KU LEUVEN

DEPARTMENT OF ECONOMICS

Cohort Catch-Up: Exploring Trends in Student Achievement Post-Pandemic in Flanders, Belgium

Letizia Gambi, Kristof De Witte

FACULTY OF ECONOMICS AND BUSINESS

DISCUSSION PAPER SERIES DPS 24.04

January 2024

Cohort Catch-Up: Exploring Trends in Student Achievement Post-Pandemic in Flanders, Belgium[1]

Letizia Gambi[2] Kristof De Witte[3]

Abstract: This paper takes stock of the long-term evolution of the trend in student achievement in the final year of primary education in Flanders, Belgium. Using panel data with standardised test scores, we assess the average change in students' test scores between-cohorts, by comparing most recent cohorts with either the pre-pandemic levels (the year 2019), or the year 2021. We show that while the 2023 cohort underperform compared to pre-pandemic levels (-0.27 SD in 'reading comprehension', -0.18 SD in 'geometry', -0.12 SD in 'technology', and -0.17 SD in 'social science – time'), a catch-up in learning between-cohorts takes place, with the 2023 cohort catching up with the 2021 cohort, regardless of the base year. Furthermore, we find that high-performing students (p75-p95), who had previously experienced an especially pronounced negative trend in test scores in 2022, now perform at levels observed in previous years. Furthermore, we observe an improvement in test scores across all schools, irrespective of their SES composition, in all subjects and content subdomains, except for listening. The trend analysis of specific content subdomains within multiple subjects, which is lacking in the literature, provide a nuanced understanding of the dynamics at play, offering valuable insights for education systems beyond that of Flanders.

Keywords: COVID-19; Learning deficits; Standardised tests; Trends in education
JEL-classification: I21, I24
Declarations of interest: none

[1] The data for this study are protected by a confidentiality agreement and we are precluded from sharing the data with others. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. We would like to thank Marijke De Meyst, Anton Derks, Teun Pauls, and participants to the LEER seminar for assistance with the data and/or useful insights and comments. The authors acknowledge financial support from the Flemish Science Organisation (grant G0D2722N) and Horizon Europe Grant EFFEct (grant 101129146).

[2] Letizia Gambi, Leuven Economics of Education Research (LEER), KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium, E-mail: letizia.gambi@kuleuven.be

[3] Kristof De Witte, Leuven Economics of Education Research (LEER), KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium, E-mail: kristof.dewitte@kuleuven.be; United Nations University Maastricht Economic and Social Research Institute on Innovation and Technology (UNU-MERIT),
Boschstraat 24, 6211 AX Maastricht, the Netherlands, E-mail: k.dewitte@maastrichtuniversity.nl

1. Introduction

The pervasive and disproportionate increase in learning deficits among students and the dispersion in test scores in the wake of the COVID-19 outbreak have been widely documented in the literature (Betthäuser et al., 2023; De Witte & François, 2023). However, over time, the observed decline in student achievement since 2020 can only partly be attributed to the COVID-19 school closures, as student performance on international education assessments (e.g., PIRLS, PISA, TIMSS) in many Western countries have shown a declining trend in the past decade prior to the pandemic (Gambi & De Witte, 2023; OECD, 2023). Moreover, many education systems have been suffering from teacher shortages (De Witte et al., 2023; De Witte & François, 2023), declining performance of top-performing students and a deterioration in mental health (HBSC Vlaanderen, 2023). However, regardless of the underlying causes of declining test scores, it is essential to monitor the trends in student achievement so that remedial actions can be taken. Extending earlier findings by Gambi & De Witte (2021, 2023), this paper assesses the long-term evolution of the trend in student achievement in the final year of primary education in Flanders, Belgium. Enriching the earlier literature, this paper distinguishes the trend in various content subdomains within the subjects of Dutch (reading and listening), math (geometry and arithmetic), science (technology and nature), social science (time and space), and foreign language French (reading and listening).

The literature on global student achievement trends finds downwards patterns in the data at both primary and secondary education levels. According to PISA, a decline in the average mean performance of 15-year-old students in mathematics across OECD countries has been observed since 2009, and in reading and science since 2012 (OECD, 2020). Among others, PISA mathematics scores have decreased in countries like Belgium, Canada, the Czech Republic, Finland, France, Hungary, Iceland, the Netherlands, New Zealand, the Slovak Republic, and the United States (OECD, 2016, 2023). The 2022 results from PISA paint an even grimmer picture. Compared with the 2018 PISA results, on average, there was a ten-points drop in the average mean performance in reading and an almost 15-point decrease in mathematics, which can be interpreted as students losing three-quarters of a year's worth of learning. The decrease in mathematics performance is three times larger than any previous change recorded in PISA (OECD, 2023). Notably, this negative trend is also reflected in the increase in the share of low-performing students observed between 2009 and 2018 in the OECD (Schleicher, 2019), which has further intensified according to the PISA 2022 results. Currently, on average across OECD countries, one in four 15-year-old is considered a low performer in mathematics, reading, and science (OECD, 2023).

Similarly, a negative trend in average reading achievement is also observed in primary school. According to PIRLS, the reading achievement of grade 4 students has declined in countries such as Canada, Finland, France, Germany, New Zealand, the Netherlands (between 2006 and 2021), Belgium (both Flanders and Wallonia), Czech Republic, Denmark, Finland, Israel, Portugal (between 2011 and 2021) (Mullis et al., 2023). Furthermore, while the achievement distribution shows large within-country variation, this decline often goes hand in hand with an increase in the share of low-performers in all countries that showed a declined reading achievement between 2006 (or 2011) and 2021 in

PIRLS (Mullis et al., 2023). Simultaneously, students' attitudes and behaviours toward reading declined from 2001 to 2016 in the majority of countries participating in PIRLS (Hooper, 2020). This finding is not surprising as, on average, students who responded that they "do not like reading" had lower average reading achievement (491score points) than students who either "very much like reading" (513 score points) or who "somewhat like reading" (501 score points) (Mullis et al., 2023). In Flanders, a similar relationship is observed, though less pronounced, as the average achievement for Flemish students who "do not like reading," "somewhat like reading" or "very much like reading" was 505, 512 and 516 score points, respectively (Mullis et al., 2023).

Our study contributes to the growing literature on global student achievement trends and educational inequality. In particular, while large-scale assessments provide a cross-country comparative picture of trends in learning outcomes in the main subjects (reading, mathematics, and science) (Hooper, 2020; OECD, 2023), evidence on specific content domains within these subjects is lacking. Our study contributes to filling this gap by relying on a unique set of standardised tests in grade 6 of primary education in the Flemish region of Belgium. We provide evidence on the competencies of students in two subdomains for the Dutch language (reading comprehension and listening), two for mathematics (geometry and arithmetic) and two for science (technology and nature). For the Dutch language, we have the test scores of two test versions, each with a different emphasis on reading comprehension (exactly the same test between 2019 and 2023) and listening (exactly the same test between 2021 and 2023), respectively. For math, two subdomains are also tested through two test versions: geometry (2019-2023) and mental arithmetic and data visualisation (2021-2023), respectively. Similarly, for science the test scores for the subdomains technology and nature come from the test versions administered between 2019 and 2023, and between 2021 and 2023, respectively.

