PEERS WITH SPECIAL NEEDS: EFFECTS AND POLICIES*

Simone Balestra[1], Beatrix Eugster[1], and Helge Liebert[2]

[1] *University of St. Gallen, Switzerland*

[2] *Harvard University, USA*

Abstract

In light of the debate over inclusive education, this paper evaluates the impact of exposure to special needs (SN) peers. More classroom peers with SN lower performance, the probability of entering post-compulsory education, and earnings at ages 17-25. SN students and students at the lower end of the achievement distribution suffer most from higher inclusion. We analyze reallocation policies and government interventions to alleviate negative externalities. We demonstrate that inclusion is preferable to segregation in terms of maximizing average test scores and that teacher quality is key to alleviating negative classroom externalities, while financial resources are not.

Keywords: class composition, special needs, peer effects, mainstreaming.

JEL Classification: I21, I28, J14.

* We are grateful to Massimo Anelli, Joshua Angrist, Eric Bettinger, Anne Ardila Brenøe, Christina Felfe, Joshua Goodman, Bryan Graham, Edward Lazear, Rafael Lalive, Michael Lechner, Geoff Lindsay, Nicole Maestas, Claudia Persico, Christopher Rauh, Carmit Segal, Ludger Woessmann, and Ulf Zölitz for their constructive comments. For the data, we thank the School Psychological Service, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office. The paper benefited from valuable feedback at the following meetings: the CESifo Conference on Employment and Social Protection, Education Workshop in Catanzaro, Education Workshop in Leuven, European Association of Labor Economics, European Economic Association, German Economic Association, Society of Labor Economists, and Swiss Society of Economics and Statistics, and departmental seminars at the universities of Göttingen, Groningen, Lucerne, St. Gallen, and Zürich. We acknowledge financing from the Swiss National Science Foundation (grant no. 100018_176381). Corresponding author: Beatrix Eugster, University of St. Gallen, Rosenbergstrasse 51, CH-9000 St. Gallen, beatrix.eugster@unisg.ch, +4171 224 31 79.

1 Introduction

Following the UN Convention on the Rights of the Child in 1989 and the UNESCO Salamanca Statement in 1994, countries worldwide have adopted more inclusive education practices for students with special needs. Special needs (SN) refer to the requirement of assistance for medical, mental, or psychological disabilities. In the United States, the Individuals with Disabilities Education Act (IDEA) mandates that students with special educational needs have the legal right to be educated in the least restrictive environment. The practice of *mainstreaming*, i.e., schooling SN students in regular classrooms among students without SN, is also pervasive in Europe, where the enrollment rate in mainstream education is approximately 97% (European Agency for Special Needs and Inclusive Education, 2018; European Commission, 2017).[1]

Despite the policy trend towards inclusive education, little is known about how inclusion affects the achievement of students and their long-run outcomes (Lindsay, 2007). While advocates of inclusion claim that all students have the right to be educated in regular schools, opponents argue that inclusive education may adversely affect students without special needs. These diverging perspectives have made inclusion one of the major topics in education policy for the last 20 years, raising debates on whether SN students should be mainstreamed and, if so, how classroom composition should be performed (Ainscow and César, 2006). Given the public good nature of classroom education production, understanding whether differences in classroom composition generate negative externalities becomes crucial. If negative externalities exist it becomes a key policy question to know whether teacher training, per pupil spending or different ability grouping is able to compensate for these externalities.

In this paper we first investigate how exposure to special needs students affects achievement, subsequent educational choices, and labor market outcomes. To answer this research question we compile a novel data set containing administrative health records, student test scores, and educational transitions as well as labor market outcomes from social security data in Switzerland. These data allow us document the direct,

[1] Throughout the paper, we use the terms *mainstreaming*, *inclusive education*, and *inclusion* interchangeably.

delayed, and long run externalities generated by the inclusion of special needs students. For identification, we exploit variation in classroom composition arising from within-school quasi-random assignment to classes when students transition from primary school to secondary school. Secondary schools are organized separately from primary schools and classes remain unchanged within each level but are reshuffled when transitioning between levels. For equity reasons and to avoid stigma, special needs information is not shared between primary and secondary schools. This implies that SN students can be neither identified nor assigned to specific classrooms or teachers in secondary school and that class composition is quasi-random in this respect. We test this assumption using balancing tests and simulations, and the results indicate that assignment of SN students to classes is random. The Swiss setting provides us with a unique opportunity to study the impact of special needs peers because of randomness in class assignment, but actually special education classification rates and programs are very similar to those in other European countries and the U.S.

The results show that having higher proportions of SN students in a class creates negative achievement externalities at the end of compulsory schooling. One additional SN student in a class of 20 reduces test scores by 2.5% of a standard deviation. There is substantial effect heterogeneity. The negative externalities are driven primarily by students with severe special needs, and the largest negative externalities are found among students at the lower end of the achievement distribution and special needs students themselves. Threshold effects are also relevant, because we find that negative spillovers occur only after SN students exceed 15-20% of students in a classroom.

Exposure to SN peers has lasting effects. Lower educational achievement perpetuates as a reduced propensity to pursue post-compulsory education and, ultimately, lower income upon labor market entry. Having higher proportions of SN peers in secondary school decreases the probability of enrolling in one of the two post-compulsory education tracks available in Switzerland (the vocational track and the academic track). While we observe no peer effect on the choice between vocational or academic, we find that within the vocational track, students with comparatively more SN peers are less likely to enter the high-quality vocational program.[2] The peer effect is still detectable in the labor market, two to eight years after exposure to classmates with special needs. A higher share of SN peers causes a reduction in earnings at ages 17-25, but this effect is relatively small in size. For example, one additional special needs student in a class of 20 during secondary school causes a 0.6% decrease in earnings (180 USD per year).

Second, given these empirical results, we analyze the implications of reallocation policies and consider the scope for government intervention to potentially reduce negative spillovers. Using Graham's (2011) technique to analyze optimal assignment, we find that despite its negative spillovers, mainstreaming clearly dominates an alternative education policy of segregation. This result is driven by the fact that under segregation, negative externalities for special needs students are not compensated by the gains from reallocation for students without special needs. The optimal assignment rule implies that special needs students should be allocated evenly across classes. Having found that inclusive education is optimal for maximizing average achievement, we further examine the role of teachers and education spending as moderating factors. We find that good teachers – in a value-added sense – are able to mitigate negative spillovers, implying that reducing classroom spillovers is possible and that teachers might play an important role in such reduction. In contrast, higher per student spending does not have a substantial influence on the size of the peer effect parameter.

The present study is the first comprehensive evaluation of the effect of mainstreaming in education over the life of young adults. We evaluate the impact of inclusive education on educational achievement, post-compulsory education and career choices, labor market participation and wages. We thereby contribute to two strands of literature, namely the empirical literature on peer effects in classroom education production and the literature on inclusive education. Several studies evaluate the immediate impact of potentially disruptive students on their peer group's educational outcomes (e.g., Carrell and Hoekstra, 2010;

[2] Switzerland has two quality types of vocational education, characterized by their different duration, curricula contents, and expected pay.

Carrell, Hoekstra, and Kuka, 2018; Figlio, 2007; Kristoffersen et al., 2015; Neidell and Waldfogel, 2010). These studies typically proxy for disruptiveness by considering parents' income and criminal background, students' gender, or similar approximate measures. They find detrimental effects of potentially disruptive students on peers' achievement. More generally, other papers on peer effects in education have established a negative relationship between peer quality and educational achievement (e.g., Bietenbeck, forthcoming; Burke and Sass, 2013; Feld and Zölitz, 2017; Sacerdote, 2001).

In contrast, the few studies focusing on students with special needs present a much less clear picture. While Fletcher (2010) and Kristoffersen et al. (2015) find that students with severe emotional problems negatively affect their peers' academic achievement, other papers in economics and education report no significant externalities from including students with disabilities (Friesen, Hickey, and Krauth, 2010; Kalambouka et al., 2007; Ruijs, 2017). Some studies suggest that disabled students may even improve the academic achievement of their peers (Ruijs and Peetsma, 2009). For example, Hanushek, Kain, and Rivkin (2002) present findings that students with a larger share of grade mates who receive special education have higher achievement gains than other individuals.

We extend previous findings in several dimensions. First, by using special needs examination records, we focus on a student group of specific policy relevance that is easily identified by school administrators, relying on neither imprecise proxies of peer quality or disruptiveness nor retrospective surveys. Second, the variation in classroom composition in combination with health examination records allows us to examine effect heterogeneity in unprecedented detail. Diagnoses and counseling are performed by psychologists, which increases both the reliability and depth of information. For each student with special needs, we know the type of SN and can also infer the severity of SN. The severity measure is constructed based on the number of contacts between the student and the SPS.

Third, due to data limitations, much of the previous literature defines the grade-cohort in a school as the relevant reference group and is agnostic with respect to the shared class structure. We focus on peer externalities in classroom education production and explicitly compare results for different definitions of peer environments. Theory posits that the classroom is the relevant peer environment for education production (Lazear, 2001), and we show that approximating classroom externalities at the grade-cohort level captures smaller – yet significant – peer effects. Fourth, observing the composition of both classes and school-grade cohorts allows us to employ different identification strategies. While our identification relies on within-school random assignment to classes when students transition from primary school to secondary school, we can alternatively exploit idiosyncratic variation in cohort composition, as is done in most previous studies (e.g., Carrell and Hoekstra, 2010; Carrell, Hoekstra, and Kuka, 2018; Hoxby, 2002). We obtain similar peer effect parameters under the two identification strategies.

Fifth, we study the persistent effects of SN classroom peers by examining long-run outcomes such as post-compulsory education choices and labor market outcomes. Our analyses reveal that negative externalities from SN peers are still statistically significant in the long run but small in economic magnitude. Sixth, we discuss different educational policies in light of their ability to alleviate the burden of negative spillovers. The small size and substantial heterogeneity in the peer effect give rise to the result that inclusion is preferable to segregation in terms of average test scores. We also find that good teachers are able to ensure a learning environment in which spillovers are reduced, whereas per student spending has no significant impact on the size of the spillovers.

The paper proceeds as follows: Section 2 introduces the institutional setting, section 3 presents the data, section 4 explains the empirical strategy, and section 5 reports the results. Section 6 discusses the implementable educational policies related to SN students, and Section 7 concludes the paper.

2 Institutional Background

2.1 Education System

In Switzerland, the primary responsibility for education lies with the 26 cantons (states). Schools are administered by municipalities, and the majority of students in Switzerland complete compulsory education at a public school in their municipality of residence. Approximately 5% of children in a cohort attend a private school. Inclusion is an important policy objective of public schools. Children of different gender, ethnic, social and economic backgrounds as well as children with disabilities are educated in an inclusive setting whenever possible. Only 2% of all children in Switzerland are educated in segregated special education schools for children with severe disabilities (e.g., Down syndrome, autism, blindness).

Our analysis considers the universe of secondary schools (both public and private) from the canton of St. Gallen, the fifth largest state in Switzerland, with a population of approximately 500,000. In St. Gallen, compulsory education begins for children at age four when they enter kindergarten. The total compulsory school period amounts to eleven years, divided into the primary level (kindergarten and primary school, two and six years, respectively) and secondary level (secondary school, three years). Secondary schools are organized separately from primary schools and have a larger catchment area. Assignment to schools is based on district of residence. This assignment is strict, and parents have no influence on the decision other than moving permanently to another school district. Appeals challenging the assignment decision are usually dismissed, and rulings regarding exceptions from residence-based assignments are rarely granted (Weder, 2016). Important for our identification strategy, classes remain unchanged within each stage but are reshuffled when transitioning between stages. As opposed to typical settings in the US with varying class composition for different subjects, class composition in St. Gallen is not only stable during the entire secondary school but also across subjects.

Tracking occurs upon completing primary school. There are two main tracks, secondary schools belong to either an upper tier *(Sekundarschule)* or a lower tier *(Realschule)*. Assignment to either track is based on student performance and the primary school teacher's recommendation. While the assignment to secondary school tracks is based on performance, there is no systematic assignment of students to classes within each school-track. When classes are composed at the beginning of secondary school, the administrative staff has no prior knowledge about the students other than administrative data on gender, place of residence, primary school attended, and nationality. Anecdotal evidence suggests that schools deliberately attempt to mix their intake cohorts based on these observable characteristics. For equity reasons, special needs information is not shared between schools when transitioning to secondary school. This implies that SN students can be neither identified nor assigned to specific classrooms or teachers in secondary school and that class composition is quasi-random in this respect. We test this implicit randomization assumption in section 4.

Towards the end of the second year of secondary school, all children take a mandatory standardized test in core subjects. The test, named "Stellwerk 8," is computer-based and administered by the Ministry of Education of the canton of St. Gallen. Stellwerk 8 is a norm-referenced, self-scoring, adaptive exam similar in spirit to the Graduate Record Examination (GRE). Students do not necessarily face the same set of questions. The software chooses questions such that they correspond to an ability measure computed based on the answers to all previous questions. The final test score thus reflects the difficulty level of the questions that the student was able to answer correctly. Upon completion of the test, students and their teachers receive feedback on the achieved score. Although the test was developed as an instrument for standardized performance feedback, the test results are important for students. After the test, students receive a certificate with their Stellwerk 8 results. This certificate is usually provided to potential employers when students apply for an apprenticeship position.

After completing compulsory schooling, approximately two-thirds of students enter vocational education and training (VET) by applying for an apprenticeship position and signing a training contract with a firm. VET combines part-time formal education with training and experience at the workplace.[3] Switzerland has a strong tradition of vocational education, as reflected in the low proportion of students choosing the academic track. In St. Gallen, approximately 15% of students per cohort complete the academic track, allowing them to pursue post-secondary education, compared to a national average of 20%. The residual category consists of individuals who do not pursue post-compulsory education (approximately 10% for our cohorts, Swiss Federal Statistical Office, 2017) or postpone their secondary education.

