The Labor Hub

Job Market Paper Series No. 16

Advisor: Andrea Carlo Michele Ichino

THE UNEQUAL SPILLOVERS OF CHILD DISABILITY IN INCLUSIVE EDUCATION

Massimo Anelli, Nicoletta Balbo, Alice Dominici, and Sofia Sierra Vasquez

The Unequal Spillovers of Child Disability in Inclusive Education*

Massimo Anelli[1,2], Nicoletta Balbo[1], Alice Dominici[†1,2,3], and Sofía Sierra Vásquez[4]

[1]Bocconi University
[2]IZA@LISER
[3]IMT School for Advanced Studies Lucca
[4]Inter-American Development Bank

Abstract

We study the effect of having a classmate with a severe disability on peers' achievement in an inclusive education regime. Using administrative data on the universe of Chilean primary school students, we compare cohorts within the same school-section that differ in whether a child with a severe disability was present in first grade. On average, exposure lowers peers' test scores by 2.1% of a standard deviation in Mathematics and 1.7% in Spanish, effects comparable in size to those caused by classmates with learning disabilities or from households with domestic violence, who are routinely included in mainstream education. The effects, however, are disproportionately borne by classmates with learning disabilities, who lose about two and a half times as much in Mathematics and four and a half times as much in Spanish as other classmates. They also originate from students with intellectual disabilities, while the mix of severe disability types in the classroom is irrelevant. Finally, spillovers are substantially smaller when students keep the same teacher across consecutive grades. The findings are consistent with a mechanism of competition for teacher attention and point to teacher continuity as a promising intervention, a low-cost and general-purpose policy.

JEL codes: I24, I28, J14
Keywords: inclusive education, disability, peer effects, special educational needs, teacher continuity

*We thank Vincenzo Bove, Thomas Crossley, Adriano De Falco, Beatrix Eugster, Benjamin Hattemer, Erik Lindqvist, Dan-Olof Rooth, Seth Sanders, Jenny Säve-Söderbergh, Francesco Serti and Giulia Vattuone for helpful comments and suggestions, as well as participants at the 29th SOLE Meeting in Denver (2026), the 50th SAEe Meeting in Barcelona, the 4th PhD and PostDoc Workshop in Naples, the 3rd CESifo/ifo Junior Workshop on the Economics of Education in Munich, the BAEM Meeting on Gender, Diversity and Human Capital in Bristol, PAA 2025, AlpPop 2025, RC28 2025, PopDays 2025 and the European Population Conference 2026, and seminar participants at the University of St. Gallen, SOFI (Stockholm University), the Dondena Centre at Bocconi University and the University of Alicante. Nicoletta Balbo and Alice Dominici acknowledge financial support from the European Union (ERC FRAILIFE, Child Disability and Family Life, G.A. 101077533) under the European Union's Horizon 2021–2027 research and innovation program. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. The opinions expressed in this study do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.

[†]Corresponding author: alice.dominici@imtlucca.it. IMT School for Advanced Studies Lucca, Piazza San Francesco 19, 55100 Lucca (LU), Italy.

1 Introduction

Inclusive education is the practice of teaching children with disabilities in mainstream classrooms, rather than in separate schools or classes. Since 2006, 191 countries have ratified the United Nations Convention on the Rights of Persons with Disabilities, committing to move children with severe disabilities out of segregated settings and into regular classrooms, and international organizations have promoted and financed inclusive education policies (e.g., UNICEF, 2012; World Bank, 2019; OECD, 2023; UNESCO, 2020). The spread of inclusion policies, however, has met resistance in several countries from political actors and education professionals, who point to competing student needs, possible negative spillovers on peers, and costly support measures that divert resources from mainstream schooling.[1] Despite the debate, sharing the classroom with a child with a severe disability has become an increasingly common experience for children without disabilities, and yet whether and how this experience affects their learning remains unclear.

The answer is not obvious ex ante. On the one hand, some students with disabilities may cause disruption and reduce peer learning (Lazear, 2001); the composition of the class may also shape how teachers target their instruction, lowering the level of instruction when some children face greater barriers to learning (Duflo et al., 2011). On the other hand, the additional support resources typically introduced by inclusion programs may generate positive spillovers (e.g., Andersen et al., 2020), and exposure to a disadvantaged group may also positively impact students' learning experience through improved pro-social skills (Boisjoly et al., 2006; Rao, 2019). This paper focuses on cognitive outcomes: we study whether and how exposure affects peers' learning, measured by standardized test scores.

Consistent with these competing mechanisms, existing estimates range from negative (Fletcher, 2010; Balestra et al., 2022; Rangvid, 2019) to null (Ruijs, 2017) and positive spillovers (Caron, 2026). Most of the evidence, however, concerns children with learning or milder sensory disabilities, who have long been taught in mainstream classrooms (Balestra et al., 2022; Huang

[1] See, for instance, The Economist (2026) and Gilmour (2018) for the United States, Latimer et al. (2025) and The Economist (2025) for England, and Libération (2025) for France. In Germany, the leader of the AfD in Thuringia called inclusion an "ideological project" (Der Standard, 2023), and in Italy a Lega candidate for the European Parliament suggested separate classes for children with greater difficulties (ANSA, 2024).

et al., 2021), whereas children with severe disabilities are still often taught in separate schools despite inclusion commitments (e.g., UNESCO, 2020). Even less is known about how these spillovers are *distributed* among classmates. Two obstacles explain these gaps: the non-random sorting of children into schools and classes complicates the identification of any spillover, and severe disabilities are rare enough that the data needed to condition on the determinants of sorting and to study how spillovers are distributed, linking each child's disability status to that of every classmate in a representative population, are scarce.

We address both obstacles using administrative records on the universe of Chilean primary school students. Chile is a particularly suitable setting: since 2009, its School Inclusion Program (PIE) has distinguished between the severe disabilities that define our exposure treatment (intellectual, sensory, motor and multiple disabilities, autism and severe speech disorders) and the other disabilities that receive only transitory support, including attention deficit disorders, dyslexia, dyscalculia and stuttering. Our sample follows the cohorts that entered first grade between 2009 and 2014, about 1.1 million children in 4,581 schools, linked to national standardized test scores in Mathematics and Spanish in grades 2, 4, 6 and 8. The data also cover a rich set of student and family characteristics, including household income, parental education, and eligibility for a targeted school subsidy, allowing us to study the distribution of spillovers across socioeconomic groups.

Identification rests on comparing cohorts within the same school-section (e.g., Hoxby, 2000), that is, the classes identified by the same letter in the same school, which share the same set of possible teachers. These cohorts differ in exposure to severe disability in first grade, while the school and section characteristics that typically raise endogeneity concerns about the assignment of students to classrooms are held fixed.[2] Our estimand is an intention-to-treat (ITT) effect of assignment in first grade.[3]

[2]Exposure varies across classes, that is, across cohorts within a school-section, so class fixed effects would absorb it. We do not hold the teacher fixed either, since teacher assignment may itself respond to exposure and is a possible channel, which we study in Section 7.

[3]Children with disabilities repeat grades more often than their classmates, especially in the first two years of school, and a repeater moves into the classroom of the following cohort. Later exposure therefore depends on the assignment of the adjacent cohort, which generates defiers and rules out instrumenting contemporaneous exposure with first-grade assignment. Section 5 discusses the estimand in detail. Importantly, in Section 5, we show that exposure does not significantly increase the probability of changing class or school (Table B4 in the Appendix).

Our main finding is that first-grade exposure to any classmates with severe disabilities lowers peers' standardized test scores by 2.1% of a standard deviation in Mathematics and 1.7% in Spanish, in a class of 33 students on average. The Mathematics penalty is stable or slightly increasing across grades, consistent with mathematics building on skills acquired in earlier grades, whereas the Spanish penalty attenuates over the first grades and then stabilizes. However, these magnitudes are comparable to the spillovers generated by groups that are routinely included in mainstream education. Depending on how we account for the difference between estimands, our estimates range between about 65% and 150% of the spillovers from an additional classmate receiving psychological support for learning and behavioral difficulties in Switzerland (Balestra et al., 2022), and of those from an additional classmate from a household with domestic violence in the United States (Carrell and Hoekstra, 2010).[4]

Average effects, however, conceal three results that redefine the scope of the policy debate around inclusion. First, when we allow the effect of exposure to depend on the classmate's own special-needs status, the penalty concentrates on children with a learning disability label. For them, the effect is 3.7% of a standard deviation in Mathematics and 4.2% in Spanish, against 1.4% and 0.9% for classmates without such a label. Socioeconomic disadvantage does not produce the same pattern: in Spanish, children eligible for the targeted school subsidy, children in the bottom income quintile and children whose mother did not attend college lose more than other children, but far less than children with learning disabilities, and in Mathematics the gradients are small and insignificant.

Second, the source of the spillovers is specific. Exposure to a classmate with an intellectual disability accounts for the whole average effect: once we control for it, the residual effect of exposure to any other type of severe disability is not distinguishable from zero. Combining different types of disability in one classroom imposes no additional cost, as a Herfindahl index of the diversity of the classroom's disability profile has no significant additional effect on test scores.

Third, the penalty is much smaller under teacher continuity. The effect of exposure among

[4]Both these studies find a reduction in test scores of 2.5% of a standard deviation in a class of twenty students. They both estimate the effect of contemporaneous exposure. The lower end of the range is our intention-to-treat estimate in Spanish; the upper end rescales our Mathematics estimate for the quarter of the comparison group that acquires a classmate with a severe disability by the time it is tested (Section 6).

students who keep the same teacher as in the previous year is 72% smaller in Mathematics and 57% smaller in Spanish than among students who change teacher, in the main specification, and exposure does not significantly predict teacher retention. Schools with an established practice of teacher looping, in which teachers follow the same class group across grades as a school-wide policy unrelated to disability, also show penalties about 40% smaller in Mathematics. This is reassuring, since looping schools serve poorer families and are more often public and rural, which should, if anything, work against finding smaller penalties there. By contrast, general teaching experience, which raises teacher effectiveness in other settings (Rockoff, 2004; Papay and Kraft, 2015), is associated with only a marginally smaller penalty: at the estimated gradient, offsetting it would take more than forty years of teaching. Experience in teaching children with severe disabilities is instead associated with a much smaller penalty, which falls to zero after about three years of such experience, although teachers with this experience are not randomly assigned to classes. Teacher continuity is not a disability policy but an organizational practice already recommended on general educational grounds (Hill and Jones, 2018), which requires neither additional staff nor disability-specific resources.

We interpret these results as evidence of competing needs: children with severe disabilities and children with learning disabilities draw on the same finite pool of teacher attention. Specialized support covers about a third of class time, is shared by the whole class, and adds little time for each child with a severe disability (Section 2), so that for much of the week the main teacher manages the class alone. If spillovers arise from competition for this teacher's time, they should come from the disabilities whose support relies most on it, and fall on the classmates who rely on it most.

Our results match both predictions. Support for sensory and motor disabilities relies more heavily on specialized inputs that do not draw on the main teacher's instruction, such as Braille, sign language and assistive devices, whereas intellectual disability requires adapting the curriculum and continuous individual instruction in the core subjects by the same teacher. Intellectual disability is associated with the smallest increase in support spending, and it is the only type with a significant negative effect on peers, an effect that grows with

the number of children with a severe disability and does not depend on the mix of types, as expected if children compete for one resource rather than for many specialized ones. The losses, in turn, concentrate on the children who rely most on adult attention to learn, such as those with learning disabilities, rather than on disadvantaged children in general; distributional estimates also show larger spillovers at lower score quantiles. Finally, the penalty is smaller when the teacher already knows the class or has experience with severe disabilities, consistent with disability-specific learning by teachers. The same logic implies that children with learning disabilities also compete with each other, consistent with the spillovers documented for them in previous work (Balestra et al., 2022); our design isolates the additional demand created by a child with a severe disability.

Disruption alone is unlikely to explain this pattern. Student, parent and teacher surveys report that exposed classrooms have only a moderately worse behavioral climate and somewhat more bullying, but these associations are similar for all classmates, whereas the learning loss is concentrated on children with learning disabilities. Disruption may still weigh more on children who are more prone to distraction, which suggests that interactions between categories of special needs deserve attention when classes are formed. Suggestive evidence on the timing of labels, however, points to competing needs: the losses concentrate on children with learning disabilities who are labeled only from fourth or fifth grade onwards.[5] These children went unrecognized and without dedicated support through the first years of school, relying more heavily on the main teacher's attention to learn.

This paper contributes to a large literature on peer effects in education (e.g., Bifulco et al., 2011; Anelli and Peri, 2019; Anelli et al., 2023; Angrist, 2014), and in particular to work on peers who place additional demands on the classroom, such as children exposed to domestic violence (Carrell and Hoekstra, 2010; Carrell et al., 2018), disruptive peers (Figlio, 2007; Kristoffersen et al., 2015) and low achievers (Lavy et al., 2012).

Within the economics of special education, our contribution is the study of cognitive spillovers from severe disabilities in an inclusive education regime. Several papers focus on learning

[5]The observed timing pattern goes against a simple labelling explanation. Exposure raises the probability that a classmate is labeled with a learning disability by 0.4 percentage points, but, in the 2013 and 2014 cohorts for which the timing of labels can be observed, only among labels recorded by third or fourth grade, whereas the spillovers on test scores concentrate on children labeled later.

disabilities and behavioral issues, typically in segregated contexts where these special needs are the only ones subject to inclusion policies (de Croes et al., 2025; Balestra et al., 2022; Huang et al., 2021; Fletcher, 2009, 2010).[6] Three papers are particularly relevant to our work. Caron (2026) shows positive spillovers on the educational attainment and employment of non-disabled students from the historical state mandates that required U.S. public schools to educate disabled children; we instead estimate individual-level effects on learning under a modern inclusion program and modern definitions of disability. Ruijs (2017) and Rangvid (2019) study partially segregated systems, the Netherlands and Denmark, in which the included students are selected by parents or experts and disability types are only partly observed.[7] Our setting and data allow us to document individual-level externalities within the classroom, and to break down the effects by disability type, showing that they depend on which needs share the same classroom.[8] More broadly, we contribute to the growing economics of disability, a field that spans labor, health and public economics and studies how disability shapes employment, earnings and reliance on social insurance, and how public policies and the organization of work affect the inclusion of people with disabilities (e.g., Maestas et al., 2013; Dahl et al., 2014; Bloom et al., 2026). We show that disability policy also reaches peers without disabilities, through the classroom.

Finally, our results on teacher continuity and disability-specific experience extend the study of teacher continuity and teacher human capital to the inclusion of children with disabilities. Keeping the same teacher, which raises achievement on average (Hwang et al., 2021), is also associated with much smaller spillovers from disability, which makes it a promising low-cost practice. Our disability-specific results are also consistent with the finding of Ost (2014)

[6] Ballis and Heath (2021a) and Ballis and Heath (2021b) assess the effects and spillovers of reduced access to special education for students with mostly mild disabilities in the U.S., where students with more severe disabilities are often taught in separate classes within mainstream schools. In Chile, one quarter of children with disabilities are enrolled in subsidized private special schools because their needs cannot be adequately supported in an inclusive context: the segregation rate of roughly 0.8%, however, is still far below other institutional regimes in the literature, such as Denmark (4–6%; Rangvid, 2019) or the Netherlands (2.1%; de Croes et al., 2025). Robustness checks that include children who attended mainstream schools before moving to special schools show quantitatively consistent results.

[7] Gottfried (2014), in the U.S., identifies spillovers on interpersonal socialization and behavioral outcomes among teacher-identified first-graders. Hanushek et al. (2002) find that special education raises the achievement of the students with disabilities who receive it.

[8] Contreras et al. (2020) study the same Chilean program, comparing outcomes before and after the 2009 regulation and pooling all categories of special needs.

that teachers improve with grade-specific experience; with that of Hill and Jones (2018) that repeated student-teacher matches are particularly beneficial to minority students; with broader models in which teachers allocate a limited amount of attention across students with different needs (Lazear, 2001; Duflo et al., 2011); and with evidence that teachers target their effort towards specific students (Neal and Schanzenbach, 2010).

The remainder of the paper is organized as follows. Section 2 describes the Chilean setting, Section 3 the data, Section 4 the measurement of disability exposure, and Section 5 the empirical strategy. Section 6 reports the average effects, Section 7 examines heterogeneity and mechanisms, and Section 8 concludes.

2 The Chilean setting

Chilean primary schools last for eight grades. Class groups are formed in first grade and remain largely unchanged throughout the entire cycle: for instance, a child who begins primary school in class 1A, where letter A denotes the section, continues in the next year in 2A with the same classmates, net of repeaters and transfers. Schools are of three kinds: public schools run by municipalities, voucher schools that are publicly funded and privately managed, and private schools funded by fees. In our sample, 34% of children attend public schools, 57% voucher schools and 8% private schools. Only publicly funded schools can take part in the School Inclusion Program described below, so that children with severe disabilities are recorded almost exclusively in public and voucher schools, which provide nearly all of our identifying variation.[9] We exclude secondary schools for two reasons: first because, even controlling for school-section characteristics, the pool of students with disabilities that reaches secondary school is positively selected; second, because exposure in secondary school correlates with exposure in primary school, reflecting likely ad-hoc class composition.

Admission to primary school has historically been decentralized, and schools were allowed to set their own criteria, typically based on expected achievement and family background (Carrasco et al., 2017). For this reason, our analysis uses only within-school variation.

Our outcomes are scores in the SIMCE, a national standardized test administered at the

[9]School types are absorbed by school fixed effects in our specification; removing private schools does not alter the results.

end of the school year to all students in the tested grades. While the test does not count towards students' grades, it matters for schools: results are published and used to rank schools (Mizala and Urquiola, 2013), and they determine performance bonuses for teachers (Contreras and Rau, 2012), so that schools and teachers have strong incentives to prepare students for the test. Our cohorts are tested in Spanish in second grade and in Mathematics and Spanish in fourth, sixth and eighth grade, between 2012 and 2019. Because not every grade is tested every year, each cohort is observed in two or three tested grades (Table A1 in the Appendix).

Since 2008, the Preferential School Subsidy (*Subvención Escolar Preferencial*, SEP) has provided an additional per-student voucher for socioeconomically disadvantaged children, labeled as *priority* students, who are also exempt from school fees in participating schools (Neilson, 2021; Feigenberg et al., 2019). Priority status is assigned on the basis of household income, enrollment in social programs and parental education, and its coverage expanded rapidly as the program was rolled out. We use priority status measured in the first-grade year, which is predetermined with respect to exposure, as a marker of socioeconomic disadvantage that is distinct from the special-needs categories described below.

Reforms for inclusive education Inclusive education in Chile dates to the 1990s, when a series of decrees created subsidies and training programs for schools enrolling students with intellectual, sensory and motor disabilities and with speech disorders. By the early 2000s about 8% of primary schools had students with special needs, though these students were usually taught in segregated groups (Godoy et al., 2004).

The national inclusion program in its current form (*Programa de Integración Escolar*, PIE) rests on Law 20.201 of 2007, which introduced differentiated subsidies for students with special educational needs, and on Supreme Decree 170 of 2009, which set the requirements for schools to participate in PIE and the rules for diagnostic assessment, and introduced for the first time the distinction between permanent and transitory needs (Ministerio de Educación de Chile, 2009).

PIE still coexists with a small network of subsidized private special schools: the law guarantees children with disabilities access to regular or special schools as appropriate (República

de Chile, 2010), implying that special schools serve severe cases that regular integration measures cannot adequately support. In the years in which our cohorts attend primary school, special schools enroll 0.8% of primary-age children, whereas three quarters of primary-age children with severe disabilities attend regular schools.[10] The children we study are therefore those included in regular classrooms, the margin that inclusion policies expand.[11]

Two features of the PIE program matter for our analysis. First, the program distinguishes permanent from transitory needs. *Permanent* special needs include intellectual, sensory, motor and multiple disabilities as well as autism and severe speech disorders, whereas *transitory* special needs cover learning difficulties such as attention deficit disorders, dyslexia, dyscalculia, and stuttering. Support for transitory needs is granted for a limited period and must be renewed through an annual reevaluation, whereas support for permanent needs continues throughout schooling (Ministerio de Educación de Chile, 2009). Throughout the paper, we define children with severe disabilities (CWD) as the children recorded with permanent special needs, and children with learning disabilities as the children recorded with transitory special needs. The latter is an institutional label rather than a clinical category: it is assigned after a professional diagnosis to children enrolled in schools that participate in PIE, it groups specific learning disorders such as dyslexia and dyscalculia with attention deficit and language disorders, and it entitles children to support that must be renewed every year. Children whose difficulties have not been diagnosed, or who attend schools without PIE, therefore carry no label.

Second, the program regulates classroom composition. A classroom participating in PIE may enroll at most two students with permanent needs (severe disabilities) and five with transitory needs. These caps constrain the number of children with each kind of need, but do not regulate which needs are combined in the same classroom. Specialized support is also assigned to the class as a whole and consists mostly of co-teaching between a special-education teacher and the main teacher(s). It covers about a third of weekly class time, so

[10] Our calculations from the national enrollment records and the special-needs registers, for children aged 6 to 13 in 2016 and 2019.