The relevance of our results directly pertains to the evaluation of countries' resilience in their education systems. Through continuous close monitoring of the evolution of Flemish students' achievement (and, when available, other school outcomes, e.g., school dropout rates, participation in higher education), we aim to prevent the further accumulation of learning deficits and greater inequality in learning. To do so, we assess the between-cohort average change in students' test scores, comparing most recent cohorts to either pre-pandemic levels (the year 2019), or the year 2021, which is the first year when the test versions for listening, arithmetic and nature where administered.

The results indicate a noticeable improvement relative to 2022 in all subjects and all content subdomains, except for the Dutch language subdomain 'listening'. In particular, the 2023 change in test scores has nearly returned to 2021 levels, using 2019 as the base year. Similarly, compared to the base year 2021, we observe that the 2023 change in test scores is approximately equal to that observed in 2021 for all content domains except for 'listening'. These findings indicate a strong catching-up in learning between-cohorts – particularly evident as the 2023 cohort caught up with the 2021 cohort. When focusing on trends in the distribution of test scores within schools – i.e. distinguishing between low- and high-performing students, we note that high-performing students (p75-p95), who had experienced a particularly pronounced negative trend in test scores in 2022, are now performing at levels observed in previous years. Furthermore, we observe an improvement in test scores across all school, irrespective of their SES composition, in all subjects and content subdomains, except for

listening. Schools with both high and low share of low SES students showed improved change in test scores in 2023 compared to 2022.

The remainder of the paper is structured as follows. Section 2 presents the data and the applied empirical strategy, while Section 3 presents the main results on the trends in standardised test scores by subject domain and the trends in the distribution of test scores within schools stemming from a quantile analysis. Section 4 concludes.

2. Data and Methods

Our analysis of trends in test scores is based on a unique dataset covering a five-year period from 2019 to 2023, bringing together administrative data and standardised tests in grade 6 of primary education in the Flemish region of Belgium. More information on the education system in Flanders is presented in Appendix C. In particular, standardised tests data are provided by the network of Flemish Catholic schools (Katholiek Onderwijs Vlaanderen, KOV), the largest education provider in Flanders.[4] The standardised test scores are collected at the student level from a curriculum-based test administered every year in June in the final year of primary school (grade 6), and are available for multiple subjects, i.e. native language (Dutch), mathematics, sciences, social sciences, and foreign language French. Furthermore, we can distinguish between the competencies of students in two subdomains for Dutch (reading and listening, respectively), two for mathematics (geometry and arithmetic, respectively), two for science (technology and nature, respectively), two for social science (time and space, respectively), two for French[5] (reading and listening, respectively). Additional information about the differences between test versions and the tested competencies is presented in Appendix C. Different cohorts of students take the test in grade 6 each year, which allows us to shed light on changes in educational achievement between-cohorts. Since 2019, two different test versions have been administered either as the primary official test of the specific year, or as an additional test taken along with the official test version: a test version administered between 2019 and 2023 (with a focus on the content subdomains: reading comprehension, geometry, and technology), and a test version introduced in 2021 and then administered every year till 2023 (with a focus on the content subdomains: listening, mental arithmetic and data visualisation, and nature). To enable comparisons over time and with other studies, test scores were standardised to have a mean of 0 and a standard deviation (SD) of 1 per test version. Given the different test versions, the main results report separately the 2019-2023 sample (Table 2) from the 2021-2023 sample (Table 3).

[4] KOV provides education to 65% of all Flemish students in school year 2023-2024 (Ministry of Education and Training, 2023).

[5] In contrast to the other four subjects and their respective content subdomains, the two competencies for French are tested in the same test version 2021, with 10 questions each for listening and reading. This version was introduced in 2021 and maintained as the official test in both 2022 and 2023. Therefore, the main results report 'French (listening)' and 'French (reading)' exclusively for the 2021-2023 sample.

The test scores data are complemented with administrative data, encompassing various school characteristics such as the share of students with special needs and school size (proxied by the total number of enrolled students). Students' characteristics in grade 6 (the share of girls, three[6] indicators of socio-economic background (SES) of the student, and the share of slow learners), and teachers' characteristics (the share of teachers with less than five years of experience, with more than 20 years of experience, the share of school staff absences and the share of teacher shortages), both aggregated at the school level, are also controlled for in the regression model. Finally, we include a control for school district (or school board) size, proxied by the number of schools in a school district.

To investigate the evolution of test scores, we estimate a fixed effects panel regression which takes the following specification:

$$y_{i,t} = \alpha + \beta(2020)_{i,t} + \gamma(2021)_{i,t} + \delta(2022)_{i,t} + \theta(2023)_{i,t} + X_{i,t} + v_i + \epsilon_{i,t} \qquad (1)$$

where $y_{i,t}$ denotes the analysed outcome variable (i.e. test score in each content domain) of school i at time t. The outcome is regressed on a constant α, and four-year dummies (2020, 2021, 2022, and 2023). Each dummy takes on the value of 1 in the year of reference (2020, 2021, 2022 and 2023, respectively), and 0 otherwise. Thus, β, γ, δ and θ represent the main coefficients of interest, as each identifies the yearly change in test scores relative to 2019 (pre-pandemic levels), outlining the evolution in test scores. Similarly, we estimate the same fixed effects panel regression when analysing the shorter sample 2021-2023. In this instance, however, the model incorporates solely two dummy variables – 2022 and 2023. In turn, these two dummies identify the yearly change in test scores relative to 2021 (the first year when the test versions for listening, arithmetic and technology were administered). Given that the panel spans only four (or two) years in addition to the reference year (2019 or 2021, depending on the sample used), the inclusion of four (or two, in the case of the 2021-2023 sample) dummy variables in the model capture also year fixed effects while estimating the changes relative to the baseline levels of 2019 (or 2021).

The school fixed effects (v_i) allow us to control for a series of unobserved characteristics that vary across schools but are invariant over time, such as school location, school reputation, school board and principal's management ability. $X_{i,t}$ denotes a vector of time-varying control variables reflecting school district, school, grade 6 and teacher characteristics. Finally, the standard errors $\epsilon_{i,t}$ are clustered at the school level to account for serial autocorrelation in the error terms. Differing from Gambi & De Witte (2023), this paper reports results at the student level. This approach allows us to leverage the greater variation in the student dataset, providing evidence on how the test scores of the average student in the sample have changed.

