the null hypotheses that these coefficients are jointly zero. Hence, once controlled for school fixed effects, the within schools cohort-to-cohort variation in the percentage of SN students is not associated with observed characteristics of students without SN.

6 Main results

Table 5 shows the main results from our baseline estimations of students fixed effects models and school fixed effects models with and without a school-specific time trend in the number of SN students. Each column represents a different group of students with SN. These results indicate that the number of SN students in the course or cohort is unrelated to a students' exam grade in secondary education and test score and test participation in primary education.

Out of the 28 estimated coefficients, there are two coefficients significant at the 5% level, one positive and one negative. As we perform multiple tests, this concurs with what we would expect to find by chance (type I error) when all null-hypotheses would be true. When accounting for multiple testing

with the Bonferroni-Holm method, the null hypothesis cannot be rejected for any of the 28 estimated coefficents.

Further, the practical relevance of the coefficients is low. The positive one is 0.002 sd for peers with Youth Care in the student fixed-effect strategy. The other (and largest) one is -0.034 sd of the final test score in primary education and only found for peers with emotional problems in the model including the school-specific time trend. The rest of the estimates are statistically insignificant and very small.

[Insert Table 3]

7 Robustness checks

Jacob et al. (2019) propose a conceptual framework for defining and interpreting null results and suggest a method for systematically examining a set of potential reasons for observing null findings. They define a null result as one in which the estimated impact is both small and has a tight confidence driven by enough precision. Next, when interpreting null findings they suggest considering several factors, for example, whether the expectations were realistic, based on what was known about the intensity of the intervention. When taking these factors into account and the result is considered truly null, they suggest researchers must explore potential causes of the findings. We use these criteria from Jacob et al. (2019) to determine if we can interpret the main results presented in the previous paragraph as a true null finding. First, according to this framework the expected size of the estimated effect is a factor to consider when designing a study. We compare if our intervention intensity is a least equal to other studies where negative effects were found (section 7.1). Second, we test for non-linear effects and test the main results on proportion instead of the number of SN students as a robustness check (section 7.2). Finally, we address the interpretation concerns resulting from the so-called proximity of the outcome by discussing less distal outcomes than peers' cognitive outcomes (section 7.2).

After this, we use the framework to explore potential reasons for not finding a negative effect of SN students on their peers' cognitive outcomes. We perform robustness checks utilizing neighbourhood variation in the number of SN students, restrict our samples to courses and cohorts with

30 or fewer students, explore heterogeneous effects and investigate whether the results are robust to other definitions of SN students.

7.1 The expectation of impact and effect size

To consider if the null result is truly null, one of the criteria of the framework is to compare expected or estimated impacts to other programs to determine what might be reasonable to expect. In our fixed effect model with panel data, we apply this criterion by comparing our results to that of similar studies. Table 4 gives an overview of the findings of recent studies. The table illustrates that the selection of the different groups of SN students is comparable with previous studies. Our proxies cover a broad group of SN students with emotional or behavioural problems. The type of intervention is similar, as the intervention is the adding of SN students to the class. Our point estimates are mostly above (closer to zero) the upper bounds of the confidence intervals estimated in these studies, supporting that our results are indeed less negative.

In the studies in table 4, the class size – when known and part of the estimate – is between 20 and 25. In the Netherlands the class size is on average 24 in primary education and on average between 21 and 27 in secondary education, depending on the track and course. In all of these studies, the outcome is a measure of academic achievement, such as (final) test scores in math or reading. We use comparable standardized final test scores in primary education and exam scores in secondary education. As we use data on all Dutch students leaving primary and secondary education from 2015 until 2018 (table 1 and table 2), our sample is sufficiently powered to detect even small effects of SN students on their peers. Based on a comparison of sample size, the definition of SN, the type of intervention, and the design we conclude it is realistic to assume that if there was an impact, our empirical estimation strategy would be able to find it.

[Insert Table 4]

7.2 Non-linear effects

In the main results, we estimate the impact of an additional SN student on the academic achievement of their peers. However, there may be non-linear effects, such as a tipping point with negative effects after including a few SN students. We, therefore, check whether the null finding is robust to non-linearities by separately estimating the effect of the presence of one, two, three, four, and five or more SN students. Figures 2 and 3 show the results from school fixed effects models and final test scores in primary education and exam grades in secondary education. Figure A.2 and A.3 present results of final test participation in primary education and the student fixed effects models in secondary education.

[Insert Figure 2 here]

[Insert Figure 3 here]

In most panels, the estimates are comparable: the results are not significantly different from zero and the results do not systematically indicate that the effects are more or less negative when including a larger number of SN students. As there are fewer schools with a large number of SN students, the confidence intervals get wider and the estimates less precise. In some cases, for students with emotional problems in primary education, for instance, the number of schools with three or more students with emotional problems is too small to estimate the effects. Overall, we conclude there is no evidence for non-linear trends that explain overall null results.

In some papers, the impact of SN students on their peers is measured by using the proportion of SN students related to class size (e.g. Balestra, Eugster and Liebert 2020; Carrell, Hoekstra and Kuka 2018) As a robustness check, table A.3 shows our main results, using proportion (percentage) of SN students instead of a number as intervention. Again, no significant effects are found at the 1% level.

7.3 The proximity of the outcome to the intervention

The proximity of the outcome can pose a concern in the context of this study if the mechanism that drives our null results is that the outcomes that we evaluate are too distal from the actual effects. In this case, potentially negative effects on more proximal outcomes may be suppressed by other mechanisms.

We investigate the effect of SN students on their peers' cognitive outcomes. As shown in table 6, academic achievement is a widely used outcome in peer effect studies (Ahn and Trogan 2017; Balestra, Eugster and Liebert 2020; Carrell, Hoekstra and Kuka 2018; Contreras et al. 2020; Kristoffersen et al. 2015; Rangvid 2016), and these studies generally do find negative effects of including SN students. We, therefore, conclude that it is unlikely that our outcome measures are too distal to pick up on potential negative effects.

An interesting alley for further research is the impact of special needs students on teachers. Teachers play a crucial part in the interaction between students in class dynamics. When teachers are compensating for potential negative externalities of students with SN, the presence of SN students may not affect the achievement of peers but may affect teacher workload. Indeed, since the Dutch policy reform, a majority of teachers report an increase in work stress (Smeets, Ledoux, and Van Loon-Dikkers 2019). Ultimately, work stress could lead to sickness absence or (propensity) to leave or burnout. A more proximal outcome could therefore be the teacher's well-being, work stress, or maybe even turnover. Unfortunately, these data are not available at the cohort level.

7.4 Unobserved factors related to within-school cohort quality

So far, we've established that our research design is quite similar to other studies and that it is unlikely that our null results can be explained by non-linearities or being underpowered. In the next paragraphs, we explore some methodological factors that may explain the null results.

A possible unobserved factor related to the presence of SN students is within-school cohort quality. In the school fixed effects strategies, we assume teachers and principals do not take characteristics of the whole cohort into account in their placement decision. However, SN students might

be more easily referred to special education when the general quality of the school cohort is lower. The estimates could then be biased as we might find no effects of the presence of SN students, while the achievement of inclusive cohorts might have been better without the presence of SN students. In secondary education, we account for this concern by using the student fixed effects strategy.

In primary education, this is not possible, as the class composition does not vary per course. We, therefore, perform a robustness check by using a strategy that does not depend on within-school cohort-to-cohort variation but exploits neighbourhood variation in the number of SN students instead. We assume the presence of an additional SN student in a district to be exogenous: parents are not likely to move because of the presence of an additional SN student in the neighbourhood. Further, schools cannot influence the number of SN students in their locality, but most primary school students go to a school in their 4-digit post code area (Ruijs 2017).[7]

This strategy investigates whether variation in the number of SN students in a neighbourhood affects the achievement of non-SN students in that area. At the neighbourhood level, the percentage of SN students in a cohort is fixed. As these results are unaffected by possible cohort-dependent choices to refer students to special education, observing similar results would corroborate our results. The equation we estimate is:

$$y_{ina} = X'_{ina}\beta_4 + P'_{ina}\gamma_4 + \delta_4 sn_{na} + \eta_{4,n} + \upsilon_{4,a} + \varepsilon_{4,ina} \qquad (4)$$

Here, y_{ina} indicates final test score or final test participation for student i in neighbourhood n in neighbourhood cohort a. X'_{ina} is a vector of student covariates, P'_{ina} is a vector of peer characteristics. sn_{na} denotes the percentage of students with SN in neighbourhood n in cohort a, which makes δ_4 the parameter of interest. $\eta_{4,n}$ and $\upsilon_{4,a}$ are neighbourhood and cohort fixed effects.