2.2 The School Psychological Service

The SPS is a centralized service provider for all schools in the canton of St. Gallen. It is divided into separate administrative units for the main city St. Gallen and the rest of the canton, which is served by seven regional offices. The SPS provides diagnosis and counseling for children, parents, and teachers for school-related problems. This includes diverse issues such as general counseling, diagnosis of learning disabilities and correspondence with physicians, monitoring developmental deficiencies, conflict mediation, and developing schooling strategies for children with special needs. Similar services exist in other European countries, such as Germany or the UK, and in US states, albeit often organized at the school or district level. In Switzerland, SPS services are organized at the state level.

The standard procedure is that the teacher registers a child for counseling with the SPS. In a few cases, requests for counseling are filed directly by the child's parents. Teachers have to provide one or several reasons for registering the child with the SPS. Typically, these reasons are learning problems or emotional-behavioral problems, sometimes both. A counseling unit usually begins with an assessment session that may include several diagnostic tests, followed by discussions with the relevant stakeholders, and concludes with a recommendation for further action, e.g., delaying school entry, additional tutoring, or therapy. Although

[3] Students who attend a VET program study part time at school for 1 to 1.5 weekdays per week. For the remaining time (3.5 to 4 weekdays), students work as apprentices in host companies with which they have an employment contract for their entire two- to four-year training period. See Oswald and Backes-Gellner (2014) for an overview of the Swiss VET system. For a broader perspective on VET across countries, see Wolter and Ryan (2011).

schools have some freedom to implement the recommendations made by the SPS, they rarely deviate.

SPS interventions generally occur during primary school, when children are still at a formative age. Vulnerable children or those with developmental delays are screened at primary school entry and can be offered the possibility of a transition year between kindergarten and primary school to foster school readiness. In primary school, students with severe special needs might be accompanied by a SN teacher either full time or during specific lessons. Because such measures are costly, they are only implemented upon recommendation of the SPS following an examination of the child. These measures are often complemented by individually assigned therapies, e.g., speech, dyslexia/dyscalculia, or behavioral therapy. Although children may still have SN and school problems beyond primary school, support during secondary school is considerably reduced. Assistance in secondary school consists primarily of additional tutoring outside the classroom, which is provided to less than 1.5% of students.

Because of the diverse issues the SPS treats, nearly one-third of all children are in contact with it at least once. For example, between 2004 and 2015, an average of approximately 3,300 children per year were registered with the SPS. The age at first registration varies but is on average eight years, typically when children are in the third year of primary school. Note that this coincides roughly with the time when children start receiving school grades (second semester of second grade) and thus potential learning disabilities become more apparent.

3 Data

Data for this study stem from four administrative sources. First, we use test score data for the population of students enrolled in eighth grade in the canton of St. Gallen during the years 2008 to 2017. Besides achievement, this data contains information on individual characteristics and a class and teacher identifier. Second, we merge the test score data with the administrative records from the SPS to obtain information on students' special needs. Third, we supplement these data with information on students' career trajectories after compulsory education. Fourth, we link our data set to the records of the social security administration to obtain individual employment and income histories.

The test score records are our primary data source, containing the following information for each eighth grade cohort between 2008 and 2017: student achievement (standardized test score in Math and German), student characteristics (date of birth, gender, and native German speaker), and composition of secondary school classes (school, track, classroom ID, and teacher ID). In contrast to many other studies on peer effects in education (e.g., Angrist and Lang, 2004; Carrell and Hoekstra, 2010; Carrell, Hoekstra, and Kuka, 2018; Hanushek et al., 2003; Kristoffersen et al., 2015; Lavy, Silva, and Weinhardt, 2012), our data allow us to identify classroom composition. Being able to identify the class constitutes an advantage of our data, because classroom peers (compared to grade-cohort peers) are the relevant peer group for the production of educational outcomes (Lazear, 2001).

Data on achievement are based on the Stellwerk 8 test, and our analysis focuses on the composite score in Math and German, which are compulsory subjects in all tracks.[4] Test scores range between 0 and 1,000, and the average composite test score across all schools is 534. For the analysis, we standardize all test scores to ease interpretation and comparability with other studies (i.e., mean zero and standard deviation one).

We impose four sample restrictions on the school data, which combined exclude approximately 8% of the raw data. First, we drop special education institutions, because students attending those schools are not required to take the Stellwerk test, and we thus do not observe complete classes. Second, we exclude segregated classes with only SN students. Third, we drop implausibly small or large classes (i.e., smaller than ten and larger than 31). Fourth, we exclude students with missing data or data entry mistakes such as test scores that exceed the possible score range or implausible age. Supplementary Table A.1 shows the sample construction in detail. Our final sample comprises 49,961 students in 2,723 classes from 155

[4] Using a composite score has the advantage of increasing precision by reducing measurement error in the dependent variable (West and Peterson, 2006). In a robustness check, we estimate the main models separately for Math and German scores and find that the peer coefficients lie within each other's confidence bounds.

school-tracks.

To construct our treatment variables, we merge the school data with administrative records from the SPS. The SPS records contain students' diagnosis and counseling history, recorded by the caseworkers in the SPS database. Students enter the database if they had contact with the SPS before or during primary school. We observe the date of each contact per student, the reason for registration with the SPS, and information on diagnoses, treatment, therapy, and other case-related data through caseworkers' comments. We classify students as special needs students if they were in contact at least twice with the SPS and were not explicitly dismissed without any diagnosis. We use the threshold of two contacts because diagnoses are recorded with error in the data and children with a single SPS contact are likely false positives. Since the decision to refer a child to the SPS is made by teachers or parents, there are cases in which children are sent to the SPS, examined by the SPS professionals, and then receive no special needs diagnosis. We thus use the two-contact rule to filter out these over-referrals. In the robustness checks, we test the sensitivity of the definition of SN status by gradually increasing the threshold number of contacts.

Using administrative data from the SPS has several advantages. First, we observe directly whether a child has special needs, without relying on survey data or proxies for SN status such as gender, ethnicity, or parental background. Second, being an SN student is measured during primary school and is thus independent of class composition in secondary education (within school-tracks). Third, the data permit the measurement of different types and severity of special needs. We observe not only whether a student has, for example, a learning impairment or a behavioral problem, but also the number of contacts between the SPS and each SN student. According to the SPS staff, children who are observed to have had more contacts with the SPS are likely to have more severe special needs than those with fewer contacts. For example, children with acute learning difficulties undertake several tests to assess their cognitive development, resulting in multiple contacts with the SPS. Similarly, children with major behavioral problems have longer histories of contacts with potentially more complex therapy settings. Following this reasoning, we use the number of contacts for each child as a measure of the severity of special needs.

To investigate the persistent effect of SN peers, we collect data on students' career trajectories after compulsory education. First, we link our data to information on all VET contracts signed in the canton of St. Gallen between 2008 and 2016. The Ministry of Vocational Education of St. Gallen collects all the VET contracts that are signed within the canton, along with begin and end dates of the contract, type of VET (i.e., occupation), and quality of VET.[5] Second, for students who enter the academic track, we link our data to the administrative high school records for the years 2008-2016. The high school data are provided by the Ministry of Education of St. Gallen and comprise all high schools within the canton. Note that for both VET and high school data, we lose the 2017 cohort due to children not yet having completed compulsory schooling.

The final piece of information we add to our data set is provided by the Swiss Federal Statistical Office and comes from the Swiss Central Compensation Office (CCO). The CCO is the central body for the first-pillar of social security in Switzerland, which covers old-age and survivors' insurance, disability insurance, and compensation for loss of earnings. To qualify for these benefits, workers pay social security taxes on their earnings, and the claimant's benefits are based on the wage earner's contributions. The CCO thus possesses monthly employment and earnings data for all registered individuals. For individuals in our data who ever worked between 2008 and 2016, we observe how many months per year they worked and their (net) earnings.[6] These data allow us to investigate the labor market effects of SN peers on individuals' early careers.

While we have information on post-compulsory education for 82% of all students, we have labor market outcomes for 53% of the sample. Higher attrition for later outcomes can almost exclusively be attributed to cohorts that have not yet completed compulsory education or entered into the labor market. While we do

[5] Switzerland offers two quality types of apprenticeships, characterized most notably by their different durations (2 vs. 3 or 4 years) and curricula contents. For many occupations, both a high- and a low-quality track are available.

[6] The only sample restriction we impose on the CCO data is to exclude 1% of each tail end of the earnings distribution. This restriction is intended to attenuate the impact of outliers, remove implausible values, and filter out summer jobs.

not have information on post-compulsory education or VET contracts outside of the canton of St. Gallen, information for income and earnings is available independent of the canton of residence but only until 2016 due to a data provision time-lag of the CCO of approximately two years. The last birth cohort we can reliably observe in the income registers is 1998, and the cohorts up to 1998 account for approximately 60% of the original sample. Attrition in the cohorts before 1998 is small (approximately 5%), which might simply be due to individuals who have no registered income (yet).[7]

Table 1 reports the descriptive statistics for our sample of 49,961 individuals, divided by information on special needs (panel A), achievement (panel B), post-compulsory education (panel C), labor market outcomes (panel D), and covariates (panel E). On average, 25.5% of all students and approximately four students per class in eighth grade had been in contact with the SPS during primary school. Supplementary Figure A.1 shows the distribution of class size, the number of SN peers per class, and the proportion of SN peers per class. The average age at first registration with the SPS is 8.35 years, and the leading reasons for a referral to the SPS are learning impairments and behavioral problems. Because it is possible that a child is registered with the SPS for both learning impairments and behavioral problems, we categorize these two SN types as follows: learning impairments without behavioral problems and behavioral problems (potentially with learning impairments). By using these mutually exclusive categories, we find that 18% of all children suffer from a learning impairment, whereas 6% have a behavioral problem. Overall, our figures on the incidence of SN are in line with the national average. Moreover, the baseline rates in our sample are comparable to those of other OECD countries (OECD, 2005). Nevertheless, cross-country comparisons should be made with caution due to institutional differences.

[Table 1 here]

To investigate the relationship between student achievement and the proportion of SN peers, Figure 1

[7] In the analyses on post-compulsory education and labor-market outcomes, we drop the corresponding younger cohorts completely to minimize any issues regarding selective attrition.

plots the average test score against the proportion of special needs peers. Although purely descriptive and contaminated by correlated background characteristics, Figure 1 reveals two important features of the data. First, SN students consistently perform worse than non-SN students on standardized tests (see also Figure A.2 in the supplementary material). Second, we observe that average achievement decreases with the proportion of SN peers. This negative relationship is present for both SN and non-SN students.

[Figure 1 here]

After compulsory education (Table 1, panel C), 67% of the students enter VET and 15% enter the academic preparation track. The residual category comprises students who are retained, postpone their post-compulsory education decision, follow post-compulsory education in a canton other than St. Gallen, or enter the labor market directly. The average worker in our data has a net income of 1,864 Swiss Francs per month (in 2018, 1 Swiss Franc was valued at 1.01 USD) and is employed for approximately ten and a half months per year (Table 1, panel D). The net earnings we have in our data are lower than the Swiss average, because we observe individuals in their early career. The income reported in the present study is earned when the workers are between 17 and 25 years of age. Most of the earnings we have are thus apprenticeship wages, not only because two-thirds of each Swiss cohort follows the vocational track but also because very few university graduates enter the labor market before age 25.[8] By restricting the sample to individuals who have completed a VET degree, we observe an average net income of 3,400 Swiss Francs, which is consistent with the national averages for early career net earnings.

4 Empirical Strategy

The aim of this paper is to evaluate the impact of differential exposure to SN peers on student achievement, educational choices, and labor market outcomes. Although studying peer effects has a long tradition in

[8]Note that some students work part time and are thus registered at the CCO. We cannot distinguish part-time from full-time work, which further compresses average wages in an age group that is likely to contain many part-time workers.

economics, measuring such effects has proven challenging for three main reasons (Manski, 1993). First, individuals in a group tend to behave similarly because they face similar environments. These correlated effects confound peer effect parameters, hindering the researcher from distinguishing true peer effects from common shocks shared by a group. The use of large data sets may resolve this problem by including a series of fixed effects that control for unobserved heterogeneity at multiple levels (Burke and Sass, 2013). Second, it is difficult to separate the effect of the group on the individual from the effect of the individual on the group. This reflection problem exists because individual and peer outcomes are determined simultaneously. For this reason, the present paper focuses on a characteristic (SN status) that is determined before entering secondary school, when achievement is measured.

The third problem is that individuals tend to self-select into peer groups with specific characteristics. If this type of selection occurs, one cannot determine whether a difference in outcome is a causal peer effect or a reason that individuals joined a group. In the context of peer effects in education, this self-selection issue can be resolved through quasi-randomization (Chetty et al., 2011; Duflo, Dupas, and Kremer, 2011; Feld and Zölitz, 2017; Lyle, 2007; Sacerdote, 2001) or by exploiting the natural variation in cohort composition over time (Carrell and Hoekstra, 2010; Carrell, Hoekstra, and Kuka, 2018; Hanushek et al., 2003; Hoxby, 2002). The majority of papers exploiting the natural variation in cohort composition defines peers as the group of grade-level individuals from the same school. However, this definition is usually data driven and not necessarily optimal. A natural approach is to define the peer group at the classroom level, because classroom peers exert a greater influence on individual outcomes (Burke and Sass, 2013; Lazear, 2001).

The data allow us to identify classrooms, and we use within-cohort quasi-randomization to resolve the self-selection issue. In the canton of St. Gallen, the transfer from primary to secondary school is regulated, in the sense that students go to a given school depending on the district they live in. Schools may have different education tracks, and classes are strictly separated between tracks. Within each school-track, classroom formation is quasi-random, as students from different primary school districts are mixed and their SN status is not observed by secondary school administrators. Neither primary schools nor the SPS shares information with secondary schools for equity reasons and to avoid stigma when transitioning between schools. Beyond this anecdotal evidence, we formally test the validity of the identification strategy with three balancing test. First, we examine whether the proportion of SN peers predicts individual baseline characteristics (gender, native speaker, and age). The aim of this test is to detect potential selection into classrooms. Second, we regress the indicator for SN status on class fixed effects, which should be jointly insignificant if assignment to classroom is random with respect to SN status (Chetty et al., 2011). Third, we conduct a simulation exercise in the spirit of Carrell and West (2010). We re-sample 10,000 classes, randomly assign SN students to classes, and test whether the observed distribution of SN students differs from the simulated one. These three balancing test are presented and discussed extensively in Appendix B, and they all indicate that the key identification assumption of quasi-random assignment of SN students to classes is valid.

