Economics of Education Review 103 (2024) 102582

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Are separate classrooms inherently unequal? The effect of within-school sorting on the socioeconomic test score gap in Hungary

Zoltán Hermann [a,b,*], Hedvig Horváth [c,d], Dorottya Kisfalusi [e,f]

[a] Institute of Economics, Centre for Economic and Regional Studies, Tóth Kálmán u. 4, Budapest 1097, Hungary
[b] Institute of Economics, Corvinus University Budapest, Fővám tér 8, Budapest 1093, Hungary
[c] University College London–Social Research Institute, 55-59 Gordon Square, London WC1H 0NU, UK
[d] CESifo Research Network, Germany
[e] HUN REN Centre for Social Sciences, Tóth Kálmán u. 4, Budapest 1097, Hungary
[f] Institute for Analytical Sociology, Linköping University, Norrköping 601 74, Sweden

ARTICLE INFO

JEL codes:
I21
I24

Keywords:
Achievement gap
Inequalities
Non-merit-based sorting
Segregation
Within-school sorting

ABSTRACT

This study investigates whether within-school sorting increases socioeconomic test score inequalities. Using universal test score data on 6th- and 8th-grade students in Hungary, we document the extent of within-school sorting in an institutional context where sorting based on ability or prior achievement is rare. We identify sorting schools as schools that systematically assign students with low and high socioeconomic status into different classrooms within the school. Then, exploiting school fixed effects and quasi-exogenous variation in sorting induced by enrollment and class size rules, we show that sorting has a significant and economically meaningful effect on test score inequalities between students from different socioeconomic backgrounds. Sorting harms low-status students, while high-status students gain much less, if anything, from attending sorting schools. We attribute our findings to the within-school reallocation of educational resources and differences in educational practices.

1. Introduction

Segregation of students with respect to social, racial or ethnic background is a persistent problem in education, often associated with far-reaching implications for inequalities and social mobility. The prevalence of between-school segregation (e.g., Gutiérrez et al. 2020) and its effects on educational inequalities (e.g., Reardon et al. 2022) has been widely documented. At the same time, the separation of students within schools is also common. One specific form of this is formal tracking or ability grouping within schools. Higher track placement is widely associated with higher achievement gains, and tracking is closely related to inequalities (see the review by Betts, 2011).

Ability grouping is not, however, the only source of within-school sorting; there is ample anecdotal evidence that non-merit-based sorting across classrooms does happen, as well. Some parents are very assertive to get the teacher they want for their kids, and in the absence of pay differentiation, some teachers will be rewarded by classroom assignments with easier-to-teach children. Such practices result in separate learning environments for students with different backgrounds, yielding unequal educational outcomes to the detriment of children who start out disadvantaged anyway.

Non-merit-based sorting raises even stronger fairness concerns than tracking on ability or achievement. To the extent that sorting goes together with unequal access to educational resources, the socioeconomic achievement gap will widen. Since the non-merit-based, typically non-transparent processes through which students are sorted within schools are more likely to occur at the early stages of education, any resulting achievement gains or losses will accumulate over time and have long-lasting effects.

In this paper, we investigate whether within-school sorting on socioeconomic status increases 8th-grade achievement inequalities in a European setting, Hungary. In our institutional context, formal ability tracking before 8th grade is rare. Students are assigned to classes in 1st grade and most often remain in the same class until 8th grade.[8] We

* Corresponding author at: Institute of Economics, Centre for Economic and Regional Studies, Tóth Kálmán u. 4, Budapest 1097, Hungary.
E-mail address: hermann.zoltan@krtk.hun-ren.hu (Z. Hermann).

[8] A caveat to this is that some academic secondary schools provide highly selective eight- or six-year-long academic programs, starting in 5th or 7th grade. Less than 10% of students attend these schools though, and we exclude them from our analysis.

https://doi.org/10.1016/j.econedurev.2024.102582
Received 29 September 2023; Received in revised form 16 August 2024; Accepted 29 August 2024

document that sorting on socioeconomic status is prevalent in schools where classrooms are notionally the same, and show that sorting on socioeconomic status results in widening achievement gaps.

Using a universal dataset covering nine recent cohorts of 6th and 8th grade (12–15 year-old) students in Hungary, we first measure the prevalence of within-school sorting. Building on Lefgren (2004), Clotfelter et al. (2006) and Horváth (2015), we use a data-driven statistical procedure to classify schools, separately in every year, into two groups: those who sort students into different classrooms based on socioeconomic status (sorting schools) and those who randomly assign them (non-sorting schools). To identify sorting schools, we regress students' socioeconomic status on a set of dummy variables indicating classrooms, separately for each school. If the joint test of the classroom dummies is statistically significant, we classify the school as a sorting school in the given year. As a robustness check, we also use the adjusted R-squared of the above regression as a continuous measure of sorting.

We construct the sorting metric using 6th-grade classrooms, which are mostly the same as initial classrooms at starting school, at age 6. We measure socioeconomic background using a continuous index of parental education and the financial circumstances of the family that have likely been stable since school entry. We show that within-school sorting is prevalent in Hungarian schools with multiple classrooms per year: about 28 % of all 6th-grade students attend schools in each year that assign higher- and lower-status kids on average into different classrooms.

Once equipped with measures of schools' sorting practices, we turn to our main focus: the effect of sorting on educational inequalities. We compare the math and reading test score gap between high and low socioeconomic status students in a sorting and non-sorting school environment in 8th grade. To acknowledge and handle the potential endogeneity of why and when schools engage in sorting, we embellish our models in two ways. First, we control for unobserved, time-invariant heterogeneity at the school level using fixed effects. Second, in an instrumental variable framework, we exploit quasi-random variation in schools' sorting practices over time induced by changes in enrollment. We use two variants of the instrument, both driven by enrollment: the actual number of classrooms on the one hand, and the predicted number of classrooms derived from enrollment and legally binding class size rules on the other. Both of these instruments are strong, suggesting that schools where the number of students force them to open (close down) classrooms will be more likely to start (stop) sorting students with different socioeconomic backgrounds into separate classrooms.

Our instrumental variable approach exploits the fact that with a larger number of classes, schools have more opportunity for sorting. A potential concern about this strategy is that school size might have an effect on test scores independently from sorting. We provide several robustness checks to address this concern; we exclude one-classroom schools, which, by definition, cannot sort, and control for total enrollment or the actual number of classrooms in fixed effects models. These robustness checks confirm that our IV results are not driven by school size effects.

Our main results show significant and economically meaningful sorting effects on test score inequalities between students with different socioeconomic backgrounds. Students with a 1 standard deviation higher socioeconomic status score 11–38 % (7–23 %) higher in math (reading) than their low-status peers if learning in a sorting rather than a non-sorting school environment. It is reassuring that we see the same qualitative pattern of results for all specifications. Our results are robust to alternative sorting classifications or for a continuous measure of sorting, measuring family background by parental education instead of a composite index, and to an extended set of controls. In addition, heterogeneity analysis shows that low-status students are significantly harmed by sorting, while the top of the socioeconomic status distribution benefit much less, if anything, from it. Finally, we provide suggestive evidence that within-school reallocation of educational resources (e.g. matching higher quality teachers to higher status classrooms) and differences in educational practices (e.g. more and more challenging homework) underlie our findings.

We interpret the results as primarily reflecting the impact of non-merit-based sorting. Although in our main dataset, we cannot empirically separate out non-merit-based sorting from merit-based sorting, we believe that what we measure is mostly non-merit-based for two reasons. First, the legal environment does not allow for merit-based selection into classes in 1st grade. Since entrance exams are banned, schools can hardly observe ability, while information on family background is more easily available. Second, although additional survey evidence suggests that merit-based sorting, i.e. ability grouping, occurs in some schools in higher grades, we show that our results are not driven by that.

Our paper is most closely related to the few previous studies that address the effects of informal within-school sorting, that is, sorting when classrooms are notionally the same in terms of curriculum and educational practices. Collins and Gan (2013) have found that elementary schools in Dallas Independent School District use various practices to sort students between classes based on previous test scores, gifted or special educational needs status, or limited English proficiency. The authors have shown that both low- and high-achieving students benefit from homogeneous classes based on previous test scores. Similarly, Ferrer-Esteban (2016) and Agasisti and Falzetti (2017) have found that some Italian junior secondary schools tend to sort students between classes based on socioeconomic status. In contrast to the results from Dallas, however, informal sorting within Italian schools has a negative effect on student achievement and contributes to educational inequalities. Overall, the results on the effects of informal within-school sorting are mixed. At the same time, none of the papers above directly estimate the unconditional effect of sorting on socioeconomic inequalities.[h]

Our contributions are twofold. First, we estimate the effects of within-school sorting in an institutional context where there is no formal sorting on ability or prior achievement, since it is legally not allowed when classrooms are formed in first grade. In addition, there is essentially no room for informal ability grouping since ability is not observed at starting school.[i] Second, we use various identification strategies, exploiting school fixed effects and novel instrumental variables, to address the endogeneity of schools' sorting decisions. Previous papers exploited essentially between-school variation in sorting and did not deal with unobserved differences across schools,[j] potentially correlated with both sorting and student achievement. We build on the approach of Colllins and Gan (2013) and Agasisti and Falzetti (2017) but our multifaceted identification strategies use within-school variation in sorting practices and novel IVs, eliminating the omitted variable bias between schools.

[h] Collins and Gan (2013) do not observe socioeconomic status, they estimate the heterogeneous effect of sorting on high and low achievers only. Agasisti and Falzetti (2017) estimate heterogeneous effects by prior achievement and socioeconomic status simultaneously. Therefore, they estimate the heterogenous effect of sorting by socioeconomic status *conditional on* its heteronegenous effect by prior achievement. What they find is not easy to interpret and calls for more investigation: higher status students benefit from sorting, while those with higher prior achievement lose out.

[i] The institutional contexts are considerably different in previous studies. In the US, classrooms are tested and reshuffled every year, while in Italy, students transfer schools at age 11 and new classrooms are formed. Therefore, in both settings, there is room for formal or informal ability grouping. In contrast with common claims for the US (see e.g., Clotfelter et al. 2021), in Hungary, sorting on socioeconomic status is not simply a byproduct of ability grouping.

[j] Collins and Gan (2013) and Agasisti and Falzetti (2017) use sorting in adjacent year/grade as an instrumental variable to address the endogeneity of sorting. This way they remove one endogeneity problem: schools' sorting decision based on composition of the entering cohort. However, this way they rely on between-school variation in sorting, as the compliers are essentially schools always and never sorting.

More broadly, our work is also related to other strands of the literature. As mentioned earlier, an extensive line of literature studies between-school segregation and its impact on educational inequalities (see e.g. Benito et al. 2014 on OECD countries; and the detailed US literature review in Reardon et al. 2022). At the same time, consequences of within-school segregation have received less attention, despite the fact that evidence for within-school sorting based on students' race, immigrant status, and socioeconomic status has been observed in the US, as well as in European countries (Clotfelter et al., 2002, 2006; Conger, 2005; Engzell & Raabe, 2023; Morgan & McPartland, 1981). Our results can be interpreted in a within-school segregation framework. The major difference is that the segregation literature usually focuses on a particular disadvantaged group of students, like racial minorities or poor students, while we measure sorting along the entire spectrum of socioeconomic status.

Second, our work is also related to the comparative analysis of educational institutions and their role in educational inequalities. Focusing on the macro-level association between tracking[k] and educational inequalities, this literature finds that early tracking increases the effect of family background on student achievement (Ammermueller, 2005; Horn, 2009; Schütz et al., 2008) and educational attainment (Brunello & Checchi, 2007). Engzell and Raabe (2023) focus on four European countries and find that at age 15, within-school sorting is larger in comprehensive systems than in systems characterized by early tracking. This suggests that non-merit-based, typically informal within-school sorting might be an important sorting mechanism in educational systems where tracking on ability is introduced at later stages.

The rest of our paper is organized as follows. Section 2 lays out the conceptual framework: after pinning down definitions of different types of sorting practices that are key to this paper, we discuss rationales for sorting and through what mechanisms it may result in unequal educational achievement. Section 3 gives an overview of the Hungarian institutional context. In Section 4, we describe the data and thoroughly explain our 2-step empirical strategy. First, how we classify schools based on their sorting practices, and second, the estimation of the effects of sorting on test scores. Section 5 presents the main results and their robustness, Section 6 discusses whether ability grouping can explain our results, while Section 7 explorers a potential mechanism behind the sorting effect. Section 8 concludes.

2. Conceptual framework

2.1. A typology of sorting practices

At the focus of this paper is within-school sorting and its effect on test score inequalities in an institutional context where sorting is mostly informal and non-merit-based. For the purposes of this focus, we consider two dimensions of within-school student sorting practices. These are summarized in Table 1, along with some widely familiar examples.

The first dimension differentiates whether the sorting takes place on some *merit-based* vs. *non-merit-based* student characteristics. We define sorting as merit-based if it takes into account academic achievement or academic skills (e.g., sorting is based on a test or prior academic achievement). In contrast, non-merit-based sorting occurs when students already admitted to a school are allocated to different classrooms based on characteristics other than their current academic achievement or skill level. Commonly used such characteristics range from some easily observed characteristics such as age, gender or ethnicity to hard-to-judge traits such as past behavior or some element of socioeconomic status (Paufler & Amrein-Beardsley, 2014).

Table 1
A typology of sorting practices and examples.

	Merit-based	Non-merit based
Formal	Gifted classrooms; test-based ability grouping (e.g. advanced class admission)	(Historical) de jure racial/ethnic segregation; single-sex education
Informal	Unannounced re-mixing of classrooms based on past achievement test/GPA/teacher assessment	De facto (but now mostly illegal) racial/ethnic segregation; parents choosing teachers; other non-transparent, non-random classroom allocation

Notes: Merit-based sorting takes into account academic achievement or academic skills (e.g., sorting is based on a test or prior academic achievement). Non-merit-based sorting occurs when students already admitted to a school are allocated to different classrooms based on characteristics unrelated to their current academic achievement or skill level. Formal sorting is the within-school allocation of students into classrooms that are labelled differently in published school documents. Informal sorting is the non-random and non-transparent allocation of students into classrooms that are notionally the same in terms of curriculum and resources.

The second dimension, analogously to Triventi et al.'s (2020) typology for differentiation in secondary education, distinguishes between *formal* and *informal* sorting. By formal sorting, we mean the within-school allocation of students into classrooms that are labelled differently in publicly available school documents. These classrooms may have slightly different curricula and teaching practices, and their distinctiveness is transparent to parents before the placement of their children. Real-world examples include gifted classrooms and classrooms with special curricula. In contrast, informal sorting is the non-random and non-transparent allocation of students into classrooms that are notionally the same in terms of curriculum and resources (except for who the actual teachers are).

The above two dimensions yield a conceptually clear typology of sorting practices. However, in practice, the four cells are often muddled for researchers: researchers may not observe the test results of a merit-based system or even whether there had been an achievement test administered to allocate students. Also, although in contexts of formal sorting, the distinctiveness of classrooms may be transparent to parents but not observed by researchers. Therefore, institutional features may play an important role in determining what research design is the most appropriate to identify the effects of what kind of sorting.

2.2. Why does within-school sorting occur?

Why do schools sort students across classrooms even in the absence of formal tracking or ability grouping? In many educational contexts, as in Hungary, formal and merit-based student sorting at the start of primary education is against the law. Moreover, in the absence of preschool testing, schools cannot observe the students' academic skills when they decide about classroom assignments, while they usually have some, even if vague, information on applicants' social backgrounds. Due to the correlation between socioeconomic status (SES) and achievement, informally sorting based on SES still results in more homogenous classrooms in terms of student achievement.

Homogenous classrooms can be beneficial for schools for several reasons. First, they allow teachers to tailor the instruction to the specific needs of the students (Collins & Gan, 2013; Duflo et al., 2011). Following a single curriculum in a class might be more effective than differentiating the curriculum based on the needs of low-and high-ability students. Second, most teachers prefer classrooms with high-ability students, where teaching is less challenging. School principals can reward their most experienced teachers and decrease their turnover by assigning them to high-ability classrooms (Kalogrides et al., 2013; Kalogrides & Loeb, 2013). Third, by creating homogeneous

[k] Tracking in this literature refers to the selection of students into different education programs (e.g. academic, vocational or some combination of these), following different curricula and leaving very limited opportunity to move from one track to another.

classrooms based on students' SES, schools can avoid high-SES parents moving to another school district or enrolling their children into another school (Clotfelter et al., 2005; Kalogrides & Loeb, 2013; Berényi et al., 2008).

Advantaged parents might also exert influence on class assignments to ensure that their child is assigned to a class with the desired teacher or student composition (Agasisti & Falzetti, 2017; Clotfelter et al., 2006; Ferrer-Esteban, 2016; Kalogrides & Loeb, 2013; Player, 2010). For instance, in the US, in many schools, parents have the opportunity to write classroom placement letters. Even if an explicit teacher request is not allowed, many internet resources provide tips about how to write the letter in a way that one gets the teacher they want for their child (The Educators Spin On It, 2024). In general, socioeconomically advantaged parents tend to ensure that their children receive higher quality education than children of less advantaged families, even if education becomes universal. This is achieved by choosing a better school, more advanced courses within the school, or influencing the class assignment process to have the most qualified teachers and high-achieving peers for their children (Lucas, 2001).

Whether schools engage in sorting thus depends on the heterogeneity of the student population, the number of other available schools in the local school market, and the extent to which school principals comply with the law. Schools with a more heterogeneous student population and those facing stronger competition for students and teachers might be more inclined to sort students across classrooms (Clotfelter et al., 2021; Player, 2010).

2.3. *How might within-school sorting increase educational inequalities?*

Within-school sorting based on student SES can increase the socioeconomic gradient in academic achievement through three main mechanisms: peer effects, educational resource allocation, and tailored instruction effects. First, due to the high correlation between SES and achievement, low-SES students are more likely to attend classes with low-ability peers, while high-SES students are more likely to have high-ability classmates. In case of sorting, therefore, high-SES students might benefit from their peers' higher average achievement levels and the higher academic standards set by the teachers, while low-SES students might lack these positive effects (e.g., Duflo et al. 2011, Lefgren 2004, and Sacerdote 2011 for a review). Furthermore, low-SES students might face the adverse effects of a higher concentration of disruptive classroom behavior (Lazear, 2001) if teachers are not prepared to tailor the instruction to the needs of students from disadvantaged social backgrounds.