In the neighbourhood analysis, we investigate students who would be in the final grade of primary education when following the typical primary school path. We define a cohort as all students in a 4-digit

[7] A 4-digit post code area is an administrative neighbourhood area with an average of 49456 inhabitants (Statistics Netherlands, 2020).

postal code area who are 11 years old on October 1st. SN students are defined by the five groups and include students attending special education, such that we cover all 11-year olds in a neighbourhood.

Table 5 shows the results of neighbourhood fixed effects model. The results are similar to the results for the school fixed effect models and show only two significant effects for students under supervision orders and final test participation. Thus, the neighbourhood fixed effects approach does not provide evidence that the aforementioned null findings are due to selective referral to special education.

[Insert Table 5 here]

7.5 Cohort size

As we investigate the presence of SN students at the level of the grade, it is possible that students in bigger schools are not exposed to the SN students when the SN student is in the same cohort (primary education) or chose the same subject (secondary education) but attends another class. Also, within cohorts, the presence of SN students may be non-random. Their presence, might, for instance, be related to teacher quality. We perform robustness checks restricting the sample to include only small schools to investigate this potential bias. In smaller schools, all students are exposed to a SN student when a SN student is present in their cohort.

Table 6 shows the results similar to Table 3 but restricted to cohorts and courses with 30 or fewer students. The coefficients are again generally insignificant and inconsistent in sign. Compared to the main results in table 5, there is a change from the significant positive effect of students with youth care in de student fixed effects model to a negative one; yet both estimates are very small in magnitude. The two significant coefficients (one positive, one negative) in the school fixed effects model in primary education are not present when including a school-specific time trend. For the models including the school-specific time trend, we find one negative effect of peers with Youth Care on final test participation at the 1% level out of the 16 estimated coefficents.

[Insert Table 6 here]

7.6 Heterogeneous effects

Next to the methodological concerns, we use the criteria of Jacob et al. (2019) to address concerns about other potential causes of the findings related to the underlying mechanisms of peer effects. The framework suggests the assumed mechanism might be generally right, but only for certain populations. For that reason, we need to rule out possible subgroup impacts. As peer effects are often heterogeneous (Balestra, Eugster and Liebert 2020; Booij, Leuven and Oosterbeek 2017; Feld and Zölitz 2017; Hoxby, 2000), the zero results might be explained by differential effects for students at different points of the achievement distribution.

For instance, classmates might learn less if SN students are misbehaving or in another way causing the teacher to reallocate time to the SN students instead of instruction (Gottfried 2013; Lazear 2001). The disruptive behaviour might directly be distractive for non-SN students and therefore slow down their learning process, or the teacher could pay less attention to non-SN students. This might especially affect low-achieving students, whereas high achieving students might be less dependent on the teacher's instruction. The null effect could be the result of the average between a positive effect on high- and negative on low-achieving students. To investigate these differential effects, the data is aggregated to the cohort level. We compute different measures of student achievement at the cohort level, such as the mean, standard deviation, and different percentiles of the exam grades and test scores. We use the following equation to estimate differential effects:

$$y_{st} = P'_{st}\gamma_5 + \delta_5 SN_{st} + \theta_5 PREDSN_{st} + \mu_{5,s} + \nu_{5,t} + \varepsilon_{5,st} \quad (5)$$

y_{st} are the outcomes of a cohort of students in school s in year t, for example, the standardized exam grade of students in the 90th percentile. P'_{st} is a vector or peer characteristics in school s in year t such as the percentage of students with low parental education level. SN_{st} denotes the number of SN students in school s in year t, which makes δ_3 the parameter of interest. $\mu_{5,s}$ are the school fixed effects which will pick up time-invariant differences between schools, $\nu_{3,t}$ indicates the year fixed effects and $\varepsilon_{5,ist}$ is an individual-specific error term. We assume $\varepsilon_{5,ist}$ to be exogenous when controlled for school

and year-fixed effects. To account for differences in school size, the regressions are weighted by the number of observations per school.

Table 7 presents the results of the differential effects on students with different abilities for primary and secondary education. Out of the 48 coefficients, there are two significant coefficients at the 5% level, and one at the 1% level, which is close to what we would expect by chance. Also, there are no clear patterns in sign or for students with different abilities. The coefficient for an additional student with emotional problems is negative for the 25th percentile in primary education, whereas it is positive for students with supervision orders in secondary education. Overall, these results do not yield support for differential effects of the presence of SN students in the Netherlands.

[Insert Table 7 here]

In secondary education, the null effect might be explained by another source of heterogeneous effects: maybe the effect depends on the ability track. It is conceivable that in the lowest track students are more dependent on the teacher's instruction and therefore the effect is more negative. Similarly, there may be less additional support available in the higher tracks, which may create more negative effects there. Therefore, we also estimate the student fixed effects and the school fixed effects models per school track in secondary education. Table 8 presents the results of the differential effects per track in secondary education. Again, the results are similar: the coefficients are small and mostly insignificant. From the 40 coefficients, 3 are significant, but they are inconsistent in sign.

[Insert Table 8 here]

7.7 Untreated disruptive peers

Even though our proxies to identify SN students are comparable to those used in the existing literature, the absence of negative effects may be related to the definition of SN students. As we use prescripted medication as one of our proxies, it is imaginable that the assumed type of disruptive behaviour is

weakened. Aizer (2008) concludes that the presence of students with ADD harms peers' reading and math test scores before the diagnosis, but the achievement of peers improves after the diagnosis. As a robustness check, we, therefore, define 'untreated disruptive peers' as peers who will be prescribed the medication for externalizing behaviour problems in the next but not the current school year. As we use data from the next school year, we can only define untreated disruptive pears in 2016 and 2017. If the null result would be related to treatment results, this robustness check would reveal negative effects. Table 9 shows no evidence for the effect of 'untreated' students, which supports our conclusions that there are no negative peer effects of disruptive students in the Netherlands.

[Insert Table 9 here]

8 Conclusions

This paper presents evidence that the presence of students with special needs has no impact on the academic achievement of students without special needs in the Netherlands. Using student-level population data, we investigate four groups of students with SN: disruptive peers, peers with emotional problems, peers who receive youth care, and peers who are under supervision orders. The null finding is consistent across these different definitions of SN students and is robust to different ways of modelling the relationship between achievement and peers with SN. These results concur with the null effects found in an earlier Dutch study on students with diagnosed severe special educational needs (Ruijs 2017). In addition, by using four broad groups of special needs students, we demonstrate the null effect is not related to a specific kind of severe diagnosed educational needs and the associated fundings.

Using criteria from the framework for learning from null results (Jacob et al. 2019), we systematically examine whether the results can be interpreted as a true null result and explore the potential causes of the findings. Using three complementary fixed effect estimation strategies, the empirical results are unambiguous: estimates across all specifications show no significant effect on their peers' cognitive outcomes for all four types of SN students. We investigate whether the relation between the presence of SN students and the achievements of non-SN students is non-linear and check for heterogeneous effects of SN students on high and low achieving students. Overall, the results are similar. These results contribute to the overall international results, in which overall more negative results on

the peer effects of SN students are reported (Ahn & Trogan 2017; Balestra, Eugster and Liebert 2020; Carrell, Hoekstra and Kuka 2018; Contreras et al. 2020; Kristoffersen et al. 2015). Below, we outline two distinguishing elements in the Dutch educational context which might be potential explanations for these findings: extra resources and teacher professional development activities.

There is a long history of additional resources for students with special needs in the Dutch context. In 2002, a pupil-bound budget was introduced which was granted by a regional committee based on national eligibility criteria. Dependent on the type of SN, schools could receive up to three times the regular funding for including a SN student. Even though the SN students studied in this paper are not the same students as the ones covered by this arrangement, the current null might imply that additional support in classes is sufficient to avoid negative externalities of the selected SN students. Whereas this specific arrangement was terminated in 2014, the additional funds are still available through regional alliances. Dutch teachers might therefore be better supported, which may prevent negative externalities. In support of this hypothesis, Contreras et al. (2020) found that with resources and adequate treatment the negative effect of SN students vanishes. Further, in Canada, where schools also receive substantial additional funding, insignificant effects were found as well (Friesen, Hickey, and Krauth 2010).

Second, to raise the share of teachers with a Master's degree the Dutch government started to provide teacher vouchers in 2008. These vouchers compensate teachers for tuition fees and yield additional funds to their employer to arrange a substitute teacher. In the first five years after the introduction of the voucher, nearly 400 million euros have been granted to about 40 thousand teachers and schools (Heijnen, Van der Steeg, and Ooms 2015). A vast majority (77%) of the teachers in primary education enrol in a program targeted at the mastery of specific pedagogical skills, the Master Special Educational Needs (Heijnen, Van der Steeg and Ooms 2015). These professional development activities might contribute to the quality of teachers and might prevent negative effects of the presence of students with special needs on students without these needs in their classroom.