[11] In 2018, 2,027 special schools operated alongside 5,662 schools with a PIE, but almost three quarters of special schools served only preschool children with specific language disorders, a transitory condition; 402 served children with intellectual disabilities (Holz, 2018).

that for much of the week the class teacher is responsible for all included children alone.[12] At the school level, the program's spending rises more with each enrolled child with a sensory disability than with a child with an intellectual disability, which is associated with the smallest increase among severe disabilities, consistent with intellectual disabilities having no one-size-fits-all auxiliary means (Appendix H). Importantly, our data have two limitations: we observe the presence of a learning disability label without knowing the underlying condition, and support is observed at the school level (expenditure) and at the school-grade level (specialist staff), so that we cannot match each child with a disability to the amount and type of support they receive.

3 Data and sample

Data sources The dataset combines several administrative sources. Enrollment records identify each child's school, grade and classroom in every year from first grade, between 2009 and 2014, to the last tested grade, in 2019; records from earlier years are used only to determine whether a child enrolled in first grade is repeating it. Restricted-access registers from the Ministry of Education, available from 2016 to 2023, provide each child's special-needs status, distinguishing permanent needs (severe disabilities) from transitory needs (learning disabilities) and reporting the type of severe disability in seven broad categories: intellectual, autism, severe speech disorder, motor, visual, hearing and multiple.

SIMCE test scores, provided by the Agency for the Quality of Education, give standardized Mathematics and Spanish results for our cohorts in grades 2, 4, 6 and 8, between 2012 and 2019. Mathematics is not tested in second grade, so second-grade results exist for Spanish only (Table A1). Survey modules administered alongside the SIMCE to students, teachers and parents in the tested years allow us to construct classroom-level indices of behavioral climate, bullying and openness to diversity, which we use in Section 7.

Administrative teacher records identify the teachers of each class in the years in which

[12] A full-time school week consists of 38 periods of 45 minutes, or 28.5 hours (República de Chile, 1997). The program prescribes at least 10 hours a week of specialized support, 8 periods of which are co-taught in the classroom, for up to five children with transitory needs (learning disabilities), and recommends only three more hours for each child with permanent needs, the severe disabilities behind our exposure treatment (Ministerio de Educación de Chile, 2009, 2013).

our cohorts attend primary school, which allows us to measure whether a class keeps the same teacher from one year to the next and the experience of each first-grade teacher. Unfortunately, data on support expenditures only exist at the school level and do not allow distinguishing which type of support was associated with which students, preventing us from assessing the moderating role of support expenditures.

Sample construction We follow the cohorts of children who entered first grade between 2009 and 2014 through primary school. The choice of cohorts is dictated by the institutional setting: 2009 is the first cohort subject to the PIE inclusion program since first grade; cohorts from 2015 onwards are dropped because they were tested in a single grade: the second-grade test was discontinued after 2015 and the 2020 test was canceled during the COVID-19 pandemic.[13]

Two restrictions serve the identification strategy directly. We exclude school-sections whose smallest class over the period falls below ten students or whose largest exceeds forty-six, respectively the third and ninety-ninth percentiles of the class-size distribution, because comparisons within very small or very large groups are unstable and anecdotally relate to very specific settings, such as particularly remote rural areas or particularly dense urban ones, that differ from other settings in various ways beyond the application of the PIE inclusion reform. Moreover, we restrict the sample to school-sections observed in at least two of the cohorts we study, the only ones that contribute identifying variation.

The resulting sample contains 2,516,981 student-by-grade observations on 1,117,224 children in 4,581 schools and 8,039 school-sections, each observed in up to six cohorts. Because covariates are missing for some records, the estimating samples are 1,527,433 observations for Mathematics and 2,056,979 for Spanish.

Our sample covers about 70% of Chilean first graders in these cohorts but only half of the schools. The excluded schools are mostly small and rural, with a median of seven first graders against forty-two in retained schools, and often have missing cohorts or single class groups combining several grades.

[13] We further exclude adult primary education and schools that serve special populations (e.g., special classes in hospitals and jails).

4 Disability measures and exposure treatment

Our analysis sample consists of peers, namely children who are not CWD (Section 2). Children with learning disabilities were included in mainstream education before inclusion policies, and remain in the analysis sample as peers. Exposure is a student-level binary indicator equal to one if the student had a classmate with a severe disability in their first year of first grade, provided neither the student nor the CWD was repeating it.

CWD repeat grades more often than their classmates, particularly in the first years of school (see Figure A1). Exposure in later grades is therefore endogenous to the CWD's own characteristics. Similarly, a CWD who is repeating first grade was assigned to that classroom on the basis of prior performance, which is exactly the kind of selection our design is meant to avoid. Measuring exposure at the point of entry, before school performance is observed, avoids both problems, at the cost of making our estimand an ITT effect.

We use a binary indicator rather than a share. Disability is rare enough that most classrooms have none, and where a classroom has any, the share is low, and its variation is dominated by class size (see Figure A2). Nevertheless, we report specifications using the count and the share as robustness. We return to these choices in Section 5.

Measurement of disability relies on ministerial information available since 2016: we retroactively code students who appear as having permanent special needs as CWD. The same applies to students with a learning disability label. This can lead to conservative estimates if CWD leave the school before being observed in 2016, so that exposed students are classified as unexposed. We provide a robustness check that counts as exposed the classmates of children who attended a special school at any point between 2009 and 2019, whom we expect to be the most severe cases: consistent with this hypothesis, estimates increase by 7–9% in magnitude, remaining overall quantitatively similar given the small number of cases (Table D4). With respect to learning disabilities, the learning disability label could in principle respond to exposure itself (e.g., Elder, 2010; Dhuey and Lipscomb, 2010): the checks discussed in Section 7.2 show that exposure raises the probability of being labeled only slightly (Table E5 and Table E6), and this cannot plausibly account for our incidence results.

Measured exposure rises across the cohorts we study, from 52% of children in the 2009

cohort to 63% in the 2014 cohort. This is due to two interrelated causes. The first is a genuine increase in diagnosis. Holding grade fixed and comparing calendar years, the recorded prevalence of permanent needs among fourth graders rises from 2.8% in 2016 to 4.1% in 2023, and among first graders it rises by more than 80%. This mirrors a broad increase in awareness of autism and other neurodevelopmental conditions over the period.[14] The second is a feature of data construction: because the registers begin in 2016, later cohorts are observed in primary school for more years, so that their CWD are more likely to be recorded. All our specifications include cohort fixed effects, which absorb this common time trend.

Among exposed children, the classroom contained on average between 1.9 and 2.3 CWD classmates, as the PIE cap of two per classroom could be exceeded in exceptional cases, and the composition is dominated by intellectual disability, present in the classrooms of 81% of exposed children; autism is present for 29% and a visual disability for 20%, while motor, hearing, multiple and severe speech disorders each appear in fewer than 5%. Figure A3 shows how the composition evolves over the period. In the analysis sample, 15.6% of peers are ever recorded with a learning disability, and 44.6% were SEP priority students in their first-grade year. The average classroom has 33 students.

5 Empirical strategy

Our specification compares cohorts within the same school-section at the time of testing:

$$Score_{i,l,s,t} = \phi_{l,s} + \eta_t + \beta \, \text{Exp}_{i,l,s,t} + \boldsymbol{\delta_1} \boldsymbol{X}_{i,l,s,t} + \boldsymbol{\delta_2} \boldsymbol{Z}_{l,s,t} + \boldsymbol{\delta_3} \boldsymbol{\Omega}_{s,t} + \varepsilon_{i,l,s,t} \tag{1}$$

Subscript i denotes a peer, s a school, l a section (l stands for "letter"), and t the first-grade cohort. Chilean schools identify the parallel classes of each grade with a letter (A, B, and so on), and a school-section is the class identified by the same letter in the same school across cohorts, observed in the tested grade. $\text{Exp}_{i,l,s,t}$ indicates individual exposure in the first-grade class, measured as described in Section 4. The school-section fixed effects $\phi_{l,s}$ hold fixed the institutional aspects that determine the set of teachers a class can be assigned to and the

[14] This was also confirmed by the Chilean Ministry of Education.

consequent selection at the time of class group formation, while the cohort fixed effects η_t absorb common trends (e.g., nationwide policies or increase in autism diagnoses), so that identification comes from variation in exposure across cohorts within the same school-section (e.g., Hoxby, 2000).[15] The vectors X, Z and Ω hold individual, class and school covariates: gender, parental education and family income quintile; class size and the income composition of the class; school size and the parental education of the school (Appendix B). We estimate results for Mathematics and Spanish separately, in each observed grade and also pooling the tested grades with grade fixed effects, and cluster standard errors on the school-section in each cohort (i.e., the class group), which groups the children of a cohort observed in the same section.

Table B1 in the Appendix tests whether exposure is related to the composition of school-sections in each cohort. Without fixed effects, exposure is far more common in classes with fewer college-educated parents and fewer families in the top income quintile, and the joint test of all composition variables returns an F statistic above 2,400: this is the sorting that motivates a within-school design. Comparing cohorts within the same school-section, the joint test no longer rejects ($F = 1.34$, $p = 0.21$), and only the mother's vocational education, which all our specifications control for, remains individually significant. Comparing parallel classes within the same school and cohort, instead, the joint test rejects ($F = 3.17$, $p = 0.001$), mostly because exposed classes have fewer families in the top two income quintiles, consistent with the non-random allocation of children to parallel classes in some Chilean schools (Valenzuela Barros and Toledo Román, 2012). The estimate of Equation 1 is also stable across control sets: once individual covariates are included, adding class and school covariates leaves it essentially unchanged. Figure A4 illustrates the identifying variation.

Estimand Our estimand is the intention-to-treat (ITT) effect of assignment to a first-grade classroom with a CWD, regardless of subsequent exposure. CWD frequently repeat grades, especially at the beginning of the cycle: about five times as often as their peers in first grade, and more than twice as often in second grade (Figure A1). As a result, 84% of exposed children remained exposed in second grade and 66% in eighth grade, against

[15] Alternative first-grade fixed effects are estimated and discussed in Section 7.5 to deal with children who change school or section after first grade.

19% and 31% of unexposed children (Table B2). We do not instrument contemporaneous exposure with first-grade assignment, however, because the monotonicity condition fails in a way that is intrinsic to the setting. When school policies or the severity of disability cause CWD to repeat grades, peers become defiers that instrumental variables approaches cannot account for: initial peers assigned to exposure become unexposed, whereas next-year peers not assigned to exposure become exposed. In addition, class groups in our sample routinely undergo reshuffling due to transfers across sections and schools, regardless of the presence of CWD. Exposure barely affects mobility: exposed children are if anything slightly *less* likely to change school or to repeat a grade, and marginally more likely to change section, with differences in estimated magnitudes not economically significant (0.1 percentage points, see Table B4).

6 Results: average effects

Table 1 reports estimates of Equation 1. Exposure to a classmate with a severe disability in first grade lowers standardized Mathematics scores by 2.1% of a standard deviation and Spanish scores by 1.7%. Both estimates are precise, with standard errors of 0.2% of a standard deviation, and both are significant at the 1% level.

The penalty persists throughout primary school (Figure D1 in the Appendix). In Mathematics, the estimates by grade are 2.2% of a standard deviation in fourth grade, 2.6% in sixth and 2.6% in eighth grade. In Spanish, the effect attenuates over the first half of primary school, from 2.5% in second grade to 1.9% in fourth grade, and then stabilizes at 1.5% in both sixth and eighth grade. The contrast between subjects is consistent with the different role of cumulative content: Mathematics builds on earlier material, so that a deficit acquired early is carried forward, while in Spanish it is easier to catch up. In neither subject, however, does the gap close over the six years in which we observe the cohorts.

The ITT estimand requires two observations on how to interpret these results. First, assignment in first grade predicts later exposure less and less as cohorts progress: the difference in the probability of having a classmate with a severe disability between children exposed and not exposed in first grade falls from 65 percentage points in second grade to 35 points

Table 1: Effect of exposure on standardized test scores

	(1) Pooled	(2) 2nd grade	(3) 4th grade	(4) 6th grade	(5) 8th grade
Panel A. Mathematics					
Exposed in first grade	-0.021***		-0.022***	-0.026***	-0.026***
	(0.002)		(0.003)	(0.004)	(0.004)
Observations	1,527,433		849,761	406,334	271,318
Panel B. Spanish					
Exposed in first grade	-0.017***	-0.025***	-0.019***	-0.015***	-0.015***
	(0.002)	(0.004)	(0.003)	(0.004)	(0.004)
Observations	2,056,979	538,192	845,599	404,029	269,136

Notes: Estimates of Equation 1 on peers, cohorts 2009 to 2014. Coefficients are in standard deviations of the test score distribution. *Exposed in first grade* equals one if the student had a non-repeating classmate with a severe disability in their first year of first grade. All specifications include school-section and cohort fixed effects, school size, school shares of parental education, class size, class shares of family income quintiles, and indicators for being female and for own parental education and income quintiles. Column 1 (pooled estimates) adds grade fixed effects. Mathematics is not tested in second grade. Standard errors clustered on the school-section in each cohort (i.e., the class group). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

in eighth grade (Table B2 in the Appendix). The ITT estimates in later grades therefore compare groups whose current exposure differs much less than in the early grades, so that part of the attenuation in Spanish may reflect this convergence, while the stable penalty in Mathematics is consistent with a persistent effect of early exposure rather than with current exposure alone.

As discussed in Section 5, exposure does not significantly increase the probability of leaving one's class; the ITT estimates are nonetheless further diluted by class group reshuffles that alter contemporaneous exposure regardless of first year assignment to CWD. 25.1% of un-exposed children have a non-repeating CWD in the tested grade. Excluding them from the comparison group roughly doubles the estimated penalty, from 2.1 to 4.1% of a standard deviation in Mathematics and from 1.7 to 3.3% in Spanish (Table B3), which we report as a bounding exercise rather than a preferred estimate, since which classrooms later receive a CWD is not random.

The penalty grows with the intensity of exposure. Table 2 replaces the exposure indicator with the number or the share of CWD in first grade, together with its square. Each additional CWD lowers Mathematics scores by 1.4% of a standard deviation and Spanish scores by 1.2%, and, in the count specification, small positive quadratic terms indicate a declining marginal

cost. In shares, moving from no exposure to a class in which 10% of children have a severe disability corresponds to a loss of about 3.9% of a standard deviation in Mathematics and 3.3% in Spanish. These estimates are only indicative of the relation between intensity and the penalty, however. About 40% of first-grade classes have no CWD, and most exposed classes have between one and three, so that the distributions of the number and of the share have a large spike at zero and much of their variation comes from the comparison between no exposure and some exposure (Figure A2 in the Appendix).

Table 2: Intensity of exposure

	Mathematics		Spanish	
	(1)	(2)	(3)	(4)
Exposure measured as	Number	Share	Number	Share
Children with a severe disability	-0.014***	-0.392***	-0.012***	-0.315***
	(0.002)	(0.040)	(0.002)	(0.039)
Squared	0.001***	-0.017	0.001*	-0.163
	(0.000)	(0.163)	(0.000)	(0.164)
Observations	1,527,433	1,527,433	2,056,979	2,056,979

Notes: Columns 1 and 3 replace the exposure indicator with the number of non-repeating classmates with a severe disability in the first-grade class and its square; columns 2 and 4 use the corresponding share of the class and its square. All grades pooled, with grade fixed effects. Other controls, fixed effects and clustering as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The size of these effects matters for the policy debate, because the inclusion of CWD is contested in a way that other sources of classroom heterogeneity are not. Schools routinely enroll children with learning disabilities, behavioral problems, and difficult home circumstances: while these may enter parents' choices of schools and classes, they are not contested in the public debate to the same degree.

Two studies provide benchmarks. Balestra et al. (2022) study peers with special needs in Switzerland, defined by repeated contact with the school psychological service: the category covers learning impairments and behavioral problems, and excludes children with severe disabilities, who are placed in separate schools in the Swiss context. Carrell and Hoekstra (2010) study children from families with a record of domestic violence, who also raise their classmates' disciplinary infractions by 17%. In both studies, adding one such child to a class of twenty lowers classmates' achievement by about 2.5% of a standard deviation.

Both benchmarks express treatment as a share of the class and report the effect of one

additional child in a class of twenty. Evaluating our specification in shares at the same point, one additional CWD lowers test scores by 2.0% of a standard deviation in Mathematics and 1.6% in Spanish,[16] about three quarters of the benchmarks. This comparison understates our effects, because both benchmarks measure classroom composition in the year in which the outcome is observed, whereas our estimates are diluted by the later exposure of the comparison group discussed above. Adjusting for this difference, one additional child in a class of twenty lowers achievement by 3.8% of a standard deviation in Mathematics and 3.1% in Spanish, up to one and a half times the benchmarks.[17] Like the estimated dilution bound on which it is based, this number is not a causal estimate. Moreover, classmates with learning disabilities, behavioral problems or disadvantaged backgrounds are more numerous and less likely to repeat grades than CWD, which makes the dynamics of later exposure, and therefore the estimates, not directly comparable. Overall, exposure to a classmate with a severe disability produces spillovers of the same order as both benchmarks, between about two thirds and one and a half times as large depending on how estimation differences are handled.

Robustness The estimates are robust to clustering standard errors at the school level (Table D1), thus allowing wider spillovers; to randomization inference, which yields p-values below 0.001 (Young, 2019); to controlling for the share of SEP priority students in the school-section in each cohort and for the student's own priority status; to retaining students with missing family background (Table D2); to excluding private schools; and to counting children who attended special schools as CWD (Table D4). Exposure does not affect the probability of taking part in the standardized tests (Table D3). Appendix D discusses these checks in detail.

An alternative design, reported for completeness, compares parallel classes within the same

[16] One additional child in a class of twenty raises the share by $1/20 = 0.05$. With coefficients of -0.392 on the share and -0.017 on its square in Mathematics, the implied effect is $-0.392 \times 0.05 - 0.017 \times 0.05^2 \approx -0.020$; in Spanish, with coefficients of -0.315 and -0.163, it is $-0.315 \times 0.05 - 0.163 \times 0.05^2 \approx -0.016$.

[17] Excluding from the comparison group the children who have a classmate with a severe disability in the tested grade raises the estimate of Equation 1 from 2.1% to 4.1% of a standard deviation in Mathematics, by a factor of $4.1/2.1 \approx 1.95$, and from 1.7% to 3.3% in Spanish, by a factor of $3.3/1.7 \approx 1.94$ (Table B3). We assume that dilution reduces the coefficients of the specification in shares by the same factor, and multiply them by it. In Mathematics, the effect of one additional child in a class of twenty becomes $1.95 \times (-0.392 \times 0.05 - 0.017 \times 0.05^2) \approx -0.038$; in Spanish, $1.94 \times (-0.315 \times 0.05 - 0.163 \times 0.05^2) \approx -0.031$.

school and cohort, replacing the school-section and cohort fixed effects with school-by-cohort fixed effects (Equation 2 in the Appendix). By construction, this comparison uses only schools with more than one section in the cohort, which account for 72% of our sample and are typically larger: their median size is 605 students, against 254 in schools with a single section. It yields slightly larger estimates (Table C1), consistent with selected class composition based on the inclusion of a CWD: balance holds less well across parallel classes than across cohorts, as exposed classes have fewer families in the top two income quintiles (Table B1, column 5).

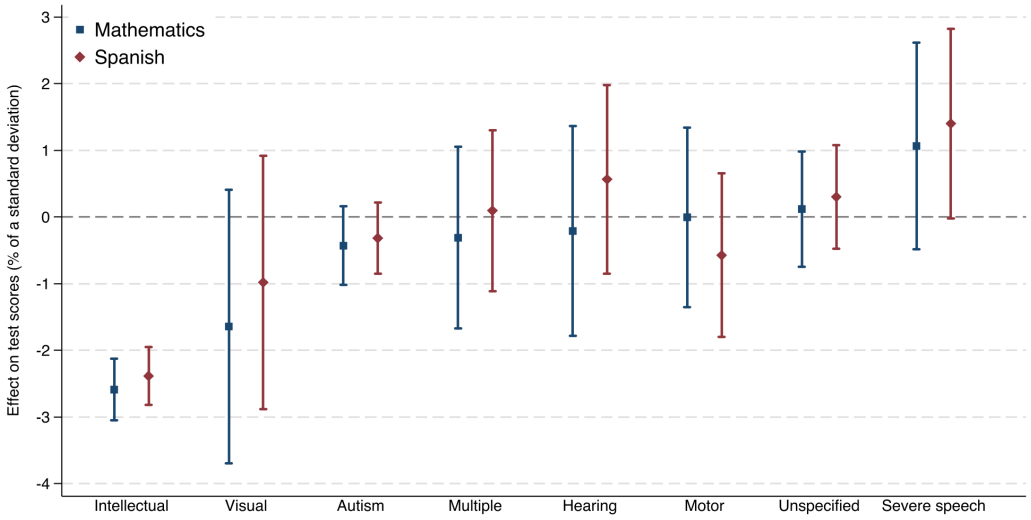
7 Heterogeneity and mechanisms

7.1 The source of spillovers: disability type and the mix of disabilities

We run two tests to identify which types of disability, and which mix of types, explain our main results. First, we test heterogeneous effects of exposure by type of disability (all grades pooled together shown in Figure 1; estimates by grade are in Table E1 in the Appendix). Exposure to a classmate with an intellectual disability lowers Mathematics scores by 2.6% of a standard deviation and Spanish scores by 2.4%, both significant at the 1% level, while all other categories return coefficients close to zero. The only exception is severe speech disorders, whose positive point estimates, 1.1% in Mathematics and 1.4% in Spanish, may reflect spillovers from language-specific support; however, only the Spanish estimate is significant, at the 10% level; it does not survive the adjustment for multiple testing (Table E4 in the Appendix), and it rests on the fewer than 5% of exposed children who have a classmate with a severe speech disorder. Intellectual disability therefore appears to drive the whole negative effect. To confirm this, we add an intellectual disability dummy to the main specification: the residual effect of exposure to the other types falls to 0.4% of a standard deviation in Mathematics and to zero in Spanish, neither significant (Table E2 in the Appendix).