Second, sorting might influence student achievement through differential resource allocations within the school (Betts & Shkolnik, 2000), such as student-teacher matching. Empirical studies from the US have found that within schools, more qualified and experienced teachers are more likely to be assigned to classes with higher-achieving students (Clotfelter et al., 2006; Kalogrides et al., 2013). Moreover, Isenberg et al. (2022) find similar positive matching between students with socioeconomic characteristics traditionally associated with achievement and teacher effectiveness measured by value-added scores. Some have speculated that what underlies such patterns is that more effective or experienced teachers are in a more powerful position to enforce their desired classroom assignments, or because high-SES parents are more likely to intervene in the assignment process (Kalogrides et al., 2013; Player, 2010).

Third, sorting allows for instruction tailored to the specific needs of low- and high-achieving students (Collins & Gan, 2013; Duflo et al., 2011). If tailored instruction is similarly beneficial for both groups of students, this mechanism can increase overall performance but has no effect on inequalities. If tailored instruction benefits either high- or low-achieving students more, it may increase or decrease inequalities. The results of the field experiment of Duflo et al. (2011) show that low-achievers gained at least as much as high-achievers by merit-based sorting into separate classrooms, through tailored instruction effects outweighing peer effects among low-achievers in an environment where educational resources were otherwise randomly allocated.

3. Institutional setting

In Hungary, compulsory education consists of two main phases. While primary schools provide education in 1st to 8th grade (from age 6/7 to 14/15), upper secondary education encompasses 9th to 12th (or 13th) grades. The upper secondary level is characterized by a stratified between-school tracking system.[l] Admission to higher-prestige secondary tracks and schools is merit-based and depends on students' academic achievement on the one hand and the results of an admission exam on the other. Because of merit-based selection into secondary schools, parents perceive the quality of primary education and, thus, primary school choice as an important determinant of their children's educational success.

Primary education in Hungary is characterized by a mixture of residence-based catchment areas and free school choice. This means that public schools[m] are required to enroll all students living in their catchment area, but they can enroll additional students from other catchment areas provided there are free places. Parents can also choose church and private schools, which do not have a catchment area. Free school choice is argued to result in a high level of between-school segregation already at the primary education phase, based on both SES and ethnicity (Hajdu et al., 2021, 2022; Hermann & Kisfalusi, 2023; Kertesi & Kézdi, 2012).

Public education is organized under school districts, the local units of the national education authority.[n] The school district hires school principals and teachers, and pays for school expenditures. It decides about the number of teachers in the schools, and also about the number of classrooms starting in first grade. However, by law, it is the school principal's responsibility to assign students and teachers to these classrooms. A maximum of 27 (30 before the academic year of 2013–14) students can be assigned to a single classroom (this can be exceeded by 20 % with the school district's approval). Similar to many other European systems, classrooms are relatively fixed units throughout the years, and in most cases, students attend all courses together with their classmates. The main subjects are taught by the same teacher or the same two teachers in the first four years, whereas in the second half of primary education, specialized teachers are responsible for each subject. Therefore, class assignment has serious consequences for student achievement because classmates face the same teachers and peers for a long period of time.

[l] Students can choose from three different tracks: 1) Academic secondary schools offer the academic track, which prepares students for tertiary education (4-5 years); 2) Vocational secondary schools offer a mixed track, which provides vocational training as well as access to tertiary education (4-5 years); 3) Vocational schools focus on vocational training and offer general education with a limited scope, with no access to tertiary education (3 years). Some academic secondary schools also provide highly selective eight- or six-year-long academic programs, starting in 5th or 7th grade. Less than 10% of students attend these schools, and we exclude them from our analysis.

[m] The majority of Hungarian primary schools belong to the public sector (85% of schools and 88% of students), the share of church schools (12% of schools and 11% of students) and private schools (2.5% of schools and 1% of students) is much lower.

[n] Before 2013, local governments were responsible for the provision of primary and secondary education in Hungary (Hermann & Semjén, 2021).

Officially, primary schools are not allowed to organize admission exams, and students are not tested before primary education starts.[o] As a result, schools cannot observe applicants' academic skills. Therefore, in the first grade of primary education, merit-based within-school sorting is ruled out by law.

Instead of merit-based sorting, schools can use other sorting practices. Though schools are also not allowed to select among applicants or sort students between classrooms based on characteristics such as ethnicity or social background, they do not always comply with the law.[p]

Why is non-merit-based sorting more prevalent if both merit-based and non-merit-based sorting are equally illegal? With no admission exam, information on student abilities is hardly available for the schools, while family background characteristics are easily observable. In addition, informal sorting on socioeconomic background is less transparent or detectable for the education authority or the public than ability grouping, which requires schools to administer some tests. Since academic preparedness and socioeconomic background are likely highly correlated even at the age of starting school, even if the aim is to create academically homogenous classrooms due to reasons described in Section 2.2., it is less costly for schools to gauge directly SES, rather than risk testing students.

In fact, qualitative sociological studies show that primary schools use various practices for selecting among the applicants and sorting them between the classrooms (Berényi et al., 2008; Erőss, 2008). This is in the interest of both the schools and high-status parents. Schools have the incentive to enroll a higher share of high-SES students to enhance the schools' prestige (Zolnay, 2018). Moreover, many schools try to prevent high-SES students in the catchment area from migrating to another school or even attract high-SES students from other catchment areas by sorting students into different classrooms based on social background.

The following selection and sorting practices have been described in the literature. First, most primary schools provide the opportunity for applicants to indicate to which first-grade teachers they would like to apply.[q] More motivated parents collect more information about their students' future teachers and might be more aware of which teachers are perceived better (Lucas, 2001). Therefore, more qualified teachers might be matched to students of more motivated families simply because of applicants' preferences. However, school principals can also intervene in this process because they decide about how students are assigned to the classes. These practices are categorized as informal and non-merit-based in Table 1.

Second, primary schools can launch classrooms with formally differentiated curricula (e.g., specialized classrooms for math, foreign languages, music, arts, or sports). For specialized classrooms in sports and arts, but not for the other subjects, schools are allowed to test applying students' aptitude for the given subject. That is, even in the case of classrooms with differentiated curricula, schools do not observe academic skills, and hence, sorting is non-merit-based in our definition. However, these aptitude tests provide the opportunity for the schools to observe the students' family background and general skills, and sort the applicants based on these. Therefore, primary schools often offer different types of classrooms: one (or more) with the regular curriculum in which they enroll low-ability, low-SES, and minority students living in the catchment area, and one (or more) with advanced curricula in which they enroll high-ability, high-SES students (Bajomi et al., 2008). These specialized classrooms are located in the formal and non-merit-based cell in Table 1. Table 2 shows that 16 % of schools provide classrooms with a specialized curriculum in any subject, and 26 % of students attend these schools.

Table 2
Incidence of formal or merit-based sorting practices.

	All schools		Schools based on school-year observations with multiple classrooms only	
	schools	students	schools	students
Schools with				
specialized curriculum classroom	0.161	0.266	0.295	0.350
gifted classroom	0.014	0.024	0.025	0.031
remedial classroom	0.007	0.008	0.009	0.009
Schools reporting ability sorting	0.098	0.151	0.175	0.197
N	22,031	628,929	10,476	455,464

Notes: The table shows the proportion of school-years in the 2011–19 period when the school used some formal or merit-based sorting. Students refer to the proportion of students enrolled in these schools in 8th grade.

Third, even in the absence of specialized curriculum classrooms, many schools organize informal events where applicants and their parents can meet their future teachers. At these events, applicants participate in different games or solve playful exercises, in which teachers can observe their social background and general skills. While formal admission tests are prohibited, school administrators can use the observations made during these events to sort applicants into classrooms. In the absence of standardized admission tests, however, the admission process is not transparent, and parents do not have a straightforward legal pathway if they dispute the school's admission decision (Berényi et al., 2008). Therefore, this practice also counts as informal and non-merit-based sorting.

Fourth, in schools with less advantaged student composition, another way of sorting low-ability students into separate classes is to organize small-sized remedial classrooms with a special curriculum for students with severe learning difficulties. This way, students with extreme disadvantages are separated from other students (Bajomi et al., 2008; Berényi et al., 2008), albeit the incidence of this is rare (see Table 2). It is not exactly clear how schools sort students into remedial classrooms; therefore, we consider this as a borderline case between merit-based and non-merit based formal sorting.

In sum, in the first four-year cycle of the Hungarian primary education system, sorting across classes is mostly non-merit-based and occurs as a result of informal sorting practices or parental choice of teachers in the first grade.

Potential for ability grouping arises later, at the beginning of 5th grade, when some schools reshuffle classes based on prior GPA or a school-year-specific placement test. This practice, belonging to either formal or informal merit-based type of sorting in Table 1, is rare, however, according to a small-scale survey (Kisfalusi et al., 2023). In a sample of 132 schools with multiple classrooms, only 5.3 % reports reshuffling classes in 5th grade. In a yearly survey covering all schools in the country (see next section), school principals are asked whether ability sorting across classes is present in their school or not. The

[o] The only exceptions are special curriculum classrooms. Most of these provide advanced education in music or sports, and schools are allowed to select students based on aptitude, but not on academic skills.

[p] This is demonstrated by several segregation lawsuits showing that ethnic Roma students had been educated in separate classrooms even though it is legally prohibited. Sandor-Szalay et al. (2019) summarize Hungarian domestic law relevant for ethnic segregation in education, as well as brief about segregation lawsuits in front of Hungarian courts between 2005 and 2015. Most of these were successfully litigated by an NGO against local governments or schools maintaining segregation, and were a mix of cases concerning both between- and within-school segregation of Roma students.

[q] Formally, students apply to classrooms, but teachers are assigned to classrooms before the application and this information is available for parents.

question refers to all grades together, and no definition of ability sorting is provided. Therefore, it is not clear what principals mean by ability sorting in this context. This can include both formal ability grouping (e.g. from Grade 5 or later) or informal ways to sort students based on ability. Altogether, around 10 percent of schools report that they employ ability sorting of students across classes, and this share is around 17 % among schools with multiple classrooms (see Table 2). Ability sorting is thus not widespread and cannot be the main mechanism of within-school sorting.

Altogether, sorting within primary schools is mostly non-merit based in Hungary due to the prohibition of entrance exams testing academic skills in the first grade and the practice of keeping the classrooms unchanged through the 8 grades in most schools.

4. Data and methods

4.1. Data

We use the data of the National Assessment of Basic Competencies (NABC). NABC is an annually administered, standardized, low-stake blind test similar to PISA (OECD's Programme for International Student Assessment), measuring reading literacy and mathematics skills for the full population of sixth-, eighth-, and tenth-grade students in Hungary.[r] NABC is complemented with a student questionnaire focusing on students' socioeconomic background and cultural resources. In addition, school principals are asked to fill out a questionnaire about the characteristics of the school.

NABC data have been available since 2006 (in 2020, the assessment did not take place due to the COVID-19 pandemic). From 2008 onwards, student test scores and background data in 6th-, 8th-, and 10th grades can be longitudinally linked through a unique student identifier. Test scores are scaled to be comparable across years and grades. School-level data can also be linked across the years through a unique school identifier. Furthermore, the dataset contains information on each student's classroom assignments. Therefore, we do not only have information on the number of classrooms in the school at a given grade but also on student composition and other characteristics of each classroom.

We use the 8th-grade NABC data from years between 2011 and 2019 because these are the cohorts we can link to their 6th-grade test scores for the value-added calculations (808,553 students in 3182 schools in total). To create our analytical sample, we have made the following restrictions. First, school-years in which the school provides six- and eight-year-long secondary academic programs are excluded because these are highly selective programs, in which enrollment is based on admission tests. Second, we exclude students with missing information on SES. Third, classes in which more than 50 % of the students have special educational needs (SEN) are excluded from the analysis because these are programs with special curricula for SEN students. Finally, students with missing information on 8th-grade math or reading scores are not included in the analysis (though they are taken into account for the measurement of within-school sorting, see below). The final analytical sample consists of 630,111 students (2721 schools).

Descriptive statistics on the analytical sample and the main variables are presented in Table 3. The average school size is relatively low: the mean number of classes is 1.7 and many schools have only one classroom per grade.

Our key variables are test scores and a single composite measure of family socioeconomic background. Math score and reading score are the test scores measured in the NABC. Both math and reading scores are standardized by grade (6th and 8th) and academic year using the complete NABC dataset to have a mean of 0 and a standard deviation of 1. Students' socioeconomic status is measured by a standardized socioeconomic status (SES) index, developed by Hermann et al. (2024) for the NABC dataset. The SES index is a weighted average of various measures of family background characteristics such as parental education and the income and financial situation of the family.[s] The weights are calculated by regressing the 6th-grade reading test score on the measures of family background using a pooled sample of students participating in the NABC between 2006 and 2019. The weights reflect the estimated contribution of each measure to the predicted reading and literacy performance in this regression model.[t] The SES index is standardized by year and grade to have a mean of 0 and a standard deviation of 1.

Table 3
Summary statistics in the analytical sample.

	N	Mean	Std. Dev.	Min	Max
Student level variables					
math test score, 8th grade	629,625	−0.100	0.985	−3.750	3.548
reading test score, 8th grade	629,998	−0.111	0.988	−4.083	3.523
SES index	630,111	−0.059	0.961	−3.244	2.430
gender: female	630,111	0.494	0.500	0	1
special education needs status	630,111	0.044	0.204	0	1
math test score, 6th grade	607,026	−0.057	0.957	−3.492	4.029
reading test score, 6th grade	607,156	−0.060	0.961	−3.792	3.669
School level controls					
mean SES	22,074	−0.291	0.719	−2.478	1.512
sd SES	22,064	0.802	0.176	0.008	1.664
Instruments (school-level)					
number of classes, 6th grade	22,074	1.677	0.841	1	8
predicted number of classes, 6th grade	22,074	1.811	0.916	1	8
total enrollment, 6th grade	22,074	36.959	23.056	1	207

Notes: The table shows the student- and the school-level summary statistics of key variables used in the analysis. Math and reading scores, and the SES index are standardized by grade and academic year using the complete NABC dataset to have a mean of 0 and a standard deviation of 1.

4.2. Methods

To estimate the effect of sorting on test score inequalities, we use a 2-step empirical strategy. First, we classify schools in each year into one of two categories, whether they assign students within the same cohort but with different socioeconomic backgrounds into separate classrooms ("sorting" schools) or not. We use a data-driven approach to do this following Lefgren (2004), Clotfelter et al. (2006) and Horváth (2015). Then, we use this year-specific school classification to estimate the effect

[r] NABC covers the full population of students with two restrictions: First, test score data are missing for students who are absent on the day of the test (due to illness or other reasons). Second, most students with special educational needs (SEN) are not required to complete the test.

[s] The variables include the mother's and father's educational attainment, social transfers in school (subsidized lunch, free lunch, free textbooks), social transfers outside school (regular child protection benefit), family assets (number of mobile phones, PCs, cars, bathrooms), family vacation ("How often has the student been on holiday with their family during the summer holidays in the last four years?"), internet connection at home, number of books at home, subjective standard of living, and the typical standard of living in the neighborhood.

[t] The weights are calculated with a regression model because this way, the weights represent the social significance (relative importance) of the single items for the socially relevant outcome (test scores) (Kertesi & Kézdi, 2016). Reading test score is used as a dependent variable in the regression model because an extensive set of literature shows that reading scores are more strongly determined by students' social background than math scores (Cooper et al., 1996; Fryer, 2014). Using data from a large-scale survey, Hermann et al. (2024) show that this SES index strongly correlates with the families' per capita income; therefore, it can be interpreted as a proxy for relative income (for which there are no data in the NABC dataset). The index also strongly correlates with parental education, the latter accounting for 75% of its variance.

of such sorting on test score inequalities by socioeconomic status. To address potential endogeneity issues, we will use two different identification strategies: school fixed effects, and instrumental variables. Next, we describe each step of our methods in detail.

4.2.1. Measuring within-school sorting

To examine the effects of within-school sorting, we first construct measures of the presence (and intensity) of sorting for each school. Following the tracking test in Horváth (2015), these measures are based on the following regression model, estimated for 6^{th}-grade students in each school and year separately:

$$SES_i = \beta_1 + \sum_{j=2}^{J} \beta_j D_{j(i)} + \varepsilon_i \tag{1}$$

where SES_i is the socioeconomic status index of student *i*, and D_j is a dummy variable indicating if student *i* is assigned to classroom *j* in 6^{th} grade. Therefore β_j represents the mean socioeconomic status index in classroom *j* relative to classroom 1. If the school allocates students to classes randomly, student composition in the classes should be similar, i.e. neither of the β_j coefficients should be significantly different from zero. Our first measure of sorting is built on this observation directly. We test the statistical significance of the β_j coefficients jointly, and if the *p*-value of the *F*-statistic is below 0.05, we classify the school as sorting in the given year.

As robustness, we also compute a continuous measure of sorting, which represents the sorting intensity, following Lefgren (2004). Sorting is stronger the larger the differences between classes are in student composition relative to the overall variance of the SES index. In statistical terms, this can be measured by the share of between-class variance in the total variance of SES in the school. This measure is provided by the R-squared of Eq. (1). However, this is not independent of the number of classes; therefore, we use the adjusted R-squared instead to measure the intensity of sorting.

Note that both sorting measures refer to the school-year level. In other words, we cannot characterize individual classes as being more or less selected in terms of student composition, just the entire school-year. In case of school-years with a single classroom in 6^{th} grade, both sorting measures get a zero value. If there is only one classroom, sorting across classes is impossible by definition. In a robustness check, we estimate our main sorting effect model for the subsample of school-years with multiple classrooms in 6^{th} grade.

4.2.2. Estimating the effect of sorting on achievement

To estimate the effects of sorting, a natural starting point would be the comparison of outcomes of sorting and non-sorting schools, while controlling for observable student and school characteristics. However, as sorting occurs as a result of school and parental decisions, sorting and non-sorting schools may differ in many respects, which are unobservable in the data. Therefore, between-school comparison likely yields biased results.[u]

In our main analysis, we rely on a different source of variation, within-school changes in sorting over time. First, we estimate school fixed effects models, eliminating the effects of all time-invariant school characteristics. However, sorting may still be an endogeneous decision of the school. Therefore, we also estimate instrumental variable fixed effects models, which rely only on exogeneous variation in sorting within schools.