2.1. Sample and attrition

Because schools have the option to voluntarily participate in the standardised tests, the number of schools taking different test versions varies yearly, leading to an unbalanced panel dataset. Table 1

[6] The fourth socio-economic background (SES) variable most commonly used in Flemish datasets – the share of students receiving financial support, is significantly highly correlated with the other three SES variables (up to r=0.80). To prevent multicollinearity issues, this variable is therefore excluded from the model.

summarises the number of participating schools and students in the different test versions per year, with an average of about 24,000 students taking at least one test version annually. In the years when the test version is also advertised by the school network as the 'official' test, participation is on average higher (i.e. in 2019, 1164 (90%) schools administered the 2019 test version to their students, whereas in 2021, 1158 (88%) schools administered the 2021 test version). In the following years, participation in the older test versions tend to drop as it is not always viable for schools to administer multiple tests along with the new, official test version due to time or resources constraints. Nevertheless, on average, 24,000 and 27,000 students take the 2019 and 2021 test versions every year, respectively. It is worth noting that, when a school opts to administer one or multiple test versions, all students in grade 6 are tested. Hence, within a school there is no attrition. Figures B1 and B2 in Appendix B show the distribution of raw test scores (not standardised) for each content domain across the years 2019-2023 (test version 2019) and 2021-2023 (test version 2021).

Concerns could arise about the external validity of our results due to the voluntary participation of schools in the test versions. Appendix A compares the school characteristics of those schools that participated in the test version of interest (2019 or 2021) for at least one content domain, and those schools of the same school network that did not participate in any of this specific test version (but could, potentially, have also participated in the other test version) in 2023. The descriptive statistics in Appendix A reveal that the schools administering test version 2019 have a smaller proportion of students with special needs, a higher share of teachers with 20 or more years of experience, and a more advantaged student population in terms of neighbourhood of residence. Schools administering test version 2021 have a larger student enrolment and a lower share of students with special needs. Both schools administering test versions 2019 and 2021 are part of larger school districts. The remaining school characteristics are well-balanced, indicating that participating schools are largely representative of the schools within the school network. Another factor threatening the external validity of our results is that the network of Catholic schools generally attracts a more advantaged student population[7] (Cherchye et al., 2010). Consequently, our results should be interpreted as a "best-case scenario" estimates of between-cohorts changes in test scores for the Flemish education system.

In terms of internal validity, concerns may arise due to the low participation rates in some test version in specific years. School-level participation rates drop from 80-90% in the years when a test version is designated by the school network as the 'official' test to 30-40% in years when the test version is administrated alongside another 'official' test version (e.g. test versions 2019 and 2021 in 2022, when a new, 'official' 2022 test version was also administered). A similar concern can be raised for the year 2020, when the participation rate of schools was 31% because of the COVID-19 pandemic. When comparing schools participating in test versions 2019 and 2021 in 2023 to those administering the

[7] Tables B10 and B11 in Appendix B supports this finding by comparing the four most common types of SES indicators (i.e., the share of students with a low-educated mother, the share of students receiving financial support, the share of students that speak a different language at home, and the share of students living in a neighborhood with high grade retention rates at the age of fifteen) as measured in Flemish administrative datasets (AGODI, 2023) for in and out-of-sample schools. Tables B10 and B11 show that the demographic composition of participating schools (regardless of the test version) attract a more advantaged student population than that of the average Flemish primary school.

same tests in 2022, no strong underlying pattern in the attrition can be detected (see Tables A3 and A4). The only significant difference observed is that schools participating in 2023 experience a higher rate of staff absences but, on average, face fewer teacher shortages as compared to those schools administering the same tests in 2022. All other characteristics are balanced.

In any case, to account for any difference between schools in the analyses, we control for the full set of school characteristics available in the dataset. Furthermore, to account for changes in the participation of schools in the sample throughout the years (i.e. those causing concerns regarding internal validity), we present the findings from two robustness tests (see Tables B13 and B14 in Appendix B). In particular, in a first robustness test, we re-estimate our model restricting the sample to those schools that participated in all years. While the sample size shrinks considerably (due to the low participation rate in 2022 for test versions 2019 and 2021, and in 2020 for test version 2019), restricting the sample in such a way allows us to define a constant sample over time. Consequently, this estimation holds all school characteristics constant over time. As shown in Table B13, the estimation results are similar to the main result for all subjects, both in terms of effect sizes and significance. This yields confidence in our earlier estimates. In a second robustness test, we use Coarsened Exact Matching (CEM; Blackwell et al., 2009; Iacus et al., 2012) to control for sample imbalances in covariates. The schools participating each year are sequentially matched (based on the school district, school, grade 6 and teacher characteristics) to the schools that took part in the test in 2023. As shown in Table B14, the results prove to be robust in this alternative econometric method.

Table 1: Participating schools and students in the different test versions per year

	Total	*Participating 2019 test*		*Participating 2021 test*		*Participating 'official' test*
	N	$N_{Schools}$ *(%)*	$N_{Students}$	$N_{Schools}$ *(%)*	$N_{Students}$	$N_{Schools}$ *(%)*
2019	1289	1164 (90%)	36990			1164 (90%)
2020	1300	401 (31%)	11302			401 (31%)
2021	1313	729 (56%)	20472	1158 (88%)	37304	1158 (88%)
2022	1317	574 (44%)	15214	1056 (80%)	14153	1159 (88%)
2023	1325	1081 (82%)	33671	1033 (78%)	32032	1034 (78%)

Note: 'Total' refers to the overall number of schools in the school network that offers grade 6. 'Participating' refers to the sub-sample of schools that participated in a test version in each respective year. Schools administrating the French test version 2021 are included in both the 'participating 2021 test' and 'participating 'official' test' columns from 2021 to 2023, as the French test version 2021 was maintained as the official test during this period.

3. Results

3.1. Trends in standardised test scores by subject subdomain

In Figures B3-B4 in Appendix B, we show the visual exploration of the trend based on the student-level data analysis presented in this section. To control for observed and unobserved heterogeneity among schools, we estimate regression model (1). Table 2 presents the student-level results for all test subjects from the analysis on the 2019-2023 sample (i.e. focus on content subdomains 'reading comprehension', 'geometry' and 'technology'), whereas Table 3 presents the student-level results for all test subjects from the analysis on the 2021-2023 sample (i.e. focus on content subdomains 'listening', 'mental arithmetic and data visualisation', and 'nature'). In both tables the outcome variable is the standardised test score for each student. The results in Tables 2 and 3 are from the most saturated regression models that include robust standard errors, school fixed effects, and controls for school district, school, grade 6 and teacher characteristics. The results are robust to gradually adding these specifications (see Appendix B, Tables B1-B4 for the 2019-2023 sample, and Tables B5-B8 for the 2021-2023 sample).