Second, we test whether the mix of disabilities matters, as it would if the burden came from coordinating different kinds of support. We add to Equation 1 one minus the Herfindahl index of the types present in the class, set to zero in unexposed classes, so that it is equivalent to

Figure 1: Effect of exposure by type of disability



Notes: Pooled estimates across all tested grades and 95 percent confidence intervals, in percent of a standard deviation, from a version of Equation 1 in which the exposure indicator is replaced by binary indicators for the presence of each type of severe disability in the first-grade class. The categories are those of the special-needs register and are not mutually exclusive, since a child may carry more than one diagnosis. Estimates by grade are in Table E1 in the Appendix. Standard errors clustered on the school-section in each cohort (i.e., the class group).

an interaction term with exposure. Most exposed classrooms are homogeneous, with 76% containing a single type, and a mixed class requires at least two CWD, so we also control for their number and for the number with an intellectual disability, with their squares. Diversity does not matter (Table 3): its coefficient is −0.5% of a standard deviation in both Mathematics and Spanish, with standard errors of 0.6.[18] It turns positive only when the number of CWD is held fixed but the number with an intellectual disability is not, because 97% of homogeneous classes with two or more CWD include only children with intellectual disabilities. Once both counts are held fixed, it returns to zero, whereas each additional child with an intellectual disability lowers scores by a further 1.1% of a standard deviation in Mathematics and 1.2% in Spanish.

One type of disability, intellectual disability, is therefore responsible for the average effect, and combining different types adds no cost of its own. We interpret this as evidence that,

[18]The null also holds for a version that merges visual and hearing disabilities (Table E7 in the Appendix) and a version that counts each child with several diagnoses once (Table E8). Indicators for the number of types, which do not hold the number of CWD fixed, are small and negative for two types, at 0.6% of a standard deviation in Mathematics and 0.9% in Spanish (significant at the 10% and 1% levels), and insignificant for three or more types.

Table 3: Diversity of disability types in the first-grade class

	Mathematics			Spanish		
	(1)	(2)	(3)	(4)	(5)	(6)
Exposed in first grade	-0.021***	0.003	-0.002	-0.017***	0.006	0.001
	(0.002)	(0.004)	(0.005)	(0.002)	(0.004)	(0.004)
Diversity (1 − HHI)	-0.005	0.019***	-0.013	-0.005	0.021***	-0.006
	(0.006)	(0.007)	(0.009)	(0.006)	(0.006)	(0.008)
Number of children with a severe disability		-0.018***	-0.005		-0.017***	-0.004
		(0.003)	(0.005)		(0.003)	(0.004)
Squared		0.0013***	0.0018**		0.0011***	0.0010
		(0.0005)	(0.0007)		(0.0004)	(0.0006)
Number with an intellectual disability			-0.011***			-0.012***
			(0.003)			(0.003)
Squared			-0.0010			-0.0002
			(0.0007)			(0.0006)
Observations	1,527,433	1,527,433	1,527,433	2,056,979	2,056,979	2,056,979

Notes: *Diversity* is one minus a Herfindahl index computed over the shares of the seven disability categories among classmates with severe disabilities in the first-grade class; it equals zero when all diagnoses belong to one category and in unexposed classes. Columns 1 and 4 add the index to Equation 1; columns 2 and 5 also control for the number of classmates with a severe disability and its square; columns 3 and 6 also control for the number with an intellectual disability and its square. Controls, fixed effects and clustering as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

while support may already moderate the spillovers of other disabilities, there is room for intervention for intellectual disabilities, the most demanding in terms of the teacher's time and attention. The supplied specialist support does not seem to fill this gap: exposure brings no additional specialist teachers or hours to the grades in which the cohorts are tested, the penalty is no smaller where specialists are more abundant, and intellectual disability is associated with the lowest increase in schools' support expenditures (Appendix H).

7.2 The spillover recipients: the unequal effects of severe disability exposure

The average penalty is not evenly spread across classmates. Table 4 reports unconditional quantile effects (Firpo et al., 2009), estimated with the specification of Equation 1. In Mathematics, exposure lowers the 10th percentile of the score distribution by 3.4% of a standard deviation, the median by 2.1% and the 90th percentile by 0.9%. In Spanish, it lowers the quantiles between the 10th and the 75th percentile by 1.7% to 2.3% of a standard deviation, and the 90th percentile by 0.8%. The estimated shifts are larger at the lower and middle quantiles than at the top of the score distribution.

If exposure operates through competition for teacher attention, its cost should depend on the

Table 4: Unconditional quantile effects

	10th	25th	50th	75th	90th
Mathematics	-0.034***	-0.025***	-0.021***	-0.019***	-0.009**
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Spanish	-0.018***	-0.023***	-0.021***	-0.017***	-0.008***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)

Notes: Effects of exposure on the recentered influence function of each quantile of the pooled score distribution (Firpo et al., 2009), with the specification of Equation 1. The density at each quantile is estimated with a Gaussian kernel and Silverman's bandwidth. Standard errors clustered on the school-section in each cohort (i.e., the class group). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

classmates' own capacity to learn without that attention. Children with learning disabilities, taught in mainstream classrooms well before the inclusion of severe disabilities and among the most affected by classmates with special needs in other settings (Balestra et al., 2022), are the group most likely to bear it. Table 5 interacts exposure with the peer's own learning disability status, fully interacting all controls.

Table 5: Incidence by the peer's own learning disability status

	(1) Pooled	(2) 2nd grade	(3) 4th grade	(4) 6th grade	(5) 8th grade
Panel A. Mathematics					
Exposed in first grade	-0.014***		-0.014***	-0.020***	-0.018***
	(0.002)		(0.003)	(0.004)	(0.004)
Learning disability	-0.676***		-0.722***	-0.679***	-0.532***
	(0.014)		(0.019)	(0.024)	(0.027)
Exposed in first grade × Learn. disab.=1	-0.023***		-0.034***	-0.008	-0.019**
	(0.005)		(0.006)	(0.008)	(0.009)
Observations	1,527,433		849,761	406,334	271,318
Panel B. Spanish					
Exposed in first grade	-0.009***	-0.015***	-0.011***	-0.007*	-0.007
	(0.002)	(0.004)	(0.003)	(0.004)	(0.004)
Learning disability	-0.629***	-0.628***	-0.651***	-0.648***	-0.523***
	(0.014)	(0.024)	(0.019)	(0.028)	(0.031)
Exposed in first grade × Learn. disab.=1	-0.033***	-0.048***	-0.034***	-0.026***	-0.021**
	(0.005)	(0.009)	(0.006)	(0.009)	(0.010)
Observations	2,056,979	538,192	845,599	404,029	269,136

Notes: *Learning disability* equals one for peers ever recorded with a learning disability. Learning disability corresponds to the Chilean transitory special educational needs label, which entitles children to temporary support. The total effect for such a peer is the sum of the exposure coefficient and the interaction. All listed controls are interacted with *Learning disability*. Sample, fixed effects and clustering as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Both terms are negative. Peers never labeled lose 1.4% of a standard deviation in Mathematics and 0.9% in Spanish, less than the spillovers generated by classmates with learning

and behavioral difficulties (Balestra et al., 2022) or from households with domestic violence (Carrell and Hoekstra, 2010), but significant at the 1% level. For labeled peers, the interaction adds 2.3% and 3.3%, for a total of 3.7% of a standard deviation in Mathematics and 4.2% in Spanish, about 2.6 and 4.5 times as much as other classmates.[19] The interaction is negative and significant in every tested grade in Spanish and in fourth and eighth grade in Mathematics.

This is inconsistent with a uniform dilution of attention: if exposure removed the same share of attention from every child, labeled children would not lose several times as much.

Because the registers begin in 2016, the learning disability label is recorded after first grade, and diagnoses of learning and attention disorders partly depend on how a child compares with classmates (Elder, 2010; Dhuey and Lipscomb, 2010). The label could therefore respond to exposure, and it does, slightly: exposure raises the probability of ever being labeled by 0.4 percentage points, against a control mean of 9.4%, or by 0.6 points with fixed effects only (Table E5 in the Appendix). Two mechanisms could link exposure to labels, and both would operate shortly after exposure: the specialized staff that the inclusion program brings to exposed classrooms may also identify other children's learning difficulties, and schools may place CWD and children with learning disabilities in the same classrooms to exploit a joint allocation of resources. For the 2013 and 2014 cohorts, the registers begin in fourth and third grade, which allows us to distinguish children already labeled by the first standardized tests from children first labeled later (Table E6 in the Appendix). In line with our expectation, exposure is associated with more early labels, by 0.35 percentage points against a base of 3.4%, and not with more late labels. If the labeling channel drove our results, the additional losses of labeled peers should concentrate among the early labeled. Instead, we find the opposite: the effect is small and insignificant for children labeled early, and largest for children labeled later, at 3.0% of a standard deviation in Mathematics and 3.8% in Spanish.

The two natural objections to this analysis concern compositional effects. Learning disabilities recognized early are likely the more visible and possibly the more severe ones, yet the

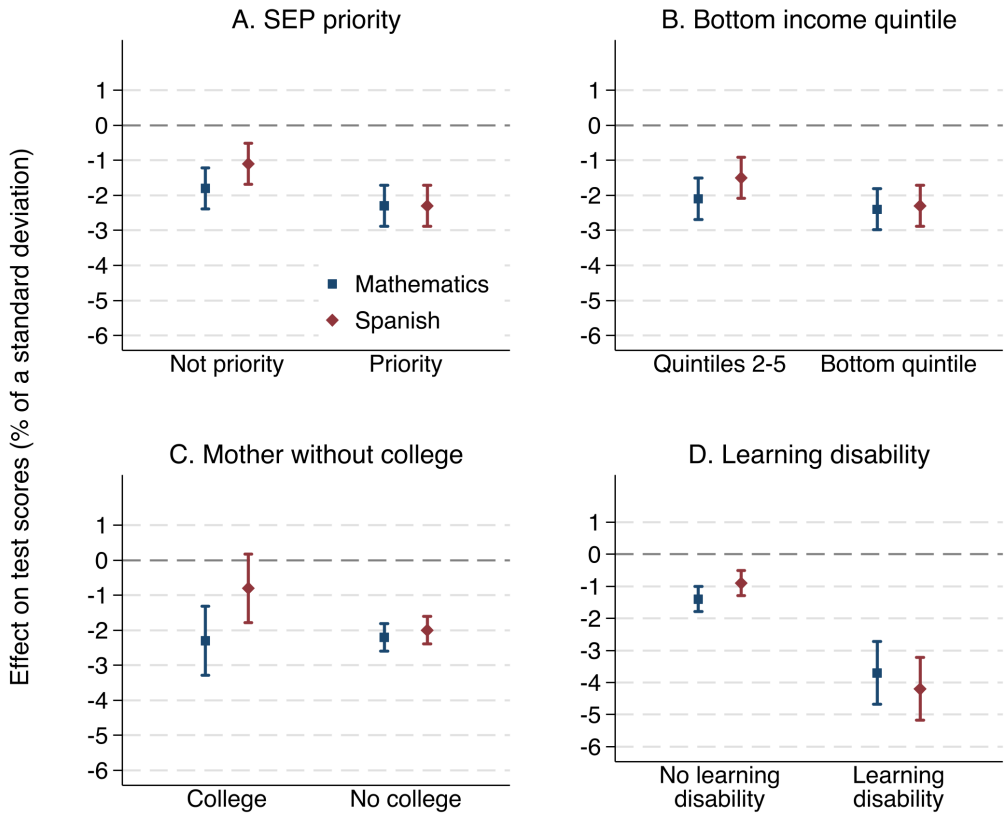
[19] Ratios computed from the unrounded estimates, $-0.0373/-0.0142$ and $-0.0424/-0.0095$.

losses concentrate on children recognized later, so that what distinguishes the most affected children is when their needs are recognized and supported rather than their severity. As for joint placement, at class formation only the children recognized early are visible, so that any deliberate joint placement would impact the early labeled: we observe instead that negative spillovers are concentrated on those labeled later. Finally, exposure does not increase late labels, as it would if the spillovers made children fall behind and led teachers to suspect a learning disability (Table E6 in the Appendix). The additional labels do not change the socioeconomic composition of labeled children: the background gap between exposed and unexposed children is very similar among labeled and unlabeled children, with differences below one percentage point that show no consistent shift towards disadvantage (Table E5). Rather than indicating compositional effects, therefore, the evidence is consistent with competing needs: the children who lose most are those whose difficulties went unrecognized, and who therefore relied on the main teacher's attention in the first years of school.

Another concern is whether learning disability labels capture other types of socioeconomic disadvantage. Labeled children are indeed more often poor: 19.8% of SEP priority students are labeled, against 11.1% of other students. Figure 2 compares the additional exposure effect of learning disability labels with the additional effect of three markers of disadvantage, each with all controls fully interacted. Socioeconomic disadvantage matters, but much less. SEP priority students lose 2.3% of a standard deviation in both subjects, against 1.8% and 1.1% for other students, 1.3 times as much in Mathematics and about twice as much in Spanish. Children in the bottom income quintile lose 1.15 times as much as others in Mathematics and 1.5 times as much in Spanish, and children whose mother did not attend college lose no more in Mathematics and more than twice as much in Spanish. All these gradients are much flatter than the gradient by label, of 2.6 in Mathematics and 4.5 in Spanish. The concentration of losses therefore falls on students with a learning disability label, which proxies greater reliance on teacher support, rather than socioeconomic disadvantage.

The larger estimated losses among labeled peers and the larger shifts at lower score quantiles are both consistent with unequal vulnerability to exposure. The common implication concerns the way classes are formed. Rather than concentrating different special needs in the

Figure 2: Effect of exposure by the peer's own status: learning disabilities and socioeconomic disadvantage



Notes: Effect of exposure, in percent of a standard deviation, with 95 percent confidence intervals, for peers outside and inside the group in each panel, from a single regression in which exposure and all controls are interacted with the group indicator. SEP priority is measured in the first-grade year. Learning disability corresponds to the Chilean transitory special educational needs label, which entitles children to temporary support. All grades pooled with grade fixed effects. Sample, fixed effects and clustering as in Table 1. Estimates are in Table E3 in the Appendix.

same classrooms, schools could distribute different sources of disadvantage across parallel classes, so that CWD, and intellectual disabilities in particular, share their classroom with peers who are better able to learn without additional attention. This would amount to a positive selection into exposure, the opposite of what we currently observe: CWD are concentrated in schools with more disadvantaged children (columns 1 and 2 of Table B1), and the additional early labels among exposed children are consistent with some schools placing CWD and children with learning disabilities, suspected or known, in the same classrooms.

7.3 The classroom environment

The concentration of losses on children with learning disabilities could also arise if CWD disrupted the classroom and some classmates were more vulnerable to disruption than others. We therefore turn to the classroom environment to assess whether disruption is an alternative mechanism, distinguishing three questions: whether exposed classrooms have a worse climate because of selection between schools, whether exposure itself is associated with a worse climate, and whether any such disruption falls on the classmates who bear the achievement cost.

Alongside the tests, students, parents and teachers answer survey modules, and participation is strongly encouraged, if not *de facto* compulsory, in most schools. Because questions vary across years, the surveys do not allow a student-level panel. We aggregate similar questions into three classroom-level indices: behavioral climate, from student and teacher reports on respect and disruption in the class; bullying, from reports on aggression among students in the school; and openness to diversity, from reports on discrimination in the school (items in Appendix tables G2 to G4). These indices provide only descriptive evidence: they aggregate the responses of everyone in the classroom or school, including repeaters and CWD, who are excluded from our analysis sample.

We estimate the associations at the level at which the indices are built, the classroom by year, with the share of the class exposed in first grade as the regressor. Table 6 shows the results with fixed effects added progressively. Without them, a fully exposed classroom scores 0.41 standard deviations worse on behavioral climate and 0.98 standard deviations higher on bullying than a fully unexposed one. Almost all of this difference reflects selection: class-level controls cut the associations by more than half, and school fixed effects remove most of what remains, consistent with the concentration of CWD in disadvantaged schools that motivates our within-school design. Descriptive comparisons of exposed and unexposed classrooms, common in studies based on small convenience samples, are therefore likely to be biased.

Within school-sections, a small association survives, in the direction predicted by the congestion of teacher attention: in the specification matching Equation 1, exposed classrooms

Table 6: Classroom environment indices and exposure. Class-level associations

	(1) Year, grade	(2) + controls	(3) + school FE	(4) Eq. (1) FE
Panel A. Behavioral climate				
Share exposed in 1st grade	-0.411***	-0.173***	-0.076***	-0.069***
	(0.024)	(0.026)	(0.019)	(0.021)
Observations	33,753	33,411	33,319	32,605
Panel B. Bullying				
Share exposed in 1st grade	0.983***	0.331***	0.069***	0.069***
	(0.029)	(0.025)	(0.017)	(0.018)
Observations	33,753	33,411	33,319	32,605
Panel C. Openness to diversity				
Share exposed in 1st grade	-0.337***	-0.074***	0.017	0.023
	(0.019)	(0.021)	(0.021)	(0.022)
Observations	33,731	33,403	33,311	32,597

Notes: Unit of observation is the classroom by year, grades 4, 6 and 8, which is the level at which the indices are constructed. The regressor is the share of the class exposed to a classmate with a severe disability in first grade. Indices are standardized. Column 2 adds class size, class income composition and class parental education. Column 3 adds school fixed effects. Column 4 uses school-section and cohort fixed effects, matching Equation 1. Standard errors clustered on the school. These associations are descriptive: the indices aggregate the responses of all children present, including repeaters and CWD, who are excluded from the achievement analysis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

score -0.07 standard deviations on behavioral climate and 0.07 on bullying, both significant. The association with openness to diversity is small and insignificant, so that reported discrimination is no higher where CWD are present.

The worse climate, however, is not concentrated on the children who bear the cost in achievement. The association between exposure and reported bullying is no larger in the classrooms of children with learning disabilities, nor in classrooms with a high share of such children (Table G1 in the Appendix). Disruption is shared by all classmates, whereas the loss of achievement is concentrated on children with learning disabilities, so that disruption alone cannot explain the incidence of the penalty. The pattern is instead consistent with a common shock producing unequal losses, because children differ in their capacity to learn without the teacher's attention.

7.4 Teacher continuity

If the mechanism is competition for teacher attention, arrangements that make teaching a class easier should be associated with smaller penalties. About half of the students in our

sample have the same teacher as in the previous year. Table 7 reports the effect of exposure separately for students who change teacher and for students who keep their teacher, with all controls fully interacted with continuity. Exposure lowers the scores of students who change teacher by 3.3% of a standard deviation in Mathematics and 2.5% in Spanish, and those of students who keep their teacher by 0.9% and 1.1%: under continuity, the penalty is 72% smaller in Mathematics and 57% smaller in Spanish. The pattern holds in every tested grade, although in Spanish the effect among students who keep their teacher is no longer significant in sixth and eighth grade.

Table 7: Teacher continuity

	(1) Pooled	(2) 2nd grade	(3) 4th grade	(4) 6th grade	(5) 8th grade
Panel A. Mathematics					
Exposed, different teacher	-0.033***		-0.033***	-0.034***	-0.036***
	(0.003)		(0.004)	(0.005)	(0.005)
Exposed, same teacher	-0.009***		-0.012***	-0.016***	-0.015***
	(0.003)		(0.004)	(0.005)	(0.005)
Observations	1,525,785		848,605	406,091	271,068
Panel B. Spanish					
Exposed, different teacher	-0.025***	-0.032***	-0.027***	-0.021***	-0.020***
	(0.003)	(0.005)	(0.004)	(0.005)	(0.006)
Exposed, same teacher	-0.011***	-0.021***	-0.012***	-0.008	-0.009
	(0.003)	(0.005)	(0.004)	(0.006)	(0.006)
Observations	2,053,024	537,026	844,306	403,741	267,927

Notes: Effect of exposure separately for students taught by a different teacher than in the year before the test and for students taught by the same teacher, from a single regression in which exposure and all controls are interacted with the same-teacher indicator. Sample, fixed effects and clustering as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Importantly, exposure does not significantly predict whether students retain their teachers. If teachers were more likely to stay with classes that are easier to teach, or less likely to stay with exposed classes, the comparison above would capture that selection. Regressing an indicator for keeping the previous year's teacher on exposure, with the fixed effects and controls of Equation 1, yields a coefficient of zero for the Mathematics teacher and of 0.3 percentage points for the Spanish teacher, against a base rate of about 48%, neither significant and with confidence intervals that rule out effects larger than one percentage point (Table F1 in

the Appendix). The positive raw association between exposure and continuity, which arises because schools enrolling CWD more often keep teachers with their classes, vanishes once school fixed effects are included (Table F2 in the Appendix).