To estimate the effects of sorting on student achievement, first we consider the following model:

$$A_i^8 = \alpha + \theta S^6_{k(i),c(i)} + \varphi S^6_{k(i),c(i)} \times SES_i + \beta SES_i + \gamma X_i + \delta W_{k(i),c(i)} + \rho W_{k(i),c(i)} \times SES_i + \tau_{c(i)} + \lambda_{k(i)} + \varepsilon_i^8, \tag{2}$$

where A_i is the 8^{th}-grade test score of student *i*, SES_i is again student *i*'s socioeconomic status index, and X is a vector of student-level controls such as gender and special education needs (SEN) status in the baseline (level) specification, and cubic polynomials of prior (6^{th}-grade) math and reading test scores added in the augmented (value-added) specification. $S^6_{k(i),c(i)}$ is a measure of sorting in student *i*'s 8^{th}-grade school $k(i)$ and cohort *c(i)*, the year when student *i* was in 6^{th} grade, and W represents a set of school-level controls in student *i*'s 8^{th}-grade school $k(i)$ and cohort *c(i)*. Finally, $\tau_{c(i)}$ denotes cohort effects, while $\lambda_{k(i)}$ school fixed effects (see more on this below).

The parameters of interest are θ and φ, representing the effect of sorting on mean student achievement and social inequalities in achievement, respectively. Since sorting is measured by a dummy variable, the coefficient φ shows how much larger the test score gap is between two students with one SD difference in socioeconomic status in a sorting school environment relative to a non-sorting set-up. As we estimate all models with school fixed effects, we compare the achievement gap across years within schools. Therefore, the coefficient φ shows the difference in the test score gap across students attending the same school, but in two different cohorts, where one cohort is subject to sorting, while the other one is randomly allocated to classrooms.

The baseline (level) model reveals the effects of sorting over the eight-year cycle of primary education, under the assumption that unobserved ability and other student characteristics are not correlated with sorting and the sorting-SES interaction. This interpretation builds on the institutional detail that in the vast majority of schools in Hungary, classes are formed in 1^{st} grade and are not reshuffled as students progress through school.

In the augmented (value-added) specification, third order polynomials of prior (6^{th}-grade) test scores in both math and reading, are also included. This model is built on a less strong assumption: the independence of sorting and unobserved ability and other student characteristics conditional on prior test scores. As usual, we assume prior test scores to control for most of the effects of ability, motivation, and other omitted variables. At the same time, in this specification, the sorting effects are estimated only for the two years, between 6^{th} and 8^{th} grades.

School-level control variables represent student composition, measured by the mean and standard deviation of the SES index in the students' 8^{th}-grade school but measured two years before, when the students were in 6^{th} grade.[v] Student composition at the school level might be correlated with both sorting on the one hand, and the overall level and inequality of student achievement due to school-level peer effects on the other. Therefore, to mitigate omitted variable bias in the estimation of the sorting-SES interaction coefficient, interaction terms of SES and the school mean and standard deviation of the SES index are also included as controls.

Including school fixed effects ensures that unobserved school characteristics correlated with both sorting and achievement that tend not to vary or to sluggishly change over time (e.g. the quality of school management and teachers, specific culture and norms, differences in student composition and parental preferences) do not bias results. With the school fixed effects included, we identify the coefficients of interest, θ and φ, from the within-school, over-time variation in sorting, that is,

[u] We report the results of between-school comparisons as a robustness check, where we use propensity score matching to mitigate omitted variable bias. See more details in Section 5.3 below and corresponding results in Appendix Table A10.

[v] Although most students stay in the same school between 6^{th} and 8^{th} grade (see in the section on institutional setting), a small minority may switch schools. By measuring the school composition in the school where the student attends in 8^{th} grade but two years preceding (at the time when the student was in 6^{th} grade), we ensure that the timing of the sorting dummy, the variable of interest coincides with the timing of the student body composition control variable.

from schools that in some years practice sorting, while in others do not.

For θ and φ to causally identify the effect of sorting and its heterogeneous effects by SES in Eq. (2), the residual determinants of achievement, subsumed in ε_i^8, needs to be conditionally uncorrelated with the sorting dummy and its interaction with students' own SES.

This assumption may be violated and so sorting effects in Eq. (2) may still be biased for several reasons. Most importantly, a school's decision to sort students may be strategic and so endogeneous: when they have some difficult-to-teach students, they may allocate them into separate classes in order to maximize student achievement, on the assumption that the pace of teaching can be targeted more precisely in more homogeneous groups (Duflo et al., 2011). Similarly, allocating misbehaving students into separate classes with a smaller class size can be thought to provide better results than heterogeneous classes, as the sum of non-linear negative peer effects is minimized (Lazear, 2001).

Another concern is the self-selection of students across schools. If some students or parents prefer a particular school environment, and this is correlated with sorting, these students can be expected to vote with their feet and choose sorting schools over non-sorting ones. Furthermore, parents or students with strong preferences may also differ in other aspects that are related to achievement, as well. For example, they may attribute a large value to schooling and put extra effort in learning. In this case, estimating sorting effects by Eq. (2) is also hindered by selection bias.

We address these concerns using instrumental variables. We use two variants of the same instrument for sorting: the actually observed number of classes in the school for the given cohort, and the predicted number of classes based on total enrollment in the cohort and a maximum class size rule. We start with the observation that the larger the number of classes is, the more wiggle room there is for the school to sort students. This is obvious when comparing the scenarios with a single class and two classes, but the argument also holds when the number of classes increases further.

A potential concern is that the number of classes is not random and may be affected by the sorting decision of the school. That is, in years when sorting is deemed useful, the school launches more classes in order to have more room to sort, while in years with no intention of sorting, the number of classes is cut. In our view, this kind of reverse causation is extremely unlikely. Increasing the number of classes requires more teachers, and therefore, substantial additional resources. The overall level of school resources is determined by school districts (or local governments before 2013). If a change in total enrollment, or more precisely, the number of students applying from the school's catchment area does not justify changing the number of classes, school districts are highly unlikely to approve additional resources. However, schools may find some room to maneuver if the necessary number of classes based on total enrollment is on the margin. Therefore, we also use an alternative measure of the number of classes instrument: the number of classes predicted from total enrollment and a maximum class size regulation, which, as a general rule, does not allow classes larger than 27 (30 before the academic year of 2013–14).

As we have two endogenous variables, in Eq. (3), sorting and its interaction with SES, we estimate two first-stage equations:

$$S^6_{k(i),c(i)} = \rho Z^6_{k(i),c(i)} + \mu C_i + \nu_i \tag{3a}$$

$$S^6_{k(i),c(i)} \times SES_i = \omega Z^6_{k(i),c(i)} \times SES_i + \vartheta C_i + \xi_i \tag{3b}$$

where Z is the number of classes instrument, actual or predicted, measured in 6^{th} grade, and $C \equiv (X, W, \tau, \lambda)$ is the vector of the entire set of control variables in Eq. (1), including cohort and school fixed effects.

In order to account for any potential correlation between the individual error terms of students in the same school, either from the same or in different cohorts, we estimate all models with robust standard errors clustered at the school level.

5. Results

5.1. Incidence of sorting

Before turning to the effects of sorting on achievement, we briefly describe the prevalence of sorting. Fig. 1 shows the distribution of p-values of the joint significance of classroom effects in the regression of the SES index on classroom effects for each school and year (see Eq. (1)). It is based on these p-values that we classify school-years to be sorting or non-sorting. If there were no sorting in the Hungarian primary schools, the figure would show a uniform distribution (Horváth, 2015). This is clearly not the case. There is a large spike at or near 0: for about 40 % of school-year observations with multiple classrooms (weighted by student numbers), classroom effects are jointly significant at the 5 % level. Just by visually inspecting this figure, we can compellingly reject the null hypothesis that there is no sorting on socioeconomic status among Hungarian primary schools. In our baseline classification, we consider school-years sorting if this p-value is below 5 %, and non-sorting otherwise.[w]

Table 4 shows summary statistics for the binary and continuous sorting measures, where for the first one, we consider 5 % as the cutoff p-value for sorting. Overall, 16.5 % of all unweighted school-year observations are classified as sorting. Excluding schools with a single class only, which are not sorting by definition, this share is 34.2 %, i.e. in an average year, students are probably sorted non-randomly in one-third of the schools with multiple classes. Since sorting schools tend to be larger, on average, 39 % of the students in schools with multiple classes face sorting (see also Fig. A1). Overall, 28 % of all students are studying in a sorting environment in a year on average.

There is substantial variation in sorting within schools over time, which we will exploit in the second step of our empirical strategy. Table 5 displays the distribution of schools with respect to the frequency of classified as a sorting school in the observed 9-year period. More than half of the schools never use sorting, but most of these schools have only one classroom per grade level. Regarding schools with multiple classes only, about one-quarter never use sorting, while 7.4 % always sort. Two-thirds of schools are in between, sorting students in some years but not in others.

The continuous metric can be regarded as measuring the intensity of sorting, with higher values suggesting more intense sorting. As the adjusted R-squared of the regression of SES on classroom effects, it can be interpreted as the between-class share of variation in student SES within a school in a given year. As shown in Table 4, this is 3–8 % on average. The histogram of these adjusted R-squared indices looks very similar to Lefgren's (2004) ability tracking metric.

5.2. The effect of sorting

Now we turn to our main results, which are visually illustrated in Fig. 2. In each panel, we see a version of test scores – math or reading; 8^{th}-grade levels or 6^{th}-to-8^{th} grade value-added[x] – as a function of the SES index, separately for sorting schools (in red, full circles) and non-sorting schools (in blue, hollow circles). The panels show a binned scatter plot of raw data with no controls, except for school fixed effects. Even this descriptive graph suggests that sorting has heterogeneous effects on students with different socioeconomic backgrounds: low-status

[w] Horváth (2015) suggests, based on simulations on US data with a similarly distributed left-hand side variable and similar within-school number of classes and class sizes as here, that a 5% p-value cutoff yields reasonably high share of correctly predicted schools. Still, in robustness checks, we consider two other alternative classifications and the continuous sorting index, the adjusted R-squared of the regression in Eq. (1).

[x] Technically, value-added scores are computed by controlling for the cubic polynomial of 6^{th}-grade math and reading scores.

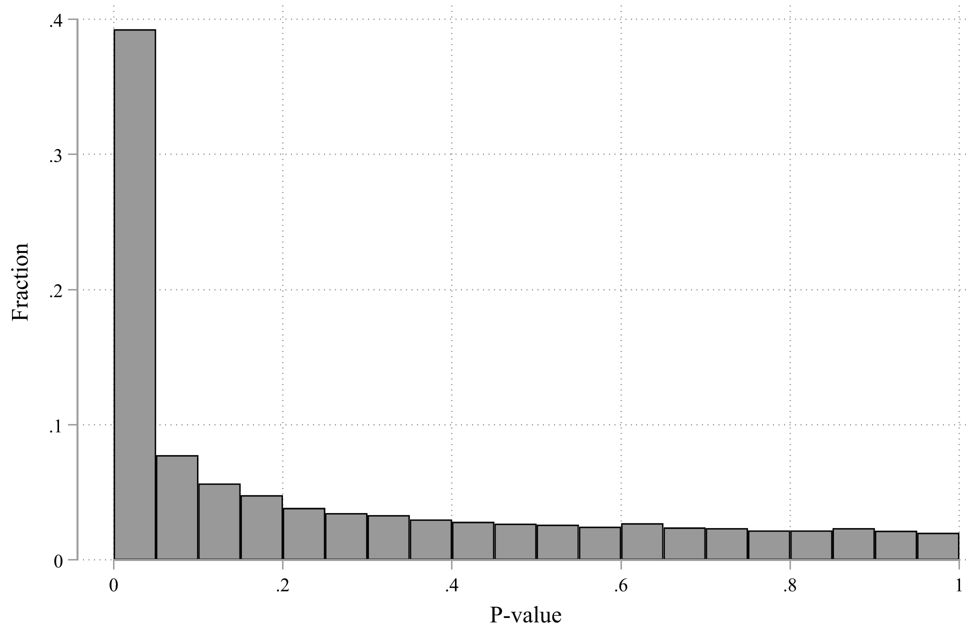

Fig. 1. Distribution of school-years with respect to sorting p-values.
Notes: P-values are calculated for the joint F-test of the coefficients of classroom indicators in the regression model of the SES index on classroom indicators, estimated for each school-year separately (Eq. (1)). School-year observations with multiple classes in the school only, weighted by the number of students.

Table 4
Summary statistics of sorting.

	N of students / schools	Mean	Std. Dev.	Min	Max
School-years weighted by the number of students					
All school-year observations					
sorting (dummy)	630,111	0.284	0.451	0	1
sorting index (adjusted R-squared)	630,111	0.055	0.105	−0.135	0.866
School-years with multiple classes					
sorting (dummy)	456,216	0.393	0.488	0	1
sorting index (adjusted R-squared)	456,216	0.076	0.117	−0.135	0.866
Schools unweighted					
All school-year observations					
sorting (dummy)	22,074	0.165	0.371	0	1
sorting index (adjusted R-squared)	22,074	0.033	0.089	−0.135	0.866
School-years with multiple classes					
sorting (dummy)	10,640	0.342	0.474	0	1
sorting index (adjusted R-squared)	10,640	0.069	0.118	−0.135	0.866

Notes: A school in a given year is considered sorting if the joint F-test of the coefficients of classroom indicators in the regression model of the SES index on classroom indicators is significant at the 5 % level (see Eq. (1)). School-year observations with a single class in school are non-sorting by definition. The sorting index is the adjusted R-squared of regression models of the SES index on classroom indicators, estimated for each school-year separately (see Eq. (1)).

Table 5
Within-school variation in sorting over time.

	All schools		Schools based on school-year observations with multiple classrooms only	
	N	%	N	%
Always sorting	63	2.3 %	119	7.4 %
Sorting in some years, not in others	1114	40.9 %	1058	66.1 %
Never sorting	1544	56.7 %	423	26.4 %
Total	2721	100.0 %	1600	100.0 %

Notes: The table displays the number and share of schools as they change their sorting practices over time. A school in a given year is considered sorting if the joint F-test of the coefficients of classroom indicators in the regression model of the SES index on classroom indicators is significant at the 5 % level (see Eq. (1)). Schools with a single classroom (in the first two columns) in a year are non-sorting by definition.

kids are harmed by sorting, while high-status students may gain, although seemingly to a much lesser extent, if to any. That is, sorting widens test score inequalities by socioeconomic status. This widening appears statistically and economically significant: The SES gradient in math (reading) test scores in sorting schools is 11–19 % (7–10 %) larger than in non-sorting schools, depending on the specification, test score levels or value-added.

Fully controlled, numerical results are displayed in Table 6, collating estimates from all three identification strategies. They show the effects of sorting on test score inequalities by SES, when sorting is measured as a binary indicator. Panel A is for math, while Panel B is for reading. The first and second columns contain the school fixed effects models, while columns 3–6 present the instrumental variable fixed effects estimates. Odd numbered columns show level models, with no controls on prior test scores, which we interpret as the cumulative effects of sorting in the first eight years of schooling. While even numbered columns show the results of the value-added models controlling for a full set of third order math and reading test scores in 6th grade, representing the sorting effects during the last two years of elementary school.

The first row in each panel presents the effect of sorting at the mean of the socioeconomic status index (*SES=0*). Since this index is standardized and has a roughly symmetric distribution, this main effect of sorting can be interpreted as an approximation of the average marginal effect of sorting on test scores. These coefficients tend to have a negative sign, but in many cases, they are statistically not significant. Overall, they suggest sorting slightly decreases, if anything, mean student achievement.

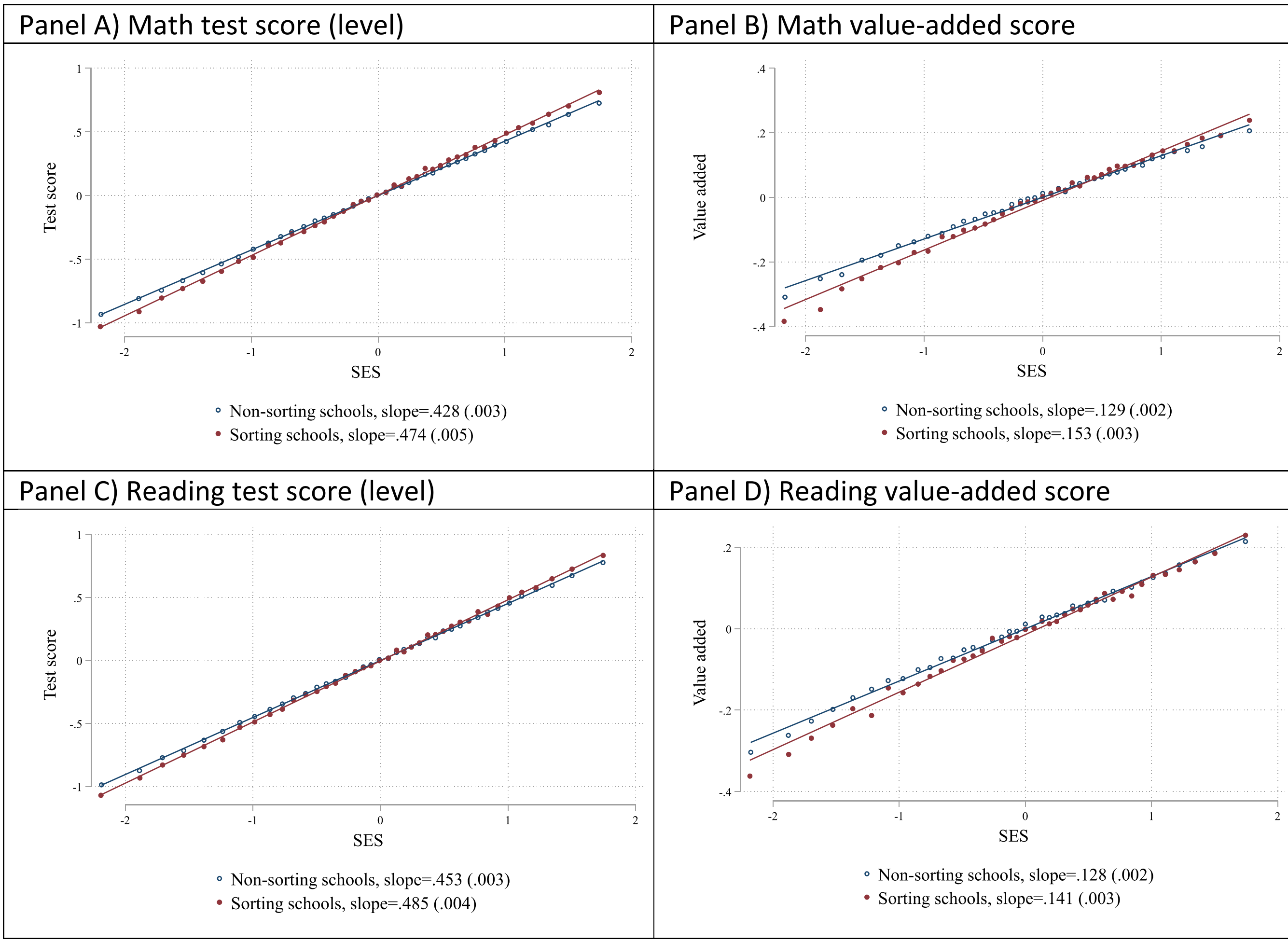

Fig. 2. SES gradient of test scores in sorting vs non-sorting schools.
Notes: Binned scatterplot of 8th-grade test scores (levels) or 6th-to-8th-grade value-added (VA) scores and SES within schools in 40 equal sized bins of the SES index. Controls: School fixed effects, and in VA plots, third-order polynomial of math and reading scores in 6th grade. Estimated SES gradients are displayed in the legend.