Our identification relies on variation between classes within school-track-years. Although families can potentially choose their district of residence and thereby influence schooling options for their children, possible selection into schools does not confound our results. In addition, municipalities within the canton of St. Gallen are very homogeneous in terms of demographics, indicating that such strategic behavior is most likely limited.[9] Mobility in Switzerland is generally low; approximately 80% of people do not move within five years (Eugster and Parchet, 2019). Moving for school choice alone is likely to be a rare occurrence. However, even if it occurs, our identification strategy is not affected by mobility between school districts.

Formally, we are interested in the following linear model:

$$Y_{icst} = \alpha + \beta\,\overline{SN}_{(-i)cst} + \gamma SN_{icst} + \lambda\,\overline{X}_{(-i)cst} + \vartheta X_{icst} + \varepsilon_{icst} \tag{1}$$

where Y_{icst} is the outcome of interest, such as the test score result of individual i in class c in school-

[9]The between-municipality variation in the unemployment rate (coefficient of variation: 0.42), the share of rich (0.19) and poor taxpayers (0.69), and the share with secondary (0.19), higher secondary (0.07), and tertiary education (0.22) is small. Data are from Eugster and Parchet (2019).

track s in cohort t. X_{icst} is a vector of observable characteristics, including gender, native German speaker, age, and class size. $\overline{X}_{(-i)cst}$ are the average characteristics of the peers of student i (proportion of girls, proportion of native speakers, and mean age). SN_{icst} is own special needs status. Our variable of interest is $\overline{SN}_{(-i)cst}$, a measure of exposure to SN students in a given class. We calculate exposure for each student i as $\overline{SN}_{(-i)cst} = (N_{cst} - 1)^{-1} \sum_{j \neq i} SN_{jcst}$, i.e., the proportion of SN students per class, excluding the SN status of i herself. The peer spillover parameter is β, which represents the impact of a marginal increase in SN peers on i's outcome.

Finally, ε_{icst} represents the error term, which we model in a components-of-variance framework. In detail, ε_{icst} is assumed to consist of two components: $\varepsilon_{icst} = \mu_{st} + e_{icst}$, where μ_{st} is a school-by-track-by-year fixed effect, and e_{icst} is an idiosyncratic error term.[10] The exogeneity of e_{icst} is a common assumption in the literature (Cooley Fruehwirth, 2009) and would be violated if there is some residual input that is correlated with observable student characteristics. As class formation within school-tracks is random and we consider only predetermined student characteristics, the assumption appears plausible in our setting. The school-by-track-by-year component exploits the natural variation in classroom composition within school-tracks for each year in the most flexible way. Throughout the paper, we refer to equation (1) as the main model. In studying effect heterogeneity, we also consider more flexible versions of $\overline{SN}_{(-i)cst}$ using cubic splines and an alternative measure of SN peer exposure based on the number of consultations with the SPS.

Being able to identify the classroom constitutes an advantage of our data, and this advantage also allows us to shed light on an additional challenge in peer effects analyses. Since correlated effects can emerge due to possible non-random assignment of teachers to classes, we specify a model that controls for time-constant heterogeneity in achievement inputs at the teacher level. Revisiting the components-of-variance framework, we specify the error term as $\varepsilon_{icst} = \pi_s + \phi_t + \eta_{cs} + u_{icst}$, where η_{cs} represents the teacher fixed effect (π_s and ϕ_t are school-track and time fixed effects, respectively). This latter specification filters out the effects of class

[10]A less flexible, but widely used, alternative specification is $\varepsilon_{icst} = \pi_s + \phi_t + v_{icst}$.

environment common to a teacher and potential matching between teachers and classes. Any potential time trends affecting schools or teachers have to be captured by the common factor ϕ_t. In sum, while our main specification relies on variation between classes within a school-track each year, the teacher-fixed-effect model is based on variation between classes of the same teacher in different years.

A trade-off exists between the identification within school-track-years and the identification within teachers. While the within-teacher identification has the potential advantage of filtering out correlated effects, this approach considerably reduces the variation that can be used for identification. Because teachers usually retain the same class throughout secondary school, they can teach an eighth grade class once every three years at most. We observe each eighth grade cohort between 2008 and 2017, and few teachers appear multiple times in the data.[11] Therefore, to estimate the effect of SN peers, the within-teacher identification exploits much less information than the within-school-track approach. It also assumes that there are no differential time/cohort effects for teachers or school-tracks. For these reasons, we retain specification (1) as the main model and defer the within-teacher specification to the sensitivity analysis.

5 Results

This section presents and discusses the results, divided into three parts. Part one focuses on the impact of SN peers on achievement, along with extensive sensitivity analyses. Part two studies heterogeneity in terms of distributional effects, severity and type of SN, and threshold effects. Part three examines the long-run impact of SN peers on post-compulsory education choices and labor market outcomes.

5.1 Effects of Classroom Composition on Achievement

The main results are presented in Table 2. Panel A shows the effect of differential exposure to SN peers on achievement for all students, while panels B and C show the results for students without and with SN,

[11] In detail, of the 1,197 teachers in our data, 16% are observed exactly once, and only 10% are observed more than four times.

respectively. All regressions control for school-track-year fixed effects and class size, and we gradually include student-level controls (column 2) and classroom-level controls (column 3) to our estimations. Standard errors are clustered at the classroom level.

The estimates of own SN status show that SN students score substantially lower than non-SN students. The results from panel A indicate that students with SN score approximately 0.1 standard deviations lower on the Math and German composite test score. This result is expected, as a large proportion of SN students have learning difficulties.

Next, we examine whether exposure to SN peers affects the academic achievement of other students in the classroom. The results from all specifications consistently indicate that exposure to SN peers reduces educational achievement. According to our preferred specification in column 3, adding one SN student to a classroom of 20 students causes the achievement of the other students to fall by 2.5% of a standard deviation (0.05 x 0.509), which is statistically significant at the 1% level. The effect corresponds approximately to one-fourth of the effect of being an SN student.

Adding covariates to the model does not substantially change the estimate. The baseline estimate without covariates (panel A, column 1) becomes slightly smaller as student-level covariates (column 2) and classroom-level covariates (column 3) are added to the model, although all confidence intervals overlap. Including individual-level covariates reduces primarily the estimate for own SN status, whereas the inclusion of classroom-level covariates reduces the peer effect estimate. This is intuitive, since SN status is correlated with other predetermined attributes – SN students are more likely to be male, older, and non-native German speakers.

Finally, the conditional results for SN status reported in panels B and C show that the negative peer externalities are driven primarily by the effects on students with special needs themselves, i.e., SN students tend to score lower if they are surrounded by other SN children in their class, compared to non-SN students. The difference in the peer effect is statistically significant. This can be tested by including an interaction between own SN status and the proportion of SN peers. As shown in Supplementary Table A.2 (column 3), the interaction coefficient is negative and statistically significant at the 1% level.

[Table 2 here]

The results suggest a moderate negative spillover effect of special needs students on the immediate educational outcomes of their peers. In the remainder of this section, we analyze the sensitivity of this result and extend the analysis in several dimensions. These checks ensure that the main results are stable and provide the groundwork for the analyses in the following sections. Supplementary Table A.2 and Figure A.3 test the sensitivity of the results with respect to the definition of outcome and treatment, whereas Supplementary Table A.3 tests the sensitivity with respect to the specification of the fixed effects and to sample restrictions.

The first two specifications of Supplementary Table A.2 perform the main analysis by subject, i.e., Math and German test scores separately. The peer effect parameter is highly significant in both specifications, but the negative externalities of SN students appear larger for German than for Math. However, the two coefficients have overlapping confidence intervals, which makes it unclear whether the peer effect for German is significantly larger than that for Math. In specifications 4 and 5, we change the definition of the treatment variable. Instead of defining SN exposure as the percentage of SN peers, we define exposure as the number of SN peers (column 4) and as at least one SN peer in the classroom (column 5). As expected, defining SN exposure as either a percentage or the number of peers with SN makes virtually no empirical difference. However, by defining exposure as at least one SN peer, the negative externalities become smaller in magnitude and significant at the 5% level. The reason for this is that specifying the treatment in this way considerably reduces the identifying variation. Only 54% of school-tracks contain classes without special needs students, and only 19.5% of the 1,398 identifying school-track-year clusters do so. Nevertheless, this result hints at non-linearities in the peer effects, which we investigate in the heterogeneity analysis in the next section.

Disruptive influences of classroom peers are likely to matter more for classroom production than influences from peers outside of the class environment. In fact, most of the interaction between students at school occurs within the classroom, while interactions with other students in a cohort are limited to the times before and after school or breaks. This is especially true in the Swiss context, where class composition remains unchanged across subjects.[12] The last specification in Supplementary Table A.2 (column 6) presents the peer effects analysis at the school-grade level, ignoring classroom composition. The purpose of this exercise is twofold. First, it uncovers whether observing the classroom rather than the grade-cohort is empirically relevant. Second, it allows us to compare the results with the literature, which generally focuses on the grade-cohort analysis. We find that the effect of own SN status remains nearly unchanged, but the peer externalities are smaller than in the main estimates. This shows that observing the relevant network structure is of key importance and that the effects differ depending on the accuracy with which the researcher identifies the set of relevant peers.

We can compare the size of our estimates to the peer effects that are obtained in studies that examine the influence of potentially disruptive children. Carrell and Hoekstra (2010) study the effect of disruptive peers on educational achievement in Florida, using variation in the share of children exposed to domestic violence in a school-grade cohort. They conclude that adding one child from a troubled family to a class of 20 reduces a composite reading and math percentile score by 2.2% to 2.7% of a standard deviation. They estimate the peer effect parameter at the school-grade level, which complicates a direct comparison with our main estimate. By repeating the main analysis and using the same aggregation level as Carrell and Hoekstra (2010), we find an effect of 0.99% of a standard deviation when adding one SN student to a class of 20 (Table A.2, column 6). This effect is smaller, suggesting that exposure to peers subjected to domestic violence may have more serious consequences.[13]

[12] The within-school identification might be more susceptible to a violation of the randomization assumption, if it is easier for parents to influence class assignment then to change schools. Note that we provide evidence for randomization in Appendix B.

[13] This hypothesis is corroborated by the data. In the SPS administrative records, we identify 558 children from troubled families (1.16% of all children). Our definition of troubled family ranges from cases of conflicts within the household to cases of domestic violence. The incidence rate we find is much lower than the 4.6% reported by Carrell and Hoekstra (2010), likely because we

It is also possible to assess the sensitivity of our results to how an SN student is defined. Using deciles of the distribution of contacts with the SPS, we begin by considering all SN students as having special needs and then gradually exclude from this definition the next lowest decile, until only children in the top decile of the number of contacts are considered to have SN. Supplementary Figure A.3 shows the parameter estimates obtained when varying the definition of SN students in this way. The estimated effects are robust to the definition of SN children but, as a general pattern, become more negative and less precise as the stringency of the definition of SN students increases. We further investigate this pattern in the heterogeneity analysis in section 5.2.

A second set of robustness checks is presented in Supplementary Table A.3, which focuses on the specification of the fixed effects and imposes some sample restrictions on our analyses. As we explained in the methods section, the fixed effects are specified at the level of randomization, i.e., at the school-track-year level. The identification thus comes from variation between classes within school-track-year cells. One alternative, less flexible specification often used in the literature is to separate the school-track fixed effects from the time fixed effect. By doing so (columns 1 and 2), we find that the peer effect parameter decreases slightly but remains statistically significant at all conventional levels.

A further alternative specification of the fixed effects is to add a teacher fixed effect. As we mentioned previously, few teachers are observed multiple times, because they retain the same class throughout secondary school. In addition, it is necessary to separate the year fixed effect from the school-track one because no teacher has two classes in the same school year. Keeping these two aspects in mind, specifications 3 and 4 in Supplementary Table A.3 perform the main analysis with the following fixed effects: year, school-track, and teacher. The findings show that including a teacher fixed effect considerably reduces the peer effect parameter, which nonetheless remains statistically significant at conventional levels. This is consistent with

only observe cases that are reported to the SPS. Nonetheless, we can estimate the negative externalities of children from troubled families at the school-grade level. We find that adding one child from a troubled family to a class of 20 reduces achievement by 2.5% of a standard deviation, which is consistent with the estimates obtained by Carrell and Hoekstra (2010) and Carrell, Hoekstra, and Kuka (2018). We also observe that children from troubled families have on average more contacts with the SPS than a typical SN student (12.3 vs. 7.97), suggesting that children from troubled families are quite severe cases.

Burke and Sass (2013), who also show that controlling for teacher impact reduces estimated peer effects. Although this reduction in the magnitude of the peer effect might be simply due to a loss of identifying variation, the reduction could also indicate that teachers have a mitigating effect on the negative externalities. Understanding what kind of teacher or teacher characteristics help buffer the negative externalities is an important policy question, which we address in section 6.

The final set of robustness checks, presented in columns 5-6 of Supplementary Table A.3, addresses sample restrictions. We perform the main analysis separately for each education track and find that negative externalities from SN students exist within both tracks. Interestingly, the peer effects are larger in the low-ability track, where the incidence of SN students is higher.

Overall, the sensitivity analyses indicate that the main results remain qualitatively unaffected by the definition of the outcome, the definition of the treatment, the specification of the fixed effects, and the sample splits. In addition, these analyses also reveal important features of and patterns in the data that we further investigate in the heterogeneity analysis and in section 6 on implementable educational policies.

5.2 Heterogeneity Analysis

This section addresses two questions related to heterogeneity. First, we analyze which students are affected the most by mainstreaming. Second, we examine which students are driving the education externalities. The results in the previous section show that achievement of SN students themselves is impaired more by higher levels of mainstreaming than is the achievement of students without special needs. The effects might also be heterogeneous across the achievement distribution. In general, low-achieving students may be more easily affected by disturbing influences in the classroom. To investigate this possibility, we transfer our main specification to a quantile regression setting and estimate the unconditional quantile treatment effects of SN peers over the distribution of test scores. We apply the unconditional quantile regression approach developed by Firpo, Fortin, and Lemieux (2009), a standard approach to investigate heterogeneous effects in a distributional framework.