In Table 2, we analyse the trends in standardised test scores between 2019 and 2023, interpreting each estimate as the change in test scores relative to the baseline year of 2019. Three distinct patterns stand out. First, consistent with findings from the literature on global trends in students' achievement, all students are found to underperform relative to the 2019 pre-pandemic levels (Table 2), as the 2023 change in test scores is still significantly negative in all content subdomains (-0.27 SD in 'reading comprehension', -0.18 SD in 'geometry', -0.12 SD in 'technology', and -0.17 SD in 'social science – time'). As about 0.4 SD corresponds to what a 12-yearsold student learns at the end of primary education, the estimated coefficients can be interpreted as a significant learning deficit of about a quarter to a half school year compared to the base year.

Second, confirming previous findings by Gambi & De Witte (2023)[8], we observe that the attainment deficits accumulate over the years 2020-2022 for all content domains except for social science (content subdomain: time). However, this is not the case anymore in 2023. For example, the negative change in test scores relative to 2019 for the Dutch language kept increasing year-on-year until 2022 (from -0.15 SD to -0.17 SD and -0.43 in 2020, 2021 and 2022, respectively) but improved again to –0.27 SD in 2023.

Third, the magnitude of the change in test scores observed for each content domain in 2023 gets closer to changes in test scores observed in 2021, showing an improvement against 2022 results and signs of a slowdown of the negative trend.

[8] Differing from Gambi & De Witte (2023), this paper presents results at the student level. Nonetheless, the results in both papers are consistent – showing an accumulation in attainment deficits in all subjects except for social science between the years 2020 and 2022. Given that test score variations are more pronounced within schools than across schools, the estimated effects are smaller in the student-level analysis as compared to the school-level analysis from Gambi & De Witte (2023).

In Table 3, we analyse the trends in standardised test scores between 2021 and 2023, interpreting each estimate as the change in test scores relative to the baseline year 2021. Consistent with the deceleration in the negative trend shown in Table 2, Table 3 suggests that, except for the 'listening' content domain, the changes in test scores for the 'mental arithmetic' and 'nature' content domains are not significantly different from the levels observed in 2021. For Social Science (specifically, the 'space' subdomain), a significant negative change is observed, though less pronounced than the negative change observed in 2022. The only subdomain that exhibits an accumulation of attainment deficits is 'listening', with the negative change in test scores in 2023 (-0.18 SD) being greater than that in 2022 (-0.11 SD). Finally, the changes in test scores for both content subdomains of French (listening and reading), are not significantly different from the levels observed in 2021.

Overall, we observe similar patterns in both Tables 2 and 3. Specifically, in all subjects and all content subdomains (except for 'listening'), we observe a noticeable improvement relative to 2022 test scores.[9] Relative to the base year 2019, we observe that the 2023 change in test scores has almost returned to the change observed in 2021. Similarly, relative to the base year 2021, we observe that the 2023 change in test scores is approximately equal to 2021 levels. A strong catch-up in learning between-cohorts – evident as the 2023 cohort caught up with the 2021 cohort, has occurred.

[9] As shown in Table B12 in Appendix B, the difference in estimated coefficients between the 'Change in 2023' and 'Change in 2022' dummies is positive and statistically significant for all subjects and content domains, except for Listening.

Table 2: Main results for all subjects, sample 2019-2023 (test version 2019)

	Dutch (Reading Comprehension)	Math (Geometry)	Science (Technology)	Social Science (Time)
Year-to-2019 change				
Change in 2020	-0.15***	-0.07***	-0.08**	-0.08***
	(0.02)	(0.03)	(0.04)	(0.02)
Change in 2021	-0.17***	-0.11***	-0.12***	-0.02
	(0.02)	(0.02)	(0.03)	(0.03)
Change in 2022	-0.43***	-0.24***	-0.21***	-0.20***
	(0.03)	(0.02)	(0.03)	(0.03)
Change in 2023	-0.27***	-0.18***	-0.12***	-0.17***
	(0.02)	(0.02)	(0.03)	(0.03)
N	92420	84174	56720	60932

Note: Robust standard errors between parentheses. Sample 2019-2023, student-level data. 'Change in 2020', 'Change in 2021', 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2020, 2021, 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2020, 2021, 2022 and 2023, relative to pre-pandemic levels. The standardised test scores (mean 0, SD 1) come from the exact same test version 2019. All regressions include school fixed effects and time varying controls at school level, reflecting several sample characteristics. School characteristics are the number ofx students in the school, the share of special needs students, the share of girls, three SES indicators (neighbourhood, mother's education, and home language), and school district size. Grade 6 characteristics are: the share of girls, four SES indicators, and the share of slow learners. Characteristics of the teachers are: the share of teachers with less than 5 years of experience, with more than 20 years of experience, the share of school staff absences and the share of teacher shortages. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Main results for all subjects, sample 2021-2023 (test version 2021)

	Dutch (Listening)	Math (Mental Arithmetic)	Science (Nature)	Social Science (Space)	French (Listening)	French (Reading)
Year-to-2021 change						
Change in 2022	-0.11***	-0.10***	-0.10***	-0.23***	0.02	0.01
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Change in 2023	-0.18***	0.01	-0.04	-0.06**	-0.01	0.02
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
N	52357	67991	41638	42850	76272	76095

Note: Robust standard errors between parentheses. Sample 2021-2023, student-level data. 'Change in 2020', 'Change in 2021', 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2020, 2021, 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2020, 2021, 2022 and 2023, relative to pre-pandemic levels. The standardised test scores (mean 0, SD 1) come from the exact same test version 2021. All regressions include school fixed effects and time varying controls at school level, reflecting several sample characteristics. School characteristics are: the number of students in the school, the share of special needs students, the share of girls, three SES indicators (neighbourhood, mother's education, and home language), and school district size. Grade 6 characteristics are: the share of girls, four SES indicators, and the share of slow learners. Characteristics of the teachers are: the share of teachers with less than 5 years of experience, with more than 20 years of experience, the share of school staff absences and the share of teacher shortages. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.2. Trends in the distribution of test scores within schools

The previous section focussed on the average student's test scores in the sample. In this section, using a quantile analysis, we explore the between-cohorts change in test scores per quintile of the distribution of students' scores within each school (i.e., per 5%). This approach allows us to investigate whether the evolution in test scores of low- and high-performing students behaves differently, and changes over time. The results are displayed in Figures 3 and 4, providing estimates for the 2019-2023 and 2021-2023 samples, respectively. In both Figures, each graph illustrates the estimates per quintile within school for the change in test scores in each year. The upper graph presents the results for the Dutch language (reading comprehension and listening in Figure 3 and 4, respectively) and math (geometry and mental arithmetic in Figure 3 and 4, respectively). In contrast, the lower graph exhibits the results for science (technology and nature in Figures 3 and 4, respectively) and social science (time and space in Figures 3 and 4, respectively). For example, the 95th percentile represents the highest-performing students within their school, signifying that only 5% of the students in the school achieved better test scores than them. Conversely, the 5th percentile corresponds to the lowest-performing students, denoting that only 5% of the students in the school attained lower test scores.