The parameters of our main interest are those sitting on interaction term of the SES index and sorting. These coefficients show the average difference in the test score SES gradients in sorting schools relative to non-sorting ones. These coefficients are consistently positive and statistically significant in each model estimated. This suggests that sorting across classes within schools amplifies the test score gap between students from disadvantaged and more affluent families.

The fixed effects models in column 1 show that comparing two students with a 1 SD difference in the SES index, attending a school with a sorting setup for 8 years increases the test score gap by 0.044 SDs (0.031 SDs) in math (in reading). We can benchmark these effects against the SES gradient in test scores in non-sorting schools (see in the row with the corresponding title in Table 6),[y] implying about a 11 % (7 %) larger SES gradient in sorting schools than in non-sorting ones. Although effect sizes of the sorting-SES interaction are smaller, as expected, in the fixed effects value-added specifications (column 2), a larger SES-gradient difference is implied in this case: in sorting schools, the test score gap is about 21 % (12 %) higher than in non-sorting ones. Both the level and value-added effects are statistically and economically significant.

Before reviewing the IV estimates in columns 3–6, let us turn to some IV diagnostics. Fig. 3 visually illustrates the strong, positive first-stage relationships between our endogenous variable, the sorting indicator and the number of class instruments. It shows how the sorting measures are related to the actual number of classes on the one hand, and the number of classes predicted by total enrollment in the grade and the maximum class size rule on the other. In a single-class case, there is no sorting by definition, while about a quarter of students in schools with two classes are taught in a sorting setup. Moreover, the probability of sorting keeps increasing when the number of classes increases further. Beyond 4 classes, the increment gets smaller, but only a small share of students attend schools that are large.

Table A1 in the Appendix then confirms these strong positive relationships by displaying the first-stage regressions, along with various first-stage F-statistics. The first-stage coefficient estimates are significant and strong by conventional standards (see multivariate F and Sanderson-Windmeijer multivariate F-tests for the two endogenous variables separately and the Kleibergen-Paap F-statistic overall, which are all well above the conventional thumb rule of 10). The table also confirms that the predicted number of classes is almost as strongly associated with sorting as the actual number of classes. This implies that the actual number of classes is determined mostly by enrollment and the maximum class size rule of 27 (30 before the academic year of 2013–14). Therefore, as long as enrollment is exogenous, the predicted number of classes instrument is exogenous, as well. The patterns of correlation between

[y] This is the average marginal effect of SES on test scores in non-sorting schools. Note that since, besides including SES main effects in our models, SES is also interacted with the school-year mean of SES and the school-year standard deviation of SES, the average marginal effect is not simply the main coefficient on SES but a linear combination of coefficients of interaction terms involving SES.

Table 6
The effect of sorting on test score inequalities.

	FE		IV FE		IV FE	
			instrument: number of classes		instrument: predicted number of classes	
	level (1)	VA (2)	level (3)	VA (4)	level (5)	VA (6)
Panel A) MATH						
sorting (dummy)	0.003	−0.011*	−0.083**	−0.043	−0.030	0.013
	(0.007)	(0.006)	(0.036)	(0.032)	(0.054)	(0.048)
sorting x SES	0.044***	0.026***	0.095***	0.045***	0.089***	0.036***
	(0.005)	(0.004)	(0.012)	(0.008)	(0.013)	(0.009)
Observations	629,607	606,512	629,605	606,510	629,605	606,510
N of schools	2721	2721	2719	2719	2719	2719
Within R-squared	0.174	0.553	0.173	0.553	0.173	0.553
SES gradient in non-sorting schools	0.402	0.124	0.387	0.118	0.389	0.121
Multivariate F-test of instruments						
first-stage F-stat for sorting			168.61	168.09	78.89	78.60
first-stage F-stat for sorting x SES			475.96	484.06	486.90	496.79
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			329.22	328.23	141.86	141.34
first-stage SW F-stat for sorting x SES			949.48	965.97	969.26	989.14
Kleibergen-Paap weak identification test F-stat			163.9	163.0	69.27	68.70
Panel B) READING						
sorting (dummy)	−0.002	−0.014***	−0.109***	−0.081***	−0.061	−0.046
	(0.005)	(0.005)	(0.030)	(0.025)	(0.045)	(0.038)
sorting x SES	0.031***	0.015***	0.074***	0.028***	0.071***	0.023***
	(0.004)	(0.003)	(0.010)	(0.007)	(0.011)	(0.007)
Observations	629,980	606,836	629,978	606,834	629,978	606,834
N of schools	2721	2721	2719	2719	2719	2719
Within R-squared	0.208	0.606	0.206	0.605	0.207	0.606
SES gradient in non-sorting schools	0.426	0.128	0.414	0.124	0.414	0.125
Multivariate F-test of instruments						
first-stage F-stat for sorting			168.48	167.97	79.01	78.70
first-stage F-stat for sorting x SES			475.73	483.80	486.82	496.60
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			329.15	328.19	142.00	141.50
first-stage SW F-stat for sorting x SES			949.18	965.60	968.96	988.67
Kleibergen-Paap weak identification test F-stat			163.9	163.1	69.34	68.79

Notes: The table shows regression estimates of math (Panel A) and reading (Panel B) test scores on the binary sorting indicator (see Eqs. (2) and (3)) and its interaction with the SES index. Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, and school and year fixed effects. Additional controls in VA models: third order polynomial of math and reading scores in 6th grade. SES gradient in non-sorting schools is computed as the average marginal effect of SES in non-sorting schools. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

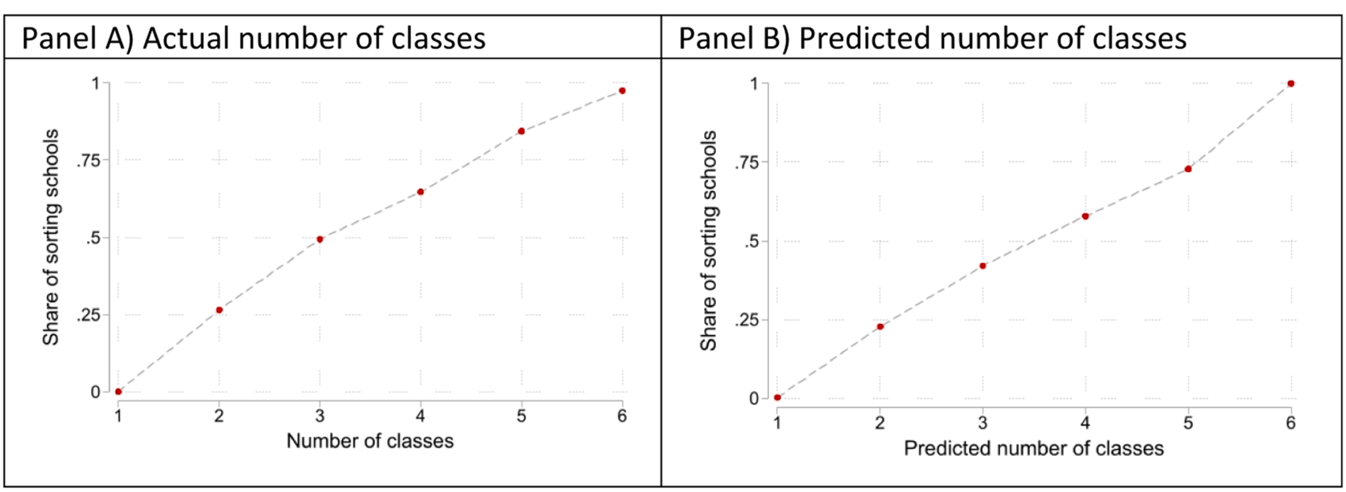

Fig. 3. Sorting and the actual and predicted number of classes in school.
Notes: The figure shows the share of sorting schools by the actual (Panel A) and the predicted (Panel B) number of classes. School-year observations with a single class in school are non-sorting by definition. Actual and predicted number of classes are top-coded as there are less than 10 school-year observations with 7 or 8 classes. Weighted by the number of students.

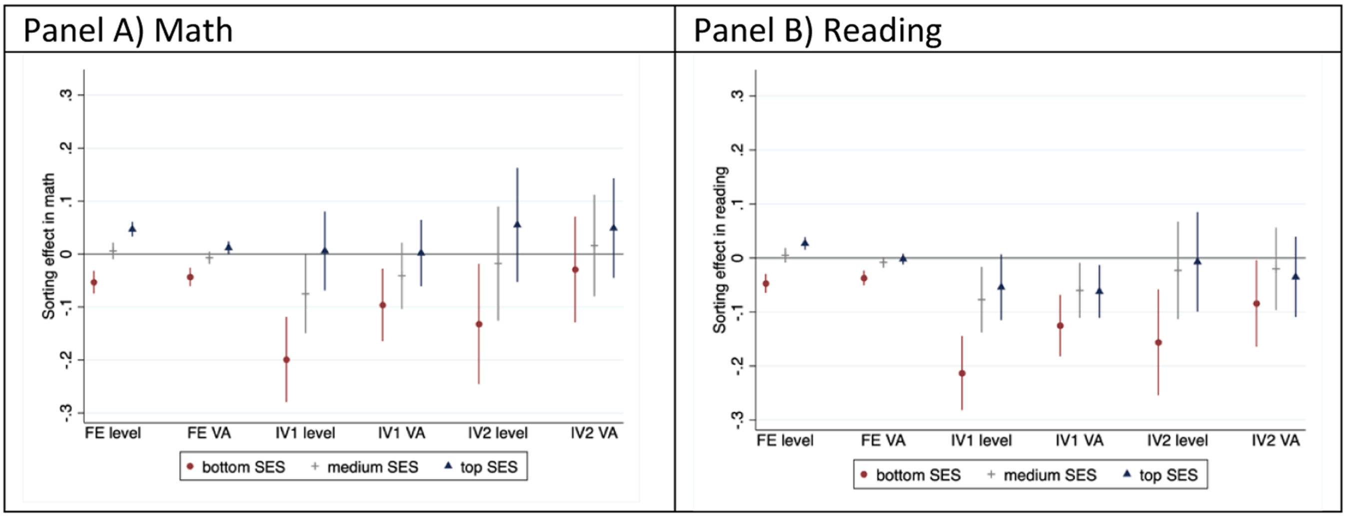

Fig. 4. Sorting effects by socioeconomic status.
Notes: The figure shows the heterogeneity of sorting effects by SES tercile in math (Panel A) and reading (Panel B). Point estimates are coefficients of the interaction terms of a binary sorting indicator and terciles of the SES index in regressions similar to Eqs. (2) and (3), but with discretized socioeconomic status (terciles of the SES index) and no sorting main effect. IV1/IV2 denote the models with actual/predicted number of classes as the instrumental variable, respectively. Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, and school and year fixed effects. Additional controls in VA models: third order polynomial of math and reading scores in 6th grade. 95 % CIs around point estimates are computed using standard errors clustered at the school level.

the actual and predicted number of classes and enrollment are shown in Fig. A2 in the Appendix, while the distributions of 6th-grade enrollment, actual and predicted number of classes are displayed in Fig. A3. The latter three highlight that average school size in Hungary is rather small, with about 60–70 % of students attending schools with at most 2 classrooms per grade level.

Now we can return to our main results. Estimates from our IV strategies are displayed in columns 3–6 in Table 6. These suggest larger effects of sorting on socioeconomic inequalities in test scores than fixed effects models, though the results are qualitatively similar. Note that the two variants of the number of classes instrument provide nearly identical results: In math, the SES gradient is 23–38 % larger in sorting school than in sorting school, while in reading, the corresponding gradient difference is 17–23 %. Larger effect sizes in the IV specifications are most likely explained by higher local average treatment effects in the complier groups. When the number of classes increases, and the school uses this opportunity to introduce or intensify sorting, this is likely to involve more substantial and intentional changes in teaching and learning than an occasional shift from a uniform to a sorted distribution of students across the same number of classes.

To explore which part of the SES distribution drives the widening test score gap finding, we re-estimate our main model in Eqs. (2) and (3) with terciles of the SES index interacted with the sorting dummy. Fig. 4 illustrates the results and confirms our impression from Fig. 2: low-status students are significantly harmed by sorting, while the sorting effect on high-status children is ambiguous; even if positive, it is statistically insignificant.

In summary, sorting appears to magnify test score gaps between high- and low-status students, in a way that low-status students lose grounds.

5.3. Robustness

To demonstrate the robustness of our main findings, we run several kinds of sensitivity checks. First, we explore the inequality effects of sorting by mother's education instead of the SES index by re-estimating the models from Eqs. (2) and (3) with four categories of mother's education (primary; vocational; secondary diploma; college/university, with the modus, secondary diploma being the reference category) in lieu of the continuous SES index. Results are shown in Table A2 in the Appendix. They confirm the patterns of our main, SES specification, as well as the non-linearities in the heterogenous effects of sorting by family background. We can see that the interaction coefficient between sorting and lower mother's education categories are always significantly negative, while the ones for college/university are positive albeit much smaller in absolute value and even insignificant in reading. This reinforces our finding that the heterogenous sorting effects are driven by harming low-status students in particular.

Second, in our main models, we used a single variable, the SES index, to account for heterogeneity in student family background. This approach was motivated by estimating an all-embracing SES gradient to which the estimated effect of sorting on the SES gradient could be directly measured. However, this single variable might be insufficient to fully account for students' family background, which, at the same time, might be correlated with sorting. Therefore, we re-estimated the main models including an extended set of family background characteristics: mother's and father's level of education, the number of books at home, socioeconomically disadvantaged status of the student, subjective affluence of the family, whether the family receives regular child protection allowance, whether the student is entitled to subsidized or free lunch, whether the student is entitled to free textbooks, and how many times the family had a holiday in the last four years. The results on the effect of sorting and the sorting-SES interaction are basically identical to the main specification (see Table A3 in the Appendix). To further expand the set of control variables, we also embellished our main specification with school-specific linear trends. Results, displayed in Table A4 in the Appendix, are essentially the same again as in our main specification.

Third, we use alternative sorting measures. On the one hand, we consider the potential sensitivity of our binary sorting classification to the p-value cutoff we used. We considered two alternatives: (1) a higher p-value cutoff (0.2, instead of 0.05) to classify schools sorting vs. non-sorting and (2) a subsample of school-years where those with $p-value \leq 0.05$ are classified sorting, but only those with $p-value \geq 0.5$ are classified non-sorting. (School-years with p-value in between were discarded.) Tables A5 and A6 show the results, which demonstrate the same pattern as our main specification.

On the other hand, as mentioned earlier, we also used the adjusted R-squared of the sorting regression (Eq. (1)) as a continuous sorting index, measuring the intensity of sorting. Results are displayed in Table A7 in the Appendix, and lead to identical qualitative conclusions (see also Fig. A4). The sorting x SES coefficients are apparently much larger, but the estimated effect sizes are similar: the sorting index (measured by the adjusted R-squared of Eq. (1)) never increases by 1, its standard deviation is approximately 0.1. In other words, in the real world, a school turning a non-sorting environment into a sorting one is likely to increase the sorting index by 0.1–0.2. This means that we should calculate the effect size by scaling the estimated coefficients of the sorting index by one-tenth to one-fifth.

Fourth, we explore whether the sorting effect is entirely driven by school size or potentially some unobserved factor correlated with size. The evidence does not support this concern. First, we exclude school-years with a single classroom to confirm that our results are not driven by classifying such schools as non-sorting by definition. The estimates become slightly larger but show identical qualitative pattern (see Table A8 in the Appendix). Second, we re-estimate our fixed effects specifications holding school size constant. In columns 1–2 and 3–4 of Table A9 in the Appendix, we restrict the sample to school-years with two and three classrooms, respectively, while columns 5–6 show estimates including the number of classes as a control variable. The results are qualitatively similar to our main fixed effects results, though the coefficients are marginally smaller.

As a final robustness check, we look at the effect of sorting in a cross-sectional comparison of schools, estimating regression models using matched samples of sorting and non-sorting schools. Note that here we use an entirely different variation in sorting: while all the other estimates use within-school changes in sorting over time, here we compare different schools within the same year.

To build a matched sample, we estimated propensity scores for sorting at the school level by regressing the school's sorting status on the schools' county, type of provider (public, church, other), type of settlement, the predicted number of classes, and the mean and SD of SES in school. We trimmed the sample at the 0.8 value of the propensity score because of a lack of common support above this threshold. Then, for each treated (sorting) school, we selected one control (non-sorting) school combining nearest neighbor matching based on the propensity score and exact matching by year, using a caliper of 0.05.

Table A10 of the Appendix shows regression estimates of the sorting effect using the matched sample. Each model includes fixed effects for the treated-control pairs. We estimated level and value-added specifications, with and without controlling for the school-level variables that were used in calculating the propensity scores. The last set of models also includes interactions between student SES and the school-level control variables. The results are similar to our main specification.