The results are shown in Figure 2, which plots quantile treatment effects and the respective 95%-confidence intervals for different percentiles of the achievement distribution. We find that low-achieving students are more strongly affected by SN peers than are students in the upper end of the achievement distribution. The heterogeneity is substantial in the lower parts of the distribution, where the peer effect parameter for a student in the tenth percentile of achievement is twice as large as that of the median student. Above the median, the peer effect parameter is stable and becomes statistically insignificant at the top quartile.

[Figure 2 here]

Having established that it is mostly the low achieving students who suffer from higher levels of inclusion, we now focus on which children within the group of SN students are driving the effect. We distinguish between two dimensions. First, we ask whether students with the most severe SN generate the same amount of externalities as the average SN student. Second, we investigate whether a specific type of SN (i.e., learning impairment or behavioral problems) is particularly detrimental for other students.

To investigate the effect of SN severity, we rely on the information on the number of contacts with the SPS. The primary definition of special needs identifies every child that has been in contact with the SPS at least twice during primary school or before, irrespective of the reason for the contact or the intensity of counseling. This definition might include very simple cases, such as requests for delaying compulsory school entry or cases where no further action was deemed necessary after diagnosing the child. Since children's contact histories are observed in the data, it is possible to construct a straightforward measure of individual *SN severity* using the number of contacts with the SPS.

Panel A of Table 3 reports estimates of peer effects exploiting the individual intensity measure, divided into three specifications. In column 1, we add the sum of contacts as an additional measure of overall classroom SN severity to the main specification. Conditional on the exposure to SN peers, increases in the severity of SN at the intensive margin have an additional negative and significant effect on achievement. A one-standard-deviation increase in individual contacts with the SPS (approximately six contacts) causes a decrease in test scores by 1.2% of a standard deviation. This effect is nearly half of that of adding one SN student to a class of 20 from the main specification. The model in column 2 considers only students with severe SN, defined as those in the highest quartile of SN intensity. Restricting the treatment measure to this group, which is likely the most disruptive to instructional quality, the peer effect estimate increases by approximately 40% compared to the corresponding baseline estimate (Table 2, panel A, column 3). Finally, in column 3, we estimate the effects separately for the share of classroom peers belonging to different quartiles of the distribution of consultations with the SPS. The results indicate that peers with a low degree of SN severity have relatively small negative effects on peer achievement. However, having a higher percentage of classmates from the top quartile of the severity distribution reduces test scores considerably for all students. These results suggest that exposure to special needs peers matters at both extensive and intensive margins but that especially students with severe special needs at the upper end of the distribution are driving the negative spillover effects.

Panel B of Table 3 examines another dimension of SN heterogeneity, i.e., the type of SN. As previously mentioned, we focus on two distinct types of SN, namely learning impairments and behavioral problems. These two categories are chosen because they are the most common SN types, accounting for nearly 95% of all the reasons for registering a child with the SPS. The models in columns 1 and 2 include the two types separately in the regression model, while the model column 3 contains both SN types in one regression.[14] There are three important findings. First, we find that both types of SN have negative externalities, and these negative effects are statistically significant. Second, as column 3 shows, both students with learning and those with behavioral problems have lower achievement levels. However, having a learning impairment

[14]This is possible because, as we mentioned in the data section, we define the two SN types to be mutually exclusive.

is worse in terms of achievement than having a behavioral problem. The coefficient of own behavioral problem is almost half of that of own learning impairment, and we reject the equality of the coefficients at all conventional significance levels. Third, the negative externalities from students with learning impairments are slightly larger than those from students with behavioral problems, although we can reject the equality of the parameters only at the 10% level of significance (p-value = 0.08).

[Table 3 here]

Overall, the findings for the type of SN are consistent with Lazear (2001), who hypothesized that both types of classroom disruption would have an effect on the other students in the class: "The impediment [to classroom learning] may take a variety of forms. Student disruption provides a concrete example. Neither the student nor his classmates can learn much when the student is misbehaving, causing the teacher to allocate her time to him. Less nefarious, but equally costly is time taken by a student who asks a question to which all other students know the answer" (Lazear, 2001, p. 780). To our knowledge, the present paper is the first to provide causal evidence that both behavioral problems and learning impediments have negative externalities.

The analyses thus far consistently show that including more SN children in a classroom can be detrimental to student achievement. However, the results do not allow us to make any statement about whether segregated education would be preferable. Assuming for now that inclusive education is a given (we will relax this assumption in section 6), we attempt to address the question of whether there is some optimal classroom composition such that negative spillover effects are mitigated. Returning to the initial definition of SN students, we examine effect heterogeneity along the dimension of how many SN students are in a classroom. That is, we abandon the linearity assumption and consider a more flexible version of the main model using (restricted) cubic splines of $\overline{SN}_{(-i)cst}$. When using a cubic spline, one obtains a continuous, smooth function that is linear before the first knot, a piece-wise cubic polynomial between adjacent knots, and a linear function again after the last knot. The locations of the knots are based on the percentiles recommended by Harrell (2001).[15]

The results are reported in Figure 3, which is divided in two panels. Panel A depicts the predicted test scores for different percentages of SN peers. As suggested by Figure 1 in the descriptive analysis, (predicted) test scores decrease as the share of SN peers grows. Regarding the marginal effect (panel B), we observe that negative spillovers worsen with the percentage of SN peers and only begin to impact student achievement after including more than 3-4 SN students in a classroom or, alternatively, having a class that is more than 15-20% SN students. The presence of these threshold effects suggests that it might be possible to achieve a potential improvement in average achievement by strategically allocating SN students homogeneously across classrooms. We further investigate this and other policies to improve overall achievement in section 6.

[Figure 3 here]

5.3 Long-Run Outcomes

In this section, we investigate the long-run effects of exposure to SN peers. Thus far, the results show significant negative effects of inclusive schooling on educational achievement. However, the effects are of moderate size. If the effects do not persist over time and are absent or significantly diminished when students enter the labor market, inclusive education may ultimately be less harmful than the previous results suggest. As special needs students potentially graduate from secondary school with a regular degree, it is possible that mainstreaming may even be beneficial overall. A regular degree is considerably more valuable and potentially enables students to enter VET and regular employment, as opposed to graduating from special education institutions, which commonly leads to supported employment and education and, often, a life-long dependence on social security.

[15] We rely on splines for flexibility and convenience. The results from a global polynomial parametric specification are similar.

For these reasons, we consider the effects of class composition on outcomes over a prolonged time horizon, focusing on the choice of post-compulsory education and labor market outcomes such as employment and income. The results in Table 4 are the peer effect estimates for post-compulsory education. The model in column 1 indicates that a higher exposure to special needs students reduces the probability of enrolling in post-compulsory education. The effect size is moderate; adding one additional SN student to a class of 20 in secondary school reduces the probability of enrolling in VET or the academic preparation track (high school) by 0.44 percentage points.[16]

In column 2, we examine the composition of post-compulsory education, comparing enrollment in the vocational track with enrollment in the academic preparation track. We find that higher levels of exposure to SN peers most likely does not impact the composition of the vocational compared to the academic track. The estimate is positive, indicating an average shift towards vocational education, but statistically indistinguishable from zero. This indicates that exposure to SN students has equal effects on the probability of enrolling in VET or high school.

Next, we examine sorting within vocational education (column 3). There are quality differences across VET programs. For many occupations, both a high- and a low-quality track is available. Tracks differ in their curricula, duration, and expected pay. The results show that exposure to SN students robustly reduces the probability of enrolling in the high-quality VET track. Adding one additional SN student to a class of 20 reduces the probability of enrolling in the high-quality VET track by 0.59 percentage points.

For all specifications, the overall effects (panel A) are qualitatively comparable to those for SN students (panel B) and non-SN students (panel C). Similar to the previous results, estimates for SN students are slightly larger. However, none of the estimates is excessively large. Overall (panel A, column 1), a one-standard-deviation increase in the proportion of special needs students in a class causes a 1.5 percentage point decrease in the propensity to pursue post-compulsory education. This amounts to a moderate 1.8%

[16]Note that the number of observations for analyses on post-compulsory education and labor-market outcomes is lower due to mechanical attrition. Younger cohorts in our data have not yet completed compulsory education. See section 3 for more details.

reduction relative to baseline enrollment and corresponds to 3.9% of a standard deviation in enrollment.

[Table 4 here]

Finally, we consider the effects of class composition on long-run labor market outcomes, focusing on employment (measured in months per year) and monthly earnings. Unlike previous outcome measures, the labor market outcomes are not observed once at a single point in time that is the same for all individuals. Test scores, for example, are sampled once in grade eight for all students. In contrast, labor market information is available for all years from 2008 to 2016. This naturally leads to different, cohort-dependent coverage lengths, as older cohorts enter the labor market earlier. We address this issue by choosing different aggregation functions (average, maximum) and focusing on the most recent observation period 2016 (for which we observe the largest number of people). Additionally, cohort fixed effects are included in all specifications to account for cohort-specific experience and coverage. The results are shown in Table 5, where wages are expressed in logs, and the coefficients in columns 4-6 can be interpreted as semi-elasticities.

Regarding employment, we do not find an effect on the average number of months employed per year (panel A, column 1), the maximum number of months employed (column 2), or employment in the last registered year (column 3). Regarding income, we find a moderate decrease in the amount of monthly income caused by differential exposure to SN students in secondary school. The estimate for log average earnings is negative but insignificant (column 4), while the estimate for highest earnings is slightly larger and marginally significant (column 5). Finally, the estimate for earnings observed in 2016 is negative and highly significant (column 6).

We prefer the specifications in columns 3 and 6, because considering outcomes at a fixed point in time together with the cohort-specific effects accounts for possible experience effects and does not artificially depress positive labor market gradients. Aggregating over labor market histories of varying length (as in columns 4 and 5) artificially reduces variation due to aggregation, effectively compressing income gradients for those with longer labor market histories. Averaging as in column 4 over-weights earnings during VET, which are regulated and heavily compressed. In contrast, considering the most recent earnings is more likely to actually capture realized post-VET competitive wages. We confirm this by conditioning the analysis on individuals with completed VET degrees and find a similarly large effect (panel B, column 6).

[Table 5 here]

In summary, the findings indicate that exposure to SN peers has lasting effects. Lower educational achievement perpetuates as a reduced propensity to pursue (high-quality) post-compulsory education and, ultimately, lower income after entering the labor market. Bearing this in mind, as with the estimates for post-compulsory education, the earnings effects are small in size. Adding one additional special needs student to a class of 20 causes a 0.6% decrease in 2016 monthly earnings. This corresponds to an absolute decrease of 15 USD per month.

6 Implementable Educational Policies

Thus far, this paper has shed light on the spillovers from children with SN within an inclusive school system. This analysis is positive and has no direct policy implications. In this section, we discuss different educational policies in light of their ability to alleviate the burden of negative spillovers. We begin by asking the general question of whether inclusion is preferable to segregation if the overall goal is to maximize average achievement. Then, we analyze the role of teachers in providing a learning environment in which spillovers are minimized. Finally, we investigate whether school financial resources help reduce spillovers.

6.1 Inclusion versus Segregation

In our main analysis, we take school inclusion as given. We now consider the possibility of segregated classrooms, where SN students are separated from non-SN students. The question is whether average test scores would be higher or lower under a segregation regime. Graham (2011) proposes a methodology to solve the problem of a social planner who maximizes average test scores by optimally allocating children to classrooms. Denote by q the share of children with SN in the classroom and by p_{SN} the share of children with SN in the population. Let $m_{SN_0}(q)$ and $m_{SN_1}(q)$ be the outcome functions for children without SN and children with SN, respectively, and $m(q)$ be the overall outcome function, a weighted average of $m_{SN_0}(q)$ and $m_{SN_1}(q)$. As in the main peer effects estimations, we assume that peers and teachers are randomly allocated to classrooms, that spillovers occur only within classrooms, and that peers of the same type are exchangeable.

The aim of the social planner is to maximize overall outcomes $m(q)$ under the restriction that the shares of children with SN in each group sum the total share of SN in the population, that is,

$$\max_{F_Q(\cdot)} \int m(q) f_Q(q)\, dq \qquad \text{s.t.} \qquad \int q f_Q(q)\, dq = p_{SN}\ . \tag{2}$$

Given knowledge of the overall outcome function and the share of SN students in the population, the planner chooses a distribution of group composition. This is a maximization problem of functions and not trivial to solve in general. However, in the special cases in which $m(q)$ is either globally concave or globally convex, the solution is straightforward. If $m(q)$ is globally concave, the optimal solution is inclusion. On the other hand, if $m(q)$ is globally convex, the optimal solution is segregation.

To obtain estimates of the type-specific outcome functions $m_{SN_0}(q)$ and $m_{SN_1}(q)$, we rely on our main specification and include a global cubic polynomial in the share of SN children per class q interacted with individual SN status. Using a simple interacted polynomial specification has the advantage that aggregation and differentiation of the outcome functions is straightforward while the estimated functions remain flexible.[17] The outcome is the standardized composite test score introduced in our main analysis, and the regres-

[17] The estimates closely resemble the cubic splines presented in Section 5.2. Alternative estimates using multivariable fractional polynomials lead to comparable outcome functions.

sion includes the same set of fixed effects and covariates as in the baseline specification. We then aggregate the estimated partial outcome functions into the overall outcome function $m(q) = (1-q)m_{SN_0}(q) + qm_{SN_1}(q)$ and derive the first and second derivative $m'(q)$ and $m''(q)$.

Figure 4(a) depicts the outcomes for students with and without SN as a function of the share of SN students in a class and the overall outcome function, which is the weighted average of the two. Consistent with our previous results, the outcome functions are decreasing and stronger peer effects can be expected if the fraction of peers with special needs is larger. This is reflected in Figures 4(b) and (c), which show the first and second derivative of the overall outcome function. The first derivative of $m(q)$ is negative for all values of q and decreasing. The curve also increases in steepness, reflecting the finding presented previously that, after a threshold value, higher shares of SN classmates lower test scores. The second derivative is negative for all values of q between 0.04 and 0.93, indicating that the outcome function is concave in any reasonable neighborhood of $p = 0.25$. This result also suggests that segregation may become the optimal policy to maximize average test scores if the share of special needs children among the student population is sufficiently large. However, this threshold value is far above the share of special needs students observed in any country.