For the Dutch language, all students across the distribution underperform relative to both pre-pandemic levels (Figure 3) and 2021 (Figure 4). Nevertheless, the results for the change in 2023 in the content domain 'reading comprehension' show a significant improvement compared to 2022 estimates across the distribution, as the results bounce back to the estimates obtained in 2020 and 2021. In the content domain 'listening', there is no evidence of further negative or positive change in 2023 relative to 2021, compared to the changes observed in 2022.

In the case of math, while all students across the distribution are found to underperform relative to pre-pandemic levels (Figure 3), the results for the change in 2023 in the content domain 'geometry' revert to the estimates obtained in 2020 and 2021. Furthermore, the change in 2023 in the content domain 'mental arithmetic and data visualisation' shows a significant improvement relative to 2022 estimates across the distribution and exhibits no significant difference with 2021 levels.

In both science subdomains ('technology' and 'nature'), the estimates for the 2023 change are above those for the 2022 change, indicating a catch-up in learning between-cohorts of students across the whole distribution. Finally, a similar pattern emerges for both social science subdomains ('time' and 'space').

Consistent with the findings from section 3.1, we note an improvement relative to 2022 in all subjects and all content subdomains, except for the content subdomain 'listening'. In particular, high-performing students (p75-p95), who had experienced a particularly pronounced negative trend in test scores in 2022, are now performing at levels observed in previous years.

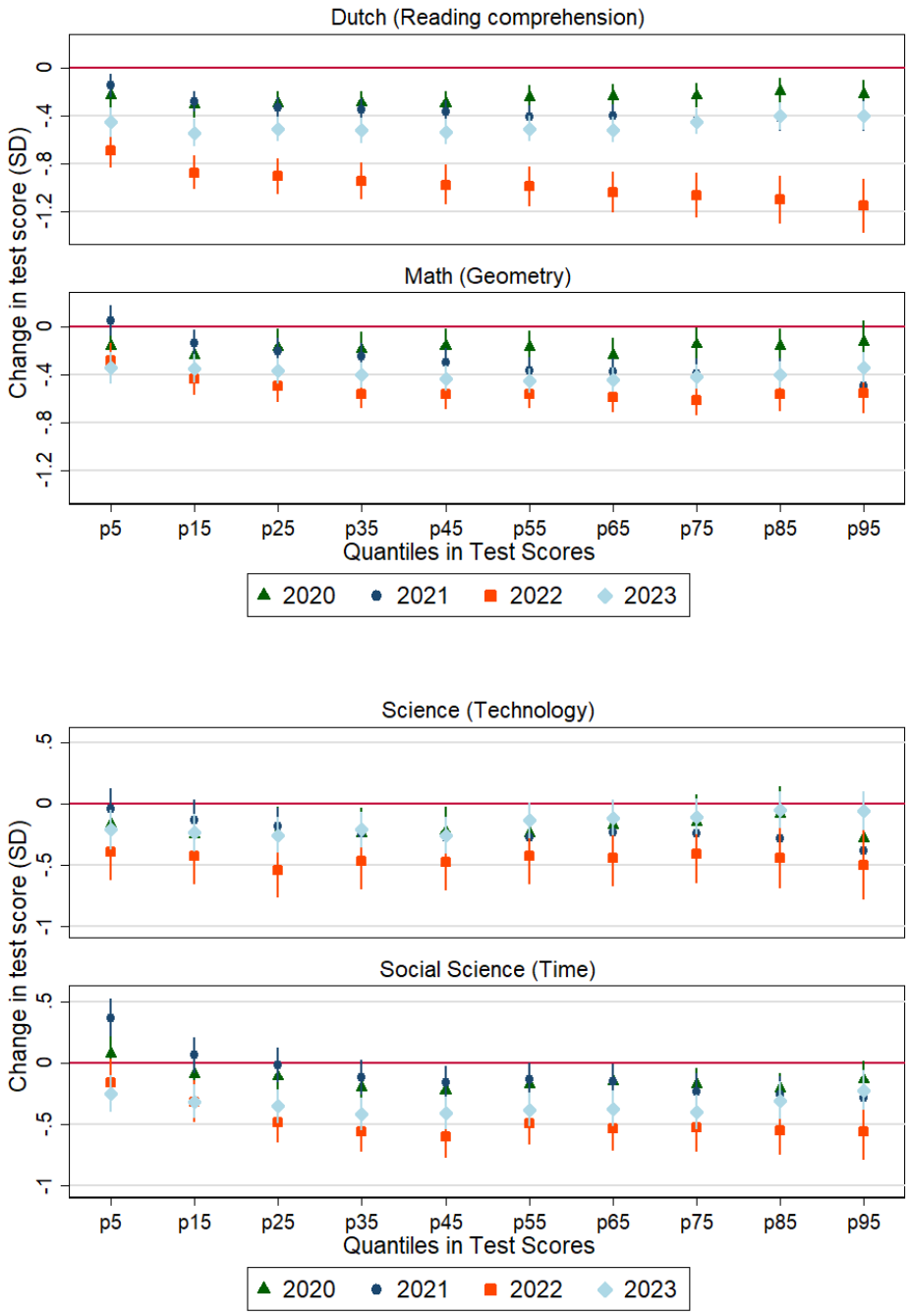

Figure 3: Estimated change per quintile of test scores within schools (test version 2019)

Note: These figures are based on a school fixed effects regression using the 2019–2023 sample (test scores from test version 2019 in every year) which includes year dummy variables and time varying controls at school level, for grade 6 characteristics and characteristics of the teacher.