Comparing schools that implement teacher looping (the practice of keeping the same teacher over multiple grades) as a school-wide policy with schools that do not points in the same direction, with penalties about 40% smaller in Mathematics in looping schools (Appendix F.1). Because looping schools are negatively selected in terms of parental socioeconomic background and schooling outcomes, we treat this comparison as a robustness check, noting that it may understate the role of teacher continuity.

We also examine the experience of the first-grade teacher, with all controls fully interacted and the fixed effects of Equation 1 (Table F5 in the Appendix). General experience is only weakly associated with the penalty: at the estimated gradients, the penalty would vanish only after 46 years of experience in Mathematics and 51 in Spanish, more than a teaching career.[20] Experience with CWD, measured over the years we observe, is associated with much smaller penalties, by about 1% of a standard deviation per year of such experience, although teachers with this experience are not randomly assigned to classes. What appears to matter is not how long a teacher has taught, but whether the teacher knows the class or has experience with the needs of CWD.

7.5 Movers and a smaller average penalty

Our preferred specification includes school-section fixed effects that hold fixed time-invariant characteristics, including school characteristics, the potential set of teachers, and section-specific policies that can determine selection across sections at the time of class formation, including support measures for CWD. What this choice leaves out is a direct analysis of the contribution of children who change section or school after first grade, 34% of observations (17% in second grade and 56% in eighth grade), who are compared with the other children of their destination section and cohort. Their first-grade schools are typically worse (Table I3 in the Appendix), and in our main specification this is only partially addressed by the

[20] When the same teacher takes first grade in the same school-section in consecutive cohorts, as in 14% of transitions, experience rises mechanically with the cohort. Allowing the effect of exposure to vary freely by cohort, or excluding these school-sections, leaves the interaction essentially unchanged.

inclusion of individual controls.

While their own exposure treatment is based on their first-grade class, and even though mobility is not systematically correlated with first-grade exposure (Table B4, discussed in Section 5), movers may carry disadvantages from their origin schools into the estimates. In other words, a child moving from a worse to a better school may score low relative to the new section while also being more likely to have been exposed in first grade.

Movers belong in the intention-to-treat analysis because the outcome of interest follows initially exposed children regardless of subsequent mobility: mobility would occur regardless of inclusion policies, so that including movers makes the estimates policy-relevant. However, understanding their contribution matters for the mechanisms. To do so, we replace the school-section fixed effects with fixed effects for the first-grade school-section, so that each child is compared with the children who started in the same first-grade section in other cohorts, regardless of the conditions experienced in subsequent grades (Appendix I).[21]

Under this alternative specification the penalty is smaller, at 1.4% of a standard deviation in Mathematics and 1.3% in Spanish, and it fades over primary school, to 0.8% in Mathematics and zero in Spanish by eighth grade. The heterogeneity results are unchanged: intellectual disability drives the effect, the mix of disabilities is irrelevant, and the cost falls mostly on classmates with learning disabilities, and spillovers are larger in the lower portion of the score distribution.[22]

Teacher continuity is a particularly important check to run, because changing school or section also implies changing teachers: restricting the main specification to children who remain in their first-grade section, continuity is associated with reductions in the penalty of about 60% in Mathematics and 35% in Spanish, somewhat smaller than the 72% and 57% in the full sample. Continuity is mechanically less frequent among movers, but changing teacher is part of the experience of exposed children who move, and therefore of the intention-to-treat

[21] We estimate the two specifications separately rather than adding both sets of fixed effects, because they answer different questions and because, in a joint specification, the effects of the current section would be identified only from movers, whose destination may itself respond to exposure or to their own performance.

[22] With first-grade fixed effects, penalties are no longer smaller in looping schools, and the association with the first-grade teacher's experience with CWD is no longer significant. However, teacher experience and looping practices are school-level policies that are more relevant to analyze at the destination school for movers, that is, under the main specification.

effect. Among children who have remained in their first-grade school-section since exposure, the average penalty falls to 1.6% of a standard deviation in Mathematics and 1.5% in Spanish (Table I3), suggesting that the direct effects of disability may be even smaller relative to the benchmark spillovers discussed in Section 6.

8 Conclusion

This paper asks whether including CWD in mainstream classrooms harms the learning of their classmates, and who bears the cost. Comparing cohorts within the same school-section in Chile, where a national program brought CWD into regular classrooms, we find that first-grade exposure lowers classmates' test scores by 2.1% of a standard deviation in Mathematics and 1.7% in Spanish. These effects are of the same order as other spillovers studied in the literature that schools already absorb from classmates with learning and behavioral difficulties (Balestra et al., 2022) or from households with domestic violence (Carrell and Hoekstra, 2010). They are not trivial, since comparable exposures in elementary school have been shown to lower adult earnings (Carrell et al., 2018), but they provide little support for treating severe-disability inclusion as uniquely costly to peers. Using alternative first-grade school-section fixed effects yields smaller average penalties, reinforcing the conclusion that average peer achievement costs are modest.

The average, however, conceals a pattern that points to competition for a fixed amount of teacher attention, a congestible input whose allocation adapts to the composition of the class (Lazear, 2001; Duflo et al., 2011). The penalty comes almost entirely from intellectual disabilities, the category that most requires continuous instructional support, while combining different types of disability adds no cost of its own. The losses concentrate on children with learning disabilities, who rely more on adult attention to keep pace and whose support is only temporary, a status that matters more than socioeconomic disadvantage; within this group, they concentrate on those whose difficulties are recognized late, and therefore went without dedicated support in the first years of school. Exposed classrooms are slightly more disorderly, but the disruption is shared by all classmates, so that it cannot explain why the losses fall on some of them. Finally, the penalty is significantly smaller when the teacher already knows the class or has experience with CWD, whereas general teaching experience

barely matters.

For policy, the results point to organizational challenges that may be mitigated by how support is provided, rather than a categorical burden. The absence of spillovers from other severe disabilities may reflect specialized support that works and that we do not observe, so that intellectual disabilities appear to be the only area in which additional mitigation is needed. Regulations cap the number of CWD and of children with learning disabilities in each classroom separately, but not their combination, although the losses concentrate on children with learning disabilities who share the classroom with CWD. We also show evidence that teacher continuity, which requires neither additional staff nor specialized training and is already recommended on general grounds, may be a promising organizational intervention to sizeably reduce the negative spillovers driven by intellectual disabilities.

Our evidence has limits that are left to future research: in particular, we cannot observe support expenditures and staffing at the classroom level, and therefore cannot estimate their mitigating role directly; due to data constraints, our outcomes do not include the attitudes and behaviors that exposure may also change, potentially in a positive direction (Rao, 2019), which might also widen the analysis to the trade-offs between losses in cognitive outcomes and gains in non-cognitive ones. However, our novel focus on the inclusion of severe disabilities highlights important results for the debates on these policies: the relevant question is whether schools are organized so that the cost of including them falls on the institution rather than on the classmates least able to absorb it.

References

Andersen, S. C., Beuchert, L., Nielsen, H. S., and Thomsen, M. K. (2020). The effect of teacher's aides in the classroom: Evidence from a randomized trial. *Journal of the European Economic Association*, 18(1):469–505.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training projects. *Journal of the American Statistical Association*, 103(484):1481–1495.

Anelli, M. and Peri, G. (2019). The effects of high school peers' gender on college major, college performance and income. *The Economic Journal*, 129(618):553–602.

Anelli, M., Shih, K., and Williams, K. (2023). Foreign students in college and the supply of STEM graduates. *Journal of Labor Economics*, 41(2):511–563.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30:98–108.

ANSA (2024). Vannacci, "classi separate per i disabili, aiuterebbe i ragazzi". ANSA, 27 April, https://www.ansa.it/sito/notizie/politica/2024/04/27/vannacci-classi-separate-per-i-disabili-aiuterebbe-i-ragazzi_30ecb4c4-3046-4c49-9c3c-51d4af82374c.html.

Balestra, S., Eugster, B., and Liebert, H. (2022). Peers with special needs: Effects and policies. *Review of Economics and Statistics*, 104(3):602–618.

Ballis, B. and Heath, K. (2021a). Direct and spillover effects of limiting minority student access to special education. EdWorkingPaper 21-364, Annenberg Institute at Brown University.

Ballis, B. and Heath, K. (2021b). The long-run impacts of special education. *American Economic Journal: Economic Policy*, 13(4):72–111.

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.

Bifulco, R., Fletcher, J. M., and Ross, S. L. (2011). The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy*, 3(1):25–53.

Bloom, N., Dahl, G. B., and Rooth, D.-O. (2026). Work from home and disability employment. *American Economic Review: Insights*, 8(2):179–195.

Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M., and Eccles, J. (2006). Empathy or antipathy? The impact of diversity. *American Economic Review*, 96(5):1890–1905.

Caron, L. (2026). The short- and long-term impacts of expanding public education for disabled students. Working Paper 2604.25767, arXiv.

Carrasco, A., Gutiérrez, G., and Flores, C. (2017). Failed regulations and school composition: Selective admission practices in Chilean primary schools. *Journal of Education Policy*, 32(5):642–672.

Carrell, S. E. and Hoekstra, M. (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, S. E., Hoekstra, M., and Kuka, E. (2018). The long-run effects of disruptive peers. *American Economic Review*, 108(11):3377–3415.

Contreras, D., Brante, M., Espinoza, S., and Zúñiga, I. (2020). The effect of the integration of students with special educational needs: Evidence from Chile. *International Journal of Educational Development*, 74:102163.

Contreras, D. and Rau, T. (2012). Tournament incentives for teachers: Evidence from a scaled-up intervention in Chile. *Economic Development and Cultural Change*, 61(1):219–246.

Dahl, G. B., Kostøl, A. R., and Mogstad, M. (2014). Family welfare cultures. *Quarterly Journal of Economics*, 129(4):1711–1752.

de Croes, L., Cornelisz, I., van Klaveren, C., and Ruijs, N. (2025). The effect of students with special needs on their peers' cognitive outcomes: Results from student-level population data in the Netherlands. *Education Finance and Policy*, 20(1):56–84.

Der Standard (2023). Höcke will Kinder mit Behinderung vom Regelunterricht ausschließen. Der Standard, August, https://www.derstandard.at/story/3000000182487/hoecke-will-kinder-mit-behinderung-vom-regelunterricht-ausschliessen.

Dhuey, E. and Lipscomb, S. (2010). Disabled or young? Relative age and special education diagnoses in schools. *Economics of Education Review*, 29(5):857–872.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101(5):1739–1774.

Elder, T. E. (2010). The importance of relative standards in ADHD diagnoses: Evidence based on exact birth dates. *Journal of Health Economics*, 29(5):641–656.

Feigenberg, B., Yan, R., and Rivkin, S. (2019). Illusory gains from Chile's targeted school voucher experiment. *Economic Journal*, 129(623):2805–2832.

Figlio, D. N. (2007). Boys named Sue: Disruptive children and their peers. *Education Finance and Policy*, 2(4):376–394.

Firpo, S., Fortin, N. M., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3):953–973.

Fletcher, J. M. (2009). The effects of inclusion on classmates of students with special needs: The case of serious emotional problems. *Education Finance and Policy*, 4(3):278–299.

Fletcher, J. M. (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.

Gilmour, A. F. (2018). Has inclusion gone too far? Weighing its effects on students with disabilities, their peers, and teachers. Education Next, 18(4), https://www.educationnext.org/has-inclusion-gone-too-far-weighing-effects-students-with-disabilities-peers-teachers/.

Godoy, M. P., Meza, M. L., Salazar, A., et al. (2004). Antecedentes históricos, presente y futuro de la educación especial en Chile. Technical report, Ministerio de Educación de Chile.

Gottfried, M. A. (2014). Classmates with disabilities and students' noncognitive outcomes. *Educational Evaluation and Policy Analysis*, 36(1):20–43.

Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (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.

Hill, A. J. and Jones, D. B. (2018). A teacher who knows me: The academic benefits of repeat student-teacher matches. *Economics of Education Review*, 64:1–12.

Holz, M. (2018). Datos de la modalidad de Educación Especial en Chile, año 2018. Informe SUP 118026, Biblioteca del Congreso Nacional de Chile, Asesoría Técnica Parlamentaria.

Hoxby, C. (2000). The effects of class size and composition on student achievement: New evidence from natural population variation. *Quarterly Journal of Economics*, 115(4):1239–1285.

Huang, B., Lu, H., and Zhu, R. (2021). Disabled peers and student performance: Quasi-experimental evidence from China. *Economics of Education Review*, 82:102121.

Hwang, N., Kisida, B., and Koedel, C. (2021). A familiar face: Student-teacher rematches and student achievement. *Economics of Education Review*, 85:102194.

Kristoffersen, J. H. G., Krægpøth, M. V., Nielsen, H. S., and Simonsen, M. (2015). Disruptive school peers and student outcomes. *Economics of Education Review*, 45:1–13.

Latimer, E., Sibieta, L., and Snape, D. (2025). Support for children with disabilities and special educational needs. IFS Green Budget, Chapter 5, Institute for Fiscal Studies, October.

Lavy, V., Paserman, M. D., and Schlosser, A. (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *Economic Journal*, 122(559):208–237.

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

Libération (2025). Scolarisation des enfants handicapés: une école à bout de souffle. Libération, 9 April.

Maestas, N., Mullen, K. J., and Strand, A. (2013). Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt. *American Economic Review*, 103(5):1797–1829.

Ministerio de Educación de Chile (2009). Decreto Supremo N. 170: Fija normas para determinar los alumnos con necesidades educativas especiales que serán beneficiarios de las subvenciones para educación especial. Diario Oficial de la República de Chile.

Ministerio de Educación de Chile (2013). Orientaciones técnicas para programas de integración escolar (PIE). División de Educación General, Unidad de Educación Especial, Santiago.

Mizala, A. and Urquiola, M. (2013). School markets: The impact of information approximating schools' effectiveness. *Journal of Development Economics*, 103:313–335.

Neal, D. and Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics*, 92(2):263–283.

Neilson, C. A. (2021). Targeted vouchers, competition among schools, and the academic achievement of poor students. Working Paper 2021-48, Princeton University, Department of Economics.

OECD (2023). Equity and inclusion in education: Finding strength through diversity. Report, OECD Publishing, Paris.

Ost, B. (2014). How do teachers improve? The relative importance of specific and general human capital. *American Economic Journal: Applied Economics*, 6(2):127–151.

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

Rangvid, B. S. (2019). Returning special education students to regular classrooms: Externalities on peers' reading scores. *Economics of Education Review*, 68:13–22.

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review*, 109(3):774–809.

República de Chile (1997). Ley n. 19.532: Crea régimen de jornada escolar completa diurna y dicta normas para su aplicación. Diario Oficial de la República de Chile.

República de Chile (2010). Ley 20.422: Establece normas sobre igualdad de oportunidades e inclusión social de personas con discapacidad. Diario Oficial de la República de Chile.

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

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

The Economist (2025). England's broken system for meeting special educational needs. The Economist, 27 October, https://www.economist.com/britain/2025/10/27/englands-broken-system-for-meeting-special-educational-needs.

The Economist (2026). America's support system for disabled pupils is broken. The Economist, 3 September, https://www.economist.com/united-states/2026/09/03/americas-support-system-for-disabled-pupils-is-broken.

UNESCO (2020). Global education monitoring report 2020: Inclusion and education, all means all. Report, UNESCO, Paris.

UNICEF (2012). The right of children with disabilities to education: A rights-based approach to inclusive education. Position paper, UNICEF Regional Office for Central and Eastern Europe and the Commonwealth of Independent States, Geneva.

Valenzuela Barros, J. P. and Toledo Román, G. (2012). Ordenamiento de profesores y estudiantes entre y dentro de los establecimientos escolares: el caso de Chile. Technical report, Universidad de Chile, Facultad de Economía y Negocios.

World Bank (2019). Every learner matters: Unpacking the learning crisis for children with disabilities. Report, World Bank, Washington, DC.

Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *Quarterly Journal of Economics*, 134(2):557–598.

A Data and sample

This appendix section describes the structure of the data and the variation used to build the treatment. Table A1 reports which cohorts were tested in which grades and years. Because not every grade is tested every year, each cohort in our sample is observed in two or three tested grades, and Mathematics is not tested in second grade.

Figure A1 compares the probability of repeating a grade for children with and without a severe disability. CWD repeat first grade in more than 20% of cases and second grade in about 12%, against 2 to 4% for other children, and the gap narrows from fourth grade onwards. This difference is the reason why we measure exposure at school entry and exclude repeating classmates from the definition of the treatment (Section 4): a CWD who repeats first grade joins the classroom of the following cohort as a consequence of their own performance, so that later exposure depends on the assignment of neighboring classrooms.

Figure A2 shows the distributions of the number and of the share of CWD in first-grade classes. About 40% of classes have none, most exposed classes have between one and three, and the share is generally small, so that its variation is largely driven by class size. For this reason, our main treatment is a binary indicator, and we report the count and the share in Table 2.

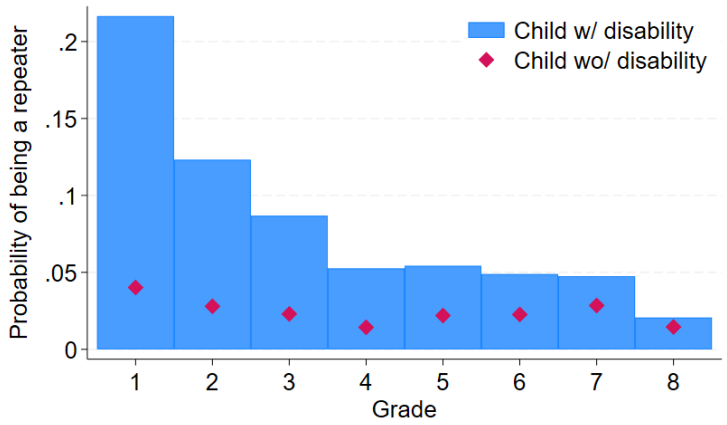
Figure A3 reports the composition of severe disabilities among CWD in first grade, by cohort. Intellectual disabilities are the most common category in every cohort, although their share declines from about 85% in the 2009 cohort to about 74% in the 2014 cohort as the share of children diagnosed with autism increases, consistent with the rise in recorded diagnoses discussed in Section 4. Finally, Figure A4 illustrates the variation exploited by Equation 1: in nine randomly selected school-sections, the number of CWD in first grade changes considerably from one cohort to the next, and several school-sections switch between exposed and unexposed cohorts.

Table A1: Tested grades and cohorts by year, SIMCE

	Cohort											
Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
2007	4											
2008		4										
2009			4									
2010				4								
2011	8				4							
2012						**4**		**2**				
2013	10		8		6		**4**		**2**			
2014		10		8		**6**		**4**		**2**		
2015			10		8		**6**		**4**		**2**	
2016				10				**6**		**4**		
2017					10		**8**				**4**	
2018						10				**6**		4
2019									**8**			

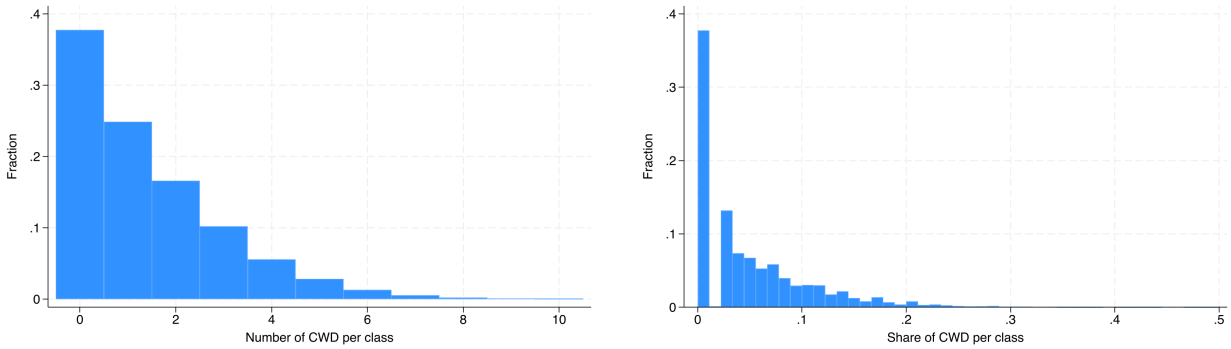
Notes: Cohorts are identified by the year in which students start first grade. Grades in bold are those used in the analysis. Second graders are tested in Spanish only.

Figure A1: Probability of repeating a grade by disability status



Notes: Share of children repeating each grade of primary school, separately for children with and without a severe disability.