[Figure 4 here]

The concavity of the outcome function over large parts of its support implies that even in the presence of negative spillovers, segregated schooling is not optimal from a social planner's perspective. The intuition behind this result is the following: while the test scores of students without SN decrease if they share a classroom with children with SN, this effect is relatively small. Conversely, students who have SN themselves suffer from stronger negative externalities due to other SN peers. Therefore, segregation would lead to a strong reduction in test scores for students with SN, while it would only yield a moderate increase in test scores for students without SN. Given the non-negligible baseline SN rate in the student population, the loss for SN students is not compensated by the gain for students without SN.

Given the overall preference for inclusion, we can quantify the potential gains from a feasible and low-cost policy, namely from composing more equal classes in terms of SN students. Using the outcome functions presented in Figure 4, we estimate the predicted outcome under the current observed class composition and compare it to the predicted outcome under equal proportions of SN students in all classes within a specific school, track, and cohort.[18] The results suggest that the overall increase in average test scores would amount to 0.6% of a standard deviation. The gains would be slightly larger for SN students (0.70%) than for students without SN (0.47%). The modest size of these potential gains is not entirely surprising, given that the balancing tests presented previously show that the assignment of SN students to classes within the same school-track-year cell is quasi-random. However, we can study the heterogeneity in the potential gains with respect to the variation between classes belonging to the same cell. By doing so, we should expect larger potential gains for those cells with more unbalanced distributions of SN students across classes. Appendix Figure A.4 presents the results and confirms this hypothesis: larger potential gains in test scores occur in cells where the variation between classes in the proportion of SN students is higher.

6.2 The Role of Teachers

As an inclusive setting is preferable in our case, the question of how to minimize negative spillovers from SN students remains. It is reasonable to expect that high-quality teachers play a significant role in handling students with SN, especially in an environment where little external support is available. This section provides suggestive evidence that teacher quality is indeed important for mitigating negative externalities within the classroom.

While teachers can be identified in the data, there is no further background information on them, and no direct measure of teacher quality is available. We construct two alternative quality measures following

[18]This simulates a within-equalization of classes and avoids comparing a potentially infeasible and expensive allocation that would imply the relocation of students to different schools.

the literature on teacher effects in education. The first measure is based on the observation that more-experienced teachers have been shown to increase student achievement (Rockoff, 2004). Although we do not observe experience, it is possible to divide the sample into teachers that are observed only once in a given school and others that are observed multiple times. Note that a teacher typically remains with the same class for three years and thus is unlikely to be observed more than four times during the period under analysis, even if he or she is continuously working at the same school. This measure is with error, because there are several reasons that a teacher is observed only once that are not necessarily linked to low experience (e.g., retirement).

As it is difficult to measure experience with the data at hand, and because some researchers find that experience is not necessarily linked to teacher quality (Rivkin, Hanushek, and Kain, 2005), we construct a second measure based on teacher value added (VA). To do so, we take the standard approach to estimate teacher VA (e.g., Chetty, Friedman, and Rockoff, 2014) and retrieve the teacher fixed effects in a regression of the composite test score on individual and class characteristics, including the share of SN students. Since we control separately for school-by-track FE and time FE, the comparison of teacher quality is within schools and tracks but potentially over different years (we use the same specification as in column 4 of Table A.3 in the supplementary material). We then divide the teachers into two groups, the first having negative FE and the second having positive FE.[19]

Table 6 shows the results of the baseline estimation in the different subsamples defined by the two measures of teacher quality. First, note that the effect of own SN status on test scores remains remarkably stable, with only a slight decrease in the effect for high-VA teachers. In contrast, the peer effect parameter changes depending on the subsample under analysis. Focusing on the experience measure (columns 1 and 2), we find that a higher proportion of SN peers significantly decreases test scores in classrooms with both an experienced teacher and inexperienced teacher. However, the effect size appears to be larger in classrooms

[19]The correlation between the two teacher quality measures is positive and significant at the 5% level but quite small (0.07).

with non-experienced teachers. This difference in magnitude should be interpreted with caution, because the two coefficients have overlapping 95% confidence bounds.

Columns 2 and 3 in Table 6 show the estimates for the VA model. We find that classes taught by teachers with low VA (i.e., those with a negative fixed effect) exhibit negative SN externalities in line with our main analysis. In classrooms with teachers with high VA (i.e., those with a positive fixed effect), we instead observe much smaller negative spillovers. The SN peers parameter remains negative but is not statistically different from zero at conventional levels of significance (p-value $= 0.16$). This result implies that reducing classroom spillovers is possible and that teachers might play an important role in such reduction. While we are unable to examine which teacher characteristic or practice is driving this result, these findings add to the recent literature on the impact of good teachers on student outcomes (e.g., Hanushek and Rivkin, 2012; Jackson, Rockoff, and Staiger, 2014).

[Table 6 here]

To summarize, our examination of the teacher effect suggests that teachers matter and that they play an important role in the emergence of spillovers in inclusive education. Nonetheless, the results should be interpreted with caution, because we do not possess comprehensive information about teachers to conclusively determine which teacher practices or teacher characteristics are associated with the quality measures.

6.3 School Resources

One additional channel potentially influencing classroom externalities is school resources. School districts with higher spending per student might be able to improve school inputs such as student-to-teacher ratios, teacher salaries, or services for special education students. The question of whether school spending affects student academic performance in general has been extensively studied since the Coleman Report (Coleman et al., 1966). The Coleman Report indicates that variation in spending per student in the U.S. is unrelated to variation in student achievement in standardized tests. In his review of the topic, Hanushek (2003) confirms the findings of the Coleman Report. In contrast, more recent studies suggest that increases in per student spending during K-12 lead to more completed years of education, higher wages in adulthood, and reductions in adult poverty incidence (Jackson, Johnson, and Persico, 2016). While we do not directly extend this strand of literature, we examine whether per student spending in secondary school is correlated with the magnitude of the classroom externalities measured in the main analysis.

To do so, we proceed in two steps. First, we collect detailed data on school spending in 2017 for each municipality in the canton of St. Gallen. The data are made available from the official accounts published by each municipality at the end of the fiscal year. We focus on secondary school spending because our analysis covers that education level, and we choose the year 2017 because it offers the highest coverage. The per student spending is calculated by dividing total secondary school spending by secondary school enrollment, which is provided by the Ministry of Education of St. Gallen. We were able to collect data on 49 municipalities, covering 128 of the 155 school-tracks. The data show that municipalities spend on average 15,520 Swiss Francs (approximately 15,675 USD) per secondary school student. This figure is higher than both the OECD average (10,106 USD) and the corresponding figure in the U.S. (12,995 USD) but consistent with the Swiss national average of 15,022 USD (OECD, 2017).

In a second step, we estimate the baseline model and include an interaction between the proportion of SN peers and per student secondary school spending. The coefficient on such interaction term is not statistically different from zero (p-value $= 0.44$). We can also calculate the correlation – weighted by district size – between the SN peer effect and spending per student at the district level. This correlation is positive (correlation of 0.21) and would suggest that as spending per student increases, the externalities from SN peers become less negative. However, this correlation is not statistically significant at conventional levels (p-value $= 0.15$). We thus conclude that if spending per student plays a role in mitigating classroom externalities, this role is likely to be marginal.

7 Conclusion

This paper evaluates the effect of SN class composition on individuals' achievement, educational choices, and labor market outcomes for the student population of the Swiss region of St. Gallen between 2008 and 2017. To do so, we utilize a unique data set in which student performance in a standardized test is matched to psychological examination records compiled by the School Psychological Service, post-compulsory education career choices, and labor market histories from the social security administration. Because SN status is determined in primary school, students have SN for a reason exogenous to their peers in secondary school. This feature allows us to estimate peer spillovers free of the reflection problem that has been difficult to overcome in the peer effects literature. Using these data, we can control for school-by-track-by-year-specific effects and thereby identify spillovers by comparing classes with idiosyncratically high proportions of SN peers to classes in the same school-track cohort with idiosyncratically low proportions of SN peers.

We consistently find that having higher levels of SN students in a class lower peers' achievement. The incidence of negative spillovers is disproportionately on other SN students and students at the bottom of the achievement distribution, while students who perform well are less affected. This partly corroborates the result obtained by Hanushek, Kain, and Rivkin (2002), who find that sharing a classroom with special education students does not necessarily impede the achievement of regular education students. In addition, we demonstrate that the intensive margin matters and that the effect is driven primarily by students with severe SN, indicating that these students generate larger negative externalities for classroom education production. There are important threshold effects of composition: negative spillovers for other class peers are contingent on there being several SN students, indicating that disruptive influences exacerbate in conjunction.

This paper provides new insights on the "bad apple" principle in education (Carrell and Hoekstra, 2010; Lazear, 2001). In classroom education, the extent to which one student is able to learn during class time depends on the behavior of others in the class. If SN students take away learning time (e.g., through disruption or their need for the teacher's attention), a single SN student can impair the learning outcomes of all other students in the classroom. The results suggest that the bad apple principle may emerge if sufficiently many SN students – 15-20% according to our estimates – are grouped in the same classroom.

The results have important implications for education policy, suggesting that mainstreaming, although it induces moderate negative externalities on other students, is preferable to a policy of full separation. This policy contribution is complemented by the ancillary findings on the role of teachers and spending per student, which show that mitigating the negative externalities is possible. Bearing this in mind, schools and service providers should improve information exchange and early screening to deliberately guide classroom composition, such that SN students are distributed evenly across classes. Where the budget allows, classes should be set such that the number of special needs students in a class never exceeds a critical threshold. Finally, since negative spillovers in our study are driven especially by those students with extreme SN, there may be potential benefits to separating students at the upper end of the SN distribution who exhibit severe behavioral disorders. Further research regarding the optimal allocation of students is needed to derive a more definitive answer.

References

Ainscow, Mel and Margarida César. 2006. "Inclusive education ten years after Salamanca: setting the agenda." *European Journal of Psychology of Education* 21 (3):231–238.

Angrist, Joshua and Kevin Lang. 2004. "Does school integration generate peer effects? Evidence from Boston's Metco Program." *American Economic Review* 94 (5):1613–1634.

Bietenbeck, Jan. forthcoming. "The long-term impacts of low-achieving childhood peers: evidence from Project STAR." *Journal of the European Economic Association* .

Burke, Mary and Tim Sass. 2013. "Classroom peer effects and student achievement." *Journal of Labor Economics* 31 (1):51–82.

Carrell, Scott and Mark 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.

Carrell, Scott, Mark Hoekstra, and Elira Kuka. 2018. "The long-run effects of disruptive peers." *American Economic Review* 108 (11):3377–3415.

Carrell, Scott and James West. 2010. "Does professor quality matter? Evidence from random assignment of students to professors." *Journal of Political Economy* 118 (3):409–432.

Chetty, Raj, John Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How does your kindergarten classroom affect your earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126 (4):1593–1660.

Chetty, Raj, John Friedman, and Jonah Rockoff. 2014. "Measuring the impacts of teachers II: teacher value-added and student outcomes in adulthood." *American Economic Review* 104 (9):2633–2679.

Coleman, James, Ernest Campbell, Carol Hobson, James McPartland, Alexander Mood, Frederic Weinfeld, and Robert York. 1966. *Equality of Educational Opportunity*. Washington, DC: U.S. Government Printing Office.

Cooley Fruehwirth, Jane. 2009. "Can achievement peer effect estimates inform policy? A view from inside the black box." *Review of Economics and Statistics* 96 (3):514–523.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya." *American Economic Review* 101 (5):1739–1774.

Eugster, Beatrix and Raphaël Parchet. 2019. "Culture and taxes." *Journal of Political Economy* 127 (1):296–337.

European Agency for Special Needs and Inclusive Education. 2018. "European Agency statistics on inclusive education: key messages and findings." Report, European Agency for Special Needs and Inclusive Education.

European Commission. 2017. "Inclusive education for learners with disabilities." Study for the European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs.

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.

Figlio, David. 2007. "Boys named Sue: disruptive children and their peers." *Education Finance and Policy* 2 (4):376–394.

Firpo, Sergio, Nicole Fortin, and Thomas Lemieux. 2009. "Unconditional quantile regressions." *Econometrica* 77 (3):953–973.

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.

Friesen, Jane, Ross Hickey, and Brian Krauth. 2010. "Disabled peers and academic achievement." *Education Finance and Policy* 5 (3):317–348.

Graham, Bryan. 2011. "Econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers." In *Handbook of Social Economics*, vol. 1, edited by Jess Benhabib, Alberto Bisin, and Matthew Jackson. Elsevier, 965–1052.

Hanushek, Eric. 2003. "The failure of input-based schooling policies." *The Economic Journal* 113 (485):F64–F98.

Hanushek, Eric, John Kain, Jacob Markman, and Steven Rivkin. 2003. "Does peer ability affect student achievement?" *Journal of Applied Econometrics* 18 (5):527–544.

Hanushek, Eric, John Kain, and Steven 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.

Hanushek, Eric and Steven Rivkin. 2012. "The distribution of teacher quality and implications for policy." *Annual Review of Economics* 4 (1):131–157.

Harrell, Frank. 2001. *Regression Modeling Strategies*. Springer-Verlag, Berlin (Germany).

Hoxby, Caroline. 2002. "The power of peers: how does the makeup of a classroom influence achievement." *Education Next* 2 (2):56–63.

Jackson, Kirabo, Rucker Johnson, and Claudia Persico. 2016. "The effects of school spending on educational and economic outcomes: evidence from school finance reforms." *Quarterly Journal of Economics* 131 (1):157–218.

Jackson, Kirabo, Jonah Rockoff, and Douglas Staiger. 2014. "Teacher effects and teacher-related policies." *Annual Review of Economics* 6 (1):801–825.

Kalambouka, Afroditi, Peter Farrell, Alan Dyson, and Ian Kaplan. 2007. "The impact of placing pupils with special educational needs in mainstream schools on the achievement of their peers." *Educational Research* 49 (4):365–382.