Table 7
Ability sorting and SES sorting.

	School-years			Students		
	SES sorting:			SES sorting:		
	no	yes	total	no	yes	total
ability sorting: no	55.94	26.53	82.47	51.3	29.04	80.34
yes	9.23	8.3	17.53	9.35	10.31	19.66
total	65.17	34.83	100	60.65	39.35	100

Notes: The table displays the share (%) of school-years and students by ability sorting and SES sorting. A school in a given year is considered using SES sorting if the joint F-test of the coefficients of classroom indicators in the regression model of the SES index on classroom indicators is significant at the 5 % level (see Eq. (1)). Ability sorting is reported by the principal in the NABC school survey. School-years with multiple classrooms in 8th grade only. N of school-years: 10,252, N of students: 446,616.

6. Non-merit-based sorting or ability grouping?

Earlier we argued that, in the Hungarian institutional context, sorting across classes is non-merit-based in most cases and occurs as a result of mainly informal sorting practices or parental choice of teachers in the first grade. Therefore, we interpret our results as the effects of mostly non-merit-based sorting.

However, ability grouping may also be present to some extent. For example, some schools reshuffle classes in 5th grade based on prior GPA or a school-year-specific placement test. Therefore, to the extent that achievement is correlated with SES and as researchers, we do not observe the occurrence and/or the result of placement tests, SES-based sorting may mask ability grouping, and our sorting measure will also capture schools that formally or informally sort on achievement. Thus, one may worry that our results are as much driven by (formal) ability sorting as by informal SES-sorting.

To minimize such concerns and support our interpretation of the results, we use NABC survey information from school principals reporting if, in a given year, they use, at any grade level, any form of ability grouping across classrooms. This is a coarse measure of merit-based sorting, as the question refers to all grades together, and no definition of ability sorting was provided to respondents. Those answering 'yes' may either use merit-based sorting in 1st grade or reshuffle classes in higher grade levels.[z] Moreover, principals may also report to use ability grouping even if it was not used in case of the 6th-grade cohort in a particular year, but some other cohorts were affected in the school. Therefore, such schools will provide an inflated measure of those sorting on ability. Consequently, we believe that our results after excluding such schools from the sample will yield a cleaner estimate of the non-merit-based sorting effect.

In the NABC school survey, principals are asked whether they use ability sorting across classes in their school or not. Altogether, almost 10 percent of schools report that they employ ability sorting of students across classes (Table 2). Reported ability sorting is positively correlated with our estimated sorting indicator (Table 7). Regarding only school-years with more than one class per grade, we find SES sorting in 32 % of schools not using ability sorting, while this proportion is 47 % for schools with ability sorting.

[z] One concern here is that principals underreport ability grouping as it is banned in the 1st grade. Note that the survey question refers to all grades together and schools may legally reshuffle classrooms later – typically in 5th or 7th grade – based on GPA or some internally administered achievement test, although this is uncommon (Kisfalusi, Hermann, & Keller, 2023). Therefore, we do not believe that schools will underreport the use of ability grouping in fear of getting caught for non-complience with the law.

Table 8
The effect of sorting on test score inequalities – subsample of schools with no ability sorting.

	FE		IV FE instrument: number of classes		IV FE instrument: predicted number of classes	
	level (1)	VA (2)	level (3)	VA (4)	level (5)	VA (6)
Panel A) MATH						
sorting (dummy)	0.010	−0.006	−0.096**	−0.036	0.005	0.069
	(0.008)	(0.007)	(0.045)	(0.040)	(0.068)	(0.064)
sorting x SES	0.034***	0.021***	0.085***	0.045***	0.077***	0.033***
	(0.006)	(0.005)	(0.013)	(0.009)	(0.014)	(0.010)
Observations	522,559	503,409	522,557	503,407	522,557	503,407
N of schools	2662	2662	2660	2660	2660	2660
Within R-squared	0.172	0.547	0.170	0.547	0.172	0.546
Multivariate F-test of instruments						
first-stage F-stat for sorting			115.57	115.95	51.16	50.70
first-stage F-stat for sorting x SES			386.87	393.09	398.71	406.77
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			218.31	218.74	91.95	91.29
first-stage SW F-stat for sorting x SES			764.76	777.49	796.70	813.04
Kleibergen-Paap weak identification test F-stat			108.0	107.7	44.48	44.00
Panel B) READING						
sorting (dummy)	−0.000	−0.013**	−0.132***	−0.082***	−0.042	−0.012
	(0.006)	(0.005)	(0.035)	(0.031)	(0.055)	(0.050)
sorting x SES	0.025***	0.014***	0.066***	0.028***	0.066***	0.024***
	(0.005)	(0.004)	(0.012)	(0.008)	(0.013)	(0.008)
Observations	522,887	503,696	522,885	503,694	522,885	503,694
N of schools	2662	2662	2660	2660	2660	2660
Within R-squared	0.206	0.602	0.204	0.601	0.206	0.602
Multivariate F-test of instruments						
first-stage F-stat for sorting			115.47	115.87	51.24	50.77
first-stage F-stat for sorting x SES			386.69	392.90	398.49	406.51
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			218.30	218.79	92.05	91.42
first-stage SW F-stat for sorting x SES			764.59	777.32	796.17	812.45
Kleibergen-Paap weak identification test F-stat			108.0	107.8	44.53	44.07

Notes: The table shows regression estimates of math (Panel A) and reading (Panel B) test scores on the binary sorting indicator (see Eqs. (2) and (3)) and its interaction with the SES index, for the subsample of schools with no ability sorting. Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, and school and year fixed effects. Additional controls in VA models: third order polynomial of math and reading scores in 6th grade. All students in school-years where principals reported in the NABC school survey that the school used ability sorting are excluded from the sample. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In order to explore heterogeneity in the effects of merit-based and non-merit-based sorting, we first re-estimate our main models for the subsample of school years with no ability sorting reported by the principal (Table 8). The results are similar, though the estimated sorting effects on test score inequalities are slightly smaller in magnitude. This suggests that our main results are not driven by the effects of merit-based sorting. As a robustness check, we also exclude schools with specialized curriculum classes, as well as schools reporting to use ability grouping. Table A11 in the Appendix shows the results, which are again qualitatively very similar to our main specification.

Second, we classified schools into four categories by combining our SES-sorting and ability grouping measures, and estimated differences in test score inequalities across the four types (Table A12 in the Appendix). Here, we use the fixed effects specification only, as we do not have separate instruments for the two different types of sorting. The results show that ability grouping with no SES-sorting is not related to inequalities. At the same time, SES-sorting on its own significantly increases test score inequalities, and this effect is further enhanced by ability grouping. Possible explanations could involve ability grouping generating stronger peer effects, or that combining the two sorting measures simply decreases measurement error in the classification of schools, resulting in coefficients less biased towards zero. In sum, these results also confirm that it is mostly non-merit-based sorting that drives our main finding about widening test score inequalities.

7. A potential mechanism: within-school reallocation of educational resources

Our main finding that sorting widens test score inequalities is intriguing in itself but what could be the underlying mechanism? In Section 2.3, we discussed three potential mechanisms. First, peer effects may be in place by the definition of our sorting measure: as a school introduces sorting, it is making its classrooms more homogenous by SES and increases the within-school variance in SES across classrooms. Therefore, high-status students will have more alike peers, while low-status students lose out on having high-status peers. As long as everyone profits from learning with high-status peers, high-status peers will gain from sorting, while low-status students will lose. Lefgren (2004) estimates such a peer effect mechanism to be positive and statistically significant, although small in magnitude. His identification strategy, however, relies on the fact that the interaction of own prior ability and the sorting status of the school “affects student achievement only through the allocation of students to peer groups” (p. 173). This precludes that schools may reallocate resources within-school and across classrooms once they introduce sorting. This assumption is overly restrictive; we argue below that such a mechanism may be a leading one

Table 9
A potential mechanism–the effect of sorting on class size, as a measure of educational resources.

	FE (1)	IV FE instrument: number of classes (2)	IV FE instrument: predicted number of classes (3)
sorting (dummy)	−0.281*** (0.074)	−9.337*** (0.678)	12.554*** (1.286)
sorting x SES	0.508*** (0.037)	0.665*** (0.094)	1.040*** (0.125)
Observations	630,093	630,091	630,091
N of schools	2721	2719	2719
R-squared (FE models: within R-squared)	0.044	−0.671	−1.407
Multivariate F-test of excluded instruments			
first-stage F-stat for sorting		168.48	79.04
first-stage F-stat for sorting x SES		475.87	487.06
Sanderson-Windmeijer multivariate F-test of excluded instruments			
first-stage SW F-stat for sorting		329.12	142.02
first-stage SW F-stat for sorting x SES		949.44	969.42
Kleibergen-Paap weak identification test F-stat		163.9	69.35

Notes: The table shows the regression estimates of class size and the binary sorting indicator (as in Eqs. (2) and (3) but with class size as the dependent variable). Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, and school and year fixed effects. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

behind the achievement inequality effects of sorting.

Second, with sorting, teachers have a more homogenous group of students to teach and so are more able to target their skill level (tailored instruction effect). We expect tailored instruction effects to be positive for all groups of students. In fact, Duflo et al. (2011) argue that in their experimental setting, where educational resources are randomly allocated, this channel is so strong that it outweighs negative peer effects among low achievers.

Third, sorting may go hand-in-hand with resource reallocation within schools. In fact, the consistently negative sorting effects we find for low-status students suggest that in our setting with presumably small peer effects (Duflo et al., 2011; Keller & Elwert, 2023; Lefgren, 2004), nonnegative tailored instruction effects, resource allocation that further widens achievement gaps must play an important role.

As we mentioned in Section 3, in Hungary it is the responsibility of the school principal to assign teachers and allocate students to classrooms. Therefore, the principal has all the power and many ways to adjust educational resources, such as class size, equipment/facilities in the physical classroom, teacher quality, in response to sorting. For instance, they may assign the computer room to higher status classes for a computer science lesson but a regular room for lower status classes. They may also allocate students unevenly across classrooms to let some students benefit from smaller classes, or on the contrary, to minimize the number of students who are exposed to disruptive peers. Last but maybe most importantly, they may assign teachers of different effectiveness to classrooms to reward or punish teachers (see also Player 2010), or even to let students (and their parents) to self-select into the class of their preferred teacher.

We provide three pieces of suggestive evidence that support the hypothesis of such within-school, across-classroom resource reallocation. First, we propose to use class size as a proxy for educational resources, and estimate analogous models to Eqs. (2) and (3) but with 8th-grade class size as the dependent variable. Table 9 displays the results.

Table 10
Share of schools reporting different educational practices across classrooms at the same school and grade level.

	non-sorting schools	sorting schools	difference
Difference across classrooms in …			
- the quantity of homework	0.20	0.39	0.19 (0.08) **
- the content of homework	0.25	0.45	0.20 (0.09) **
- the difficulty of tests	0.14	0.22	0.08 (0.07)
- the grading standards	0.16	0.14	−0.02 (0.07)
- teaching the class in separate groups	0.18	0.35	0.16 (0.08) **
- any of these	0.42	0.63	0.21 (0.09) **
Number of schools	76	49	

Note: Data based on a survey of school principals in 2021 (Kisfalusi et al., 2023). Schools with multiple classrooms per grade only. Sorting schools: classified as sorting in at least two years in the 2015–19 period. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

With the introduction of sorting, the change in class size on average is ambiguous,[aa] but unambiguously, higher-status students are placed into larger classes than their low-status peers. This differential treatment is unlikely to explain the sorting effect on the achievement gap, but it suggests that other resources may also be reallocated across classrooms sorted on SES. For instance, Barrett and Toma (2013) show that higher value-added teachers get larger classes.

Second, access to higher quality resources across sorted classrooms may be unequal. Hermann and Horváth (2022), using a smaller set of schools but in the same Hungarian context, estimate teacher value-added scores, a measure of teacher effectiveness, and look at the correlation between this and students' SES. They find that within schools, higher value-added teachers are assigned to classrooms with higher SES students on average.

Finally, the third piece of suggestive evidence for resource reallocation comes from a small-scale survey conducted among Hungarian primary school principals as part of the research project in Kisfalusi et al. (2023). In one question of this original data collection, school principals were asked about the presence of different educational practices across classrooms at the same grade level. Table 10 displays their answers, separately for schools we classify as sorting and non-sorting. The table clearly demonstrates that differentiating across classrooms in the use of resource-intensive educational practices is significantly more frequent in sorting schools than in non-sorting ones.

8. Conclusion

The effect of separating high- and low-SES students on inequalities is at the center of many academic debates focusing on school segregation

[aa] The ambiguity is due to the opposing signs of the within-school reduced form estimates, which boils down to different complier groups. The cross-sectional relationship of class size and the actual number classes is significantly positive, just as the one between the predicted number of classes and class size. This latter is unaffected by the inclusion of school fixed effects, meaning that if the predicted number of classrooms in a school is higher than average, classes will be larger than on average. This suggests a complier group that increases the likelihood/intensity of sorting even without actually opening new classes but only by sticking a couple of more students in existing classrooms. In contrast, in the case of the actual number of classes IV, the complier group consists of schools which engage in sorting once they do open a new classroom, and then they operate each classroom with somewhat fewer students.

and tracking. In this paper, we contribute to this research agenda by investigating the test score inequality effects of within-school sorting of students in the absence of formal tracking.

In line with previous studies from the US and some European countries (Agasisti & Falzetti, 2017; Clotfelter et al., 2006; Collins & Gan, 2013; Engzell & Raabe, 2023; Ferrer-Esteban, 2016), we have found that a significant proportion of Hungarian primary schools tend to sort students between classrooms based on SES. Throughout the analyzed 9-year period, 28 % of students studied in a school in which students were sorted based on SES across classrooms.

With regard to mean student achievement in the school, we find small negative or zero effects. Depending on the model specification, sorting either decreases overall student achievement or has no effect on it. In contrast, between-class sorting within schools widens the test score gap between students from disadvantaged and more affluent families. Students with a 1 standard deviation higher socioeconomic status score 11–38 % (7–23 %) higher in math (reading) than their low-status peers if learning in a sorting rather than a non-sorting school environment. The results suggest that sorting harms low-status students, while high-status students gain little from attending sorting schools. The findings are robust to alternative sorting measures, to an extended set of control variables, or for excluding schools offering classrooms with specialized curriculum in some subject(s).

We provide suggestive evidence that within-school reallocation of educational resources and differences in educational practices underlie our findings. On the one hand, in a small survey, Hungarian primary school principals are much more likely to report differential use of resource-intensive educational practices across classrooms at the same grade level in sorting schools than in non-sorting ones. On the other hand, Hermann and Horváth (2022) demonstrate that in a subset of Hungarian primary school districts, classrooms with higher SES students are assigned higher value-added math teachers on average. Alternative mechanisms that we cannot disentangle may include peer effects (Hanushek et al., 2003; Keller & Elwert, 2023; Lefgren, 2004; Sacerdote, 2011).

We interpret the results as primarily reflecting the impact of non-merit-based sorting. Our analysis showcases that even if merit-based ability grouping is rare in an education system, non-transparent, non-random allocation of students to classrooms may still result in widening achievement gaps. Several lessons follow from this result for policy makers aiming to reduce educational inequalities.

First, our setting has demonstrated that legislative tools appear not to be effective in eliminating sorting. In Hungary, although sorting on achievement in first grade or discrimination on a wide range of socio-demographic characteristics are legally banned, we documented that sorting does frequently occur. A significant portion of schools find a way to evade the rules by grouping students into different classes through informal processes, which are easily left unnoticed by education authorities. Thus, they are also harder to litigate than more transparent achievement tests. Therefore, in addition to setting legal barriers to sorting, policy makers may also want to shape schools' economic incentives to sort. This involves counteracting the reinforcing resource allocation mechanism, a main channel through which we suggest sorting widens the achievement gap.

Second, recall that Duflo et al. (2011) find that in their controlled experiment, merit-based sorting on its own - that is, in absence of the reinforcing resource allocation effect -, does not widen achievement gaps, and low-achieving student gain from it at least as much as high achievers. Squaring this result with ours further underlines the importance of equalizing the access to educational resources for all classrooms, in addition to trying to eliminate sorting using legal tools.

Through what policy practices could this be achieved? For instance, providing the necessary pedagogical support and/or financial reward for teachers who work with difficult-to-teach students could compensate them for the potentially inferior working conditions, and thus, school principals would not need to resort to sorting to retain teachers (Player, 2010). School accountability measures may also alter the incentives to assign less or inferior resources to low-status students. For instance, in the US, since the 'No Child Left Behind' Act, which, despite other criticism, reportedly has narrowed socioeconomic achievement gaps (Dee & Jacob, 2011) schools are required to report educational progress metrics not only for all students together but also for different socioeconomic subgroups, including major racial/ethnic groups, low-income students, students with disabilities, and limited English proficiency students (Stullich et al., 2006).

Finally, our results indicate that it is at the expense of low-status students that the within-school sorting practices we study widen the socioeconomic achievement gap, with high-status students gaining little, if anything. This means that even though high-status families may presumably be a stakeholder party highly interested in maintaining the status quo of sorting, they are not clear winners of the practice. Their sole undebatable "benefit" from it is not higher academic achievement that they hope for but that their children do not have to mingle with low-status students. It is for future research to investigate if these families still supported sorting should they be aware of the true benefits, that is, a more homogenous peer composition but little achievement gains.

Funding

This work was supported by the National Research, Development and Innovation Fund of the Ministry of Innovation and Technology of Hungary, NKFIH grant number K124989. Horváth acknowledges the support of the European Union's Horizon 2020 Research and Innovation Programme (grant agreement no. 949995).

CRediT authorship contribution statement

Zoltán Hermann: Conceptualization, Formal analysis, Funding acquisition, Methodology, Writing – original draft. **Hedvig Horváth:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Writing – original draft. **Dorottya Kisfalusi:** Conceptualization, Formal analysis, Methodology, Writing – original draft.

Declaration of competing interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

We thank two anonymous reviewers for their helpful suggestions on earlier drafts of the manuscript.