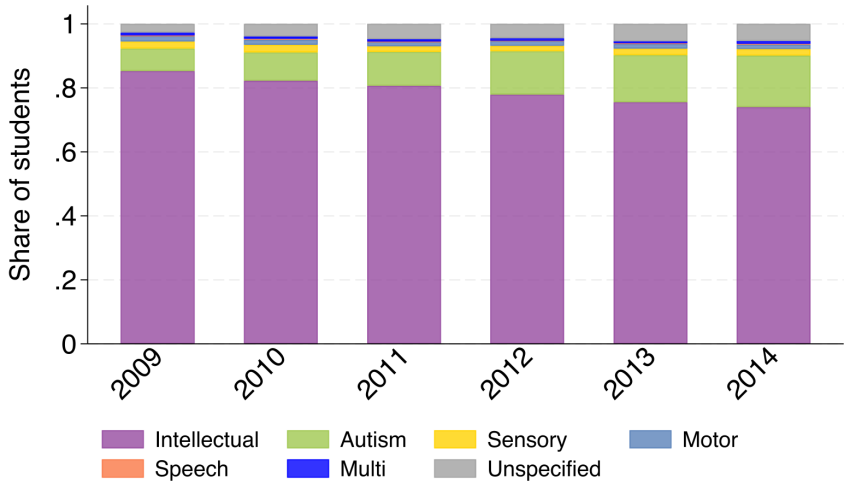
Figure A2: Exposure to classmates with severe disabilities in first grade



(a) Number **(b)** Share

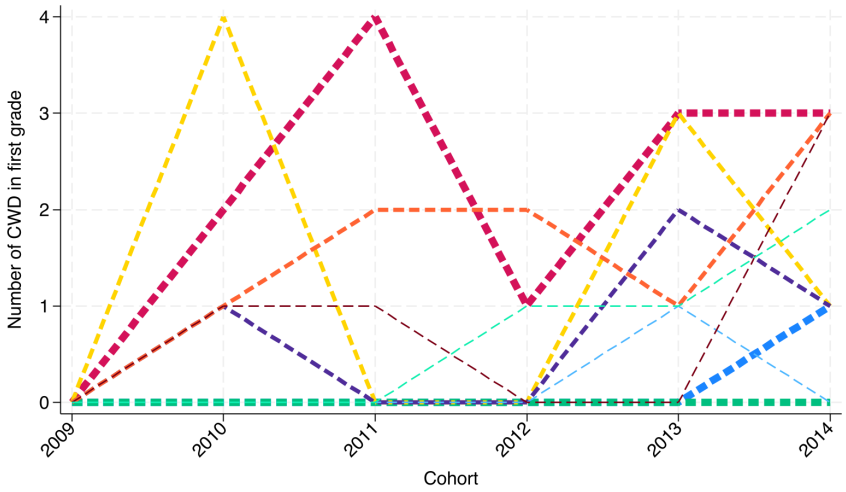
Notes: Distribution of the number and the share of non-repeating CWD in first-grade classrooms of the estimation sample, cohorts 2009 to 2014.

Figure A3: Composition of disability types over time



Notes: Share of each category among non-repeating children with a single recorded type of severe disability in first-grade classes of the estimation sample, by cohort.

Figure A4: Cross-cohort variation in exposure within school-sections



Notes: Number of CWD in first grade for nine randomly selected school-sections of the estimation sample observed in all six cohorts, illustrating the variation used by Equation 1.

B Identification and estimand

The vectors X, Z and Ω in Equation 1 hold individual, class and school covariates. X includes an indicator for being female, indicators for each parent having vocational or college education, and indicators for family income quintile. Z includes class size and the class shares of children in each income quintile. Ω includes school size and school-level shares of parental education. We estimate each subject separately and report a pooled specification with grade fixed effects alongside grade-specific estimates. Because exposure does not vary within a student once the fixed effects are in place, we cannot include student fixed effects. Within a grade, the first-grade cohort and the test year are the same variable, since a child on schedule sits the test in the year given by cohort plus grade minus one; in the pooled specification, cohort and grade fixed effects together span year effects. We cluster standard errors on the school-section in each cohort (i.e., the class group), which groups the children of a cohort observed in the same section, is coarser than the contemporaneous classroom and accounts for the fact that a student appears in the panel more than once, in different classrooms, across tested grades.

Table B1 reports the balance tests discussed in Section 5, at the level of the school-section in each cohort (i.e., the class group). Without fixed effects or with cohort fixed effects only (columns 1 and 2), exposed classes have far fewer children from families with college-educated parents and in the top income quintiles, which reflects the sorting of CWD into public and disadvantaged schools. School fixed effects remove most of these differences (column 3), and school-section fixed effects, the structure of Equation 1, remove all of them jointly ($F = 1.34$, $p = 0.21$, column 4). Across parallel classes in the same school and cohort, the structure of Equation 2, exposed classes have fewer families in the top two income quintiles, and the joint test rejects ($F = 3.17$, $p = 0.001$, column 5; see Appendix C). Comparing cohorts within school-sections therefore provides the cleaner source of variation.

Two further points support the design. Exposure rises mechanically with class size, since a larger class is more likely to contain a CWD, but controlling for class size flexibly rather than linearly leaves the balance results unchanged. Moreover, the estimate of Equation 1 is stable across control sets. Without any controls, it is 3.4% of a standard deviation in Mathematics;

adding individual covariates brings it to 2.1%, and adding class and school covariates leaves it at 2.1%. In Spanish, it moves from 2.7% without controls to 1.8% with individual covariates and 1.7% with all covariates. The covariates that matter are the individual ones, and once they are included the contextual controls change nothing.

Table B2 documents the relation between assignment in first grade and exposure in later grades, which defines the intention-to-treat estimand. Among children exposed in first grade, 84% still have a classmate with a severe disability in second grade and 66% in eighth grade; among children not exposed in first grade, the share rises from 19% to 31%. Exposure in first grade therefore raises later exposure substantially in every grade, but the difference narrows over time, which dilutes the intention-to-treat estimates in later grades. Table B3 quantifies this dilution: excluding from the comparison group the children who have a non-repeating classmate with a severe disability in the tested grade roughly doubles the estimates, whereas excluding only those exposed through a repeating classmate leaves them unchanged. The 25.1% of unexposed children who have a non-repeating classmate with a severe disability in the tested grade reach that situation through several routes: the CWD may change class within the school or enroll from another school after first grade, and the unexposed child may repeat a grade, change class within the school, or change school. A further 1.4% of the comparison group is exposed only through a classmate with a severe disability who repeated a grade and joined their class. The dilution therefore mostly comes from ordinary reshuffling, mobility and new enrollment rather than from the grade repetition of CWD. This motivates the additional check with alternative fixed effects described in Section 7.5.

Table B4 examines whether exposure changes the composition of the classroom through mobility. Exposed children are slightly less likely to change school and to repeat a grade, and marginally more likely to change class within the school, with differences of at most a few tenths of a percentage point. Mobility in response to exposure is therefore too small to affect our estimates, and exposed children are, if anything, less likely to leave their school.

Table B1: Exposure and class composition

	(1) Exposed	(2) Exposed	(3) Exposed	(4) Exposed	(5) Exposed
Share female	-0.061***	-0.070***	-0.013	0.000	-0.014
	(0.010)	(0.010)	(0.012)	(0.014)	(0.015)
Mother vocational education	-0.198***	-0.239***	-0.048***	-0.041**	-0.034
	(0.016)	(0.016)	(0.018)	(0.018)	(0.026)
Mother college education	-0.069***	-0.141***	-0.043	-0.042	0.044
	(0.026)	(0.026)	(0.028)	(0.028)	(0.038)
Father vocational education	-0.115***	-0.116***	0.002	0.011	-0.041
	(0.018)	(0.018)	(0.019)	(0.019)	(0.027)
Father college education	-0.193***	-0.154***	-0.027	-0.008	-0.055
	(0.027)	(0.027)	(0.028)	(0.029)	(0.038)
Income quintile 2	-0.175***	-0.163***	-0.012	-0.011	-0.036
	(0.015)	(0.015)	(0.018)	(0.018)	(0.025)
Income quintile 3	-0.250***	-0.206***	-0.004	-0.001	-0.019
	(0.020)	(0.021)	(0.024)	(0.025)	(0.033)
Income quintile 4	-0.194***	-0.218***	-0.050	-0.034	-0.107***
	(0.025)	(0.025)	(0.031)	(0.032)	(0.041)
Income quintile 5	-0.590***	-0.564***	-0.071*	-0.052	-0.143***
	(0.024)	(0.024)	(0.037)	(0.038)	(0.049)
Class size	0.000	0.001**	0.000	0.000	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
School size	-0.000***	-0.000***			
	(0.000)	(0.000)			
Cohort FE	No	Yes	Yes	Yes	No
School FE	No	No	Yes	No	No
School-section FE	No	No	No	Yes	No
School-by-cohort FE	No	No	No	No	Yes
Joint test, F	2472.58	2516.67	2.63	1.34	3.17
Joint test, p	0.000	0.000	0.005	0.212	0.001
School-section by cohort cells	43603	43603	43579	43465	29775

Notes: The unit of observation is a school-section in a given cohort (for example, section A of a school in the 2011 cohort). There are 43,854 such cells in 8,039 school-sections and 4,581 schools. The dependent variable is the share of children in the cell who were exposed to a classmate with a severe disability in their first-grade class, 61 percent on average. Composition is measured as the class mean of each student characteristic. Column 1 includes no fixed effects, column 2 adds cohort fixed effects, column 3 adds school fixed effects, and column 4 replaces school with school-section fixed effects, matching the structure of Equation 1. Column 5 includes school-by-cohort fixed effects, matching Equation 2, and is estimated on the classes of schools with more than one class in the cohort. School size drops out once school, school-section or school-by-cohort fixed effects are included. The joint test is a Wald test of all composition variables, excluding class and school size. Standard errors clustered on the school-section. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B2: Persistence of exposure after first grade

	(1) 2nd grade	(2) 4th grade	(3) 6th grade	(4) 8th grade
Panel A. Probability of exposure in grade g				
Not exposed in 1st grade	0.193	0.260	0.270	0.306
Exposed in 1st grade	0.839	0.740	0.681	0.655
Difference	0.646***	0.480***	0.412***	0.349***
	(0.004)	(0.003)	(0.004)	(0.005)
Panel B. Number of classmates with a severe disability in grade g				
Not exposed in 1st grade	0.32	0.47	0.50	0.60
Exposed in 1st grade	1.98	1.75	1.59	1.54
Observations	670,324	1,011,833	499,266	335,558

Notes: Panel A reports the probability of having at least one non-repeating classmate with a severe disability in the tested grade, separately for students exposed and not exposed in first grade, and the difference between them. Panel B reports the corresponding average number of such classmates. Standard errors on the difference are clustered on the school-section in each cohort (i.e., the class group). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B3: Intention to treat and later exposure of the comparison group

	(1) Mathematics: all	(2) Mathematics: excl. non-rep.	(3) Mathematics: excl. repeater-only	(4) Mathematics: excl. any	(5) Spanish: all	(6) Spanish: excl. non-rep.	(7) Spanish: excl. repeater-only	(8) Spanish: excl. any
Exposed in first grade	-0.021***	-0.041***	-0.021***	-0.041***	-0.017***	-0.033***	-0.018***	-0.034***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Observations	1,527,433	1,354,344	1,521,330	1,348,241	2,056,979	1,845,546	2,045,311	1,833,878

Notes: Column 1 reproduces the baseline. Column 2 excludes from the comparison group students who were not exposed in first grade but have a non-repeating classmate with a severe disability in the tested grade, 25.1 percent of the comparison group. Column 3 excludes only those exposed through a repeating classmate, 1.4 percent of the comparison group. Column 4 excludes both. The exercise bounds the dilution of the intention to treat and is not a preferred estimate, since later exposure is not randomly assigned. Sample, controls, fixed effects and clustering as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B4: Exposure and student mobility

	(1) Pooled	(2) 2nd grade	(3) 4th grade	(4) 6th grade	(5) 8th grade
Panel A. Changed school					
Exposed in first grade	-0.003***	-0.001	-0.006***	-0.004***	-0.001
	(0.001)	(0.003)	(0.001)	(0.002)	(0.002)
Observations	2,071,652	538,192	852,949	408,005	272,483
Control mean	0.090	0.100	0.090	0.090	0.080
Panel B. Changed class within school					
Exposed in first grade	0.001*	0.002	0.001*	0.001	-0.002*
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Observations	2,071,652	538,192	852,949	408,005	272,483
Control mean	0.030	0.040	0.030	0.030	0.030
Panel C. Repeated grade					
Exposed in first grade	-0.001***	-0.003***	-0.000	-0.001	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Observations	2,071,652	538,192	852,949	408,005	272,483
Control mean	0.010	0.020	0.010	0.010	0.010

Notes: Outcomes are indicators for having changed school since the previous year (Panel A), having changed section within the same school (Panel B), and having repeated the grade (Panel C). Coefficients are changes in probability (0.01 = one percentage point). *Control mean* is the mean of the outcome among students not exposed in first grade. Sample, controls, fixed effects and clustering as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

C Comparing parallel classes within the same school and cohort

An alternative design compares parallel classes within the same school and cohort, replacing the school-section and cohort fixed effects with school-by-cohort fixed effects $\psi_{s,t}$:

$$Score_{i,l,s,t} = \psi_{s,t} + \beta \, \text{Exp}_{i,l,s,t} + \boldsymbol{\delta_1} \boldsymbol{X}_{i,l,s,t} + \boldsymbol{\delta_2} \boldsymbol{Z}_{l,s,t} + \boldsymbol{\delta_3} \boldsymbol{\Omega}_{s,t} + \varepsilon_{i,l,s,t} \tag{2}$$

By construction, this comparison uses only schools with more than one section in the cohort. Because exposed and unexposed parallel classes are less comparable than exposed and unexposed cohorts within the same school-section (column 5 of Table B1), we report it only as a robustness check on the main estimates.

Table C1 reports estimates of Equation 2. The penalty is 2.6% of a standard deviation in Mathematics and 2.3% in Spanish, and it follows the same profile across grades as in Table 1. The two designs therefore agree on the sign, the significance and the persistence of the effect, while the slightly larger magnitude of the cross-classroom estimates is consistent with the less favorable socioeconomic composition of exposed classes within the same school and cohort.

Table C2 applies to Equation 2 the inference checks that Appendix D applies to Equation 1. Clustering standard errors on the school rather than on the school-section in each cohort (i.e., the class group) raises them from 0.002 to 0.003 in both subjects, and both estimates remain significant at the 1% level. Randomization inference permutes exposure across sections within school-cohorts (Young, 2019). In both subjects, none of 1,000 permutations produces an estimate as large in absolute value as the actual one, so that the p-value is below 0.001, and the actual estimates are 1.7 (Mathematics) and 1.9 (Spanish) times larger in absolute value than the 1st and 99th percentiles of the permutation distribution.

Table C1: Effect of exposure on standardized test scores, comparing parallel classes

	(1) Pooled	(2) 2nd grade	(3) 4th grade	(4) 6th grade	(5) 8th grade
Panel A. Mathematics					
Exposed in first grade	-0.026***		-0.027***	-0.029***	-0.028***
	(0.002)		(0.003)	(0.004)	(0.004)
Observations	1,527,329		849,664	404,329	269,972
Panel B. Spanish					
Exposed in first grade	-0.023***	-0.028***	-0.025***	-0.021***	-0.017***
	(0.002)	(0.004)	(0.003)	(0.004)	(0.005)
Observations	2,056,859	536,530	845,506	401,998	267,760

Notes: Estimates of Equation 2, which includes school-by-cohort fixed effects and compares parallel classes within the same school and cohort. Sample, controls and clustering as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C2: Inference for the comparison of parallel classes

	(1) Mathematics	(2) Spanish
Exposed in first grade	-0.026***	-0.023***
Standard error, clustered on the school-section by cohort	(0.002)	(0.002)
Standard error, clustered on the school	(0.003)	(0.003)
Randomization inference p	<0.001	<0.001
Permutations	1000	1000
Observations	1,527,329	2,056,859

Notes: Pooled estimates of Equation 2, with standard errors clustered on the school-section in each cohort, i.e., the class group (significance stars), and on the school. Randomization inference permutes exposure across sections within school-cohorts and reports the share of permutations yielding an estimate at least as large in absolute value as the actual one; the statistic is the coefficient from a regression of the outcome, residualized on controls and fixed effects and averaged by school-section, cohort and grade, on exposure. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

D Robustness of the average effects

Figure D1 plots the estimates of Table 1 by grade with their confidence intervals. Table D1 reports the estimates of Equation 1 with standard errors clustered at the school-section in each cohort (i.e., the class group) and at the school. School-level clustering raises the standard error only in Mathematics, from 0.002 to 0.003, and both estimates remain significant at the 1% level. The corresponding checks for the comparison of parallel classes are in Appendix C.

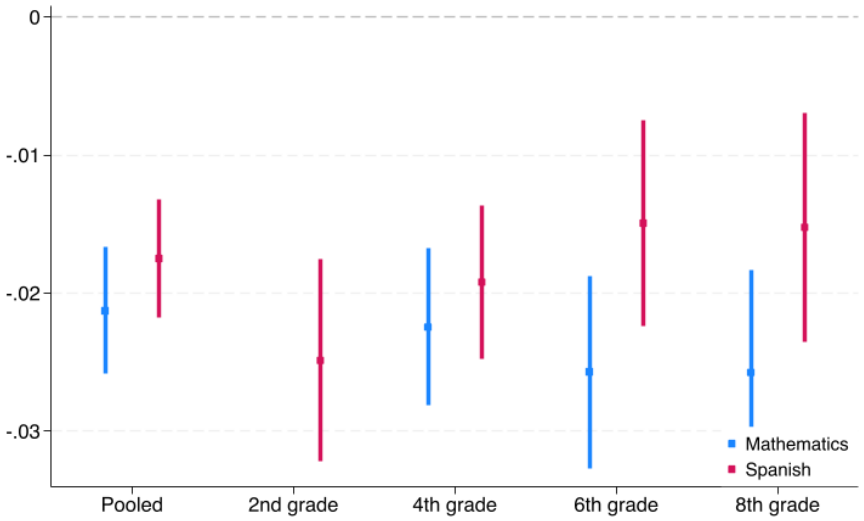
Table D2 displays further checks on the main estimates. The class-level balance test does not include SEP priority status, which is available only for the first-grade year. At the level of the school-section in each cohort (i.e., the class group), the share of priority students retains a small association with exposure once school-section fixed effects are included: a cell with ten percentage points more priority students has a share of exposed children about 0.5 percentage points higher, against a mean of 61%. Adding the class share of priority students leaves the estimates unchanged, at 2.1% of a standard deviation in Mathematics and 1.7% in Spanish, and adding the student's own priority status reduces them only marginally, to 2.0% and 1.6%. The incidence results are also unchanged when the class share of priority students is added to the fully interacted specification. Retaining all students with a test score, replacing missing covariates with zero and adding an indicator for each missing covariate, yields slightly larger estimates, 2.5% of a standard deviation in Mathematics and 2.0% in Spanish. Randomization inference permutes exposure across cohorts within school-sections (Young, 2019). In both subjects, none of 1,000 permutations produces an estimate as large in absolute value as the actual one, so that the p-value is below 0.001, and the actual estimates are 2.1 (Mathematics) and 2.6 (Spanish) times larger in absolute value than the 1st and 99th percentiles of the permutation distribution.

Table D3 shows that exposure does not affect the probability of observing a test score, so that the estimates are not driven by selective test-taking. Exposure raises the probability that family background information is missing by 0.3 percentage points, against a base of 17.5%, a difference that the missing-indicator specification of Table D2 addresses.

Table D4 concerns potential CWD who are missing from the special-needs registers, which

begin in 2016 and observe only a minority of the (few) children enrolled in special schools. A CWD who left regular schools before 2016 is never recorded, so that their first-grade classmates are coded as unexposed. Leaving the school system altogether is rare: 1.2% of the children who entered first grade between 2009 and 2014 are not enrolled anywhere between 2016 and 2019, and children who repeated first grade, a quarter of whom have a recorded disability, are no more likely to leave than other children. Moves to special schools are more relevant. Counting as CWD all first graders who attended a special school between 2009 and 2019, other than schools for specific language disorders, and dropping these children from the sample of peers, reclassifies 2.7% of peers as exposed. The estimates rise slightly, from 2.1% to 2.3% of a standard deviation in Mathematics and from 1.7% to 1.9% in Spanish, consistent with a mild attenuation of the main estimates by unrecorded CWD.

Figure D1: Effect of exposure by grade



Notes: Point estimates and 95 percent confidence intervals from Equation 1, pooled and by grade. Standard errors clustered on the school-section in each cohort (i.e., the class group).