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.

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.

Lazear, Edward. 2001. "Educational production." *Quarterly Journal of Economics* 116 (3):777–803.

Lindsay, Geoff. 2007. "Educational psychology and the effectiveness of inclusive education/mainstreaming." *British Journal of Educational Psychology* 77 (1):1–24.

Lyle, David. 2007. "Estimating and interpreting peer and role model effects from randomly assigned social groups at West Point." *Review of Economics and Statistics* 89 (2):289–299.

Manski, Charles. 1993. "Identification of endogenous social effects: the reflection problem." *Review of Economic Studies* 60 (3):531–542.

Neidell, Matthew and Jane Waldfogel. 2010. "Cognitive and noncognitive peer effects in early education." *Review of Economics and Statistics* 92 (3):562–576.

OECD. 2005. *Students with Disabilities, Learning Difficulties and Disadvantages: Statistics and Indicators*. Organisation for Economic Cooperation and Development, Paris (France).

———. 2017. *Education at a Glance 2017: OECD Indicators*. Organisation for Economic Cooperation and Development, Paris (France).

Oswald, Yvonne and Uschi Backes-Gellner. 2014. "Learning for a bonus: how financial incentives interact with preferences." *Journal of Public Economics* 118:52–61.

Rivkin, Steven, Eric Hanushek, and John Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica* 73 (2):417–458.

Rockoff, Jonah. 2004. "The impact of individual teachers on student achievement: evidence from panel data." *American Economic Review* 94 (2):247–252.

Ruijs, Nienke. 2017. "The impact of special needs students on classmate performance." *Economics of Education Review* 58:15–31.

Ruijs, Nienke and Thea Peetsma. 2009. "Effects of inclusion on students with and without special educational needs reviewed." *Educational Research Review* 4 (2):67–79.

Sacerdote, Bruce. 2001. "Peer effects with random assignment: results for Dartmouth roommates." *Quarterly Journal of Economics* 116 (2):681–704.

Swiss Federal Statistical Office. 2017. "Bildungsstand der Bevölkerung." Report, Bundesamt für Statistik.

Weder, Christina. 2016. "Nur wenige Rekurse haben Erfolg." *St. Galler Tagblatt*, NZZ Mediengruppe.

West, Martin and Paul Peterson. 2006. "The efficacy of choice threats within school accountability systems: results from legislatively induced experiments." *The Economic Journal* 116 (510):C46–C62.

Wolter, Stefan and Paul Ryan. 2011. "Apprenticeship." In *Handbook of the Economics of Education*, vol. 3, edited by Erik Hanushek, Stephen Machin, and Ludger Woessmann, chap. 11. Elsevier, 521–576.

Tables and Figures

Table 1: Descriptive Statistics

	(1) Mean	(2) SD	(3) Min	(4) Max
A. Special needs (N = 49,961)				
Special needs student	0.255	0.436	0.00	1.00
Number of special needs peers	4.395	2.742	0.00	15.0
Proportion of special needs peers	0.254	0.170	0.00	0.91
Learning impairment	0.181	0.385	0.00	1.00
Behavioral problem	0.061	0.239	0.00	1.00
Number of consultations	2.030	4.666	0.00	144
B. Achievement (N = 49,961)				
Test score: Composite (standardized)	0.000	1.000	–4.61	4.29
Test score: Math (standardized)	0.000	1.000	–4.52	3.85
Test score: German (standardized)	0.000	1.000	–4.51	4.05
C. Post-compulsory education (N = 40,730)				
No post-compulsory education	0.176	0.381	0.00	1.00
Vocational education track	0.672	0.470	0.00	1.00
Academic preparation track	0.153	0.360	0.00	1.00
D. Labor market outcomes (N = 26,429)				
Employment (mos/yr): average	10.73	2.173	1.00	12.0
Employment (mos/yr): maximum	11.54	1.722	1.00	12.0
Employment (mos/yr): last registered year	10.85	2.510	1.00	12.0
Monthly earnings: average	1,864	937.0	241	4,418
Monthly earnings: maximum	2,803	1,583	268	6,690
Monthly earnings: last registered year	2,563	1,593	179	6,148
E. Covariates (N = 49,961)				
Female	0.498	0.500	0.00	1.00
Native speaker	0.853	0.355	0.00	1.00
Age at test	14.91	0.709	12.0	17.0
Classroom size	19.17	3.671	10.0	31.0

Notes: Descriptive statistics for the main estimation sample, based upon 2,723 classes from 155 school-tracks. The last registered year in the social security data is 2016. Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Table 2: Estimates of the Effect of SN Class Composition on Student Achievement

	(1) Composite test score	(2) Composite test score	(3) Composite test score
A. All students (N = 49,961)			
Proportion of SN peers	–0.682***	–0.640***	–0.509***
	(0.079)	(0.077)	(0.070)
Own SN status	–0.119***	–0.103***	–0.097***
	(0.009)	(0.009)	(0.008)
B. Non-SN students (N = 37,236)			
Proportion of SN peers	–0.598***	–0.560***	–0.447***
	(0.080)	(0.078)	(0.073)
C. SN students (N = 12,725)			
Proportion of SN peers	–0.807***	–0.760***	–0.601***
	(0.113)	(0.108)	(0.096)
School-by-track-by-year FE	✓	✓	✓
Classroom size	✓	✓	✓
Student-level covariates		✓	✓
Classroom-level covariates			✓

Notes: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. Standard errors clustered at the classroom level are in parentheses below the coefficients. The dependent variable is the standardized composite math and German test score. The proportion of SN peers is scaled from 0 to 1, and to calculate the impact of 1 student in a class of 20, divide the estimated peer parameter by 20. Student-level covariates include indicators for gender, age, and native speaker. Classroom-level covariates include peers' average age, proportion of girls, and proportion native speakers. Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Table 3: Heterogeneity Analysis by Severity and Type of Special Needs

	(1) Composite test score	(2) Composite test score	(3) Composite test score
	A. Severity of special needs		
Peers' number of consultations	–0.002**		
	(0.001)		
Proportion of SN peers	–0.308***		
	(0.111)		
Proportion of peers with severe SN		–0.734***	
		(0.135)	
Proportion of peers (bottom SN quartile)			–0.436***
			(0.118)
Proportion of peers (second SN quartile)			–0.380***
			(0.111)
Proportion of peers (third SN quartile)			–0.384***
			(0.122)
Proportion of peers (top SN quartile)			–1.016***
			(0.137)
Own SN status	–0.085***	–0.087***	–0.066***
	(0.010)	(0.008)	(0.008)
	B. Type of special needs		
Proportion of peers with learning impairment	–0.588***		–0.615***
	(0.080)		(0.080)
Own learning impairment status	–0.113***		–0.120***
	(0.009)		(0.009)
Proportion of peers with behavioral problem		–0.255**	–0.377***
		(0.121)	(0.118)
Own behavioral problem status		–0.038***	–0.067***
		(0.014)	(0.014)
School-by-track-by-year FE	✓	✓	✓
Classroom size	✓	✓	✓
Student-level covariates	✓	✓	✓
Classroom-level covariates	✓	✓	✓
Observations	49,961	49,961	49,961

Notes: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. Standard errors clustered at the classroom level are in parentheses below the coefficients. The dependent variable is the standardized composite math and German test score. The proportion of SN peers is scaled from 0 to 1, and to calculate the impact of 1 student in a class of 20, divide the estimated peer parameter by 20. Student-level covariates include indicators for gender, age, and native speaker. Classroom-level covariates include peers' average age, proportion of girls, and proportion native speakers. Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Table 4: Estimates of the Effect of SN Class Composition on Post-Compulsory Education Outcomes

	(1) Post-compulsory education started	(2) Vocational track (vs. academic)	(3) High quality vocational track
A. All students			
Proportion of SN peers	–0.088***	0.029	–0.118***
	(0.021)	(0.028)	(0.026)
Own SN status	0.016***	0.006	–0.012**
	(0.005)	(0.004)	(0.005)
Observations	*40,730*	*33,566*	*27,354*
B. Non-SN students			
Proportion of SN peers	–0.072***	0.038	–0.080***
	(0.024)	(0.036)	(0.027)
Observations	*30,074*	*24,786*	*19,465*
C. SN students			
Proportion of SN peers	–0.097**	–0.007	–0.172***
	(0.039)	(0.031)	(0.054)
Observations	*10,656*	*8,780*	*7,889*
School-by-track-by-year FE	✓	✓	✓
Classroom size	✓	✓	✓
Student-level covariates	✓	✓	✓
Classroom-level covariates	✓	✓	✓

Notes: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. Standard errors clustered at the classroom level are in parentheses below the coefficients. The proportion of SN peers is scaled from 0 to 1, and to calculate the impact of 1 student in a class of 20, divide the estimated peer parameter by 20. Student-level covariates include indicators for gender, age, and native speaker. Classroom-level covariates include peers' average age, proportion of girls, and proportion native speakers. Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Table 5: Estimates of the Effect of SN Class Composition on Labor Market Outcomes

	(1) Employment: average	(2) Employment: maximum	(3) Employment: last registered year	(4) Monthly earnings: average	(5) Monthly earnings: maximum	(6) Monthly earnings: last registered year
A. All individuals (N = 26,429)						
Proportion of SN peers	–0.218	–0.035	–0.067	–0.037	–0.062*	–0.121***
	(0.162)	(0.118)	(0.174)	(0.034)	(0.037)	(0.046)
Own SN status	0.008	0.025	0.007	–0.019***	–0.014*	–0.008
	(0.030)	(0.023)	(0.036)	(0.007)	(0.007)	(0.009)
B. Individuals with completed VET degree (N = 13,948)						
Proportion of SN peers			0.087			–0.104**
			(0.214)			(0.047)
Own SN status			–0.057			–0.025**
			(0.044)			(0.011)
School-by-track-by-year FE	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓
Classroom size	✓	✓	✓	✓	✓	✓
Student-level covariates	✓	✓	✓	✓	✓	✓
Classroom-level covariates	✓	✓	✓	✓	✓	✓

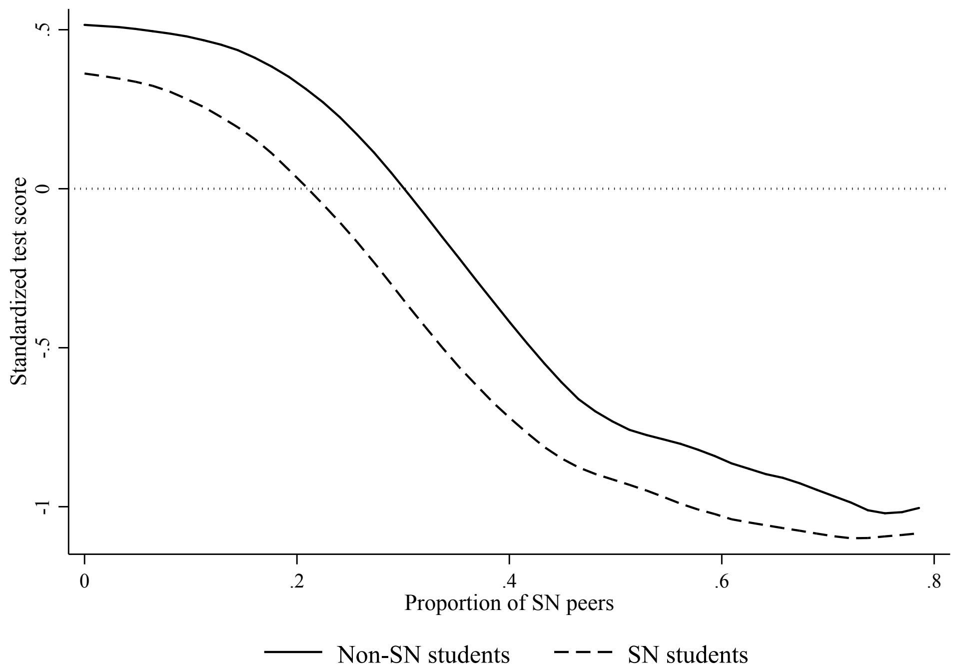
Notes: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. Standard errors clustered at the classroom level are in parentheses below the coefficients. Employment is expressed in number of months employed per year, monthly earnings are expressed on a logarithmic scale, and the last registered year is 2016. The proportion of SN peers is scaled from 0 to 1, and to calculate the impact of 1 student in a class of 20, divide the estimated peer parameter by 20. Student-level covariates include indicators for gender, age, native speaker, and year of birth. Classroom-level covariates include peers' average age, proportion of girls, and proportion native speakers. Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Table 6: Estimates of the Effect of SN Peer Externalities by Teacher Type

	(1) Not experienced teacher	(2) Experienced teacher	(3) Low VA teacher	(4) High VA teacher
Proportion of SN peers	–0.823***	–0.483***	–0.507***	–0.113
	(0.241)	(0.080)	(0.099)	(0.080)
Own SN status	–0.010***	–0.099***	–0.112***	–0.060***
	(0.024)	(0.009)	(0.012)	(0.011)
School-by-track-by-year FE	✓	✓	✓	✓
Classroom size	✓	✓	✓	✓
Student-level covariates	✓	✓	✓	✓
Classroom-level covariates	✓	✓	✓	✓
Observations	7,825	42,136	20,838	29,123