Finally, we study the impact of SN students on peers' academic performance measured by school grades or test performance. Although this is a widely used approach (Ahn and Trogan 2017; Balestra, Eugster and Liebert 2020; Carrell, Hoekstra and Kuka 2018; Contreras et al. 2020; Kristoffersen et al 2015; Rangvid 2016), it is possible that class composition also affects non-cognitive outcomes such as

socio-emotional functioning or well-being at school (Ruijs, Peetsma and Van der Veen 2010). For instance, the behaviour of SN students might affect teachers' choices in class organisation or class curriculum. In the results of the Dutch evaluation of the inclusive education policy (Smeets, Ledoux, and Van Loon-Dikkers 2019) teachers reported feeling that students without special needs get less attention due to the reform. With SN student(s) in the class, teachers might deliberately choose to focus more on the core subjects and pay less attention to non-cognitive outcomes or enrichment of the curriculum. These choices may affect the behaviour or well-being of other students. To find out the mechanisms that explain the null result, future research could focus on the interventions for teachers to effectively deal with disruptive behaviour and on the effects of SN students on curriculum choices or other student outcomes, such as well-being and socio-emotional functioning.

Appendix

[Insert figure A.1 here]

[Insert figure A.2 here]

[Insert figure A.3 here]

[Insert figure A.4 here]

[Insert table A.1 here]

[Insert table A.2 here]

References

Ahn, Tom, and Justin G. Trogdon. 2017. Peer delinquency and student achievement in middle school. *Labour Economics* 44: 192–217. https://doi.org/10.1016/j.labeco.2017.01.006

Aizer, Anna. 2008. Peer Effects and Human Capital Accumulation: the Externalities of ADD. NBER working Paper No. w14354. https://doi.org/10.3386/w14354

Balestra, Simone, Beatrix Eugster, and Helge Liebert. 2020. Peers with Special Needs: Effects and Policies. *The Review of Economics and Statistics*, 1–42. https://doi.org/10.1162/rest_a_00960

Bietenbeck, Jan. 2020. The Long-Term Impacts of Low-Achieving Childhood Peers: Evidence from Project STAR. *Journal of the European Economic Association* 18(1): 392–426. https://doi.org/10.1093/jeea/jvy049

Booij, Adam S., Edwin Leuven, and Hessel Oosterbeek. 2017. Ability Peer Effects in University: Evidence from a Randomized Experiment. *The Review of Economic Studies* 84: 547-578. https://doi.org/10.1093/restud/rdw045

Brodaty, Thibault, and Marc Gurgand. 2016. Good peers or good teachers? Evidence from a French University. *Economics of Education Review* 54: 62–78. https://doi.org/10.1016/j.econedurev.2016.06.005

Carrell, Scott E., Mark L. Hoekstra, and Elira Kuka. 2018. The Long-Run Effects of Disruptive Peers. *American Economic Review* 108(11):3377-3415. https://doi.org/10.1257/aer.20160763

Carrell, Scott E., and Mark L. Hoekstra. 2010. Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids. *American Economic Journal: Applied Economics 2*(1): 211–228. https://doi.org/10.1257/app.2.1.211

Centraal Bureau voor de Statistiek. 2019. *Jeugdhulp 2018*. Den Haag: Centraal Bureau voor de Statistiek [Youth Care 2018]

Contreras, Dante, Miguel Brante, Sebastian Espinoza, and Isabel Zuñiga. 2020. The effect of the integration of students with special educational needs: Evidence from Chile. *International Journal of Educational Development* 74: 102163. https://doi.org/10.1016/j.ijedudev.2020.102163

Feld, Jan, and Ulf Zölitz. (2017). Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects. *Journal of Labor Economics* 35(2):387-428. https://doi.org.10.1086/689472

Figlio, David N. 2007. Boys Named Sue: Disruptive Children and Their Peers. *Education Finance and Policy* 2 (4): 376–394. doi: https://doi.org/10.1162/edfp.2007.2.4.376

Fletcher, Jason. 2010. Spillover effects of inclusion of classmates with emotional problems on test scores in early elementary school. *Journal of Policy Analysis and Management 29*(1): 69–83. https://doi.org/10.1002/pam.20479

Friesen, Jane, Ross Hickey, and Brian Krauth. 2010. Disabled Peers and Academic Achievement. *Education Finance and Policy* 5(3): 317–348. https://doi.org/10.1162/EDFP_a_00003

Gottfried, Michael A. 2014. Classmates With Disabilities and Students' Noncognitive Outcomes. *Educational Evaluation and Policy Analysis 36*(1): 20–43. https://doi.org/10.3102/0162373713493130

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2002. Inferring Program Effects for Special Populations: Does Special Education Raise Achievement for Students with Disabilities? *Review of Economics and Statistics* 84(4): 584–599. https://doi.org/10.1162/003465302760556431

Heijnen, Suzanne, Marc van der Steeg and Tahnee Ooms. 2015. *Lerarenbeurs: Aanvragers, Deelname en Afronding*. Den Haag: CPB [Teacher voucher: apply, participate and finish]

Horoi, Irina, and Ben Ost. 2015. Disruptive peers and the estimation of teacher value added. *Economics of Education Review* 49: 180–192. https://doi.org/10.1016/j.econedurev.2015.10.002

Hoxby, Caroline 2000. Peer effects in the classroom: learning from gender and race variation. NBER Working paper No. 7876

Huang, Bin, Haiyang Lu, and Rong Zhu. 2021. Disabled Peers and Student Performance: Quasi-Experimental Evidence from China. Economics of Education Review 82: 102121. https://doi.org/10.1016/j.econedurev.2021.102121

Huang, Bin, and Rong Zhu. 2020. Peer effects of low-ability students in the classroom: evidence from China's middle schools. *Journal of Population Economics 33*(4): 1343–1380. https://doi.org/10.1007/s00148-020-00780-8

Inspectorate of Education (2020). *De Staat van het Onderwijs 2020.* Utrecht: Inspectorate of Education. [The State of Education in the Netherlands 2020]

Inspectorate of Education (2022). *De Staat van het Onderwijs 2022.* Utrecht: Inspectorate of Education. [The State of Education in the Netherlands 2020]

Inspectorate of Education. 2018. *Zicht op de besteding van de middelen voor passend onderwijs.* Utrecht: Inspectorate of Education. [Insights in the spending of funding for inclusive education]

Jacob, Robin T., Fred Doolittle, James Kemple, and Marie-Andrée Somers. 2019. A Framework for Learning From Null Results. *Educational Researcher* 48(9): 580–589. https://doi.org/10.3102/0013189X19891955

Kristoffersen, Jannie Helene Grøne, Morten Visby Krægpøth, Helena Skyt Nielsen, and Marianne Simonsen. 2015. Disruptive school peers and student outcomes. *Economics of Education Review* 45: 1–13. https://doi.org/10.1016/j.econedurev.2015.01.004

Lazear, Edward P. 2001. Educational Production. *The Quarterly Journal of Economics* 116(3): 777–803. https://doi.org/10.1162/00335530152466232

Lavy, Victor, Daniele Paserman and Analia Schlosser. 2012. Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The economic Journal* 122(559): 208-237

Lavy, Victor, Olmo Silva, and Felix Weinhardt. 2012. The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools. *Journal of Labor Economics* 30(2): 367–414. https://doi.org/10.1086/663592

Ministerie van Onderwijs. 2020. *OCW Rijksbegroting*. Den Haag: Ministerie van Onderwijs Cultuur en Wetenschap [Department of education State budget]

Neidell, Matthew, and Jane Waldfogel. 2010. Cognitive and noncognitive peer effects in early education. *The review of Economics and statistics* 92(3): pages 562-576 https://doi.org/10.1162/REST_a_00012

Horoi, Irene, and Ben Ost. 2015. Disruptive peers and the estimation of teacher value added. *Economics of Education Review* 49: 180–192. https://doi.org/10.1016/j.econedurev.2015.10.002

Ohinata, Asako, and Jan C. van Ours. 2013. How Immigrant Children Affect the Academic Achievement of Native Dutch Children. *The Economic Journal* 123(570): F308–F331. https://doi.org/10.1111/ecoj.12052

Rangvid, Beatrice Schindler. 2019. Returning special education students to regular classrooms: externalities on peers' outcomes. *Economics of Education Review, Elsevier* 68(C): 13-22. https://doi.org/10.1016/j.econedurev.2018.11.002

Ruijs, Nienke. 2017. The impact of special needs students on classmate performance. *Economics of Education Review* 58: 15–31. https://doi.org/10.1016/j.econedurev.2017.03.002

Ruijs, Nienke, Thea Peetsma, and Ineke van der Veen. 2010. The presence of several students with special educational needs in inclusive education and the functioning of students with special educational needs. *Educational Review* 62(1): 1–37. https://doi.org/10.1080/00131910903469551