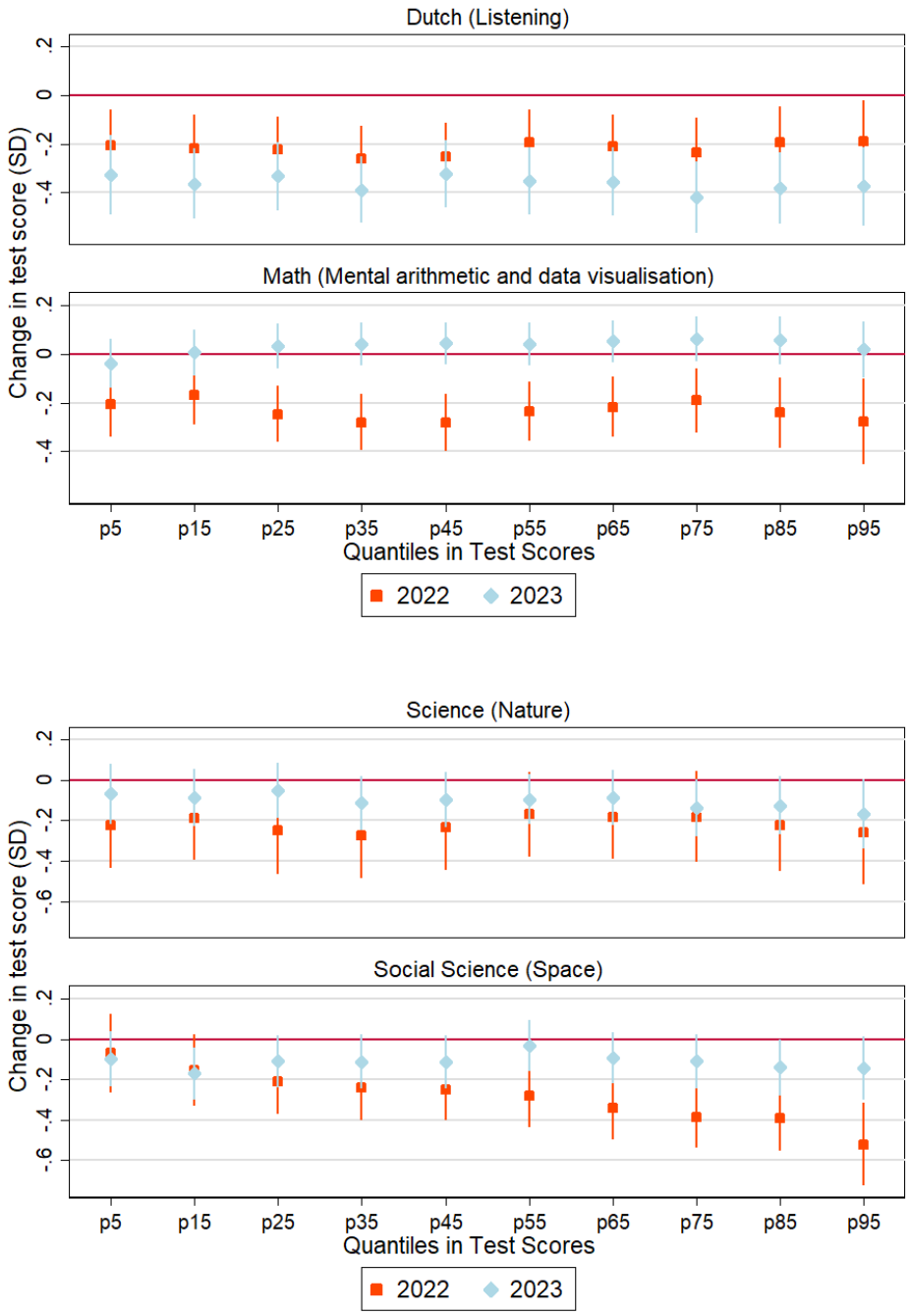

Figure 4: Estimated change per quintile of test scores within schools (test version 2021)

Note: These figures are based on a school fixed effects regression using the 2021–2023 sample (test scores from test version 2021 in every year) which includes year dummy variables and time varying controls at school level, for grade 6 characteristics and characteristics of the teacher.

3.3. Marginal change 2022 and 2023 by SES background

The socioeconomic background has been identified as a significant factor related to student achievement, playing an even stronger role in the aftermath of the COVID-19 pandemic (Betthäuser et al., 2023). To gain a deeper understanding of the interaction between SES characteristics and trends in test scores, we decompose the changes in student performance in 2022 and 2023 by SES characteristics to investigate which schools saw larger improvements between 2022 and 2023. We estimate the marginal effects based on SES indicators in Grade 6 of the school. Figures 5 and 6 show the marginal change in 2022 and 2023 based on the three SES indicators in Grade 6 of the school for Dutch and math for the sample 2019-2023 and 2021-2023. The figures are based on a linear regression (solid lines) and the dotted lines indicate the confidence intervals. On all three indicators, a higher level of the SES indicator is associated with a larger confidence interval, indicating a wider spread of scores and a lower number of observations.

Two distinct patterns stand out. First, we observe that the SES-profile of the school is less determining the decline in test scores in 2023 compared to 2022. On some dimensions of SES-composition, such as mother education for reading comprehension, the SES-composition even does not matter anymore for the change in test scores (controlled for the other variables). On other dimensions, e.g. native language, the SES-composition still influences the difference between the 2019 and the 2023 test score, although to a lesser extent than in 2022.

A similar relationship is observed between the change in 2023 for geometry, and the share of students with a different home language and living in a neighbourhood with high grade retention rates at the age of fifteen. Second, for listening, the average changes in 2022 and 2023 show similar patterns. In particular, the change in test scores becomes more negative with increasing shares of students from a neighbourhood with high grade retention rates and more students who do not speak the native language Dutch at home.

Overall, these results point towards an improvement in test scores all across the board in all subjects and content subdomains, except for listening.