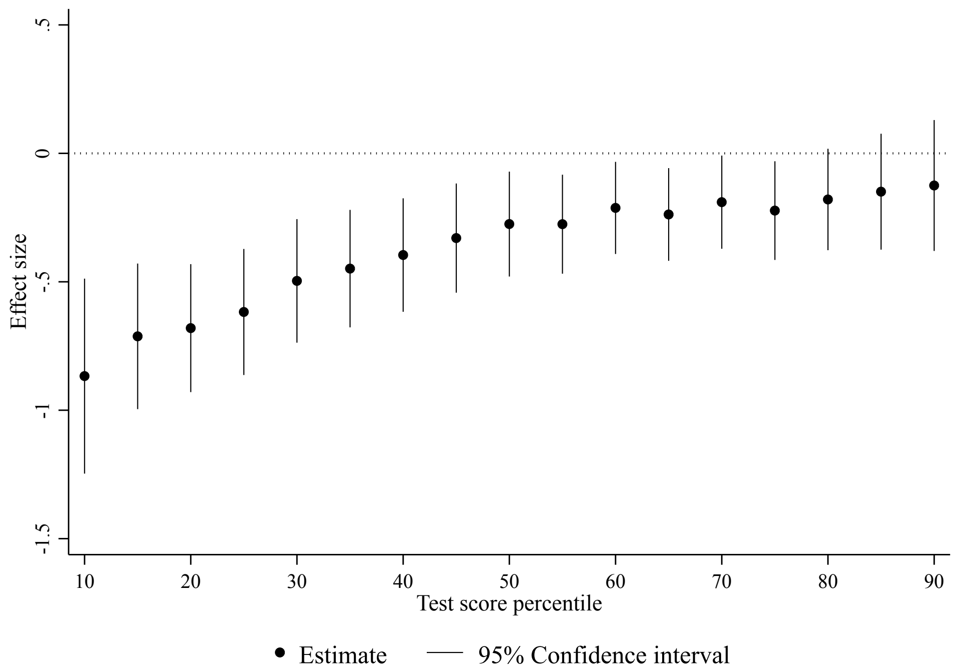
Notes: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. Standard errors clustered at the classroom level are in parentheses below the coefficients. The dependent variable is the standardized composite math and German test score. The proportion of SN peers is scaled from 0 to 1, and to calculate the impact of 1 student in a class of 20, divide the estimated peer parameter by 20. Student-level covariates include indicators for gender, age, and native speaker. Classroom-level covariates include peers' average age, proportion of girls, and proportion native speakers. Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Figure 1: Average Test Score by Proportion of Special Needs Peers



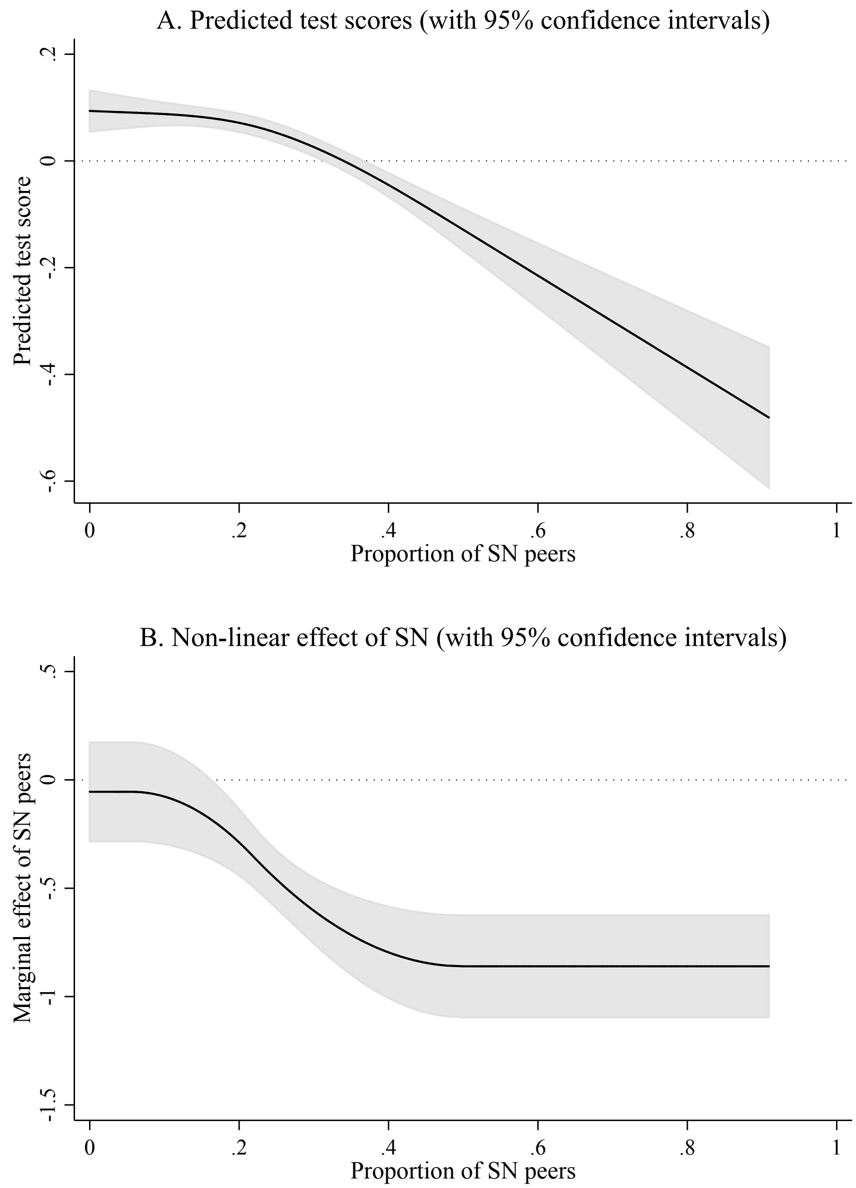
Notes: The lines are obtained by performing a kernel-weighted local polynomial regression of test scores on the proportion of SN peers. Data come from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Figure 2: Quantile Treatment Effects of Special Needs Peers



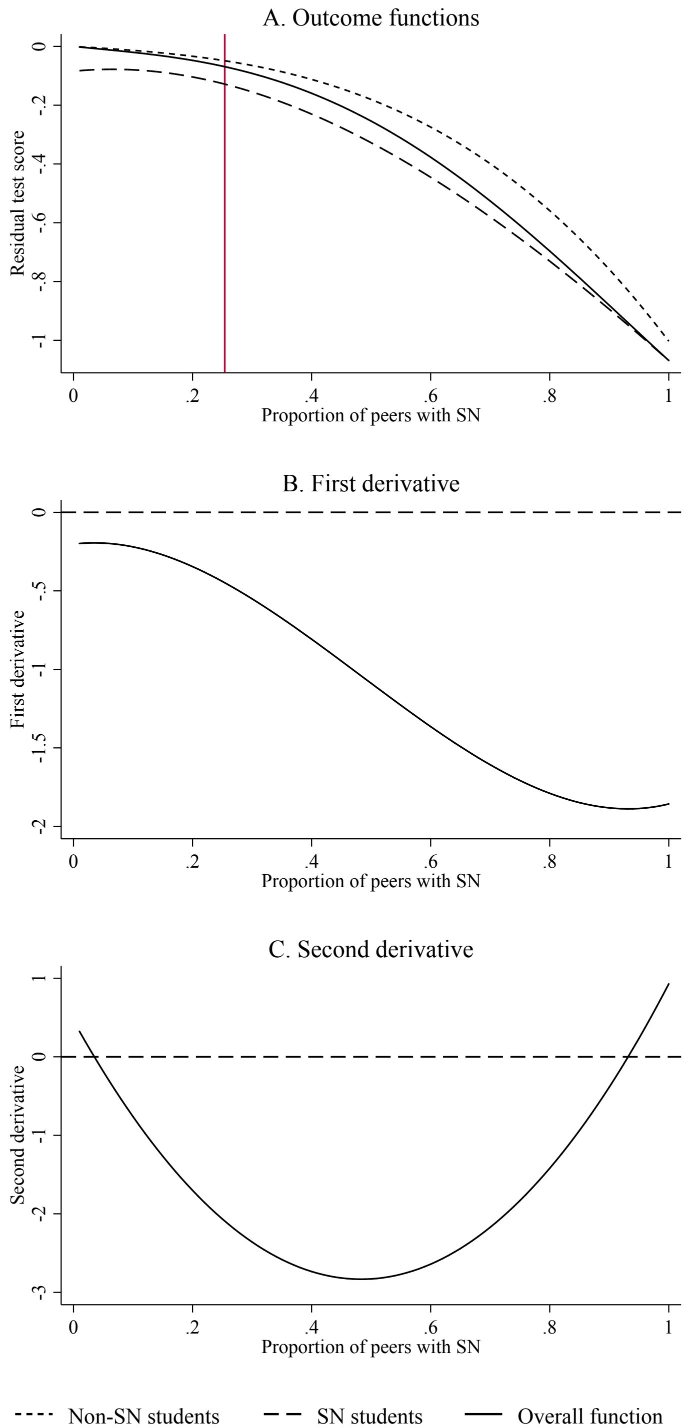
Notes: The figure depicts estimates from our main model using unconditional quantile regressions à la Firpo, Fortin, and Lemieux (2009). Data are from the School Psychological Service St. Gallen, the Ministry of Eduction of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Figure 3: Nonlinear Effects of Special Needs Peers



Notes: The estimates are based on restricted cubic splines applied to our main specification. Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Figure 4: Estimated Outcome Functions and First and Second Derivatives



Notes: The lines are obtained by performing a regression of test scores on a cubic polynomial of the proportion of SN peers interacted by SN status. Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

A Appendix: Supplementary Material

Table A.1: Construction of the Sample

		Observations
Raw data		54,451
Segregated special schools	–	1,265
Missing covariates	–	6
Missing test scores	–	464
Age restriction (12-17)	–	115
Segregated classrooms	–	275
Class size restriction (10-31)	–	2,365
Final sample		49,961

Notes: Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Table A.2: Sensitivity Analysis: Outcome and Treatment

	(1) Math test score	(2) German test score	(3) Composite test score	(4) Composite test score	(5) Composite test score	(6) Composite test score
Proportion of SN peers (classroom)	–0.405***	–0.534***	–0.466***			
	(0.069)	(0.069)	(0.070)			
Proportion of SN peers (cohort)						–0.198***
						(0.063)
Number of SN peers				–0.021***		
				(0.004)		
At least one SN peer					–0.060**	
					(0.027)	
Own SN status	–0.091***	–0.087***	–0.048***	–0.093***	–0.084***	–0.093***
	(0.009)	(0.009)	(0.016)	(0.008)	(0.008)	(0.008)
(Own SN status) x (Proportion of SN peers)			–0.167***			
			(0.046)			
School-by-track-by-year FE	✓	✓	✓	✓	✓	
School-by-track FE						✓
Year FE						✓
Classroom size	✓	✓	✓	✓	✓	
Cohort size						✓
Student-level covariates	✓	✓	✓	✓	✓	✓
Classroom-level covariates	✓	✓	✓	✓	✓	
Cohort-level covariates						✓
Observations	49,961	49,961	49,961	49,961	49,961	49,961

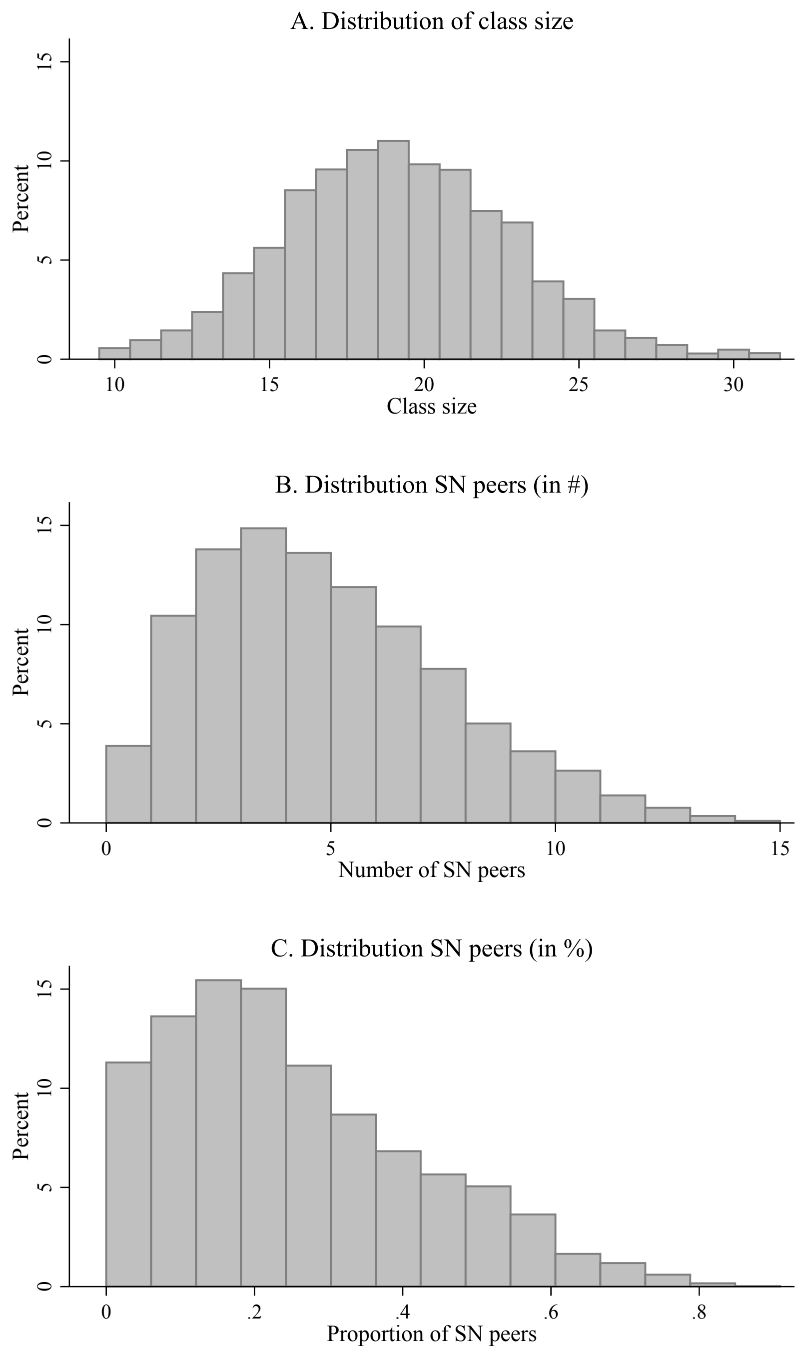
Notes: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. Standard errors clustered at the classroom level (cohort level for column 5) are in parentheses below the coefficients. The dependent variable in columns (3)-(6) is the standardized composite math and German test score. The proportion of SN peers is scaled from 0 to 1, and to calculate the impact of 1 student in a class of 20, divide the estimated peer parameter by 20. Student-level covariates include indicators for gender, age, and native speaker. Classroom-level and cohort-level covariates include peers' average age, proportion of girls, and proportion native speakers. Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Table A.3: Sensitivity Analysis: Fixed-Effects and Sample Restrictions