Smeets, Ed, Guuske Ledoux, and Liselotte van Loon-Dikkers. 2019. P*assend onderwijs in de klas. Tweede meting in het basisonderwijs en voortgezet onderwijs*. Nijmegen/Amsterdam:

KBA/Kohnstamm Instituut [Nr 56 in de reeks Evaluatie Passend onderwijs]. [Inclusive education in class. Second survey in primary and secondary education]

UNESCO. 2020. *Towards inclusion in education: status, trends and challenges: the UNESCO Salamanca Statement 25 years on*. France: United Nations Educational, Scientific and Cultural Organization

Table 1. *Descriptive statistics for students in primary education*

	Disruptive		Emotional problems		Youth Care		Supervision order	
	non-SN	SN	non-SN	SN	non-SN	SN	non-SN	SN
Boy	0.487	0.723***	0.499	0.524	0.487	0.58***	0.498	0.468***
Parental ed. unknown	0.184	0.189	0.184	0.196	0.183	0.157***	0.181	0.092***
University	0.185	0.145***	0.184	0.16	0.191	0.159***	0.188	0.059***
University of applied sciences	0.232	0.233	0.232	0.211	0.235	0.233	0.236	0.11***
Vocational ed. level 3-4	0.234	0.278***	0.236	0.265	0.23	0.275***	0.235	0.29***
Vocational ed. level 2	0.164	0.155**	0.163	0.167	0.16	0.176***	0.159	0.449***
Dutch background	0.749	0.871***	0.755	0.825***	0.747	0.789***	0.753	0.642***
Non-western migrant	0.181	0.079***	0.176	0.115***	0.183	0.143***	0.177	0.252***
Western migrant	0.069	0.049***	0.069	0.061	0.071	0.068**	0.07	0.106***
Weighted student funding	0.098	0.077***	0.097	0.08	0.096	0.095	0.094	0.202***
Age at October 1st	11.024	11.155***	11.030	11.121***	11.017	11.093***	11.024	11.228***
Backpack funding	0.010	0.050***	0.012	0.082***	0.009	0.026***	0.011	0.021***
Final test participation	0.954	0.949**	0.953	0.887***	0.952	0.935***	0.950	0.906***
Standardized test score	0.017	-0.357***	-0.001	-0.165***	0.044	-0.343***	0.003	-0.546***
N students	353918	18128	371435	611	408035	55207	459231	4011
Schools	5195		5196		5240		5240	
Mean number of non-SEN per cohort	39.664		39.672		39.934		39.937	
SD number of non-SEN per cohort	21.592		21.619		21.823		21.837	
Mean number of SEN per cohort	1.868		0.061		4.555		0.296	
SD number of SEN per cohort	1.992		0.255		3.524		0.581	

Note: This table reports descriptive statistics for the primary school samples. The columns represent the differences between SN and non-SN students per SN type. The stars refer to the *p*-values of two-group mean comparison *t*-tests. ***p<0.001 **p<0.01 *p<0.05.

Table 2 *Descriptives for students in secondary education*

	Disruptive		Emotional Problems		Youth Care		Supervision Order	
	non-SN	SN	non-SN	SN	non-SN	SN	non-SN	SN
Boy	0.490	0.666***	0.500	0.265***	0.505	0.414***	0.497	0.414***
Parental education								
Unknown	0.259	0.244***	0.259	0.23***	0.256	0.204***	0.252	0.15***
University	0.128	0.154***	0.129	0.149***	0.135	0.129***	0.135	0.032***
University of applied sciences	0.184	0.206***	0.185	0.197*	0.19	0.201***	0.191	0.081***
Vocational education level 3 -4	0.251	0.261***	0.251	0.27**	0.248	0.28***	0.25	0.265*
Vocational education level 2	0.177	0.135***	0.176	0.154***	0.172	0.187***	0.171	0.472***
No migrant background	0.782	0.887***	0.786	0.84***	0.784	0.802***	0.786	0.626***
Non-western migrant background	0.156	0.063***	0.153	0.1***	0.156	0.13***	0.153	0.281***
Western migrant background	0.061	0.05***	0.061	0.059	0.061	0.068***	0.061	0.093***
Age at October 1st	16.036	15.963***	16.029	16.402***	16.044	15.872***	16.031	15.686***
Backpack funding	0.013	0.076***	0.016	0.047***	0.011	0.034***	0.013	0.029***
Extra educational care	0,123	0,175***	0,125	0,102***	0,120	0,165***	0,122	0,271***
Weighted student funding	0.124	0.079***	0.123	0.104***	0.122	0.13***	0.123	0.191***
Practical education	0.114	0.161***	0.116	0.092***	0.108	0.147***	0.111	0.232***
Standardized test score	0.033	0.145***	0.038	0.080***	0.033	0.077***	0.037	-0.017***
N (students)	568138	26943	589861	5220	723932	66505	785638	4799
Schools	1212		1212		1232		1232	
Mean number of non-SEN per cohort	99.741		99.532		99.832		99.386	
SD number of non-SEN per cohort	50.558		50.539		50.355		50.361	
Mean number of SEN per cohort	4.221		0.859		7.683		0.472	
SD number of SEN per cohort	3.44		1.086		5.124		0.803	

Note: This table reports descriptive statistics for the secondary school samples. The columns represent per SN type the differences between SN and non-SN students. The descriptives compare background characteristics between SN and non-SN students in the population. The stars refer to the *p*-values of two-group mean comparison *t*-tests. ***p<0.001 **p<0.01 *p<0.05.

Table 3. *Main results on the impact of special needs students on the academic achievement of their peers*

Disruptive	Emotional problems	Youth care	Supervision order
Secondary education - student fixed effects - standardized exam grades			
0.000	-0.004	0.002*	0.001
(0.001)	(0.002)	(0.001)	(0.003)
3644914	3776299	4658439	5039979
Secondary education - school fixed effects - standardized exam grades			
-0.000	-0.003	-0.001	0.000
(0.001)	(0.002)	(0.000)	(0.002)
560676	582088	716923	777949
Secondary education - school fixed effects including time trend - standardized central grades			
-0.000	-0.002	-0.000	0.002
(0.001)	(0.002)	(0.001)	(0.003)
560676	582088	716923	777949
Primary education - school fixed effects - final test participation			
0.001	-0.007	-0.000	-0.003
(0.001)	(0.007)	(0.001)	(0.003)
352941	370408	407208	458285
Primary education - school fixed effects including time trend - final test participation			
0.001	0.013	-0.001	-0.003
(0.002)	(0.008)	(0.001)	(0.003)
352941	370408	407208	458285
Primary education - school fixed effects - standardized final test score			
0.002	-0.018	-0.001	0.001
(0.002)	(0.011)	(0.001)	(0.005)
336142	352750	387308	434958
Primary education - school fixed effects including time trend - standardized final test score			
0.000	-0.034*	-0.001	0.003
(0.003)	(0.015)	(0.001)	(0.005)
336142	352750	387308	434958

Note: Each column reports results on a different group of SN students. Each coefficient represents a separate regression with the number of SN students as independent variable. Standard errors are clustered at the school level and reported in parentheses. The number of students without SN are reported below the standard errors. ***p<0.001 **p<0.01 *p<0.05.

Table 4. *Overview of type of SEN and effect in similar studies*

Paper	Type of SEN	Level	Outcome	Effect	Range c.i.
Ahn and Trogdon (2016)	Delinquent peers: offenses recorded	Grade- and school level	Reading or mathematics test score	-0.037 SD	-0.01 -0.07
Balestra et al. (2020)	Diagnosed special needs: learning problems or emotional-behaviour problems	Classroom	Achievement, subsequent educational choices and labour market outcomes	-0.025 SD	-0.02 -0.03
Carrell et al. (2018)	Disruptive peers: children whose families are characterized by domestic violence	School-cohort	High school test scores (math and reading), college attendance and degree attainment and earnings at ages 24 - 28	-0.024 SD	0.00 -0.04
Contreras et al. (2020)	Diagnosed educational needs: hearing, visual, intellectual and multiple disabilities, autism, dysphasia and deafness-blindness, deficit hyperactivity disorder, specific language disorders, borderline performance	Classroom	Mathematics test	-0.038 SD	-0.03 -0.04
Rangvid (2016)	Returning SN students (from segregated education to a regular classroom)	Grade-cohort	Reading test score gains from National test	-0.036 SD	0.00 -0.07

Note: The interventions differ from adding one SN students to a class of 20 (Balestra et al. 2020) or 25 (Carrell et al. 2018) to an increase of one SD in exposure to peer misbehaviour (And and Trogdon 2016). The last column represent the range of the confidence intervals (at 95% level) found in the studies. For some studies, the point estimate and confidence intervals were recalculcated to enable a comparison of effect at 95% confidence interval.