Table D1: Robustness to the level of clustering

	Mathematics		Spanish	
	(1)	(2)	(3)	(4)
Clustering	School-section by cohort	School	School-section by cohort	School
Exposed in first grade	-0.021***	-0.021***	-0.017***	-0.017***
	(0.002)	(0.003)	(0.002)	(0.002)
Observations	1,527,433	1,527,433	2,056,979	2,056,979

Notes: Pooled estimates of Equation 1 under two clustering levels, the school-section in each cohort (i.e., the class group) and the school. Point estimates are identical by construction. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table D2: Robustness of the main estimates

	(1) Mathematics	(2) Spanish
Baseline	-0.021***	-0.017***
	(0.002)	(0.002)
Observations	1,527,433	2,056,979
Adding the class SEP priority share	-0.021***	-0.017***
	(0.002)	(0.002)
Observations	1,526,953	2,056,413
Adding the class SEP share and own SEP priority	-0.020***	-0.016***
	(0.002)	(0.002)
Observations	1,526,953	2,056,413
Full sample, missing-value indicators	-0.025***	-0.020***
	(0.002)	(0.002)
Observations	1,796,838	2,469,188
Randomization inference p	<0.001	<0.001
Permutations	1000	1000

Notes: Pooled estimates of Equation 1. Row 1 reproduces Table 1. Row 2 adds the share of SEP priority students in the school-section in each cohort, and row 3 also adds the student's own SEP priority status, both measured in the first-grade year. Row 4 keeps all students with a test score, replacing missing covariates with zero and adding an indicator for each missing covariate. Randomization inference permutes exposure across cohorts within school-sections and reports the share of permutations yielding an estimate at least as large in absolute value as the actual one; the statistic is the coefficient from a regression of the outcome, residualized on controls and fixed effects and averaged by school-section, cohort and grade, on exposure. Standard errors clustered on the school-section in each cohort (i.e., the class group). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table D3: Exposure and attrition

	(1) Missing Mathematics score	(2) Missing Spanish score	(3) Missing background
Exposed in first grade	0.0005	0.0001	0.0030***
	(0.0003)	(0.0002)	(0.0009)
Control mean	0.025	0.018	0.175
Observations	1,846,656	2,516,980	2,516,980

Notes: Outcomes are indicators for a missing Mathematics score (grades 4, 6 and 8), a missing Spanish score (grades 2 to 8), and missing family background information, among students enrolled in the tested grade. School-section, cohort and grade fixed effects, no controls. *Control mean* is the mean among students not exposed in first grade. Standard errors clustered on the school-section in each cohort (i.e., the class group). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table D4: Children with severe disabilities missing from the registers

	Mathematics	Spanish
Panel A. Main exposure: recorded children with disabilities		
Exposed in first grade	-0.0213***	-0.0175***
	(0.0023)	(0.0022)
Observations	1,527,433	2,056,979
Share of peers exposed	58.6%	
Panel B. Exposure also counting special-school attendees		
Exposed in first grade	-0.0228***	-0.0191***
	(0.0024)	(0.0022)
Observations	1,524,644	2,052,552
Share of peers exposed	61.2%	

Notes: Pooled estimates of Equation 1. Panel A reproduces Table 1. In Panel B, a first-grade class is exposed if it includes a non-repeating child who is either recorded with a severe disability or attended a special school (excluding schools for specific language disorders) at any point between 2009 and 2019, and peers who attended a special school are dropped. Special-school enrollment is taken from the national enrollment records. *Share of peers exposed* is computed over student-grade observations. Standard errors clustered on the school-section in each cohort (i.e., the class group). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

E Robustness of the heterogeneity results

Table E3 reports the estimates plotted in Figure 2, the effect of exposure for peers outside and inside each group defined by special needs or socioeconomic disadvantage. Table E1 reports the effect of exposure by type of disability, pooled and by grade; the pooled estimates are plotted in Figure 1. Table E2 adds exposure to a classmate with an intellectual disability to the main specification, so that the general exposure indicator measures the residual effect of all other types.

Table E4 adjusts the p-values of the two families of heterogeneity tests, the seven types of disability plus the unspecified category and the differences between groups of peers, for multiple testing. The effect of intellectual disabilities and the concentration of losses on children with a learning disability label survive both the family-wise and the false discovery rate adjustments, whereas the positive point estimate for severe speech disorders does not survive either. Among socioeconomic gradients, the larger effect on SEP priority students in Spanish survives both adjustments, the Spanish gradients by income and mother's education survive only the false discovery rate adjustment at the 5% significance level (Benjamini et al., 2006; Anderson, 2008), and none of the Mathematics gradients is significant before adjustment.

Table E5 addresses whether the learning disability label responds to exposure, discussed in Section 7.2. Panel A shows that exposure raises the probability of ever being labeled by 0.4 to 0.6 percentage points. Panel B asks whether these additional labels change who is labeled. If they went disproportionately to a particular kind of child, for instance to more disadvantaged children, exposed labeled children would differ in background from unexposed labeled children, and their larger loss could reflect their composition rather than exposure. We therefore regress each predetermined characteristic on exposure separately for never and ever labeled children, with school-section and cohort fixed effects, and test whether the two coefficients differ. Exposed children are more often SEP priority students and from the bottom income quintile, and less often have a college-educated mother, in both groups, consistently with the small association between exposure and the share of priority students discussed in Appendix D. What matters is whether this gap is larger among labeled children.

Table E1: Heterogeneity by type of disability

	(1) Pooled	(2) 2nd grade	(3) 4th grade	(4) 6th grade	(5) 8th grade
Panel A. Mathematics					
Motor disability	-0.000		0.002	-0.017	-0.003
	(0.007)		(0.008)	(0.010)	(0.011)
Visual disability	-0.016		-0.008	-0.033**	-0.026
	(0.010)		(0.014)	(0.015)	(0.017)
Hearing disability	-0.002		-0.012	0.005	0.016
	(0.008)		(0.010)	(0.012)	(0.014)
Intellectual disability	-0.026***		-0.028***	-0.030***	-0.035***
	(0.002)		(0.003)	(0.004)	(0.004)
Autism	-0.004		-0.010***	0.000	0.004
	(0.003)		(0.004)	(0.005)	(0.005)
Multiple disabilities	-0.003		-0.000	-0.008	-0.008
	(0.007)		(0.008)	(0.011)	(0.012)
Severe speech	0.011		0.010	0.025**	0.007
	(0.008)		(0.010)	(0.012)	(0.013)
Unspecified	0.001		0.003	-0.005	0.010
	(0.004)		(0.005)	(0.007)	(0.008)
Observations	1,527,433		849,761	406,334	271,318
Panel B. Spanish					
Motor disability	-0.006	-0.000	-0.006	-0.022*	-0.017
	(0.006)	(0.010)	(0.008)	(0.011)	(0.013)
Visual disability	-0.010	-0.034**	0.000	-0.007	-0.025
	(0.010)	(0.016)	(0.013)	(0.017)	(0.019)
Hearing disability	0.006	0.013	0.001	-0.001	0.012
	(0.007)	(0.011)	(0.010)	(0.013)	(0.015)
Intellectual disability	-0.024***	-0.032***	-0.025***	-0.019***	-0.025***
	(0.002)	(0.004)	(0.003)	(0.004)	(0.004)
Autism	-0.003	-0.006	-0.006	0.003	0.002
	(0.003)	(0.004)	(0.004)	(0.005)	(0.006)
Multiple disabilities	0.001	-0.005	0.011	-0.005	-0.000
	(0.006)	(0.010)	(0.008)	(0.012)	(0.013)
Severe speech	0.014*	0.020*	0.011	0.010	0.009
	(0.007)	(0.012)	(0.009)	(0.013)	(0.014)
Unspecified	0.003	-0.002	0.002	0.009	0.018**
	(0.004)	(0.006)	(0.005)	(0.007)	(0.009)
Observations	2,056,979	538,192	845,599	404,029	269,136

Notes: The exposure indicator is disaggregated into indicators for the presence of each type of severe disability in the first-grade class. The categories are those reported by the special-needs register and are not mutually exclusive, since a child may carry more than one diagnosis. Sample, controls, fixed effects and clustering as in the main analysis in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table E2: Exposure to intellectual disability and residual exposure

	(1) Pooled	(2) 2nd grade	(3) 4th grade	(4) 6th grade	(5) 8th grade
Panel A. Mathematics					
Exposed in first grade	-0.004		-0.004	-0.007	0.003
	(0.004)		(0.004)	(0.006)	(0.006)
Intellectual disability	-0.023***		-0.025***	-0.025***	-0.037***
	(0.004)		(0.004)	(0.006)	(0.006)
Observations	1,527,433		849,761	406,334	271,318
Panel B. Spanish					
Exposed in first grade	-0.000	-0.006	-0.001	-0.001	0.010
	(0.003)	(0.005)	(0.004)	(0.006)	(0.007)
Intellectual disability	-0.024***	-0.028***	-0.024***	-0.018***	-0.033***
	(0.003)	(0.005)	(0.004)	(0.006)	(0.007)
Observations	2,056,979	538,192	845,599	404,029	269,136

Notes: *Exposed in first grade* is the general exposure indicator and *Intellectual disability* indicates the presence of a classmate with an intellectual disability in the first-grade class. The first coefficient therefore measures the residual effect of exposure to any other type of severe disability. Sample, controls, fixed effects and clustering as in the main analysis in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

It is not for SEP priority status, maternal education and gender, where the differences are 0.3, 0.3 and -0.3 percentage points, none of them significant. It is slightly larger for the bottom income quintile, by 0.8 percentage points, while the point estimate for maternal education goes in the opposite direction: compared with exposed unlabeled children, exposed labeled children are a little more often poor but, if anything, also a little more often have a college-educated mother. All differences are below one percentage point, and those for income and maternal education point in opposite directions, so that the additional labels do not shift the labeled group towards more or less disadvantaged children. Panel C reports trimming bounds on the effect for labeled children: because the excess share of labels, 3.2%, is large relative to the difference in effects between groups, the worst-case bounds are wide and include zero. These bounds assume that all additional labels concern children at the very bottom, or at the very top, of the score distribution, which is the most extreme case.

Table E6 provides a less extreme check, based on the timing of the learning disability label. The special-needs registers begin in 2016, so that the label is never observed in first grade, and for each child we observe only the first year in which a label is recorded between 2016 and 2023. For the cohorts that entered first grade in 2013 and 2014, 2016 corresponds to

Table E3: Incidence by the peer's own status: learning disabilities and socioeconomic disadvantage

	(1) Mathematics	(2) Spanish
Panel A. SEP priority in first grade		
Exposed, not in group	-0.018***	-0.011***
	(0.003)	(0.003)
Exposed, in group	-0.023***	-0.023***
	(0.003)	(0.003)
Observations	1,527,433	2,056,979
Panel B. Bottom income quintile		
Exposed, not in group	-0.021***	-0.015***
	(0.003)	(0.003)
Exposed, in group	-0.024***	-0.023***
	(0.003)	(0.003)
Observations	1,527,433	2,056,979
Panel C. Mother without college		
Exposed, not in group	-0.023***	-0.008*
	(0.005)	(0.005)
Exposed, in group	-0.022***	-0.020***
	(0.002)	(0.002)
Observations	1,527,433	2,056,979
Panel D. Learning disability		
Exposed, not in group	-0.014***	-0.009***
	(0.002)	(0.002)
Exposed, in group	-0.037***	-0.042***
	(0.005)	(0.005)
Observations	1,527,433	2,056,979

Notes: Each panel reports the effect of exposure for peers outside and inside the group defined in the panel heading, from a single regression in which exposure and all controls are interacted with the group indicator. SEP priority is measured in the first-grade year. Learning disability corresponds to the Chilean transitory special educational needs label, which entitles children to temporary support. Controls that coincide with the group indicator are dropped in Panels B and C. All grades pooled with grade fixed effects. Sample, fixed effects and clustering as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

fourth and third grade, respectively, so that we can distinguish children already labeled by the first observed test score (*early*) from children first labeled from 2017 onwards (*late*). Panel A shows that exposure is associated with a higher probability of an early label, by 0.35 percentage points against a base of 3.4%, but not of a late label. Panel B estimates the effect of exposure separately for never, early and late labeled peers, with exposure and all controls interacted with the three groups. The effect is small and insignificant for early-labeled peers, and largest for late-labeled peers, at 3.0% of a standard deviation in Mathematics and 3.8%

in Spanish; the difference between the two groups is significant in Spanish. If the additional labels associated with exposure drove the concentration of losses on labeled peers, it would appear among the early labeled, which is not the case. Several caveats apply. The sample includes only two cohorts and, in Mathematics, two tested grades, so that the estimates for the early group are less precise. Labeled children whose first year with a label cannot be matched in the yearly records are excluded. The interpretation is also speculative: learning disabilities that are recognized early may be more visible, and schools may place CWD and children with learning disabilities in the same classrooms together with additional resources, which would explain the additional early labels. The finding that the largest losses are borne by children whose needs are recognized later, and who are therefore likely to have received less support in the first years of school, is nonetheless consistent with the competing-needs interpretation.

Tables E7 and E8 report alternative measures of the diversity of disability types: the index estimated within exposed classes only, an indicator for two or more types, separate indicators for two and for three or more types, an index with sensory disabilities collapsed into one category, and an index in which each child is counted once. All indices of diversity are close to zero and insignificant, so that the null result of Table 3 does not depend on how diversity is measured. The indicators for the number of types do not hold the number of CWD fixed and are small and negative: relative to classes with a single type, two or more types lower scores by 0.5% of a standard deviation in Spanish (significant at the 10% level) and 0.4% in Mathematics (insignificant), and two types by 0.6% in Mathematics and 0.9% in Spanish (significant at the 10% and 1% levels), while three or more types have no significant effect.

Table E9 reports heterogeneity by gender. In Mathematics, girls lose 0.9% of a standard deviation more than boys, a difference that survives the false discovery rate adjustment but not the family-wise adjustment at the 5% significance level, while in Spanish the effect is identical for girls and boys. We have no prior on a gender difference in this setting, and we do not interpret it further.

Tables E10 and E11 ask whether the penalty depends on the presence of classmates with learning disabilities in the class group rather than on the peer's own status. Neither the

contemporaneous presence of a labeled classmate, available only for the test years 2016 to 2019, nor an above-median share of labeled classmates in first grade significantly changes the effect of exposure, for labeled and unlabeled peers alike. The incidence of the penalty therefore depends on the child's own reliance on teacher attention rather than on how many children in the classroom share it.

Table E4: Adjustment for multiple hypothesis testing

		Estimate	Unadjusted p	Holm p	Sharpened q
Panel A. Type of disability in the first-grade class					
Intellectual	Mathematics	-0.026	0.000	0.000	0.001
	Spanish	-0.024	0.000	0.000	0.001
Autism	Mathematics	-0.004	0.155	1.000	0.631
	Spanish	-0.003	0.245	1.000	0.834
Severe speech	Mathematics	0.011	0.178	1.000	0.631
	Spanish	0.014	0.053	0.744	0.330
Motor	Mathematics	-0.000	0.995	1.000	1.000
	Spanish	-0.006	0.360	1.000	0.908
Visual	Mathematics	-0.016	0.117	1.000	0.616
	Spanish	-0.010	0.313	1.000	0.884
Hearing	Mathematics	-0.002	0.793	1.000	1.000
	Spanish	0.006	0.433	1.000	0.908
Multiple	Mathematics	-0.003	0.656	1.000	1.000
	Spanish	0.001	0.875	1.000	1.000
Unspecified	Mathematics	0.001	0.782	1.000	1.000
	Spanish	0.003	0.445	1.000	0.908
Panel B. Difference in the effect between groups of peers					
Learning disability	Mathematics	-0.023	0.000	0.000	0.001
	Spanish	-0.033	0.000	0.000	0.001
SEP priority	Mathematics	-0.005	0.202	0.807	0.131
	Spanish	-0.011	0.002	0.018	0.007
Bottom quintile	Mathematics	-0.003	0.417	1.000	0.264
	Spanish	-0.007	0.033	0.164	0.031
Mother without college	Mathematics	0.001	0.844	1.000	0.406
	Spanish	-0.012	0.014	0.097	0.022
Female	Mathematics	-0.009	0.015	0.097	0.022
	Spanish	-0.000	0.946	1.000	0.406

Notes: Each panel is one family of hypotheses (16 and 10 tests). Panel A reports the coefficients of Table E1, pooled across grades. Panel B reports the difference in the effect of exposure between peers inside and outside each group, from fully interacted regressions. Holm p-values control the family-wise error rate; sharpened q-values control the false discovery rate (Benjamini et al., 2006; Anderson, 2008). Learning disability corresponds to the Chilean transitory special educational needs label, which entitles children to temporary support.

Table E5: Exposure and the learning disability label

Panel A. Exposure and the probability of a learning disability label

	With controls	Fixed effects only
Exposed in first grade	0.0042***	0.0063***
	(0.0009)	(0.0008)
Control mean	0.094	0.094
Observations (students)	912,569	1,117,142

Panel B. Exposure and predetermined characteristics, by label

	SEP priority	Bottom quintile	Mother college	Female
Exposed, no learning disability	0.0189***	0.0120***	-0.0097***	-0.0009
	(0.0012)	(0.0012)	(0.0009)	(0.0012)
Exposed, learning disability	0.0220***	0.0195***	-0.0073***	-0.0044
	(0.0030)	(0.0031)	(0.0017)	(0.0030)
Difference	0.0031	0.0076**	0.0025	-0.0033
	(0.0031)	(0.0032)	(0.0018)	(0.0032)

Panel C. Incidence by label, trimming bounds

	Mathematics		Spanish	
	No learning disability	Learning disability	No learning disability	Learning disability
Baseline	-0.014***	-0.037***	-0.009***	-0.042***
	(0.002)	(0.005)	(0.002)	(0.005)
Trimming the bottom of exposed with learning disability	-0.014***	0.003	-0.009***	0.005
	(0.002)	(0.005)	(0.002)	(0.005)
Trimming the top of exposed with learning disability	-0.014***	-0.098***	-0.010***	-0.112***
	(0.002)	(0.005)	(0.002)	(0.005)

Notes: One observation per student in Panels A and B. Panel A regresses an indicator for ever being recorded with a learning disability on exposure, with school-section and cohort fixed effects. Panel B regresses predetermined characteristics on exposure separately for students without and with a learning disability, and reports the difference between the two coefficients, estimated in a pooled regression in which exposure and the fixed effects are interacted with the label. Panel C re-estimates the fully interacted incidence regression of Table 5 after trimming, within each grade, the exposed students with a learning disability by the excess share of such students among the exposed implied by the fixed-effects-only estimate of Panel A (3.2%), once from the bottom and once from the top of the score distribution. Learning disability corresponds to the Chilean transitory special educational needs label, which entitles children to temporary support. Standard errors clustered on the school-section in each cohort (i.e., the class group). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table E6: Exposure and the timing of the learning disability label, cohorts 2013 and 2014

Panel A. Exposure and the timing of the learning disability label		
	Learning disability by 2016	Learning disability from 2017
Exposed in first grade	0.0035***	0.0009
	(0.0010)	(0.0014)
Control mean	0.034	0.057
Panel B. Effect of exposure by the timing of the peer's own label		
	Mathematics	Spanish
No learning disability	-0.0252***	-0.0171***
	(0.0045)	(0.0040)
Learning disability by 2016	-0.0105	-0.0043
	(0.0145)	(0.0130)
Learning disability from 2017	-0.0304**	-0.0382***
	(0.0118)	(0.0111)
Difference, late minus early	-0.0198	-0.0339**
	(0.0184)	(0.0163)
Observations	400,836	679,289
of which learning disability by 2016	25,155	42,720
of which learning disability from 2017	34,550	59,606

Notes: Cohorts entering first grade in 2013 and 2014, for whom the special-needs registers begin in fourth and third grade. *Learning disability by 2016* identifies peers already recorded with a learning disability in 2016; *learning disability from 2017* identifies peers first recorded from 2017 onwards. Learning disability corresponds to the Chilean transitory special educational needs label, which entitles children to temporary support. Panel A uses one observation per student, with school-section and cohort fixed effects; the control mean is the share among students not exposed in first grade. Panel B reports the effect of exposure for each group from a single regression in which exposure and all controls are interacted with the three-group indicator, with school-section, cohort and grade fixed effects. Standard errors clustered on the school-section in each cohort (i.e., the class group). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table E7: Diversity of disability types, alternative measures

	(1) Mathematics: within exposed	(2) Mathematics: 2+ types	(3) Mathematics: N types	(4) Mathematics: sensory coll.	(5) Spanish: within exposed	(6) Spanish: 2+ types	(7) Spanish: N types	(8) Spanish: sensory coll.
Diversity (1-HHI)	-0.007				-0.006			
	(0.006)				(0.006)			
Exposed in first grade		-0.020***				-0.016***		
		(0.002)				(0.002)		
2+ disability types		-0.004				-0.005*		
		(0.003)				(0.003)		
2 types			-0.006*				-0.009***	
			(0.003)				(0.003)	
3+ types			-0.001				0.005	
			(0.006)				(0.005)	
Diversity (1-HHI), sensory collapsed				-0.007				-0.006
				(0.006)				(0.006)
Observations	880,805	1,527,433	880,805	880,805	1,202,319	2,056,979	1,202,319	1,202,319

Notes: Columns 1, 3 and 4 (Mathematics) and 5, 7 and 8 (Spanish) are estimated within exposed classes only; columns 2 and 6 use the full sample. Columns 1 and 5 estimate the diversity gradient. Columns 2 and 6 use an indicator for two or more types, set to zero for unexposed classes. Columns 3 and 7 use indicators for two and for three or more types, with single-type classes as the reference. Columns 4 and 8 use the diversity index computed after collapsing visual and hearing into one sensory category. Sample, controls, fixed effects and clustering as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table E8: Diversity of disability types, index computed over children