Appendix

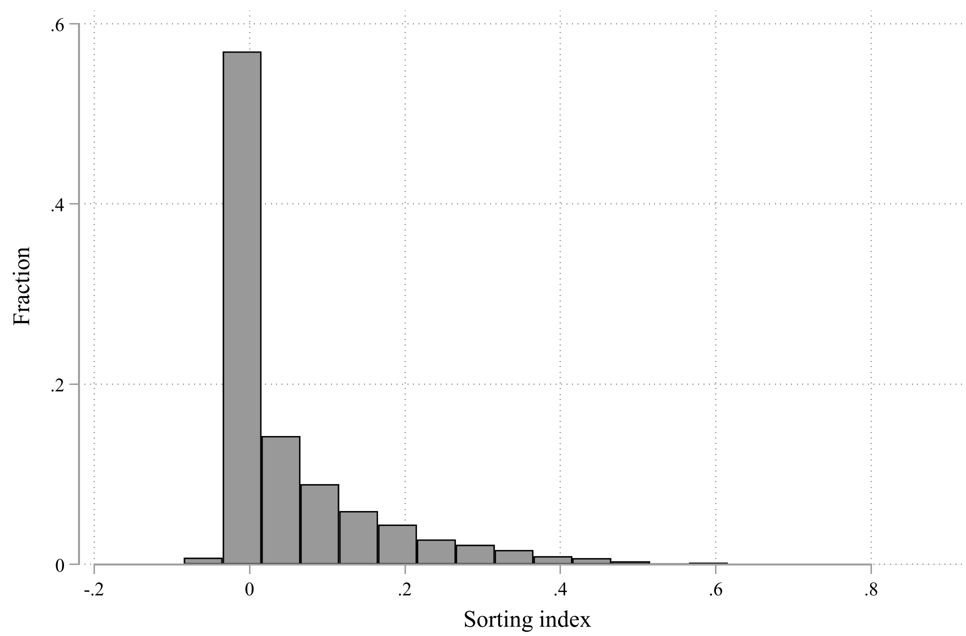

Fig. A1. Distribution of school-years with respect to the sorting index.
Notes: The figure shows the distribution of the continuous sorting index, the adjusted R-squared of regression models of the SES index on classroom indicators, estimated for each school-year separately (see Eq. (1)). School-year observations with multiple classes in the school only, weighted by the number of students.

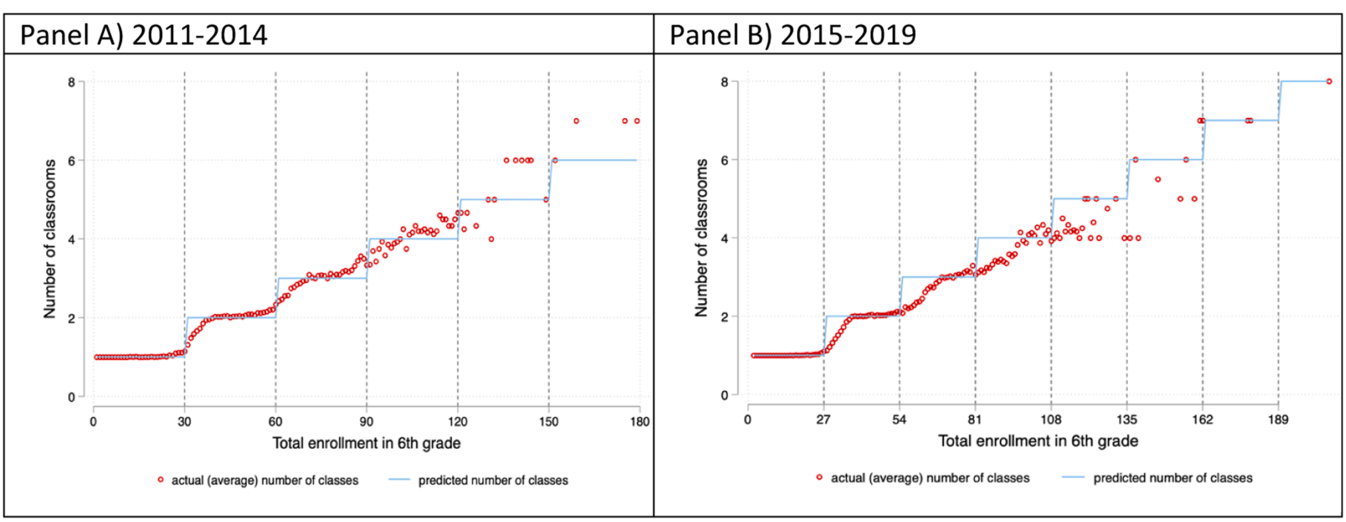

Fig. A2. Enrollment and the actual and predicted number of classes.
Notes: The figure shows relationship between enrollment and the observed average (“actual”) and predicted number of classes during the period 2011–14 (Panel A, when maximum class size rule was 30) and 2015–19 (Panel B, when class size rule was 27). Each dot represents an average number of classes/class size at the given value of enrollment. Dashed vertical lines at nominal class size rule cutoffs.

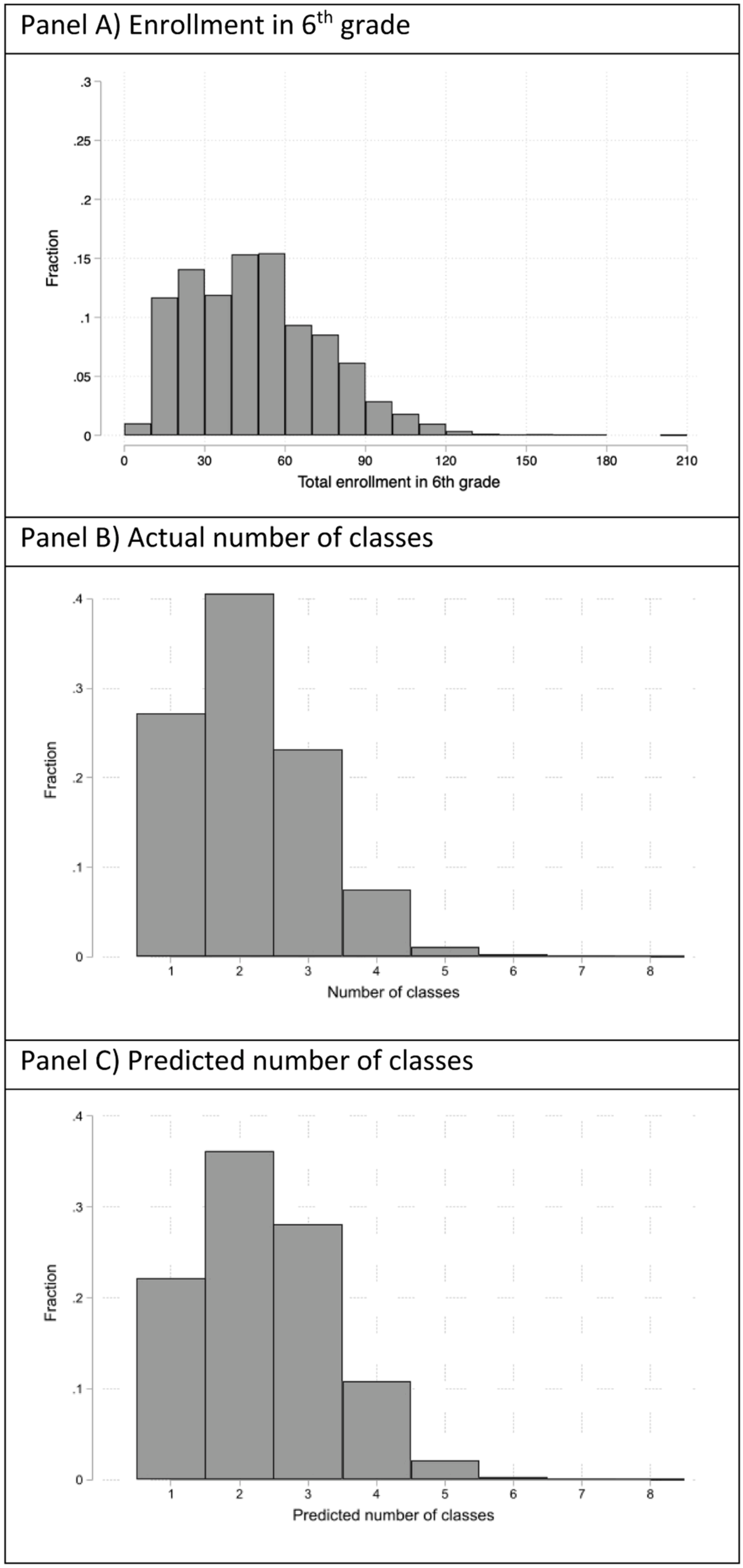

Fig. A3. Distribution of school-year observations with respect to total enrollment, actual and predicted number of classes in 6th grade.
Notes: The figure shows the fractional distribution of school-year observations with the given number of students enrolled (Panel A), actual (Panel B) and predicted (Panel C) number of classes in 6th grade. School-year observations are weighted by the number of students.

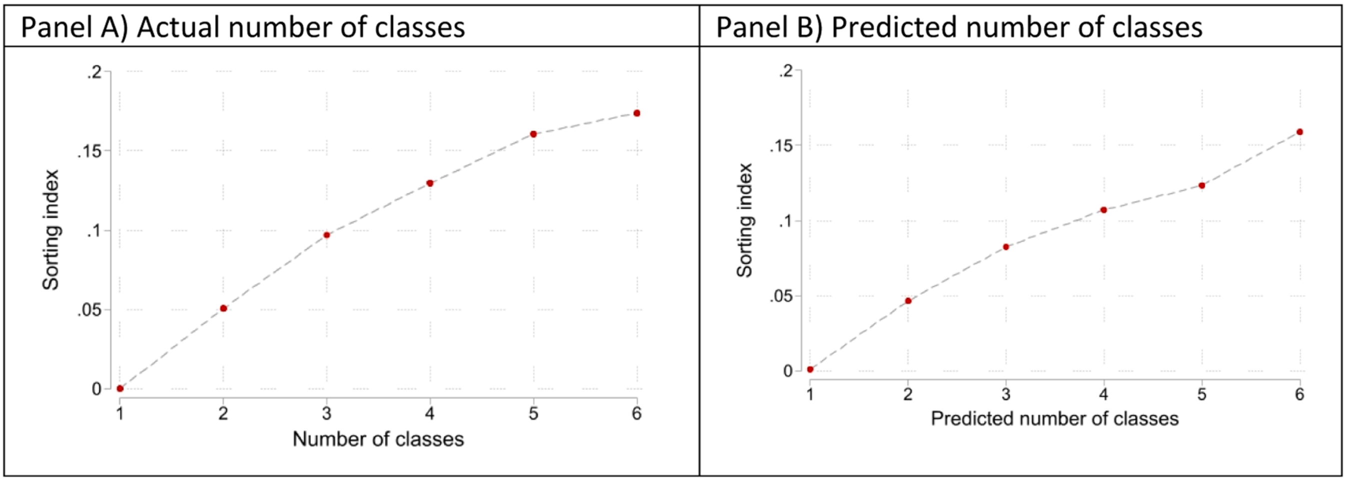

Fig. A4. Sorting index and the actual and predicted number of classes in school.
Notes: The figure shows the mean of the continuous sorting index by the actual (Panel A) and the predicted (Panel B) number of classes. The sorting index is the adjusted R-squared of regression models of the SES index on classroom indicators, estimated for each school-year separately (see Eq. (1)). School-year observations with a single class in school has a sorting index of zero. Actual and predicted number of classes are top-coded at 6 as there are less than 10 school-year observations with 7 or 8 classes. School-year observations are weighted by the number of students.

Table A1
First-stage regressions – The relationship between the actually observed and predicted number of class instruments and the binary sorting indicator in the school.

	level		VA		level		VA	
Dependent variable	sorting (1)	sorting*SES (2)	sorting (3)	sorting*SES (4)	sorting (5)	sorting*SES (6)	sorting (7)	sorting*SES (8)
Panel A) MATH								
actual number of classes	0.178***	−0.001	0.177***	−0.001				
	(0.010)	(0.005)	(0.010)	(0.005)				
actual number of classes*SES	−0.001	0.215***	−0.001	0.215***				
	(0.001)	(0.007)	(0.001)	(0.007)				
predicted number of classes					0.097***	−0.009**	0.097***	−0.010**
					(0.008)	(0.004)	(0.008)	(0.004)
predicted number of classes*SES					−0.003**	0.187***	−0.003**	0.187***
					(0.001)	(0.006)	(0.001)	(0.006)
Observations	629,605	629,605	606,510	606,510	629,605	629,605	606,510	606,510
N of schools	2719	2719	2719	2719	2719	2719	2719	2719
Multivariate F-test of excluded instruments								
first-stage F-stat	168.61	475.96	168.09	484.06	78.89	486.90	78.60	496.79
Sanderson-Windmeijer multivariate F-test of excluded instruments								
first-stage SW F-stat	329.22	949.48	328.23	965.97	141.86	969.26	141.34	989.14
Kleibergen-Paap weak identification test F-stat	163.9		163.0		69.27		68.70	
Panel B) READING								
actual number of classes	0.178***	−0.001	0.177***	−0.001				
	(0.010)	(0.005)	(0.010)	(0.005)				
actual number of classes*SES	−0.001	0.215***	−0.001	0.215***				
	(0.001)	(0.007)	(0.001)	(0.007)				
predicted number of classes					0.097***	−0.009**	0.097***	−0.010**
					(0.008)	(0.004)	(0.008)	(0.004)
predicted number of classes*SES					−0.003**	0.187***	−0.003**	0.187***
					(0.001)	(0.006)	(0.001)	(0.006)
Observations	629,978	629,978	606,834	606,834	629,978	629,978	606,834	606,834
N of schools	2719	2719	2719	2719	2719	2719	2719	2719
Multivariate F-test of excluded instruments								
first-stage F-stat	168.48	475.73	167.97	483.80	79.01	486.82	78.70	496.60
Sanderson-Windmeijer multivariate F-test of excluded instruments								
first-stage SW F-stat	329.15	949.18	328.19	965.60	142.00	968.96	141.50	988.67
Kleibergen-Paap weak identification test F-stat	163.9		163.1		69.34		68.79	

Notes: The table shows first-stage regression estimates of the sorting binary indicator and its interaction with the SES index on number of classes instruments for the math (Panel A) and reading (Panel B) level- and value added samples (see Eqs. (3a) and (3b)). Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, and school and year fixed effects. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2
The effect of sorting on test score inequalities by mother's education.

	FE		IV FE instrument: number of classes		IV FE instrument: predicted number of classes	
	level (1)	VA (2)	level (3)	VA (4)	Level (5)	VA (6)
Panel A) MATH						
sorting (dummy)	0.013*	−0.005	−0.059	−0.027	−0.006	0.025
	(0.007)	(0.006)	(0.039)	(0.032)	(0.055)	(0.049)
sorting * mother educ: elementary	−0.079***	−0.043***	−0.182***	−0.078***	−0.168***	−0.058***
	(0.012)	(0.010)	(0.026)	(0.018)	(0.029)	(0.020)
sorting * mother educ: vocational	−0.031***	−0.015***	−0.072***	−0.039***	−0.055***	−0.027**
	(0.007)	(0.005)	(0.018)	(0.011)	(0.019)	(0.012)
sorting * mother educ: college/university	0.046***	0.019***	0.080***	0.031**	0.090***	0.033**
	(0.008)	(0.005)	(0.020)	(0.012)	(0.021)	(0.013)
Observations	625,048	602,487	625,046	602,485	625,046	602,485
N of schools	2721	2721	2719	2719	2719	2719
Within R-squared	0.128	0.549	0.126	0.549	0.127	0.549
Multivariate F-test of excluded instruments						
first-stage F-stat for sorting			86.21	85.71	39.96	39.85
first-stage F-stat for sorting x mother educ: low			262.53	261.47	240.11	241.62
first-stage F-stat for sorting x mother educ: voc			293.59	294.54	268.18	267.42
first-stage F-stat for sorting x mother educ: high			223.59	221.42	237.35	235.96
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			750.42	742.26	439.70	426.79
first-stage SW F-stat for sorting x mother educ: low			911.16	931.43	876.72	884.29
first-stage SW F-stat for sorting x mother educ: voc			1132.82	1111.07	1023.66	1015.18
first-stage SW F-stat for sorting x mother educ: high			727.65	718.84	805.54	793.90
Kleibergen-Paap weak identification test F-stat			82.17	81.73	34.76	34.49
Panel B) READING						
sorting (dummy)	0.012*	−0.006	−0.073**	−0.055**	−0.014	−0.017
	(0.006)	(0.005)	(0.032)	(0.026)	(0.047)	(0.039)
sorting * mother educ: elementary	−0.070***	−0.034***	−0.186***	−0.076***	−0.189***	−0.068***
	(0.011)	(0.009)	(0.026)	(0.017)	(0.028)	(0.018)
sorting * mother educ: vocational	−0.031***	−0.014***	−0.067***	−0.035***	−0.066***	−0.034***
	(0.007)	(0.005)	(0.017)	(0.010)	(0.018)	(0.011)
sorting * mother educ: college/university	0.024***	0.002	0.030	−0.015	0.033*	−0.014
	(0.007)	(0.004)	(0.019)	(0.011)	(0.020)	(0.012)
Observations	625,411	602,804	625,409	602,802	625,409	602,802
N of schools	0.156	0.601	0.154	0.600	0.155	0.601
Within R-squared	2721	2721	2719	2719	2719	2719
Multivariate F-test of excluded instruments						
first-stage F-stat for sorting			86.09	85.62	39.98	39.87
first-stage F-stat for sorting x mother educ: low			261.91	260.84	239.99	241.52
first-stage F-stat for sorting x mother educ: voc			293.45	294.41	267.92	267.25
first-stage F-stat for sorting x mother educ: high			223.76	221.66	237.35	236.04
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			751.21	742.94	441.39	428.17
first-stage SW F-stat for sorting x mother educ: low			911.64	931.79	877.42	884.76
first-stage SW F-stat for sorting x mother educ: voc			1130.77	1109.06	1021.51	1013.10
first-stage SW F-stat for sorting x mother educ: high			728.12	718.97	805.36	793.32
Kleibergen-Paap weak identification test F-stat			82.17	81.74	34.80	34.54

Notes: The table shows regression estimates of math (Panel A) and reading (Panel B) test scores on the binary sorting indicator (see Eqs. (2) and (3)) and its interaction with three categories mother's education (with secondary diploma as the reference category). Control variables in all models: gender, special education needs status, mother's education main effect, mean and SD of SES in school and their interaction with the categories of mother's education, and school and year fixed effects. Additional controls in VA models: third order polynomial of math and reading scores in 6th grade. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3
The effect of sorting on test score inequalities, with extended set of family controls.