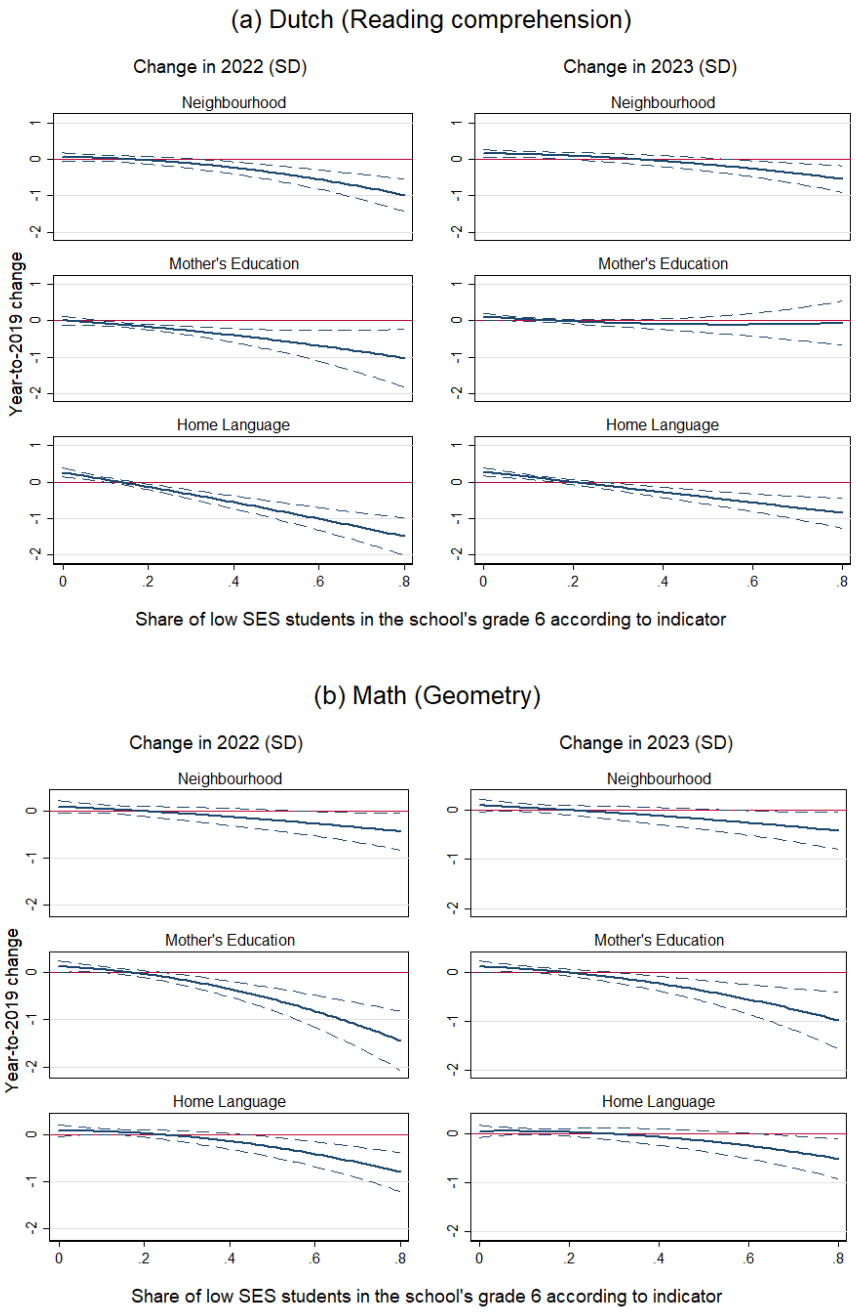

Figure 5: Marginal change by socioeconomic background: (a) Dutch; (b) math (test version 2019)
Note: These figures are based on a school fixed effects regression using the 2019–2023 sample (test scores from test version 2019 in every year) which includes year dummy variables and time varying controls at school level, for grade 6 characteristics and characteristics of the teacher.

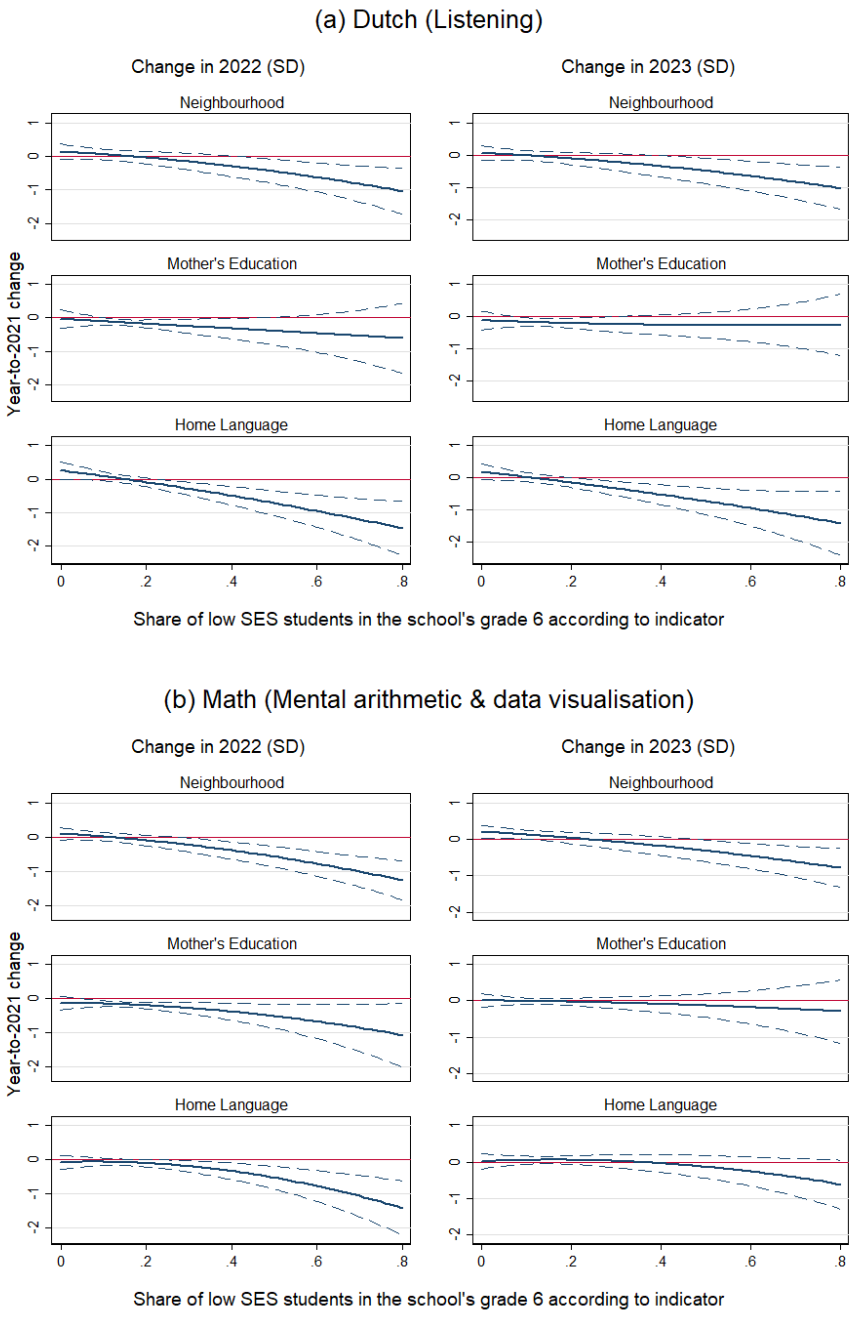

Figure 6: Marginal change by socioeconomic background: (a) Dutch; (b) math (test version 2021)
Note: These figures are based on a school fixed effects regression using the 2021–2023 sample (test scores from test version 2021 in every year) which includes year dummy variables and time varying controls at school level, for grade 6 characteristics and characteristics of the teacher.

4. Conclusion and discussion

This paper provides evidence of the long-term evolution of the trend in student achievement at the end of primary education in Flanders, Belgium. We utilise a comprehensive dataset comprising administrative data and standardised test scores from schools pertaining to the same private Catholic school network, which serves as the largest primary education provider in the Flemish region of Belgium. Employing two versions of a curriculum-based test (one administered since 2019 and one since 2021, respectively), we assess the trends in various content subdomains within the subjects of Dutch, math, science, social science, and foreign language French. Specifically, this study separately explores the trends in students' competences in reading comprehension and listening (two subdomains of the subject Dutch language), in geometry and arithmetic (two subdomains of the subject math), in technology and nature (two subdomains of the subject science), time and space (two subdomains of the subject social science), and in listening and reading (two subdomains of the subject French).