	(1) Mathematics: main	(2) Mathematics: +diversity	(3) Spanish: main	(4) Spanish: +diversity
Exposed in first grade	-0.021***	-0.020***	-0.017***	-0.017***
	(0.002)	(0.002)	(0.002)	(0.002)
Diversity (1-HHI)		-0.007		-0.006
		(0.006)		(0.006)
Observations	1,527,433	1,527,433	2,056,979	2,056,979

Notes: As Table 3, with an index in which each child contributes once. Children recorded in two or more categories, 7.4 percent of CWD, are assigned to the multiple-disability category before the shares are computed. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table E9: Heterogeneity by gender

	(1) Pooled	(2) 2nd grade	(3) 4th grade	(4) 6th grade	(5) 8th grade
Panel A. Mathematics					
Exposed in first grade	-0.017***		-0.018***	-0.023***	-0.018***
	(0.003)		(0.004)	(0.005)	(0.005)
Female	-0.054***		-0.011	-0.052***	-0.185***
	(0.012)		(0.015)	(0.020)	(0.023)
Exposed in first grade × Female	-0.009**		-0.009**	-0.006	-0.015**
	(0.004)		(0.004)	(0.006)	(0.007)
Observations	1,527,433		849,761	406,334	271,318
Panel B. Spanish					
Exposed in first grade	-0.017***	-0.028***	-0.020***	-0.013**	-0.013**
	(0.003)	(0.005)	(0.004)	(0.005)	(0.006)
Female	0.219***	0.188***	0.222***	0.174***	0.240***
	(0.011)	(0.019)	(0.016)	(0.022)	(0.026)
Exposed in first grade × Female	-0.000	0.007	0.002	-0.004	-0.004
	(0.004)	(0.006)	(0.005)	(0.006)	(0.008)
Observations	2,056,979	538,192	845,599	404,029	269,136

Notes: All controls are interacted with the female indicator. Sample, fixed effects and clustering as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table E10: Presence of classmates with learning disabilities

	(1) Pooled	(2) 4th grade	(3) 6th grade	(4) 8th grade
Panel A. Mathematics				
Exposed in first grade	-0.028***	-0.029***	-0.048***	-0.021***
	(0.004)	(0.007)	(0.009)	(0.005)
Any classmate with learning disability	-0.112***	-0.109	-0.103	0.005
	(0.035)	(0.069)	(0.170)	(0.052)
Exposed in first grade × Any classmate with learning disability	0.009	0.009	0.020	-0.009
	(0.006)	(0.009)	(0.012)	(0.008)
Observations	705,638	283,596	150,692	271,300
Panel B. Spanish				
Exposed in first grade	-0.019***	-0.017***	-0.040***	-0.014**
	(0.004)	(0.007)	(0.010)	(0.006)
Any classmate with learning disability	-0.094***	0.003	-0.107	-0.102*
	(0.034)	(0.065)	(0.189)	(0.057)
Exposed in first grade × Any classmate with learning disability	0.008	0.011	0.019	-0.002
	(0.006)	(0.010)	(0.014)	(0.008)
Observations	700,805	281,969	149,665	269,119

Notes: *Any classmate with learning disability* indicates that the contemporaneous classroom contains at least one student recorded with a learning disability. Learning disability corresponds to the Chilean transitory special educational needs label, which entitles children to temporary support. Because the register of learning disabilities begins in 2016, this measure is available only for test years 2016 to 2019, which is why the sample is smaller. Sample, controls, fixed effects and clustering otherwise as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table E11: Exposure and the first-grade share of classmates with learning disabilities

	(1) Mathematics: no learning disability	(2) Mathematics: learning disability	(3) Spanish: no learning disability	(4) Spanish: learning disability
Exposed in first grade	-0.017***	-0.022***	-0.011***	-0.016**
	(0.003)	(0.007)	(0.003)	(0.008)
Above-median learning disability share, first grade	-0.015***	-0.014*	-0.011***	-0.005
	(0.004)	(0.008)	(0.004)	(0.008)
Exposed in first grade × Above-median learning disability share, first grade	0.007	-0.001	0.001	-0.012
	(0.005)	(0.010)	(0.005)	(0.010)
Observations	1,311,955	213,851	1,757,866	297,413

Notes: Estimated separately for peers without a learning disability (columns 1 and 3) and peers with one (columns 2 and 4). *Above-median Learn Disab share, first grade* is an indicator for an above-median leave-one-out share of first-grade classmates ever recorded with a learning disability, excluding the student's own status. Learning disability corresponds to the Chilean transitory special educational needs label, which entitles children to temporary support. Sample, controls, fixed effects and clustering as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

F Teacher continuity and experience

Table F1 tests whether exposure predicts teacher continuity, regressing an indicator for having the same teacher as in the previous year on exposure with the fixed effects and controls of Equation 1. The coefficient is zero to four decimal places for the Mathematics teacher and 0.3 percentage points for the Spanish teacher, against a base rate of about 48%, and neither is significant. The 95% confidence intervals rule out effects of exposure on continuity larger than 0.5 percentage points in either direction for the Mathematics teacher and 0.8 percentage points for the Spanish teacher. By grade, only one of seven coefficients is marginally significant, 0.7 percentage points for the Spanish teacher in sixth grade. Within the variation our design uses, exposure therefore does not change who keeps their teacher, so that teachers selecting into or out of exposed classes cannot explain the smaller penalty under continuity. When the same teacher takes first grade in the same school-section in consecutive cohorts, as in 14% of transitions, the teacher's experience rises mechanically by one year per cohort, so that the experience interaction of Table F5 could capture a trend in the exposure effect; allowing the effect of exposure to vary freely by cohort, or restricting the sample to school-sections whose first-grade teacher changes between consecutive cohorts, leaves the interaction essentially unchanged, at 0.08% and 0.07% of a standard deviation per year in Mathematics and 0.05% in Spanish.

Table F2 decomposes the selection test of Table F1 by adding fixed effects and controls step by step, on a common sample. Without controls, exposed students are 4.1 percentage points more likely to keep their Mathematics teacher and 4.7 points more likely to keep their Spanish teacher, against a base rate of 48%. Individual controls halve this difference, and class and school controls reduce it further, but the association disappears only once school fixed effects are included. The raw association therefore reflects differences between schools: schools that enroll CWD more often keep teachers with their classes, consistently with the higher prevalence of looping among them (Appendix F.1). Within schools, and within school-sections across cohorts, exposure does not predict teacher continuity.

Table F5 reports the heterogeneity by the experience of the first-grade teacher discussed in Section 7.4, with exposure and all controls interacted with experience. General teaching

Table F1: Selection into teacher continuity

	(1) Pooled	(2) 2nd grade	(3) 4th grade	(4) 6th grade	(5) 8th grade
Panel A. Mathematics teacher					
Exposed in first grade	-0.0000		-0.0036	-0.0036	0.0023
	(0.0024)		(0.0036)	(0.0039)	(0.0029)
Observations	2,068,930		851,789	407,762	272,229
Control mean	0.4770		0.4851	0.4735	0.4683
Panel B. Spanish teacher					
Exposed in first grade	0.0030	0.0040	-0.0002	0.0074*	-0.0024
	(0.0024)	(0.0057)	(0.0036)	(0.0039)	(0.0030)
Observations	2,067,667	537,026	851,648	407,713	271,256
Control mean	0.4743	0.4827	0.4840	0.4545	0.4593

Notes: The dependent variable is an indicator for having the same teacher as in the previous year. Coefficients are changes in probability (0.01 = one percentage point). *Control mean* is the share of students not exposed in first grade who keep their teacher. School-section, cohort and grade fixed effects and the controls of Equation 1; standard errors clustered on the school-section in each cohort (i.e., the class group). A null coefficient indicates that, within the variation used by Equation 1, exposure does not predict teacher continuity. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table F2: Selection into teacher continuity, adding fixed effects and controls

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Mathematics teacher						
Exposed in first grade	0.0412***	0.0400***	0.0218***	0.0112***	-0.0006	-0.0000
	(0.0029)	(0.0029)	(0.0029)	(0.0031)	(0.0025)	(0.0024)
R^2	0.002	0.002	0.004	0.007	0.137	0.159
Observations	2,068,930	2,068,930	2,068,930	2,068,930	2,068,930	2,068,930
Panel B. Spanish teacher						
Exposed in first grade	0.0466***	0.0450***	0.0263***	0.0149***	0.0019	0.0030
	(0.0029)	(0.0029)	(0.0029)	(0.0030)	(0.0025)	(0.0024)
R^2	0.002	0.004	0.006	0.009	0.136	0.158
Observations	2,067,667	2,067,667	2,067,667	2,067,667	2,067,667	2,067,667
Cohort and grade FE	No	Yes	Yes	Yes	Yes	Yes
Individual controls	No	No	Yes	Yes	Yes	Yes
Class and school controls	No	No	No	Yes	Yes	Yes
School FE	No	No	No	No	Yes	No
School-section FE	No	No	No	No	No	Yes

Notes: The dependent variable is an indicator for having the same teacher as in the previous year. All columns are estimated on the same sample of students with non-missing controls. Column 6 corresponds to Table F1. Standard errors clustered on the school-section in each cohort (i.e., the class group). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

experience moderates the penalty only weakly in both subjects, whereas each year of experience with CWD is associated with a penalty smaller by about 0.7 to 1.1% of a standard deviation. The estimates are unchanged when the effect of exposure is allowed to vary by cohort, which excludes that the interaction captures the mechanical increase in experience when the same teacher takes first grade in consecutive cohorts.

F.1 Teacher looping across schools

This section compares schools with and without an established practice of teacher looping, in which teachers follow the same class across grades as a matter of school-wide practice. The comparison complements the student-level evidence on teacher continuity in Section 7.4, but it is a comparison between schools, and looping schools are a selected group.

Selection. Looping is the norm rather than the exception: 74% of the schools in our sample loop in at least one grade. Looping schools differ from other schools in ways that point to harder conditions. They are more often public (50% against 27%) and rural, and they serve poorer families: 12% of mothers have college education, against 17% in non-looping schools, and 9% of families are in the top income quintile, against 15%. They also report worse behavioral climate and more bullying. Exposure is also more common in looping schools (a mean of 0.65 against 0.57), but this association disappears once school socioeconomic composition, type and size are accounted for, whereas these school characteristics continue to predict looping. Looping is therefore a practice of poorer and more often public schools rather than a response to the arrival of CWD, but it cannot be treated as independent of school characteristics.

Results. Table F3 estimates Equation 1 separately in the two groups of schools, so that each estimate retains the within-school variation of the main design. In Mathematics, the penalty is 2.9% of a standard deviation in schools without looping and 1.8% in schools with it, about 40% smaller; in Spanish, it is 2.0% and 1.6%, about 20% smaller. In a single regression in which exposure, all controls and the cohort and grade fixed effects are interacted with the looping indicator, the difference is significant in Mathematics (1.1% of a standard deviation, $p = 0.03$) but not in Spanish (0.4%, $p = 0.40$). The penalty remains significant in looping schools, so it is smaller, but not absent, where looping is practiced. Because looping schools

face harder conditions on average, selection on these characteristics should work against finding smaller penalties there. Exposure also affects mobility differently in the two groups of schools: in schools without a looping tradition it raises the probability of changing school by 0.7 percentage points, whereas in looping schools it lowers it by 0.6 points (Table F4).

Table F3: Heterogeneity by the school's looping practice

	Pooled		2nd grade		4th grade		6th grade		8th grade	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Mathematics										
Exposed in first grade	-0.029***	-0.018***			-0.027***	-0.012**	-0.030***	-0.013*	-0.027***	-0.023***
	(0.004)	(0.003)			(0.003)	(0.005)	(0.004)	(0.007)	(0.005)	(0.007)
Observations	433,463	1,093,970			578,629	271,132	297,309	109,025	179,967	91,351
Panel B. Spanish										
Exposed in first grade	-0.020***	-0.016***	-0.026***	-0.023***	-0.023***	-0.012**	-0.015***	-0.015*	-0.010**	-0.027***
	(0.004)	(0.003)	(0.005)	(0.006)	(0.003)	(0.005)	(0.004)	(0.008)	(0.005)	(0.008)
Observations	589,307	1,467,672	310,615	227,577	558,213	287,386	311,060	92,969	186,709	82,427
School practice	No looping	Looping	No looping	Looping	No looping	Looping	No looping	Looping	No looping	Looping

Notes: The last row indicates whether each column covers schools with or without an established practice of teacher looping. A school grade is classified as looping when at least 70% of its classes keep, for most of their students, the teacher of the subject from the previous year, over the test years 2012 to 2019. In the pooled columns, a school is classified as looping if any of its grades is; in the grade-specific columns, the classification refers to that grade, so that the samples of the two sets of columns do not add up. Mathematics is not tested in second grade, so those columns are empty. Sample, controls, fixed effects and clustering as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table F4: Exposure and student mobility, by the school's looping practice

	Pooled		2nd grade		4th grade		6th grade		8th grade	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Changed school										
Exposed in first grade	0.007***	-0.006***	0.006*	-0.010***	-0.006***	-0.007***	-0.005***	-0.002	0.000	-0.002
	(0.003)	(0.001)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Observations	388,651	1,683,001	301,441	236,751	502,897	350,052	246,343	161,662	139,749	132,734
Control mean	0.090	0.090	0.100	0.100	0.100	0.090	0.090	0.080	0.080	0.080
Panel B. Changed class within school										
Exposed in first grade	0.005***	0.000	0.002	0.002	0.002**	-0.000	0.001	0.001	-0.001	-0.002*
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	388,651	1,683,001	301,441	236,751	502,897	350,052	246,343	161,662	139,749	132,734
Control mean	0.060	0.020	0.050	0.020	0.030	0.010	0.030	0.020	0.040	0.020
Panel C. Repeated grade										
Exposed in first grade	-0.000	-0.001***	-0.002**	-0.003***	-0.000	-0.000	-0.000	-0.001	0.001	-0.001
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	388,651	1,683,001	301,441	236,751	502,897	350,052	246,343	161,662	139,749	132,734
Control mean	0.010	0.010	0.020	0.020	0.010	0.010	0.010	0.010	0.020	0.010
School practice	No looping	Looping	No looping	Looping	No looping	Looping	No looping	Looping	No looping	Looping

Notes: As Table B4, by the school's looping practice, indicated in the last row. Here a school is classified as looping if it loops in either subject, whereas the classification of Table F3 is subject-specific, so that the two tables split the sample differently. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table F5: Teacher experience

	Mathematics			Spanish		
	(1)	(2)	(3)	(4)	(5)	(6)
Exposed in first grade	-0.0312***	-0.0228***	-0.0320***	-0.0249***	-0.0251***	-0.0318***
	(0.0037)	(0.0040)	(0.0049)	(0.0035)	(0.0038)	(0.0046)
× Years of experience	0.0007***		0.0006***	0.0005**		0.0004**
	(0.0002)		(0.0002)	(0.0002)		(0.0002)
× Years with CWD		0.0075**	0.0072**		0.0104***	0.0108***
		(0.0034)	(0.0035)		(0.0032)	(0.0032)
Observations	1,454,415	1,525,843	1,454,415	1,961,787	2,054,237	1,961,787

Notes: *Years of experience* is the first-grade teacher's total years of teaching. *Years with CWD* is the teacher's previous years of experience with CWD, measured over the years we observe and therefore a lower bound. Exposure and all controls are interacted with the experience measures. Columns 2 and 5 use all observations with non-missing experience with CWD, so that their samples differ from the other columns. School-section, cohort and grade fixed effects, as in Equation 1, all grades pooled. Standard errors clustered on the school-section in each cohort (i.e., the class group). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

G Classroom environment

Table G1 asks whether the association between exposure and reported bullying is larger in the classrooms of children with learning disabilities, who bear most of the achievement cost. Because the bullying index is built from items about the school, this comparison concerns the classrooms of labeled and unlabeled children rather than the experience of individual children within a classroom. The association is smaller in the classrooms of labeled children than in those of other children, and it does not differ between classrooms with a low and a high share of labeled classmates in first grade. The worse classroom environment reported in exposed classrooms is therefore similar for all children rather than concentrated on the children whose achievement falls most, which is consistent with a loss of teacher attention rather than with targeted disruption. Tables G2 to G4 list the survey items entering each classroom environment index, by grade.

Table G1: Exposure, learning disabilities and reported bullying

	(1) Own learning disability interaction	(2) Below-median learning disability share	(3) Above-median learning disability share
Exposed in first grade	0.028***	0.031***	0.027***
	(0.006)	(0.008)	(0.009)
Learning disability	0.162***		
	(0.024)		
Exposed in first grade × Learning disability	-0.013**		
	(0.006)		
Observations	720,743	405,145	314,910

Notes: The dependent variable is the standardized bullying index of the student's classroom. Column 1 interacts exposure with the student's own learning disability status and is estimated at student level, since own status varies within classrooms; since the index is built from items about the school, it compares the classrooms of labeled and unlabeled children. Columns 2 and 3 split classrooms by whether the share of classmates ever recorded with a learning disability in the first-grade class is below or above the median. Learning disability corresponds to the Chilean transitory special educational needs label, which entitles children to temporary support. As with all index regressions, these associations are descriptive. Fixed effects and clustering as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table G2: Survey items in each index, 4th grade

Index	Respondent	Questions
Class environment	Student	There is respect among classmates
		Classmates respect rules/order/instructions
	Spanish teacher	Frequency students follow rules
		Frequency listen to peers with respect
		(Inverse) Frequency disorder/fights disrupt class
		(Inverse) Frequency students interrupt class
School bullying	Student	Frequency students push/hit others
		Frequency students threaten others
		Frequency students coerce others
		Frequency students mock others
	Spanish teacher	Presence of Physical bullying in school
		Presence of Verbal bullying in school
		Presence of Media bullying in school
		Presence of Social bullying in school
		Frequency mockery between students in school
		Frequency knife assault between students in school

Table G3: Survey items in each index, 6th grade

Index	Respondent	Questions
Class environment	Student	There is respect among classmates
		Classmates respect rules/order/instructions
	Spanish/Mathematics teacher	Frequency students follow rules
		Frequency listen to peers with respect
		(Inverse) Frequency disorder/fights disrupt class
		(Inverse) Frequency students interrupt class
		Frequency students pay attention
School bullying	Student	Frequency students push/hit others
		Frequency students threaten others
		Frequency students coerce others
		Frequency students mock others
	Spanish/Mathematics teacher	Presence of Physical bullying in school
		Presence of Verbal bullying in school
		Presence of Media bullying in school
		Presence of Social bullying in school
		Frequency mockery between students in school
		Frequency knife assault between students in school
		Frequency fights between students in school
School openness	Spanish/Mathematics teacher	(Inverse) Discrimination on appearance in school
		(Inverse) Discrimination on indigenous in school
		(Inverse) Discrimination on gender in school
		(Inverse) Discrimination on immigrant in school
		(Inverse) Discrimination on physical in school
		(Inverse) Discrimination on SES in school
		(Inverse) Discrimination on religion in school
		(Inverse) Discrimination on grades/learning speed in school
		(Inverse) Discrimination on personality in school
	Parent	(Inverse) Discrimination on appearance in school
		(Inverse) Discrimination on indigenous in school
		(Inverse) Discrimination on gender in school
		(Inverse) Discrimination on immigrant in school
		(Inverse) Discrimination on physical in school
		(Inverse) Discrimination on SES in school
		(Inverse) Discrimination on religion in school
		(Inverse) Discrimination on grades/learning speed in school
		(Inverse) Discrimination on personality in school

Table G4: Survey items in each index, 8th grade

Index	Respondent	Questions
Class environment	Student	There is respect among classmates
		Classmates respect rules/order/instructions
	Spanish/Mathematics teacher	Frequency students follow rules
		(Inverse) Frequency disorder/fights disrupt class
		(Inverse) Frequency students interrupt class
		Frequency students pay attention
School bullying	Student	Physical bullying in school
		Verbal bullying in school
		Media bullying in school
		Social bullying in school
		Frequency harassment between students in school
		Frequency fights between students in school
		Frequency insults between students in school
	Student	Felt scared of harm in classroom
		Felt scared of harm in entrance
		Felt scared of harm in bathroom
		Felt scared of harm in courtyard
		Felt scared of harm in hallway
	Spanish/Mathematics teacher	Presence of Physical bullying in school
		Presence of Verbal bullying in school
		Presence of Social bullying in school
		Frequency mockery between students in school
		Frequency knife assault between students in school
		Frequency fights between students in school
		Frequency threats between students in school
School openness	Student	(Inverse) Felt discriminated on personality in school
		(Inverse) Felt discriminated on appearance in school
		(Inverse) Felt discriminated on indigenous in school
		(Inverse) Felt discriminated on gender in school
		(Inverse) Felt discriminated on immigrant in school
		(Inverse) Felt discriminated on physical in school
		(Inverse) Felt discriminated on SES in school
		(Inverse) Felt discriminated on religion in school
		(Inverse) Felt discriminated on grades in school
		(Inverse) Felt discriminated on LGBTQ in school
		(Inverse) Felt discriminated for having a baby in school
		(Inverse) Felt discriminated on political views in school
	Teacher	(Inverse) Discrimination on appearance in school
		(Inverse) Discrimination on indigenous in school
		(Inverse) Discrimination on gender in school
		(Inverse) Discrimination on immigrant in school
		(Inverse) Discrimination on physical in school
		(Inverse) Discrimination on SES in school
		(Inverse) Discrimination on religion in school
		(Inverse) Discrimination on grades in school
		(Inverse) Discrimination on personality in school
		(Inverse) Discrimination on LGBTQ in school
		(Inverse) Discrimination on having a baby in school
		(Inverse) Discrimination on political views in school

H Specialist staff and support expenditure

Two data sources describe the support that accompanies children with severe disabilities. First, the administrative teacher records report, for each teacher, the grades taught and the weekly hours, together with special-education degrees and, from 2012, contracts under the inclusion program. We define as specialists the teachers with a special-education degree or a contract under the program, and measure the specialists assigned to each school, grade and year and their weekly hours, splitting each teacher's hours equally across the grades taught. Specialists can be assigned to grades but not to classes, about 10% of them have missing grade information, and other professionals of the program, such as psychologists and speech therapists, are not covered.