	FE		IV FE instrument: number of classes		IV FE instrument: predicted number of classes	
	level (1)	VA (2)	level (3)	VA (4)	level (5)	VA (6)
Panel A) MATH						
sorting (dummy)	0.003	−0.011*	−0.081**	−0.042	−0.030	0.012
	(0.007)	(0.006)	(0.036)	(0.032)	(0.054)	(0.048)
sorting x SES	0.044***	0.026***	0.095***	0.046***	0.089***	0.037***
	(0.005)	(0.004)	(0.012)	(0.008)	(0.013)	(0.009)
Observations	629,607	606,512	629,605	606,510	629,605	606,510
N of schools	2721	2721	2719	2719	2719	2719
Within R-squared	0.177	0.554	0.175	0.554	0.176	0.554

(continued on next page)

Table A3 (*continued*)

	FE		IV FE		IV FE	
			instrument: number of classes		instrument: predicted number of classes	
	level (1)	VA (2)	level (3)	VA (4)	level (5)	VA (6)
Multivariate F-test of instruments						
first-stage F-stat for sorting			168.63	168.13	78.90	78.65
first-stage F-stat for sorting x SES			476.83	484.71	487.47	497.21
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			329.05	328.08	141.66	141.14
first-stage SW F-stat for sorting x SES			951.07	967.12	970.16	989.69
Kleibergen-Paap weak identification test F-stat			163.9	163	69.21	68.65
Panel B) READING						
sorting (dummy)	−0.002	−0.014***	−0.106***	−0.079***	−0.057	−0.044
	(0.005)	(0.005)	(0.030)	(0.025)	(0.045)	(0.038)
sorting x SES	0.031***	0.015***	0.073***	0.028***	0.069***	0.022***
	(0.004)	(0.003)	(0.010)	(0.007)	(0.011)	(0.007)
Observations	629,980	606,836	629,978	606,834	629,978	606,834
N of schools	2721	2721	2719	2719	2719	2719
Within R-squared	0.210	0.606	0.208	0.606	0.209	0.606
Multivariate F-test of instruments						
first-stage F-stat for sorting			168.50	168.01	79.03	78.75
first-stage F-stat for sorting x SES			476.62	484.45	487.40	497.02
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			328.98	328.04	141.80	141.31
first-stage SW F-stat for sorting x SES			950.77	966.75	969.87	989.23
Kleibergen-Paap weak identification test F-stat			163.9	163	69.29	68.74

Notes: The table shows regression estimates of math (Panel A) and reading (Panel B) test scores on the binary sorting indicator (see Eqs. (2) and (3)) and its interaction with the SES index, with an extended set of family controls. Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, mother's and father's level of education, the number of books at home, socio-economically disadvantaged status of the student, subjective affluence of the family, whether the family receives regular child protection allowance, whether the student is entitled to subsidized or free lunch, whether the student is entitled to free textbooks, how many times the family had a holiday in the last four years, and school and year fixed effects. Additional controls in VA models: third order polynomial of math and reading scores in 6^{th} grade. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4
The effect of sorting on test score inequalities, with school-specific linear trends.

	FE		IV FE		IV FE	
			instrument: number of classes		instrument: predicted number of classes	
	level (1)	VA (2)	level (3)	VA (4)	level (5)	VA (6)
Panel A) MATH						
sorting (dummy)	0.002	−0.011*	−0.106***	−0.070*	−0.028	0.007
	(0.007)	(0.006)	(0.038)	(0.036)	(0.057)	(0.055)
sorting x SES	0.043***	0.025***	0.096***	0.047***	0.090***	0.039***
	(0.005)	(0.004)	(0.012)	(0.008)	(0.013)	(0.009)
Observations	610,178	587,913	610,178	587,913	610,178	587,913
N of schools	2392	2392	2392	2392	2392	2392
Within R-squared	0.336	0.643	0.172	0.556	0.174	0.556
Multivariate F-test of instruments						
first-stage F-stat for sorting			112.43	111.79	54.70	54.52
first-stage F-stat for sorting x SES			447.51	455.33	461.11	470.77
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			217.98	216.74	96.33	95.97
first-stage SW F-stat for sorting x SES			890.93	907.23	916.52	936.39
Kleibergen-Paap weak identification test F-stat			108.5	107.6	47.19	46.87
Panel B) READING						
sorting (dummy)	−0.001	−0.014***	−0.057*	−0.041	0.011	0.010
	(0.005)	(0.005)	(0.032)	(0.030)	(0.048)	(0.045)
sorting x SES	0.031***	0.014***	0.073***	0.027***	0.070***	0.023***
	(0.004)	(0.003)	(0.010)	(0.007)	(0.011)	(0.007)
Observations	610,536	588,223	610,536	588,223	610,536	588,223
N of schools	2392	2392	2392	2392	2392	2392
Within R-squared	0.359	0.682	0.207	0.608	0.208	0.608
Multivariate F-test of instruments						
first-stage F-stat for sorting			112.34	111.73	54.76	54.57
first-stage F-stat for sorting x SES			447.46	455.27	461.15	470.72
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			217.92	216.75	96.33	96.02
first-stage SW F-stat for sorting x SES			890.89	907.17	916.41	936.14
Kleibergen-Paap weak identification test F-stat			108.5	107.6	47.21	46.91

Notes: The table shows regression estimates of math (Panel A) and reading (Panel B) test scores on the binary sorting indicator (see Eqs. (2) and (3)), with school-specific linear trends. Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, school and year fixed effects and school specific trends (school fixed effect and year interactions). Additional controls in VA models: third order polynomial of math and reading scores in 6th grade. Sample: Schools with at least 6 school-year observations. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5
The effect of sorting on test score inequalities, with alternative sorting threshold ($p - value \leq 0.2$).

	FE		IV FE instrument: number of classes		IV FE instrument: predicted number of classes	
	level (1)	VA (2)	level (3)	VA (4)	level (5)	VA (6)
Panel A) MATH						
sorting (dummy)	0.008	−0.002	−0.069**	−0.035	−0.024	0.011
	(0.006)	(0.006)	(0.031)	(0.027)	(0.044)	(0.040)
sorting x SES	0.034***	0.017***	0.080***	0.038***	0.074***	0.030***
	(0.005)	(0.004)	(0.010)	(0.007)	(0.011)	(0.007)
Observations	629,607	606,512	629,605	606,510	629,605	606,510
N of schools	2721	2721	2719	2719	2719	2719
Within R-squared	0.174	0.553	0.173	0.553	0.173	0.553
Multivariate F-test of excluded instruments						
first-stage F-stat for sorting			200.27	199.18	110.47	109.94
first-stage F-stat for sorting x SES			458.43	467.74	674.81	698.12
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			397.48	394.79	206.89	205.32
first-stage SW F-stat for sorting x SES			911.90	930.15	1352.31	1399.06
Kleibergen-Paap weak identification test F-stat			199	197.7	102.8	101.8
Panel B) READING						
sorting (dummy)	−0.000	−0.008*	−0.092***	−0.068***	−0.050	−0.037
	(0.005)	(0.004)	(0.025)	(0.022)	(0.037)	(0.032)
sorting x SES	0.028***	0.013***	0.062***	0.024***	0.059***	0.019***
	(0.004)	(0.003)	(0.009)	(0.006)	(0.010)	(0.006)
Observations	629,980	606,836	629,978	606,834	629,978	606,834
N of schools	2721	2721	2719	2719	2719	2719
Within R-squared	0.208	0.606	0.206	0.605	0.207	0.606
Multivariate F-test of excluded instruments						
first-stage F-stat for sorting			200.09	199.09	110.51	109.98
first-stage F-stat for sorting x SES			458.69	468.06	675.07	698.47
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			397.31	394.76	206.90	205.36
first-stage SW F-stat for sorting x SES			912.54	930.90	1352.75	1399.67
Kleibergen-Paap weak identification test F-stat			198.9	197.7	102.8	101.9

Notes: The table shows regression estimates of math (Panel A) and reading (Panel B) test scores on the binary sorting indicator (see Eqs. (2) and (3)), with an alternative sorting classification. Schools are classified as sorting if the p-value of the joint F-test of classroom dummies is equal to or below 0.2 (Eq. (1)). Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, and school and year fixed effects. Additional controls in VA models: third order polynomial of math and reading scores in 6th grade. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A6
The effect of sorting on test score inequalities – subsample of schools with $p - value \leq 0.05$ (sorting) and $p - value \geq 0.5$ (non-sorting).

	FE		IV FE instrument: number of classes		IV FE instrument: predicted number of classes	
	level (1)	VA (2)	level (3)	VA (4)	level (5)	VA (6)
Panel A) MATH						
sorting (dummy)	0.006	−0.010	−0.104***	−0.063**	−0.063	−0.006
	(0.008)	(0.008)	(0.033)	(0.029)	(0.054)	(0.047)
sorting x SES	0.047***	0.027***	0.080***	0.040***	0.080***	0.036***
	(0.006)	(0.004)	(0.010)	(0.007)	(0.011)	(0.007)
Observations	459,403	442,290	459,401	442,288	459,401	442,288
N of schools	2708	2708	2706	2706	2706	2706
Within R-squared	0.178	0.552	0.177	0.551	0.178	0.552
Multivariate F-test of excluded instruments						
first-stage F-stat for sorting			172.88	170.82	86.00	85.18
first-stage F-stat for sorting x SES			360.98	365.74	650.39	668.07
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			352.82	347.78	158.27	156.92
first-stage SW F-stat for sorting x SES			714.55	723.70	1298.05	1333.51

(continued on next page)

Table A6 (*continued*)

	FE		IV FE instrument: number of classes		IV FE instrument: predicted number of classes	
	level (1)	VA (2)	level (3)	VA (4)	level (5)	VA (6)
Kleibergen-Paap weak identification test F-stat			173.9	171.7	78.96	78.17
Panel B) READING						
sorting (dummy)	−0.004	−0.016**	−0.111***	−0.078***	−0.102**	−0.071*
	(0.007)	(0.006)	(0.028)	(0.025)	(0.046)	(0.039)
sorting x SES	0.035***	0.017***	0.063***	0.026***	0.060***	0.020***
	(0.005)	(0.004)	(0.008)	(0.006)	(0.009)	(0.006)
Observations	459,669	442,521	459,667	442,519	459,667	442,519
N of schools	2708	2708	2706	2706	2706	2706
Within R-squared	0.212	0.606	0.211	0.605	0.211	0.605
Multivariate F-test of excluded instruments						
first-stage F-stat for sorting			172.80	170.83	86.05	85.27
first-stage F-stat for sorting x SES			361.10	365.90	650.76	668.57
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			352.80	347.88	158.33	157.04
first-stage SW F-stat for sorting x SES			714.92	724.13	1298.68	1334.30
Kleibergen-Paap weak identification test F-stat			173.8	171.7	78.99	78.23

Notes: The table shows regression estimates of math (Panel A) and reading (Panel B) test scores on the binary sorting indicator (see Eqs. (2) and (3)) and its interaction with the SES index, with an alternative sorting classification. School-years are classified as sorting if the p-value of the joint F-test of classroom dummies is equal to or below 0.05 (Eq. (1)). School-years with a p-value between 0.05 and 0.5 are excluded. Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, and school and year fixed effects. Additional controls in VA models: third order polynomial of math and reading scores in 6th grade. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A7
The effect of sorting on test score inequalities – continuous sorting index.

	FE		IV FE instrument: number of classes		IV FE instrument: predicted number of classes	
	level (1)	VA (2)	level (3)	VA (4)	level (5)	VA (6)
Panel A) MATH						
sorting index (adj. R-squared)	0.050*	−0.025	−0.349**	−0.182	−0.124	0.102
	(0.029)	(0.027)	(0.170)	(0.147)	(0.330)	(0.296)
sorting x SES	0.203***	0.098***	0.446***	0.213***	0.455***	0.190***
	(0.022)	(0.016)	(0.055)	(0.037)	(0.065)	(0.043)
Observations	629,607	606,512	629,605	606,510	629,605	606,510
N of schools	2721	2721	2719	2719	2719	2719
Within R-squared	0.174	0.553	0.172	0.553	0.173	0.553
Multivariate F-test of instruments						
first-stage F-stat for sorting			120.50	119.65	41.52	41.39
first-stage F-stat for sorting x SES			162.82	162.75	152.34	153.23
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			234.65	231.27	66.16	64.67
first-stage SW F-stat for sorting x SES			325.25	326.34	288.87	291.57
Kleibergen-Paap weak identification test F-stat			116.0	114.9	32.94	32.28
Panel B) READING						
sorting index (adj. R-squared)	0.032	−0.038*	−0.480***	−0.364***	−0.325	−0.266
	(0.025)	(0.022)	(0.140)	(0.119)	(0.277)	(0.236)
sorting x SES	0.154***	0.058***	0.341***	0.129***	0.356***	0.110***
	(0.021)	(0.015)	(0.048)	(0.031)	(0.057)	(0.036)
Observations	629,980	606,836	629,978	606,834	629,978	606,834
N of schools	2721	2721	2719	2719	2719	2719
Within R-squared	0.208	0.606	0.206	0.605	0.206	0.605
Multivariate F-test of instruments						
first-stage F-stat for sorting			120.39	119.51	41.59	41.45
first-stage F-stat for sorting x SES			162.85	162.77	152.48	153.35
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			234.59	231.19	66.21	64.74
first-stage SW F-stat for sorting x SES			325.36	326.40	288.99	291.70
Kleibergen-Paap weak identification test F-stat			116.0	114.8	32.96	32.31

Notes: The table shows regression estimates of math (Panel A) and reading (Panel B) test scores on the continuous sorting index (see Eqs. (2) and (3)). The continuous sorting index is measured by the adjusted R-squared of Eq. (1). Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, and school and year fixed effects. Additional controls in VA models: third order polynomial of math and reading scores in 6th grade. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A8
The effect of sorting on test score inequalities – subsample of school-years with multiple classrooms.

	FE		IV FE instrument: number of classes		IV FE instrument: predicted number of classes	
	level	VA	level	VA	level	VA
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A) MATH						
sorting (dummy)	0.005	−0.010	−0.125**	−0.049	−0.111	−0.030
	(0.007)	(0.006)	(0.051)	(0.044)	(0.073)	(0.064)
sorting x SES	0.037***	0.021***	0.111***	0.035**	0.110***	0.028*
	(0.006)	(0.004)	(0.020)	(0.014)	(0.023)	(0.016)
Observations	458,583	442,138	458,583	442,138	458,583	442,138
N of schools	1600	1600	1600	1600	1600	1600
Within R-squared	0.178	0.575	0.174	0.575	0.175	0.575
Multivariate F-test of instruments						
first-stage F-stat for sorting			77.88	77.72	34.02	34.10
first-stage F-stat for sorting x SES			159.60	162.36	140.53	143.92
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			156.93	157.10	66.73	66.89
first-stage SW F-stat for sorting x SES			318.85	324.24	281.10	287.81
Kleibergen-Paap weak identification test F-stat			77.35	77.09	32.73	32.63
Panel B) READING						
sorting (dummy)	0.000	−0.013***	−0.135***	−0.079**	−0.095	−0.049
	(0.005)	(0.005)	(0.042)	(0.034)	(0.061)	(0.051)
sorting x SES	0.025***	0.011***	0.093***	0.024**	0.090***	0.016
	(0.005)	(0.004)	(0.017)	(0.011)	(0.020)	(0.013)
Observations	458,845	442,362	458,845	442,362	458,845	442,362
N of schools	1600	1600	1600	1600	1600	1600
Within R-squared	0.210	0.619	0.206	0.618	0.208	0.619
Multivariate F-test of instruments						
first-stage F-stat for sorting			77.84	77.68	34.07	34.15
first-stage F-stat for sorting x SES			159.51	162.23	140.51	143.82
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			156.82	156.99	66.79	66.97
first-stage SW F-stat for sorting x SES			318.66	323.94	281.03	287.60
Kleibergen-Paap weak identification test F-stat			77.34	77.08	32.76	32.67

Notes: The table shows regression estimates of math (Panel A) and reading (Panel B) test scores on the binary sorting indicator (see Eqs. (2) and (3)), for the subsample of school-years with multiple classrooms. Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, and school and year fixed effects. Additional controls in VA models: third order polynomial of math and reading scores in 6th grade. All students in school-years where any of the classes follows a specialized curriculum are excluded from the sample. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A9
The effect of sorting on test score inequalities, fixed effects models separately for schools with 2 and 3 classrooms, and controlling for the number of classrooms.

	Two classrooms FE		Three classrooms FE		Control: number of classrooms FE	
	level	VA	level	VA	level	VA
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A) MATH						
sorting (dummy)	0.017*	−0.008	0.014	0.003	0.007	−0.010
	(0.009)	(0.009)	(0.012)	(0.011)	(0.007)	(0.006)
sorting x SES	0.024***	0.019***	0.036***	0.017**	0.043***	0.025***
	(0.007)	(0.006)	(0.009)	(0.007)	(0.005)	(0.004)
Observations	255,615	246,326	145,926	140,719	629,607	606,512
N of schools	1417	1417	692	692	2721	2721
Within R-squared	0.173	0.560	0.181	0.589	0.174	0.554
Panel B) READING						
sorting (dummy)	0.013*	−0.011	0.005	−0.006	0.003	−0.012**
	(0.008)	(0.007)	(0.010)	(0.009)	(0.005)	(0.005)
sorting x SES	0.016**	0.011**	0.026***	0.009*	0.030***	0.015***
	(0.007)	(0.005)	(0.007)	(0.005)	(0.004)	(0.003)
Observations	255,788	246,473	145,991	140,774	629,980	606,836
N of schools	1417	1417	692	692	2721	2721
Within R-squared	0.206	0.609	0.211	0.628	0.208	0.606

Notes: The table shows regression estimates of math (Panel A) and reading (Panel B) test scores on the binary sorting indicator (see Eqs. (2) and (3)) and its interaction with the SES index, using alternative model specifications. Models 1 and 2 only include schools with two classrooms, models 3 and 4 only include schools with three classrooms, models 5 and 6 include all school and control for the number of classrooms. Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, and school and year fixed effects. Additional controls in VA models: third order polynomial of math and reading scores in 6th grade. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A10
The effect of sorting on test score inequalities using matched samples of treated (sorting) and control (non-sorting) schools.