Table 5. *Effect of the number of SN students in the zip code area for primary education*

Disruptive	Emotional problems	Youth care	Supervision order
Neighbourhood fixed effects - final test participation			
0.000	0.003	0.000	-0.004**
(0.001)	(0.004)	(0.001)	(0.001)
481202	512039	575237	668451
Neighbourhood fixed effects - standardized final test score			
-0.001	0.005	-0.001	-0.001
(0.001)	(0.005)	(0.001)	(0.002)
258524	270120	298019	333146
Neighbourhood fixed effects including time trend- final test participation			
-0.000	0.002	-0.000	-0.004**
(0.001)	(0.004)	(0.000)	(0.001)
481023	511852	575103	668296
Neighbourhood fixed effects including time trend - standardized final test score			
0.002	0.001	-0.001	-0.002
(0.002)	(0.007)	(0.001)	(0.003)
258438	270029	297973	333096

Note: Each column represents a type of SEN. Each coefficient represents a separate regression with the number of SN students as independent variable. Standard errors are reported in parentheses, robust and clustered at the level of the post code area. ***p<0.001 **p<0.01 *p<0.05.

Table 6. *Student- and school fixed effects primary and secondary education<30 students*

Disruptive	Emotional problems	Youth care	Supervision order
Secondary education - student fixed effects standardized exam grades <30			
0.004	-0.006	-0.006**	-0.011
(0.004)	(0.007)	(0.002)	(0.008)
814886	846566	1030292	1125626
Secondary education - school fixed effects - standardized exam grades <30			
-0.001	0.012	0.000	-0.003
(0.007)	(0.015)	(0.004)	(0.010)
29593	31123	37099	41451
Secondary education - school fixed effects including time trend - standardized central grades <30			
0.009	0.008	0.002	0.001
(0.010)	(0.022)	(0.004)	(0.012)
29593	31123	37099	41451
Primary education - school fixed effects - final test participation <30			
0.003*	-0.010	-0.001	-0.008**
(0.001)	(0.009)	(0.001)	(0.003)
144959	152119	166686	187602
Primary education - school fixed effects including time trend - final test participation <30			
0.003	-0.004	-0.002**	-0.007
(0.002)	(0.009)	(0.001)	(0.004)
144959	152119	166686	187602
Primary education - school fixed effects - standardized final test score <30			
-0.001	0.009	-0.000	0.004
(0.005)	(0.022)	(0.003)	(0.009)
138738	145603	159555	179191
Primary education - school fixed effects including time trend - standardized final test score <30			
-0.005	-0.013	-0.003	-0.000
(0.006)	(0.031)	(0.003)	(0.010)
138738	145603	159555	179191

Note: Each column reports results on a different group of SN students. Each coefficient represents a separate regression with the number of SN students as independent variable. Standard errors are clustered at the school level and reported in parentheses. The number of students without SEN are reported below the standard errors. ***p<0.001 **p<0.01 *p<0.05.

Table 7. *Differential effects on students with different abilities*

	Disruptive	Emotional problems	Youth care	Supervision order
Primary education - school fixed effects - standardized final test score				
mean	0.002	-0.039*	0.001	0.007
	(0.003)	(0.016)	(0.001)	(0.006)
sd	-0.001	0.011	0.001	-0.000
	(0.002)	(0.008)	(0.001)	(0.003)
p10	0.002	-0.055	0.003	0.002
	(0.006)	(0.030)	(0.003)	(0.010)
p25	0.001	-0.079**	0.001	0.007
	(0.005)	(0.024)	(0.002)	(0.008)
p50	0.001	-0.035	0.001	0.011
	(0.004)	(0.019)	(0.002)	(0.007)
p90	0.003	-0.004	-0.000	-0.001
	(0.003)	(0.015)	(0.001)	(0.006)
students	346988	347360	437096	438457
schools	4252	4258	4264	4284
Secondary education - school fixed effects - standardized exam grades				
mean	-0.000	-0.001	-0.000	0.003
	(0.001)	(0.002)	(0.001)	(0.003)
sd	0.000	0.001	0.000	-0.000
	(0.001)	(0.001)	(0.000)	(0.001)
p10	-0.002	-0.002	-0.001	0.003
	(0.001)	(0.003)	(0.001)	(0.003)
p25	-0.000	-0.002	0.000	0.005*
	(0.001)	(0.002)	(0.001)	(0.003)
p50	-0.000	-0.000	0.000	0.003
	(0.001)	(0.003)	(0.001)	(0.003)
p90	0.000	-0.001	-0.001	0.003
	(0.001)	(0.003)	(0.001)	(0.003)
students	587310	587312	784000	784006
schools	2828	2829	2900	2901

Note: Each column represents a type of SEN. Each coefficient represents a separate regression with the number of SN students as independent variable. All regressions include year fixed effect, school fixed effects, cohort controls and a school specific time trend in the number of SN students. The p stands for the percentile rank. Regressions are weighted by the mean cohort size within schools. Standard errors are clustered at the school level and reported in parentheses. ***p<0.001 **p<0.01 *p<0.05

Table 8. *Estimates of the effect of types of SN students on student achievement (school fixed effect models FE3) secondary education*

	Disruptive	Emotional problems	Youth Care	Supervision order
Secondary education - student fixed effects - standardized exam grades				
Pre vocational track I	-0.009	-0.005	-0.006*	0.005
	(0.005)	(0.013)	(0.003)	(0.009)
	267818	283900	323949	363011
Pre vocational track II	-0.010*	-0.019	-0.004	-0.001
	(0.004)	(0.012)	(0.003)	(0.008)
	428056	448829	536050	591509
Pre vocational track III	-0.001	-0.005	0.002	0.001
	(0.002)	(0.005)	(0.001)	(0.005)
	1042912	1091103	1315282	1448734
Middle track	0.004*	-0.002	0.003**	-0.009
	(0.002)	(0.003)	(0.001)	(0.007)
	1053336	1086461	1361742	1459951
Advanced track	0.001	0.001	0.002	0.023
	(0.003)	(0.005)	(0.002)	(0.016)
	852792	866006	1121416	1176774
Secondary education - school fixed effects including time trend - standardized exam grades				
Pre vocational track I	0.000	0.008	-0.002	0.004
	(0.003)	(0.015)	(0.002)	(0.005)
	53452	56687	64739	72678
Pre vocational track II	-0.003	-0.001	-0.000	0.001
	(0.003)	(0.007)	(0.001)	(0.004)
	82565	86576	103935	114743
Pre vocational track III	0.002	-0.003	-0.000	-0.003
	(0.002)	(0.004)	(0.001)	(0.003)
	160707	168200	203664	224450
Middle track	-0.000	-0.003	-0.001	0.002
	(0.001)	(0.003)	(0.001)	(0.006)
	155409	160355	201470	215931
Advanced track	-0.001	0.002	0.002	0.015
	(0.003)	(0.004)	(0.002)	(0.012)
	108543	110270	143115	150147

Note: Each column reports results on a different group of SN students. Each coefficient represents a separate regression with the number of SN students as independent variable. Standard errors are clustered at the school level and reported in parentheses. The number of students without SEN are reported below the standard errors. ***p<0.001 **p<0.01 *p<0.05.

Table 9. *Main results on the impact of untreated disruptive pears on the academic achievement of their peers (all and <30 students)*

Untreated Disruptive	Untreated disruptive <30
Secondary education - student fixed effects - standardized exam grades	
-0.003	-0.017
(0.004)	(0.012)
2523820	572328
Secondary education - school fixed effects - standardized exam grades	
-0.006	-0.028
(0.003)	(0.026)
386321	21054
Primary education - school fixed effects - final test participation	
-0.003	-0.008
(0.005)	(0.005)
262737	107376
Primary education - school fixed effects - standardized final test score	
-0.012	-0.020
(0.008)	(0.019)
251513	103248

Note: Each coefficient represents a separate regression with the number of SN students as independent variable. Standard errors are clustered at the school level and reported in parentheses. The number of students without SN are reported below the standard errors. ***p<0.001 **p<0.01 *p<0.05.

Table A.1 *Descriptive statistics final test participation in primary education*

	non-SN	
	with final test	without final test
Boy	0.482	0.485
Parental ed. unknown	0.186	0.220***
University	0.191	0.141***
University of applied sciences	0.236	0.179***
Vocational ed. level 3-4	0.231	0.216***
Vocational ed. level 2	0.156	0.244***
Dutch background	0.757	0.569***
Non-western migrant	0.176	0.325***
Western migrant	0.068	0.106***
Weighted student funding	0.093	0.190***
Age at October 1st	11.012	11.104***
N students	304918	14150

Note: This table reports students in school taking the final test for at least one student in the sample. The stars refer to the *p*-values of two-group mean comparison *t*-tests. ***p<0.001 **p<0.01 *p<0.05.