Table H1 regresses three measures (the presence of any specialist, the number of specialists, the specialist hours) on first-grade exposure in the tested grades, with the fixed effects of Equation 1 and controlling for the number of sections in the grade. Exposure is not associated with more specialists or more specialist hours, whether the classmate has an intellectual disability or another type, and the estimates are precise (Panel A). In the first-grade year, exposed classes are slightly more likely to have a specialist in their grade, mainly when the classmate has an intellectual disability (Panel B), but this association does not persist into the grades in which the cohorts are tested. Nor is the penalty smaller where specialists are more abundant: splitting schools at the median of specialist hours per student in the program, observed from 2016 for 63% of the sample, the interaction with exposure is 0.2% of a standard deviation in Mathematics and 0.4% in Spanish, with standard errors of 0.6%.

Second, schools report their spending of the program's subsidies by account from 2014 onwards. Spending is pooled at the school level, so that it cannot be attributed to individual children, and between 87% and 95% of it goes to salaries. Table H2 relates school spending to the number of enrolled children with each type of need, from 2016, when types are recorded, with school and year fixed effects. Each additional child with an intellectual disability is associated with the smallest increase in salary spending among severe disabilities, about 720 thousand pesos a year, against about 1.5 million for a child with a sensory disability and 2.6 million for a child with autism, and close to the increase associated with a child with

a learning disability. With school-specific linear trends, all coefficients shrink and the difference between intellectual and sensory disabilities is no longer significant, but intellectual disability remains associated with the smallest increase. These associations are consistent with support for intellectual disabilities resting more on the class teacher than on specialized staff, but they cannot show which support each child receives.

Table H1: Specialist teachers in the cohort's grade and first-grade exposure

	Any specialist (1)	Number of specialists (2)	Specialist hours (3)
Panel A. Tested grades (2, 4, 6 and 8)			
Exposed in first grade	0.002	0.003	-0.01
	(0.001)	(0.004)	(0.05)
By type of disability:			
Intellectual disability	0.001	0.003	-0.04
	(0.002)	(0.004)	(0.05)
Other severe disabilities only	0.003	0.006	0.05
	(0.002)	(0.006)	(0.07)
Mean of the outcome	0.52	1.04	9.88
Observations (students)	2,514,865	2,514,865	2,514,865
Panel B. First grade, cohorts 2011–2014			
Exposed first-grade class	0.008	0.018*	0.12
	(0.005)	(0.011)	(0.12)
By type of disability:			
Intellectual disability	0.011*	0.020	0.10
	(0.006)	(0.012)	(0.13)
Other severe disabilities only	-0.000	0.013	0.16
	(0.007)	(0.017)	(0.19)
Mean of the outcome	0.44	0.79	6.92
Observations (classes)	34,249	34,249	34,249

Notes: Specialists are teachers with a special-education degree or a contract under the inclusion program, assigned to the school, grade and year (teacher records, 2011–2019); hours are weekly, with each teacher's hours split equally across the grades taught. Panel A: one observation per student and tested grade; fixed effects for school-section, cohort and grade, and the number of sections in the grade; standard errors clustered by first-grade class. The type rows come from one regression with indicators for exposure to a classmate with an intellectual disability and for exposure to other severe disabilities only. Panel B: one observation per first-grade class; fixed effects for school-section and cohort, and the number of sections; standard errors clustered by school and cohort. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table H2: Spending of the inclusion program and enrolled children by type of need

	Total	Salaries	Materials	Equipment	Diagnosis
	(1)	(2)	(3)	(4)	(5)
Number of enrolled children with:					
Intellectual disability	931***	722***	50***	45***	19***
	(69)	(68)	(8)	(6)	(3)
Autism	2,418***	2,563***	-27**	-48***	28***
	(120)	(120)	(13)	(10)	(6)
Sensory disability	1,411***	1,507***	-6	-8	25
	(383)	(373)	(35)	(32)	(17)
Motor disability	1,126**	1,104**	-24	-8	34
	(487)	(476)	(45)	(46)	(23)
Multiple disabilities	1,881***	1,799***	-6	-16	-10
	(396)	(393)	(47)	(47)	(19)
Learning disability	957***	859***	25***	16***	13***
	(30)	(29)	(3)	(3)	(1)
p-value, intellectual = sensory	0.217	0.038	0.122	0.100	0.702
Mean spending per school-year	110,773	103,554	1,822	1,259	1,102
School-years	42,484	42,484	42,484	42,484	42,484
Schools	5,816	5,816	5,816	5,816	5,816

Notes: One observation per school and year, 2016–2023. The dependent variables are the school's spending of the program's subsidies, in thousands of 2023 pesos, winsorized at the 1st and 99th percentiles within year. Regressors are the numbers of enrolled children with each type of need in the year, jointly; controls for enrollment and for program students outside primary and secondary education; school and year fixed effects; standard errors clustered by school. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

I Movers and the definition of school-sections

The school-section fixed effects of Equation 1 are defined on the section in which children are observed in each tested grade, whereas exposure is defined in the first-grade class. The two coincide for children who remain in their first-grade school and section, but not for children who change section or school after first grade. The share of observations no longer in their first-grade school-section is 17% in second grade, 34% in fourth grade, 44% in sixth grade and 56% in eighth grade, and the share no longer in their first-grade school rises from 13% to 48%. In the main specification, a child who moved is therefore compared with the other children of the new section and cohort, who started first grade in other classes and often in other schools. If some sections attract children who join them after first grade and carry over influences from their previous schools, and these influences are correlated with exposure in their first-grade class, the estimates reflect these correlations.

To assess this concern, we re-estimate the analyses replacing the school-section fixed effects with fixed effects for the first-grade school-section, defined by the school and the class letter in which each child attended first grade, and cluster standard errors on the first-grade class, within which exposure is constant. Each child is then compared only with the children who started in the same first-grade school-section in other cohorts, regardless of what happened to them afterwards, which is instead captured in the main analysis. All other elements of the specifications are unchanged.

We keep the tested-grade sections in the main specification because movers are part of the setting we study: they carry their own first-grade exposure into a new section, with a new teacher and the new section's arrangements for CWD, and fixed effects for the tested-grade section hold fixed the set of teachers and the section-specific policies and support that children face when they are tested. The cost of this choice is that the main specification cannot show how movers affect the estimates, which is what this section assesses.

Table I1 reports the main estimates. The penalty falls from 2.1% to 1.4% of a standard deviation in Mathematics and from 1.7% to 1.3% in Spanish, both significant at the 1% level, and it declines over primary school: in Mathematics from 1.6% in fourth grade to 0.8% in eighth grade, significant at the 10% level, and in Spanish from 1.9% in second grade to

0.2% in eighth grade, not significant. The difference between the two specifications grows with grade, as does the share of children who have moved. The number of CWD keeps a negative and significant coefficient, and the unconditional quantile effects keep their shape, with the largest losses at and below the median and no effect at the 90th percentile.

Table I2 reports the heterogeneity analyses. Intellectual disability still drives the effect, at between 1.5 and 1.6% of a standard deviation in both subjects. Autism becomes significant, at 0.6% in Mathematics (at the 10% level) and 0.7% in Spanish (at the 5% level), and the other categories remain close to zero. When exposure to intellectual disability is added to the main specification, its coefficient is 1.1% in Mathematics and 1.3% in Spanish, and the residual effect of the other disabilities is not significant. The diversity of disability types remains irrelevant. The cost still falls mostly on classmates with learning disabilities, who lose 3.0% of a standard deviation in Mathematics and 3.6% in Spanish, against 0.8% and 0.6% for other classmates, about 3.6 and 5.7 times as much. The gradients by socioeconomic disadvantage become steeper: SEP priority students and children in the bottom income quintile lose between two and three times as much as other classmates, still less than the gradient by special-needs status.

Table I3 documents what movers bring with them, splitting observations into children who remain in their first-grade school-section (66%), children who change section within their first-grade school (6%), and children who change school (28%). Panel A regresses background characteristics and the composition of the first-grade school on exposure, with the fixed effects of Equation 1. For children who remain in their section, exposure is unrelated to background, with differences of at most a few tenths of a percentage point, and the same holds, with slightly larger differences, for children who change section within the school. For children who change school, instead, exposure in first grade signals a more disadvantaged origin: exposed school movers are 5.9 percentage points more likely to be SEP priority students and 3.5 points less likely to have a college-educated mother, and their first-grade schools have 9.9 points more families in the bottom income quintile. Because the fixed effects of Equation 1 refer to the destination section, they do not absorb these differences, which are only captured by observable individual controls. Panel B estimates Equation 1

separately for the three groups: the penalty is 1.6% of a standard deviation in Mathematics and 1.5% in Spanish for children who remain in their section, similar for children who change section within the school, and 3.9% and 3.7% for children who change school, close to the estimates with first-grade fixed effects for the first group and consistent with origin disadvantage inflating the estimates for the last.

Table I4 reports the teacher analyses. With first-grade fixed effects, continuity is still associated with a much smaller penalty in Mathematics, 0.7% of a standard deviation among children who keep their teacher against 2.0% among those who change teacher, about 65% smaller, whereas in Spanish the difference shrinks to 1.1% against 1.4%, about 16% smaller. Exposure still does not predict continuity. Penalties are no longer smaller in looping schools, and the association with the first-grade teacher's experience with CWD is no longer significant, while the association with general experience is similar, at 0.05% of a standard deviation per year in both subjects. Note, however, that teacher experience and looping practices are school-level policies that are more relevant to analyze at the destination school for movers, that is, under the main specification. In addition, moving school may be the result of teacher-related policies at the origin school, so that controlling for them via first-grade fixed effects is endogenous to moving itself.

Panel D restricts the main specification to children who remain in their first-grade school-section, for whom the two definitions of the section coincide. Their penalty is 1.6% of a standard deviation in Mathematics and 1.5% in Spanish, against 3.6% and 3.0% for children who moved, and close to the pooled estimates with first-grade fixed effects, so that the larger penalty of movers in the main specification likely reflects the children with whom they are compared. Among children who remain, continuity is associated with a penalty about 60% smaller in Mathematics (0.9% against 2.3%) and 35% smaller in Spanish (1.2% against 1.9%), and penalties are about a quarter smaller in looping schools in Mathematics and about 30% smaller in Spanish.

The remaining analyses of the paper, including the balance tests, the dilution and benchmark calculations and randomization inference, use the specification of Equation 1.

Table I1: Main estimates with fixed effects for the first-grade school-section

School-section measured in	Mathematics		Spanish	
	Tested grade (1)	First grade (2)	Tested grade (3)	First grade (4)
Panel A. Exposure				
All grades	-0.021***	-0.014***	-0.017***	-0.013***
	(0.002)	(0.003)	(0.002)	(0.003)
Grade 2			-0.025***	-0.019***
			(0.004)	(0.004)
Grade 4	-0.022***	-0.016***	-0.019***	-0.013***
	(0.003)	(0.003)	(0.003)	(0.003)
Grade 6	-0.026***	-0.013***	-0.015***	-0.009**
	(0.004)	(0.004)	(0.004)	(0.005)
Grade 8	-0.026***	-0.008*	-0.015***	-0.002
	(0.004)	(0.005)	(0.004)	(0.005)
Panel B. Intensity of exposure				
Number of children with a severe disability	-0.014***	-0.010***	-0.012***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
Squared	0.001***	0.000	0.001*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Share of children with a severe disability	-0.392***	-0.328***	-0.315***	-0.278***
	(0.040)	(0.052)	(0.039)	(0.048)
Squared	-0.017	0.540**	-0.163	0.273
	(0.163)	(0.211)	(0.164)	(0.198)
Panel C. Unconditional quantile effects				
10th percentile	-0.034***	-0.023***	-0.018***	-0.011***
	(0.004)	(0.004)	(0.004)	(0.004)
25th percentile	-0.025***	-0.018***	-0.023***	-0.018***
	(0.003)	(0.004)	(0.003)	(0.004)
50th percentile	-0.021***	-0.014***	-0.021***	-0.017***
	(0.003)	(0.004)	(0.003)	(0.003)
75th percentile	-0.019***	-0.012***	-0.017***	-0.012***
	(0.003)	(0.004)	(0.003)	(0.003)
90th percentile	-0.009**	-0.001	-0.008***	-0.004
	(0.003)	(0.004)	(0.003)	(0.003)

Notes: Columns 1 and 3 reproduce the estimates of the paper, with fixed effects for the school-section in the tested grade and standard errors clustered on the school-section in each cohort (i.e., the class group). Columns 2 and 4 replace them with fixed effects for the first-grade school-section and cluster standard errors on the first-grade class. Panel A reproduces Table 1, Panel B Table 2 and Panel C Table 4. Cohort fixed effects, grade fixed effects in the pooled specifications, and controls as in Table 1. Mathematics is not tested in second grade. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table I2: Heterogeneity with fixed effects for the first-grade school-section

	Mathematics		Spanish	
School-section measured in	Tested grade	First grade	Tested grade	First grade
	(1)	(2)	(3)	(4)
Panel A. Type of disability				
Intellectual	-0.026***	-0.015***	-0.024***	-0.016***
	(0.002)	(0.003)	(0.002)	(0.003)
Autism	-0.004	-0.006*	-0.003	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)
Motor	-0.000	-0.001	-0.006	-0.009
	(0.007)	(0.008)	(0.006)	(0.007)
Visual	-0.016	-0.024**	-0.010	-0.013
	(0.010)	(0.012)	(0.010)	(0.011)
Hearing	-0.002	-0.001	0.006	0.006
	(0.008)	(0.009)	(0.007)	(0.008)
Multiple	-0.003	0.001	0.001	-0.002
	(0.007)	(0.008)	(0.006)	(0.007)
Severe speech	0.011	0.013	0.014*	0.014*
	(0.008)	(0.009)	(0.007)	(0.008)
Unspecified	0.001	-0.000	0.003	0.002
	(0.004)	(0.005)	(0.004)	(0.004)
Panel B. Intellectual disability and residual exposure				
Intellectual disability	-0.023***	-0.011***	-0.024***	-0.013***
	(0.004)	(0.004)	(0.003)	(0.004)
Any other severe disability	-0.004	-0.006	-0.000	-0.004
	(0.004)	(0.004)	(0.003)	(0.004)
Panel C. Diversity of disability types				
Exposed in first grade	-0.021***	-0.013***	-0.017***	-0.012***
	(0.002)	(0.003)	(0.002)	(0.003)
Diversity (1 − HHI)	-0.005	-0.004	-0.005	-0.008
	(0.006)	(0.007)	(0.006)	(0.006)
Panel D. Incidence: effect of exposure by the peer's own status				
No learning disability	-0.014***	-0.008***	-0.009***	-0.006**
	(0.002)	(0.003)	(0.002)	(0.003)
Learning disability	-0.037***	-0.030***	-0.042***	-0.036***
	(0.005)	(0.005)	(0.005)	(0.005)
Not SEP priority	-0.018***	-0.009***	-0.011***	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)
SEP priority	-0.023***	-0.020***	-0.023***	-0.020***
	(0.003)	(0.004)	(0.003)	(0.003)
Not bottom income quintile	-0.021***	-0.009***	-0.015***	-0.008***
	(0.003)	(0.003)	(0.003)	(0.003)
Bottom income quintile	-0.024***	-0.021***	-0.023***	-0.020***
	(0.003)	(0.004)	(0.003)	(0.004)
Mother with college	-0.023***	-0.009	-0.008*	-0.001
	(0.005)	(0.006)	(0.005)	(0.005)
Mother without college	-0.022***	-0.015***	-0.020***	-0.015***
	(0.002)	(0.003)	(0.002)	(0.003)

Notes: Fixed effects and clustering as in Table I1. Panel A reproduces Figure 1 and Table E1, Panel B Table E2 and Panel C columns 1 and 4 of Table 3. Panel D reports the effect of exposure outside and inside each group, from regressions in which exposure and all controls are interacted with the group indicator, as in Table 5 and Table E3. Learning disability corresponds to the Chilean transitory special educational needs label, which entitles children to temporary support. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table I3: Movers: background and effect of exposure by mobility status

	Stayers (1)	New section, same school (2)	Other school (3)
Panel A. Exposure and background			
SEP priority	0.002	0.006	0.059***
	(0.002)	(0.004)	(0.002)
Mother with college	-0.001	-0.008***	-0.035***
	(0.001)	(0.003)	(0.001)
Bottom income quintile	0.001	0.008*	0.041***
	(0.001)	(0.004)	(0.001)
Origin school: share SEP priority	0.003**	0.001	0.088***
	(0.001)	(0.002)	(0.001)
Origin school: share mother with college	-0.001**	-0.001*	-0.061***
	(0.001)	(0.001)	(0.001)
Origin school: share bottom quintile	-0.001	-0.000	0.099***
	(0.001)	(0.001)	(0.001)
Panel B. Effect of exposure on test scores			
Mathematics	-0.016***	-0.024**	-0.039***
	(0.004)	(0.010)	(0.003)
Observations	929,411	105,305	492,387
Spanish	-0.015***	-0.015*	-0.037***
	(0.003)	(0.009)	(0.003)
Observations	1,376,319	124,245	556,118

Notes: Observations split by whether the child is tested in the first-grade school and section (column 1), in the first-grade school but a different section (column 2), or in a different school (column 3). Panel A regresses each characteristic on exposure; origin-school shares are means over the peers of the child's first-grade school and cohort. Panel B estimates Equation 1 in each group. All regressions include the school-section, cohort and grade fixed effects of Equation 1, with standard errors clustered as in Table 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table I4: Teacher continuity and experience with fixed effects for the first-grade school-section

School-section measured in	Mathematics		Spanish	
	Tested grade (1)	First grade (2)	Tested grade (3)	First grade (4)
Panel A. Teacher continuity				
Changes teacher	-0.033***	-0.020***	-0.025***	-0.014***
	(0.003)	(0.003)	(0.003)	(0.003)
Keeps teacher	-0.009***	-0.007*	-0.011***	-0.011***
	(0.003)	(0.004)	(0.003)	(0.003)
Same teacher on exposure	-0.000	-0.001	0.003	0.000
	(0.002)	(0.003)	(0.002)	(0.003)
Panel B. Looping				
Non-looping schools	-0.029***	-0.011**	-0.020***	-0.011**
	(0.004)	(0.005)	(0.004)	(0.005)
Looping schools	-0.018***	-0.015***	-0.016***	-0.013***
	(0.003)	(0.003)	(0.003)	(0.003)
Panel C. First-grade teacher experience				
Exposed in first grade	-0.031***	-0.021***	-0.025***	-0.020***
	(0.004)	(0.004)	(0.003)	(0.004)
× years of experience	0.0007***	0.0005**	0.0005**	0.0005**
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Exposed in first grade	-0.023***	-0.010**	-0.025***	-0.016***
	(0.004)	(0.005)	(0.004)	(0.004)
× years with CWD	0.0075**	-0.0020	0.0104***	0.0042
	(0.0034)	(0.0040)	(0.0032)	(0.0036)
Panel D. Main specification, by whether the child remains in the first-grade school-section				
Remains in first-grade school-section	-0.016***		-0.015***	
	(0.004)		(0.003)	
Moved	-0.036***		-0.030***	
	(0.003)		(0.003)	
Remains, changes teacher	-0.023***		-0.019***	
	(0.005)		(0.004)	
Remains, keeps teacher	-0.009*		-0.012***	
	(0.005)		(0.004)	
Remains, non-looping school	-0.020***		-0.019***	
	(0.007)		(0.006)	
Remains, looping school	-0.015***		-0.013***	
	(0.004)		(0.004)	

Notes: Fixed effects and clustering as in Table I1. Panel A reproduces Table 7 (pooled) and the selection test of Table F1, Panel B Table F3 (pooled) and Panel C Table F5, with all controls interacted with continuity or experience as in those tables. Panel D estimates the specification of the paper separately for children who remain in their first-grade school-section and for children who moved, and reports teacher continuity and looping among children who remain. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.