	FE		FE		FE	
	level	VA	level	VA	level	VA
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A) MATH						
sorting (dummy)	0.013*	−0.015***	0.011	−0.017***	0.010	−0.017***
	(0.008)	(0.005)	(0.007)	(0.005)	(0.007)	(0.005)
sorting x SES	0.035***	0.021***	0.033***	0.020***	0.034***	0.019***
	(0.006)	(0.005)	(0.006)	(0.004)	(0.006)	(0.004)
Student SES	Yes	Yes	Yes	Yes	Yes	Yes
Student-level controls	No	No	Yes	Yes	Yes	Yes
School-level controls	No	No	Yes	Yes	Yes	Yes
Interactions of student SES and school-level controls	No	No	No	No	Yes	Yes
Third order polynomial of math and reading scores in 6^{th} grade	No	Yes	No	Yes	No	Yes
Treated-control pair FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	325,496	313,747	325,496	313,747	325,496	313,747
N of schools	1538	1538	1538	1538	1538	1538
Within R-squared	0.285	0.664	0.313	0.668	0.315	0.668
Panel B) READING						
sorting (dummy)	0.006	−0.015***	0.006	−0.016***	0.005	−0.016***
	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)
sorting x SES	0.030***	0.014***	0.028***	0.013***	0.030***	0.013***
	(0.006)	(0.004)	(0.006)	(0.004)	(0.005)	(0.004)
Student SES	Yes	Yes	Yes	Yes	Yes	Yes
Student-level controls	No	No	Yes	Yes	Yes	Yes
School-level controls	No	No	Yes	Yes	Yes	Yes
Interactions of student SES and school-level controls	No	No	No	No	Yes	Yes
Third order polynomial of math and reading scores in 6^{th} grade	No	Yes	No	Yes	No	Yes
Treated-control pair FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	325,647	313,884	325,647	313,884	325,647	313,884
N of schools	1538	1538	1538	1538	1538	1538
Within R-squared	0.291	0.687	0.335	0.693	0.337	0.694

Notes: The table shows regression estimates of math (Panel A) and reading (Panel B) test scores on the binary sorting indicator (see Eqs. (2) and (3)) and its interaction with the SES index, using matched pairs of treated (sorting) and control (non-sorting) schools. Control variables in all models: student SES and treated-control pair fixed effects. Student-level controls include gender and special educational needs status. School-level controls include county, type of provider, type of settlement, the predicted number of classes, and the mean and SD of SES in school. Additional controls in VA models: third order polynomial of math and reading scores in 6^{th} grade. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A11
The effect of sorting on test score inequalities – subsample of schools with no ability grouping and no specialized curriculum classes.

	FE		IV FE instrument: number of classes		IV FE instrument: predicted number of classes	
	level	VA	level	VA	level	VA
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A) MATH						
sorting (dummy)	0.013	−0.011	−0.050	−0.020	0.066	0.139*
	(0.009)	(0.008)	(0.049)	(0.046)	(0.087)	(0.083)
sorting x SES	0.031***	0.026***	0.088***	0.055***	0.070***	0.035***
	(0.007)	(0.006)	(0.017)	(0.012)	(0.019)	(0.014)
Observations	410,366	395,065	410,364	395,063	410,364	395,063
N of schools	2465	2465	2463	2463	2463	2463
Within R-squared	0.171	0.535	0.170	0.535	0.171	0.533
Multivariate F-test of instruments						
first-stage F-stat for sorting			104.46	103.86	39.50	38.98
first-stage F-stat for sorting x SES			253.85	262.31	215.52	222.38
Sanderson-Windmeijer multivariate F-test of instruments						
first-stage SW F-stat for sorting			204.73	204.16	68.06	67.59
first-stage SW F-stat for sorting x SES			501.73	519.41	424.08	437.93
Kleibergen-Paap weak identification test F-stat			102.3	102	33.26	33.02
Panel B) READING						
sorting (dummy)	0.001	−0.017**	−0.124***	−0.102***	−0.005	0.009
	(0.008)	(0.007)	(0.041)	(0.038)	(0.073)	(0.067)
sorting x SES	0.021***	0.017***	0.069***	0.035***	0.070***	0.030**
	(0.006)	(0.005)	(0.016)	(0.011)	(0.018)	(0.012)
Observations	410,657	395,320	410,655	395,318	410,655	395,318
N of schools	2465	2465	2463	2463	2463	2463
Within R-squared	0.206	0.595	0.204	0.594	0.206	0.595
Multivariate F-test of instruments						
first-stage F-stat for sorting			104.35	103.79	39.57	39.05
first-stage F-stat for sorting x SES			253.57	262.02	215.48	222.33
Sanderson-Windmeijer multivariate F-test of instruments						

(*continued on next page*)

Table A11 (*continued*)

	FE		IV FE instrument: number of classes		IV FE instrument: predicted number of classes	
	level	VA	level	VA	level	VA
first-stage SW F-stat for sorting			204.71	204.19	68.17	67.73
first-stage SW F-stat for sorting x SES			501.38	519.03	423.99	437.86
Kleibergen-Paap weak identification test F-stat			102.3	102	33.33	33.10

Notes: The table shows regression estimates of math (Panel A) and reading (Panel B) test scores on the binary sorting indicator (see Eqs. (2) and (3)) and its interaction with the SES index, for the subsample of schools with no ability grouping and no specialized curriculum classes. Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, and school and year fixed effects. Additional controls in VA models: third order polynomial of math and reading scores in 6th grade. All students in school-years where any of the classes follows a specialized curriculum are excluded from the sample. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A12
The effect of SES sorting and ability sorting on test score inequalities.

	Math level (1)	VA (2)	Reading level (3)	VA (4)
SES sorting: yes, ability sorting: no	0.009	−0.007	0.001	−0.013**
	(0.007)	(0.007)	(0.006)	(0.005)
SES sorting: no, ability sorting: yes	0.022**	0.017*	0.008	0.008
	(0.010)	(0.010)	(0.008)	(0.008)
SES sorting: yes, ability sorting: yes	0.001	−0.008	−0.005	−0.010
	(0.013)	(0.011)	(0.010)	(0.008)
(SES sorting: yes, ability sorting: no) x SES	0.035***	0.022***	0.026***	0.014***
	(0.006)	(0.005)	(0.005)	(0.004)
(SES sorting: no, ability sorting: yes) x SES	0.009	0.007	0.008	0.005
	(0.007)	(0.006)	(0.007)	(0.006)
(SES sorting: yes, ability sorting: yes) x SES	0.071***	0.036***	0.051***	0.019***
	(0.010)	(0.007)	(0.007)	(0.006)
SES	0.283***	0.092***	0.307***	0.097***
	(0.012)	(0.010)	(0.011)	(0.009)
Observations	628,428	605,370	628,801	605,694
N of schools	2714	2714	2714	2714
R-squared	0.174	0.553	0.208	0.606

Notes: The table shows regression estimates of math and reading test scores on combinations of the binary sorting indicator (see Eqs. (2) and (3)) and the ability sorting indicator, and their interactions with the SES index. Control variables in all models: gender, special education needs status, SES main effect, mean and SD of SES in school and their interaction with SES, and school and year fixed effects. Additional controls in VA models: third order polynomial of math and reading scores in 6th grade. Standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

References

Agasisti, T., & Falzetti, P. (2017). Between-classes sorting within schools and test scores: An empirical analysis of Italian junior secondary schools. *International Review of Economics, 64*(1), 1–45. https://doi.org/10.1007/s12232-016-0261-4

Ammermueller, A. (2005). Educational opportunities and the role of institutions (SSRN Scholarly Paper 753366). 10.2139/ssrn.753366.

Bajomi, I., Berényi, E., Erőss, G., & Imre, A. (2008). Ahol ritka jószág a tanuló: Oktatásirányítás, cselekvési logikák és egyenlőtlenségek Budapest egyik kerületében. In E. Berényi, B. Berkovits, & G. Erőss (Eds.), *Iskolarend. Kiváltásg és különbségtétel a közoktatásban* (pp. 279–327). Gondolat.

Barrett, N., & Toma, E. F. (2013). Reward or punishment? Class size and teacher quality. *Economics of Education Review, 35*, 41–52. https://doi.org/10.1016/j.econedurev.2013.03.001

Benito, R., Alegre, M.À., & Gonzàlez-Balletbò, I. (2014). School segregation and its effects on educational equality and efficiency in 16 OECD comprehensive school systems. *Comparative Education Review, 58*(1), 104–134. https://doi.org/10.1086/672011

Berényi, E., Berkovits, B., & Erőss, G. (2008). Iskolarendszer és szabad választás. A jóindulatú szegregációról. In E. Berényi, B. Berkovits, & G. Erőss (Eds.). *Iskolarend. Kiváltság és különbségtétel a közoktatásban* (pp. 15–26). Gondolat.

Betts, J. R., & Shkolnik, J. L. (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review, 19*(1), 1–15. https://doi.org/10.1016/S0272-7757(98)00044-2

Betts, J. R. (2011). The Economics of Tracking in Education. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *vol. 3. Handbook of the Economics of Education* (pp. 341–381). Elsevier. https://doi.org/10.1016/B978-0-444-53429-3.

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

Clotfelter, C. T., Ladd, H. F., Clifton, C. R., & Turaeva, M. R. (2021). School segregation at the classroom level in a southern 'new destination' state. *Race and Social Problems, 13*(2), 131–160. https://doi.org/10.1007/s12552-020-09309-w

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review, 24*(4), 377–392. https://doi.org/10.1016/j.econedurev.2004.06.008

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2002). Segregation and resegregation in North Carolina's public school classrooms do southern schools face rapid resegregation. *North Carolina Law Review, 81*(4), 1463–1512.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources, XLI*(4), 778–820. https://doi.org/10.3368/jhr.XLI.4.778

Collins, C. A., & Gan, L. (2013). Does sorting students improve scores? An analysis of class composition (Working Paper 18848). *National Bureau of Economic Research.* https://doi.org/10.3386/w18848

Conger, D. (2005). Within-school segregation in an urban school district. *Educational Evaluation and Policy Analysis, 27*(3), 225–244. https://doi.org/10.3102/01623737027003225

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66*(3), 227–268. https://doi.org/10.3102/00346543066003227

Dee, T. S., & Jacob, B. (2011). The impact of no Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30*(3), 418–446. https://doi.org/10.1002/pam.20586

Duflo, E., Dupas, P., & 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. https://doi.org/10.1257/aer.101.5.1739

Engzell, P., & Raabe, I. J. (2023). Within-school achievement sorting in comprehensive and tracked systems. *Sociology of Education.* https://doi.org/10.1177/00380407231183952, 00380407231183952.

Erőss, G. (2008). Mozaik-Magyarország felfedezése, avagy az esetlegesség uralma. In E. Berényi, B. Berkovits, & G. Erőss (Eds.), *Iskolarend. kiváltság és különbségtétel a közoktatásban* (pp. 27–76). Gondolat.

Ferrer-Esteban, G. (2016). Trade-off between effectiveness and equity? An analysis of social sorting between classrooms and between schools. *Comparative Education Review, 60*(1), 151–183. https://doi.org/10.1086/684490

Fryer, R. G., Jr (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *The Quarterly Journal of Economics, 129*(3), 1355–1407. https://doi.org/10.1093/qje/qju011

Gutiérrez, G., Jerrim, J., & Torres, R. (2020). School segregation across the world: Has any progress been made in reducing the separation of the rich from the poor? *The Journal of Economic Inequality, 18*(2), 157–179. https://doi.org/10.1007/s10888-019-09437-3

Hajdu, T., Hermann, Z., Horn, D., Hőnich, H., & Varga, J. (2022). *A közoktatás indikátorrendszere 2021 [The indicator system of the Hungarian public education 2021]*. KRTK KTI. https://kti.krtk.hu/wp-content/uploads/2022/02/A_kozoktatas_indikatorrendszere_2021.pdf.

Hajdu, T., Kertesi, G., & Kézdi, G. (2021). Ethnic segregation and inter-ethnic relationships in Hungarian schools. *On Education. Journal for Research and Debate, 4* (11). https://doi.org/10.17899/on_ed.2021.11.5

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics, 18*(5), 527–544. https://doi.org/10.1002/jae.741

Hermann, Z., & Horváth, H. (2022). Tanári eredményesség és tanár–diák összepárosítás az általános iskolákban. Empirikus mintázatok három magyarországi tankerület adatai alapján. *Közgazdasági Szemle, 69*(11), 1377–1406. https://doi.org/10.18414/KSZ.2022.11.1377

Hermann, Z., Kertesi, G., & Varga, J. (2024). A teszteredmények társadalmi egyenlőtlensége és az általános iskolai szegregáció. *Közgazdasági Szemle, 71*(4). https://doi.org/10.18414/KSZ.2024.4.353. Article 4.

Hermann, Z., & Kisfalusi, D. (2023). School segregation, student achievement, and educational attainment in Hungary. *International Journal of Comparative Sociology*. https://doi.org/10.1177/00207152231198434, 00207152231198434.

Hermann, Z., & Semjén, A. (2021). The effects of centralisation of school governance and funding on inequalities in education lessons from a policy reform in Hungary (Working Paper KRTK-KTI WP-2021/38). KRTK-KTI Working Papers. https://www.econstor.eu/handle/10419/256909.

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

Horváth, H. (2015). *Classroom assignment policies and implications for teacher value-added estimation. Job market paper*. University of California-Berkeley Department of Economics [Google Scholar].

Isenberg, E., Max, J., Gleason, P., & Deutsch, J. (2022). Do low-income students have equal access to effective teachers? *Educational Evaluation and Policy Analysis, 44*(2), 234–256. https://doi.org/10.3102/01623737211040511

Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher, 42*(6), 304–316. https://doi.org/10.3102/0013189X13495087

Kalogrides, D., Loeb, S., & Béteille, T. (2013). Systematic sorting: Teacher characteristics and class assignments. *Sociology of Education, 86*(2), 103–123. https://doi.org/10.1177/0038040712456555

Keller, T., & Elwert, F. (2023). Feasible peer effects: Experimental evidence for deskmate effects on educational achievement and inequality. *Sociological Science, 10*, 806–829. https://doi.org/10.15195/v10.a28

Kertesi, G., & Kézdi, G. (2012). Ethnic segregation between Hungarian schools: Long-run trends and geographic distribution. *Hungarian Statistical Review, 90*, 18–45. Special number 16.

Kertesi, G., & Kézdi, G. (2016). On the test score gap between Roma and non-Roma students in Hungary and its potential causes. *Economics of Transition, 24*(1), 135–162. https://doi.org/10.1111/ecot.12076

Kisfalusi, D., Hermann, Z., & Keller, T. (2023). Discrimination in track recommendation but not in grading: Experimental evidence among primary school teachers in Hungary. Manuscript under Review.

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

Lefgren, L. (2004). Educational peer effects and the Chicago public schools. *Journal of Urban Economics, 56*(2), 169–191. https://doi.org/10.1016/j.jue.2004.03.010

Lucas, S. R. (2001). Effectively maintained inequality: Education transitions, track mobility, and social background effects. *American Journal of Sociology, 106*(6), 1642–1690. https://doi.org/10.1086/321300

Morgan, P.R., & McPartland, J.M. (1981). The extent of classroom segregation within desegregated schools. https://eric.ed.gov/?id=ED210405.

Paufler, N. A., & Amrein-Beardsley, A. (2014). The random assignment of students into elementary classrooms: Implications for value-added analyses and interpretations. *American Educational Research Journal, 51*(2), 328–362. https://doi.org/10.3102/0002831213508299

Player, D. (2010). Nonmonetary compensation in the public teacher labor market. *Education Finance and Policy, 5*(1), 82–103. https://doi.org/10.1162/edfp.2009.5.1.5105

Reardon, S. F., Weathers, E., Fahle, E., Jang, H., & Kalogrides, D. (2022). Is separate still unequal? *New Evidence on School Segregation and Racial Academic Achievement Gaps*. https://cepa.stanford.edu/content/separate-still-unequal-new-evidence-school-segregation-and-racial-academic-achievement-gaps.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far?. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.). In *Handbook of the economics of education, 3* pp. 249–277). Elsevier. https://doi.org/10.1016/B978-0-444-53429-3.00004-1.

Sandor-Szalay, E., Rutai, Z., & Tordai, E. (2019). *Synergies of national and international protection of minorities living in Hungary–Activities of the ombudsman for minorityrights of Hungary 2013-2019*. OFFICE OF THE COMMISSIONER FOR FUNDAMENTAL RIGHTS OF HUNGARY. https://www.ajbh.hu/documents/14315/131278/SYNERGIES+OF+NATIONAL+AND+INTERNATIONAL+PROTECTION+OF+MINORITIES+LIVING+IN+HUNGARY+E_BOOK.pdf/20ff4ec2-eda2-8b56-8608-854c122cd051?version?1.0&t=1571209064906& viewed on 24 May 2024.

Schütz, G., Ursprung, H. W., & Wößmann, L. (2008). Education policy and equality of opportunity. *Kyklos, 61*(2), 279–308. https://doi.org/10.1111/j.1467-6435.2008.00402.x

Stullich, S., Eisner, E., McCrary, J., & Roney, C. (2006). Stullich,S., Eisner,E., McCrary, J., & Roney, C. (2006). National assessment of title I: Interim report. *Volume* 1*:* Implementation of Title I. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

The Educators Spin On It. (2024). *How to write a letter for classroom placement or teacher request with a sample letter*. Educators Spin On It. by the https://theeducatorsspinonit.com/how-to-write-placement-letter-for/. on May 8, 2024 10:36.

Triventi, M., Skopek, J., Kulic, N., Buchholz, S., & Blossfeld, H. P. (2020). Advantage 'Finds Its Way': How privileged families exploit opportunities in different systems of secondary education. *Sociology, 54*(2), 237–257. https://doi.org/10.1177/0038038519874984

Zolnay, J. (2018). Commuting to segregation. The role of pupil commuting in a Hungarian city: Between school segregation and inequality. *Szociológiai Szemle, 28* (4). Article 4.