Table A.2 *Balancing test for the number of SN students in primary and secondary education - school fixed effects including time trend*

	All primary education students				Primary education students with final test score			
	Disruptive	Emotional problems	Youth care	Supervision	Disruptive	Emotional problems	Youth care	Supervision
Boy	-0.000	-0.000	-0.013*	-0.001	-0.001	-0.000	-0.014**	-0.002
	(0.003)	(0.001)	(0.005)	(0.001)	(0.003)	(0.001)	(0.005)	(0.001)
Parental ed. Unknown	0.007	0.000	-0.000	0.001	0.008	-0.000	0.001	0.001
	(0.005)	(0.001)	(0.008)	(0.002)	(0.005)	(0.001)	(0.009)	(0.002)
University of applied	-0.007	-0.000	0.001	-0.001	-0.007	-0.000	0.000	-0.001
Sciences	(0.005)	(0.001)	(0.008)	(0.002)	(0.005)	(0.001)	(0.008)	(0.002)
Vocational education	0.002	0.000	0.006	-0.001	-0.000	0.000	0.005	-0.001
level 3 -4	(0.005)	(0.001)	(0.008)	(0.002)	(0.005)	(0.001)	(0.009)	(0.002)
Vocational education	0.001	-0.001	0.010	0.003	0.001	-0.001	0.009	0.003
level 2	(0.006)	(0.001)	(0.009)	(0.002)	(0.006)	(0.001)	(0.010)	(0.002)
No migrant background	0.003	0.000	-0.010	-0.005*	0.004	0.000	-0.008	-0.005*
	(0.005)	(0.001)	(0.008)	(0.002)	(0.005)	(0.001)	(0.008)	(0.002)
Western migrant background	0.005	-0.001	-0.011	0.000	0.007	-0.001	-0.012	-0.001
	(0.006)	(0.001)	(0.010)	(0.002)	(0.006)	(0.001)	(0.010)	(0.002)
Weighted student funding	-0.004	0.001	-0.001	-0.003	-0.007	0.001	-0.009	-0.005
	(0.005)	(0.001)	(0.009)	(0.002)	(0.005)	(0.001)	(0.009)	(0.002)
Age at October 1st	0.001	0.001	-0.010	-0.002*	0.001	0.001	-0.009	-0.002
	(0.003)	(0.000)	(0.005)	(0.001)	(0.003)	(0.000)	(0.005)	(0.001)
n (students)	352941	370408	407208	458285	336540	353150	387649	435509
n(groups)	5117	5118	5168	5168	5109	5116	5108	5144
df	9	9	9	9	9	9	9	9
dfr	5116	5117	5167	5167	5108	5115	5107	5143
F statistic	1.215	0.845	1.593	2.075	1.407	0.765	1.658	2.244
p	0.281	0.574	0.111	0.028	0.179	0.649	0.093	0.017

	Alll secondary education			
	Disruptive	Emotional problems	Youth care	Supervision
Boy	0.004	0.001	-0.005	-0.001
	(0.004)	(0.002)	(0.006)	(0.001)
Parental ed. Unknown	0.009	-0.001	0.011	-0.002
	(0.006)	(0.003)	(0.009)	(0.002)
University of applied	0.001	0.000	-0.003	-0.001
Sciences	(0.006)	(0.003)	(0.009)	(0.002)
Vocational education	0.005	0.001	0.009	-0.000
level 3 -4	(0.006)	(0.003)	(0.009)	(0.002)
Vocational education	0.009	0.003	0.003	-0.002
level 2	(0.007)	(0.003)	(0.010)	(0.002)
No migrant	0.002	0.001	-0.011	0.002
background	(0.005)	(0.002)	(0.009)	(0.002)
Western migrant	0.006	0.003	0.021*	0.000
Background	(0.006)	(0.003)	(0.010)	(0.002)
Weighted student	0.004	-0.001	0.010	-0.001
Funding	(0.006)	(0.003)	(0.010)	(0.003)
Extra educational	-0.008	-0.001	-0.011	-0.006
Care	(0.010)	(0.003)	(0.014)	(0.005)
Age at October 1st	-0.004	0.000	-0.005	0.001
	(0.003)	(0.001)	(0.004)	(0.001)
n (students)	560676	582088	716923	777949
n(groups)	2829	2829	2901	2901
df	11	11	11	11
dfr	1159	1159	1184	1184
F statistic	1.043	0.607	1.494	0.508
p	0.406	0.824	0.127	0.899

Note: Each column represents a regression of the number of SN students in the cohort on student background characteristics. *F*-statistics, *p*-values and degrees of freedom at the bottom of the table refer to *F*-test on the joint significance of gender, age, migrant background and disadvantaged neighbourhood/weighted student funding. Standard errors are robust and clustered at the school level. ***p<0.001 **p<0.01 *p<0.05

Table A.3 *Results of the student fixed (secondary education) and student- and school fixed models (primary and secondary education) with SN in percentages*

Disruptive with	Emotional problems	Youth care	Supervision order
Secondary education - student fixed effects standardized exam grades			
-0.000	-0.001*	-0.001**	-0.001
(0.000)	(0.001)	(0.000)	(0.001)
568132	589856	723930	785620
Secondary education - school fixed effects - standardized exam grades			
-0.000	-0.001	-0.001*	-0.000
(0.001)	(0.001)	(0.000)	(0.001)
560676	582088	716923	777949
Secondary education - school fixed effects including time trend - standardized central grades			
-0.000	-0.001	-0.000	-0.000
(0.001)	(0.002)	(0.000)	(0.001)
560676	582088	716923	777949
Primary education - school fixed effects - final test participation			
0.000	-0.001	-0.000	-0.001*
(0.000)	(0.001)	(0.000)	(0.001)
352941	370408	407208	458285
Primary education - school fixed effects including time trend - final test participation			
0.000	0.001	-0.000*	-0.001
(0.000)	(0.001)	(0.000)	(0.001)
352941	370408	407208	458285
Primary education - school fixed effects - final test score			
0.000	-0.006	-0.000	0.000
(0.001)	(0.003)	(0.000)	(0.001)
336142	352750	387308	434958
Primary education - school fixed effects including time trend - final test score			
-0.001	-0.005	-0.000	0.001
(0.001)	(0.005)	(0.000)	(0.002)
336142	352750	387308	434958

Note: Each column reports results on a different group of SN students. Each coefficient represents a separate regression with the percentages of SN students as independent variable. Standard errors are clustered at the school level and reported in parentheses. The number of students without SN are reported below the standard errors. ***p<0.001 **p<0.01 *p<0.05.