	(1) Composite test score	(2) Composite test score	(3) Composite test score	(4) Composite test score	(5) Composite test score	(6) Composite test score
Proportion of SN peers	–0.467***	–0.395***	–0.132**	–0.098*	–0.498***	–0.339***
	(0.061)	(0.058)	(0.052)	(0.051)	(0.095)	(0.102)
Own SN status	–0.090***	–0.089***	–0.073***	–0.072***	–0.122***	–0.063***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)	(0.012)
Low-ability track	✓	✓	✓	✓	✓	
High-ability track	✓	✓	✓	✓		✓
School-by-track FE	✓	✓	✓	✓		
Year FE	✓	✓	✓	✓		
Teacher FE			✓	✓		
School-by-year FE					✓	✓
Classroom size	✓	✓	✓	✓	✓	✓
Student-level covariates	✓	✓	✓	✓	✓	✓
Classroom-level covariates		✓		✓	✓	✓
Observations	49,961	49,961	49,961	49,961	18,151	31,810

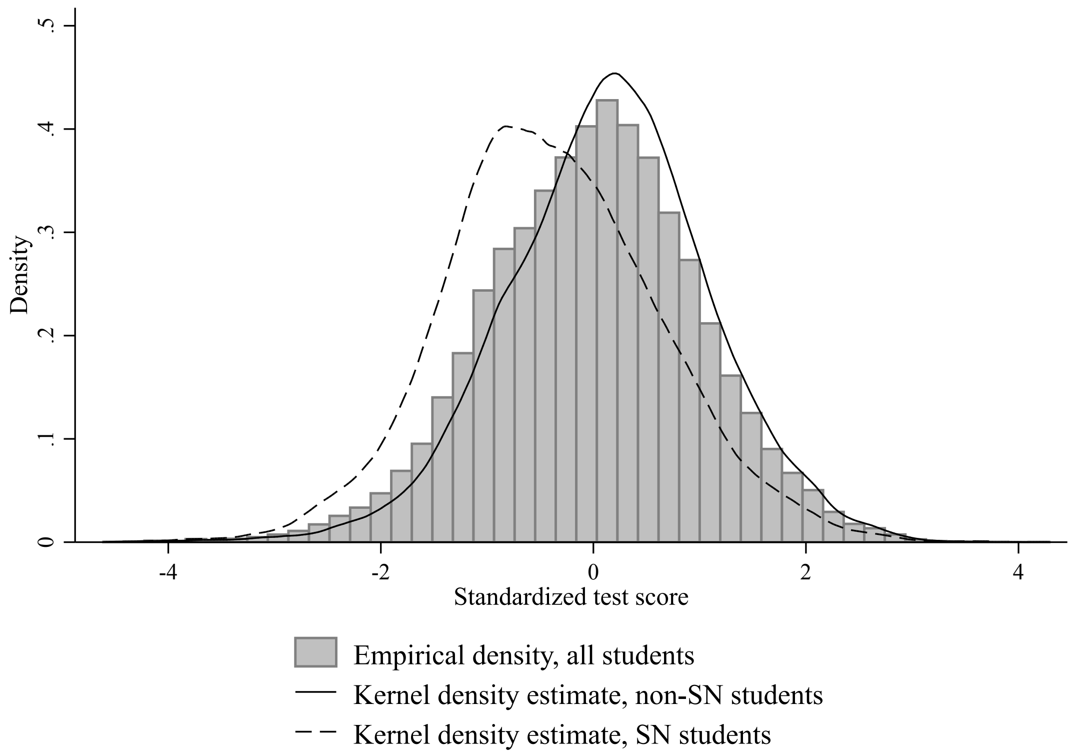
Notes: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. Standard errors clustered at the classroom level are in parentheses below the coefficients. The dependent variable is the standardized composite math and German test score. The proportion of SN peers is scaled from 0 to 1, and to calculate the impact of 1 student in a class of 20, divide the estimated peer parameter by 20. Student-level covariates include indicators for gender, age, and native speaker. Classroom-level covariates include peers' average age, proportion of girls, and proportion native speakers. Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Figure A.1: Distribution of Class Size and Special Needs Peers



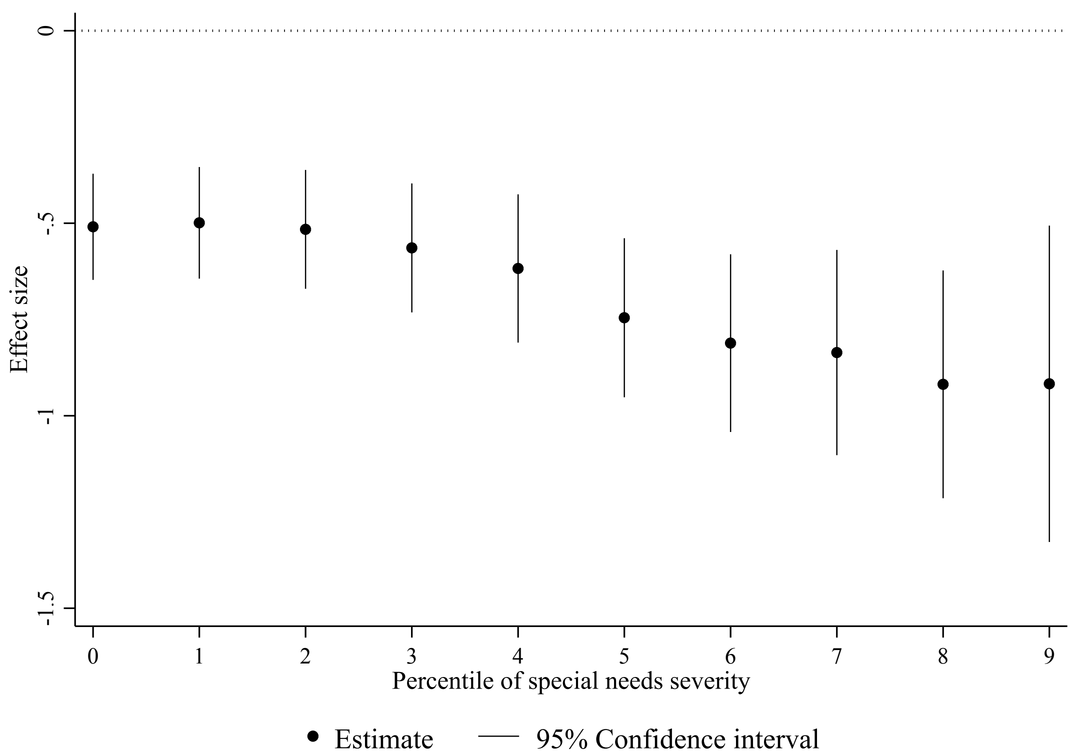
Notes: Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Figure A.2: Distribution of Test Scores



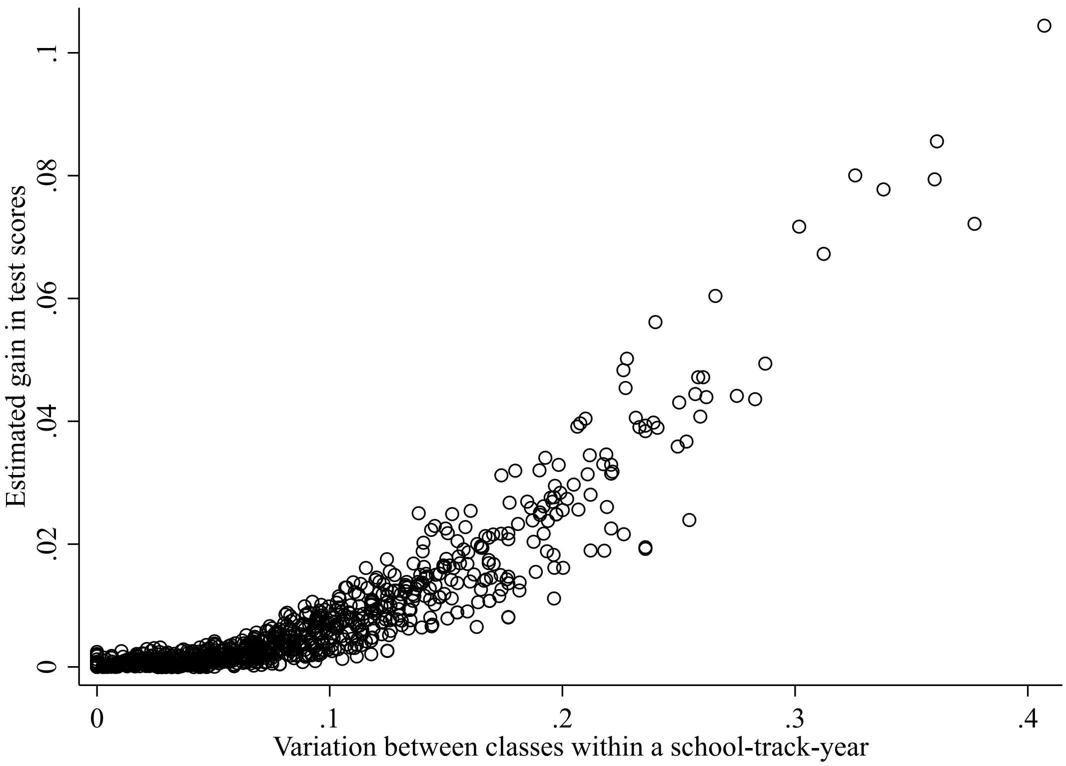
Notes: Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Figure A.3: Sensitivity to the Definition of Special Needs



Notes: Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

Figure A.4: Distribution of Potential Gains from Balanced Allocation of SN Students to Classes



Notes: Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.

B Appendix: Randomization and Balancing Tests

To test the validity of the identification strategy, we perform three sets of balancing tests. First, we test whether the proportion of SN peers predicts individual baseline characteristics (gender, native speaker, and age). The aim of this test is to detect potential selection into classrooms. Second, we regress the indicator for SN status on class fixed effects, which should be jointly insignificant if assignment to classroom is random with respect to SN status (for a detailed explanation of this method see Chetty et al., 2011). Third, we conduct a simulation exercise in the spirit of Carrell and West (2010). We re-sample classes, randomly assign SN students to classes, and test whether the observed distribution of SN students differs from the simulated one. This procedure should uncover any worrisome pattern in the assignment of SN students to classes.

Appendix Table B.1 shows the results for the first balancing test, presented for both the classroom level (panel A) and the cohort level (panel B). Each regression controls for the relevant fixed effects (school-track-year for classroom identification and school-track plus year for cohort identification), class (cohort) size, own SN status, and we progressively add student-level covariates and classroom-(cohort-)level covariates. None of the coefficients in Table B.1 is statistically significant and, in addition, the size of the effects are quite small in comparison to those presented in the main analysis. We also test for joint significance of the individual characteristics by regressing the proportion of SN peers on gender, native speaker, and age (along with the relevant fixed-effects, class/cohort size, and own SN status). We cannot reject the null hypothesis that the coefficients on gender, native speaker, and age are jointly zero, with a p-value of 0.332 using identification at the classroom level and 0.499 using identification at the cohort level.

Although we find no evidence for selection into classrooms according to observable characteristics, we might suspect selection into classroom based on unobservables. For this reason, in the second balancing test we examine whether the classroom indicator (i.e., classroom fixed effects) predicts SN status. To do so, we proceed in two steps. First, we regress SN status on the school-by-track-by-year fixed effects and retrieve the residuals from such regression. Second, we regress the residuals from the first step on classroom fixed effects and test whether these fixed effects are jointly significant. The resulting F-statistics is 0.564, well below the critical value of 1.00 required for 10% significance.

The third randomization check is simulation-based. For each class in a school-track, we re-sample 10,000 classes of the same size from the corresponding cell and calculate the average exposure to SN students. For every set of re-sampled classes, we calculate an empirical p-value as the proportion of simulations with exposure smaller than that observed in the original class. If class composition is random, the distribution of p-values within a school-track should be approximately uniform, which is testable using a one-sample Kolmogorov-Smirnov test. We reject uniformity one time out of 156 (i.e., 0.6% of cases), indicating that assignment of SN students to classes appears to be random. Overall, all the balancing tests we performed indicate that the key identification assumption of quasi-random assignment of SN students to classes is plausible.

Table B.1: Balancing Tests

	(1) Female	(2) Female	(3) Female	(4) Native	(5) Native	(6) Native	(7) Old at test	(8) Old at test	(9) Old at test
A. Classroom level									
Proportion of SN peers	–0.021	–0.015	–0.024	–0.050	–0.046	–0.047	0.005	0.004	0.007
	(0.024)	(0.024)	(0.025)	(0.032)	(0.032)	(0.033)	(0.009)	(0.009)	(0.009)
School-by-track-by-year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Classroom size	✓	✓	✓	✓	✓	✓	✓	✓	✓
Student-level covariates		✓	✓		✓	✓		✓	✓
Classroom-level covariates			✓			✓			✓
B. Cohort level									
Proportion of SN peers	0.030	0.035	0.029	0.026	0.027	0.031	0.007	0.007	0.006
	(0.027)	(0.027)	(0.029)	(0.040)	(0.040)	(0.041)	(0.008)	(0.008)	(0.008)
School-by-track FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort size	✓	✓	✓	✓	✓	✓	✓	✓	✓
Student-level covariates		✓	✓		✓	✓		✓	✓
Cohort-level covariates			✓			✓			✓
Observations	49,961	49,961	49,961	49,961	49,961	49,961	49,961	49,961	49,961

Notes: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. Standard errors clustered at the level of randomization are in parentheses below the coefficients. The dependent variable is the standardized composite math and German test score. The proportion of SN peers is scaled from 0 to 1, and to calculate the impact of 1 student in a class of 20, divide the estimated peer parameter by 20. Student-level covariates include indicators for gender, age, and native speaker. Classroom- and cohort-level covariates include peers' average age, proportion of girls, and proportion native speakers. Data are from the School Psychological Service St. Gallen, the Ministry of Education of the canton of St. Gallen, the Stellwerk test service provider, the Swiss Federal Statistical Office, and the Swiss Central Compensation Office.