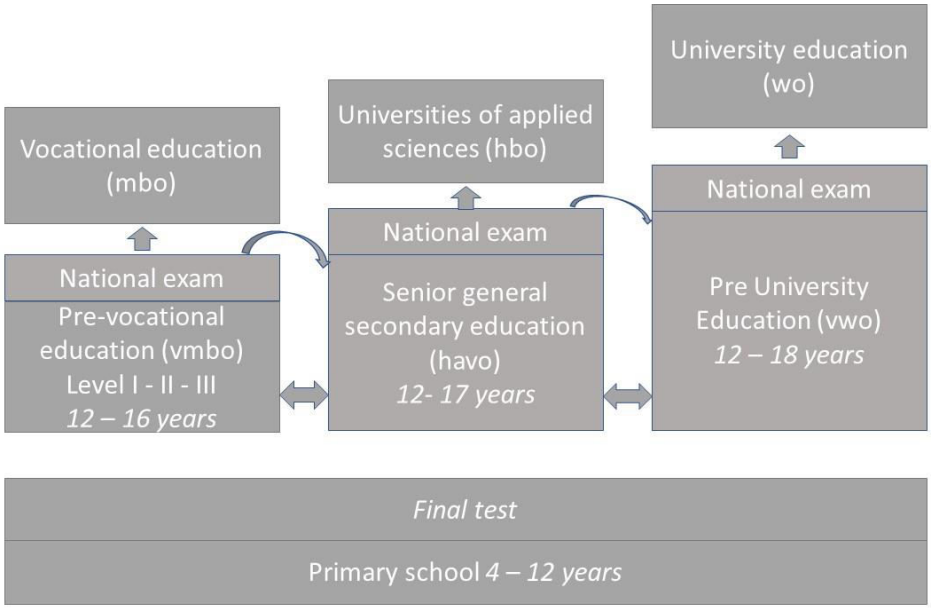

Figure 1. The Dutch school system

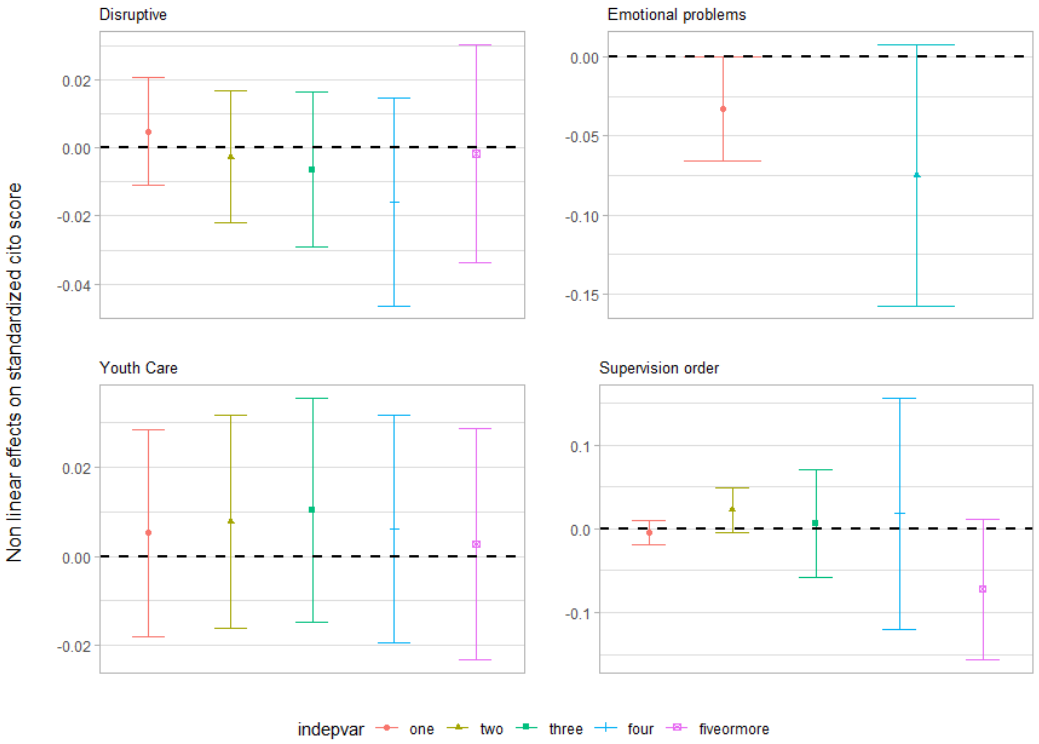

Figure 2. Non-linear effects on standardized final test score (primary education)

Note: Each panel reports the results of one regression for a specific group of SN students. The coefficients represents the point estimate and 95% confidence interval with the presence of the indicated number of SN students as independent variable. All models include school fixed effects, individual controls, cohort controls and school specific time trend in the number of SN students. Standard errors are clustered at the school level.

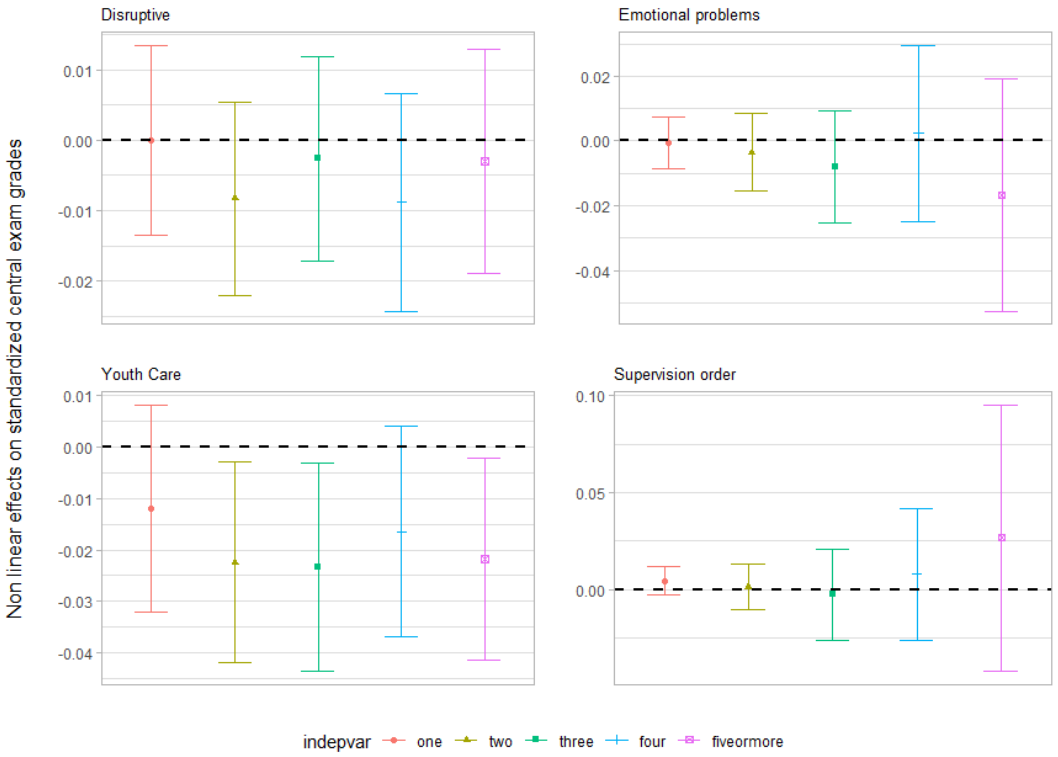

Figure 3. Non-linear effects on standardized exam grades (secondary education)

Note: Each panel reports the results of one regression for a specific group of SN students. The coefficients represents the point estimate and 95% confidence interval with the presence of the indicated number of SN students as independent variable. All models include school fixed effects, individual controls, cohort controls and school specific time trend in the number of SN students.. Standard errors are clustered at the school level.

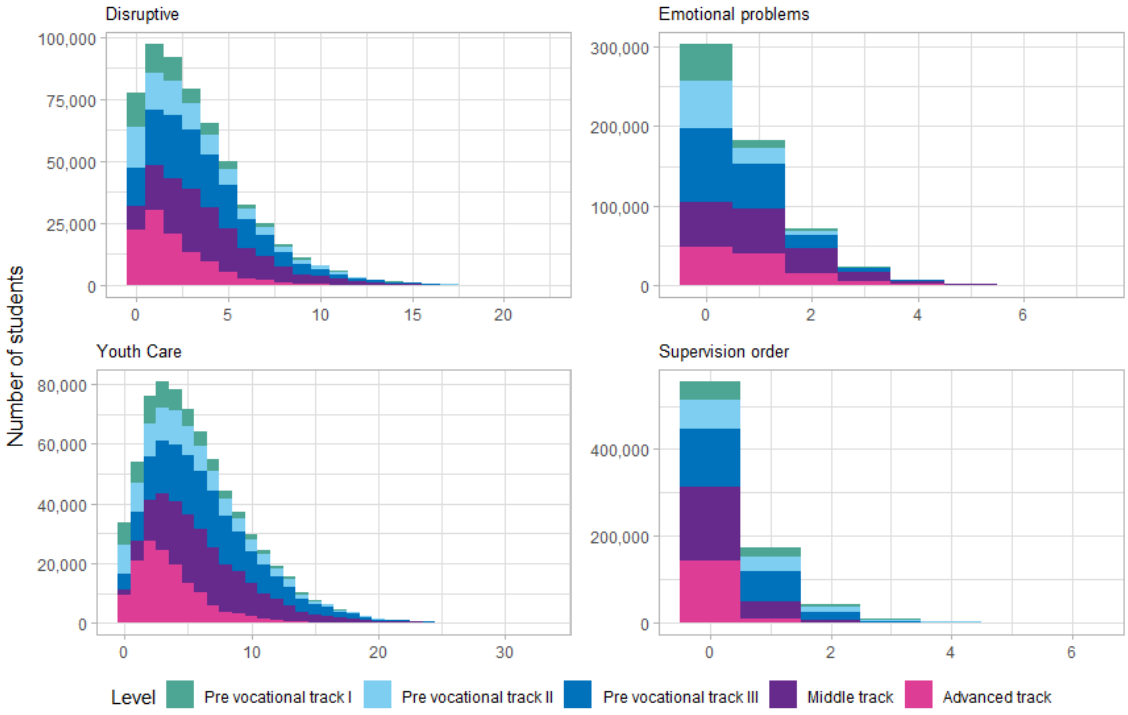

Figure A.1 Within non-SN student variation in the number of SN students between courses per tracks

Note: On the x-axis the maximum variation in the number of SN students between courses within non-SN students is displayed. On the y-axis the amount of students with this maximum variation.

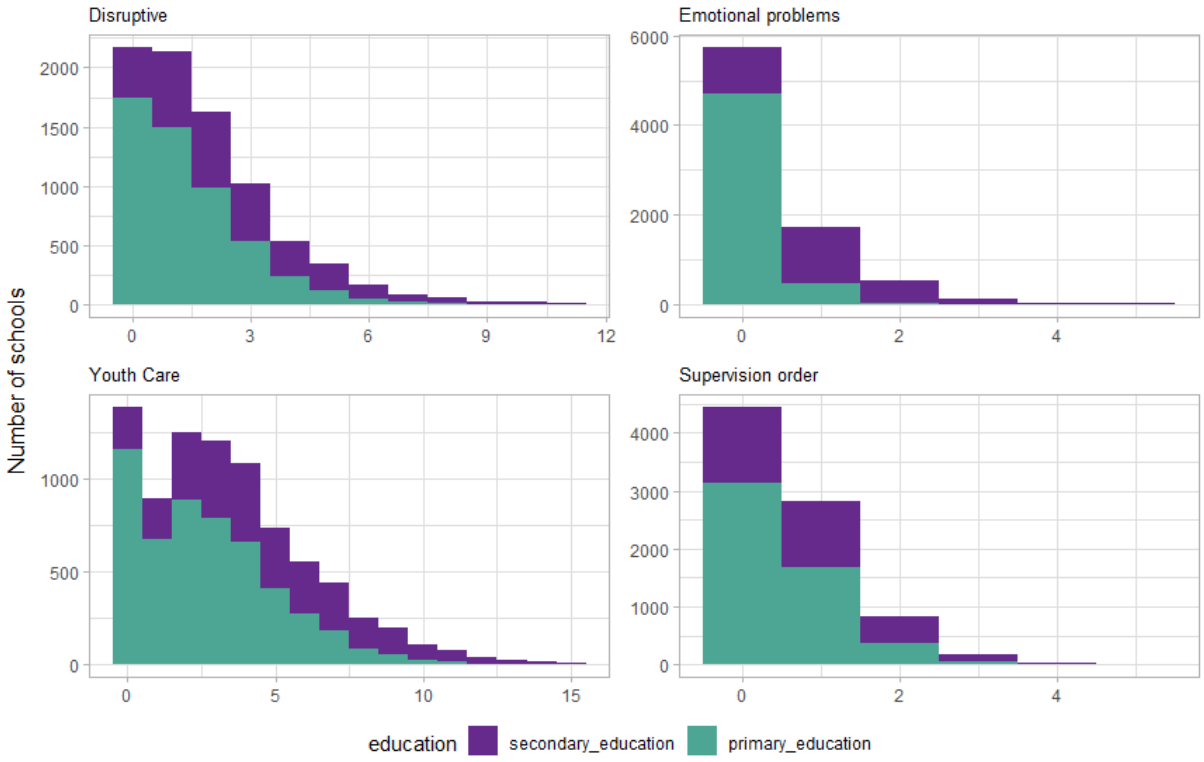

Figure A.2 Within school variation in the number of SN students between schoolyear per type of education.

Note: On the x-axis the maximum variation in the number of SN students between cohorts within schools is displayed. On the y-axis the amount of schools with this maximum variation

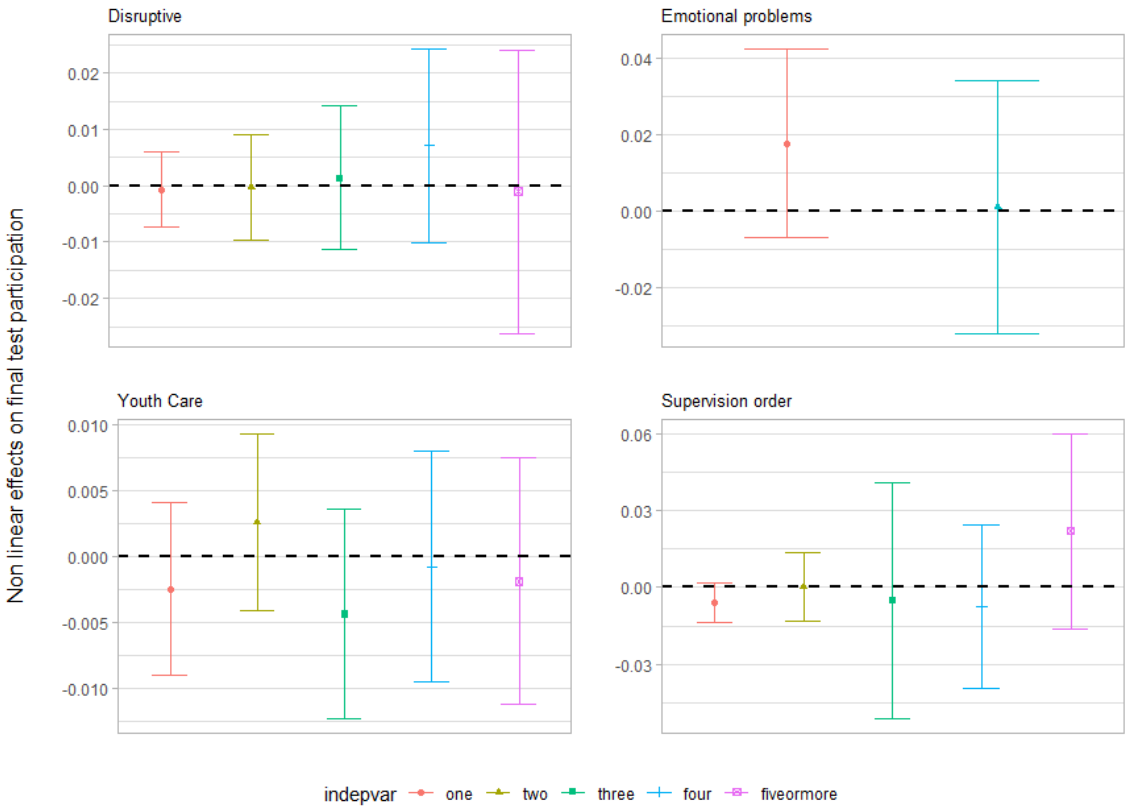

Figure A.3 Non-linear effects on final test participation (primary education)

Note: Each panel reports the results of one regression for a specific group of SN students. The coefficients represents the point estimate and 95% confidence interval with the presence of the indicated number of SN students as independent variable. All models include school fixed effects, individual controls, cohort controls and school specific time trend in the number of SN students. Standard errors are clustered at the school level.

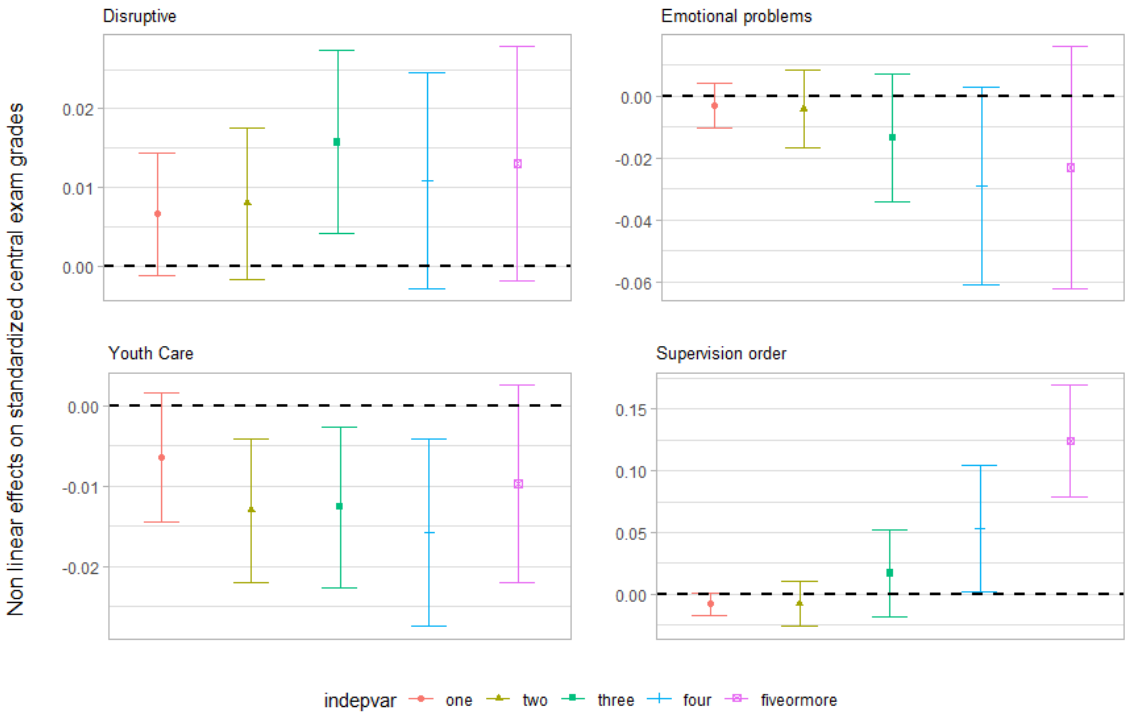

Figure A.4 Non-linear effects on standardized exam grades (student fixed effects secondary education)

Note: Each panel reports the results of one regression for a specific group of SN students. The coefficients represents the point estimate and 95% confidence interval with the presence of the indicated number of SN students as independent variable. All models include student fixed effects, individual controls and cohort controls. Standard errors are clustered at the school level.