Overall, the results indicate a strong catch-up in learning between-cohorts – particularly evident as the change in test scores of the 2023 student cohort has nearly caught up with the 2021 cohort. More in detail, there is a noticeable improvement relative to 2022 in all subjects and all content subdomains, except for the Dutch language subdomain 'listening'. In particular, the 2023 change in test scores has nearly returned to 2021 levels, using 2019 as the base year. Similarly, compared to the base year 2021, we observe that the 2023 change in test scores is approximately equal to that observed in 2021 for all content domains except for 'listening'. When focusing on trends in the distribution of test scores within schools – i.e. distinguishing between low- and high-performing students, we note that high-performing students (p75-p95), who had experienced a particularly pronounced negative trend in test scores in 2022, are now performing at levels observed in previous years. Furthermore, we observe an improvement in test scores across all school, irrespective of their SES composition, in all subjects and content subdomains, except for listening. Hence, we observe that the SES-profile of the school is less determining the decline in test scores in 2023 compared to 2022.

The notable improvement in test scores observed in 2023 relative to 2022 can be attributed to a combination of long-term, ongoing, and short-term interventions implemented within the education system. First, the is a growing consensus that the education system should be strengthened. This is reflected in upcoming long-term strategies (e.g. on teacher professionalism; on HR-policies; on curriculum goals) that are currently being implemented but received already significant attention in the media. Second, initiatives like the Reading Offensive (following the Irish example) and the emphasis on higher aspirations represent ongoing efforts to enhance overall learning environments and student engagement. Third, short-term interventions, including efforts to attract lateral entrants to the teaching profession and additional resources allocated to ICT in education, have alleviated immediate challenges. The synergistic effect of these varied interventions reflects a comprehensive and multifaceted approach, contributing to the observed positive shift in student achievement in 2023.

The findings of this paper offer valuable insights for education systems beyond the specific setting of Flanders. The use of a comprehensive dataset and the detailed analysis of multiple content subdomains

in Dutch, math, science, social science, and foreign language French provide a nuanced understanding of the dynamics at play. The observed strong catch-up in learning between-cohorts, particularly evident in the 2023 student cohort, highlights the potential for effective educational strategies and interventions to address challenges and facilitate improvement over a relatively short period. This is of significance to other education systems globally, as it suggests that targeted efforts can yield positive outcomes in student achievement. Additionally, the study's focus on the distribution of test scores within schools, especially the recovery of high-performing students from a negative trend in 2022, offers insights into strategies that can benefit different student groups. Moreover, given the observed negative trend in student achievement in many countries around the globe as shown by international assessments, our findings show the importance of rich data collections. In particular, while international assessments are crucial in cross-country comparisons, they differ from traditional assessments by not directly measuring the content taught in schools through the curriculum. Curriculum-based tests, on the other hand, provide a more accurate assessment of what students learn in school. Hence, we advocate for continuous monitoring and the implementation of national, country-specific curriculum-based tests. These tests can enable national policymakers to assess the adequacy of official curricula, standards, and procedures for their students.

References

AGODI. (2023). *Leerlingenkenmerken per school (2022–2023)*. Available at https://onderwijs.vlaanderen.be/nl/onderwijsstatistieken/themas-onderwijsstatistieken/leerlingenkenmerken

Betthäuser, B. A., Bach-Mortensen, A. M., & Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour*, *7*(3), 3. https://doi.org/10.1038/s41562-022-01506-4

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). cem: Coarsened exact matching in Stata. *Stata Journal*, *9*(4), 524–546.

De Witte, K., De Cort, W., & Gambi, L. (2023). *Evidence-based Solutions to Teacher Shortages* [EENEE Analytical report]. European Commission. https://doi.org/10.2766/475647

De Witte, K., & François, M. (2023). COVID-19 Learning Deficits in Europe: Analysis and Practical Recommendations. Analytical Report. In *European Commission* [EENEE Analytical report]. European Commission. https://doi.org/10.2766/881143

Dockx, J., Van Landeghem, G., Aesaert, K., Van Damme, J., & De Fraine, B. (2019). *Begrijpend lezen van het vierde naar het zesde leerjaar. Herhalingsmeting van PIRLS in 2018 vergeleken met PIRLS 2016. [Reading comprehension from fourth to sixth grade. PIRLS 2018 compared to PIRLS 2016]*. Centrum voor Onderwijseffectiviteit en -evaluatie. https://lirias.kuleuven.be/2907263

Gambi, L., & De Witte, K. (2021). *The resiliency of school outcomes after the COVID-19 pandemic. Standardised test scores and inequality one year after long term school closures*. KU Leuven – Faculty of Economics and Business. https://lirias.kuleuven.be/retrieve/636001

Gambi, L., & De Witte, K. (2023). *The uphill battle: The amplifying effects of negative trends in test scores, COVID-19 school closures and teacher shortages* (Discussion Paper Series). KU Leuven – Faculty of Economics and Business.

HBSC Vlaanderen. (2023). *Health Behaviour in School-aged children – Factsheet Mentaal, Sociaal en Fysiek Welzijn.* (p. 47). UGent. https://www.jongeren-en-gezondheid.ugent.be/materialen/factsheets-vlaanderen/gezondheid-en-welzijn/

Hooper, M. (2020). Troubling Trends: An International Decline in Attitudes toward Reading. IEA Compass: Briefs in Education. Number 8. In *International Association for the Evaluation of Educational Achievement*. International Association for the Evaluation of Educational Achievement. https://eric.ed.gov/?id=ED608237

Iacus, S. M., King, G., & Porro, G. (2012). Causal Inference Without Balance Checking: Coarsened Exact Matching. *Political Analysis*, *20*(1), 1--24.

Katholiek Onderwijs Vlaanderen. (2023a). *Gestandaardiseerde proeven einde bao: Nederlands*. Available at https://pro.katholiekonderwijs.vlaanderen/evaluatiebox-basisonderwijs/evaluatie-instrumenten/gestandaardiseerde-proef-eind-bao-nederlands

Katholiek Onderwijs Vlaanderen. (2023b). *Gevalideerde toets einde bao: Wetenschappen en techniek*. Available at https://pro.katholiekonderwijs.vlaanderen/evaluatiebox-basisonderwijs/evaluatie-instrumenten/gevalideerde-toets-einde-bao-wetenschappen-en-techniek

Katholiek Onderwijs Vlaanderen. (2023c). *Gevalideerde toetsen einde bao: Frans*. Available at https://pro.katholiekonderwijs.vlaanderen/evaluatiebox-basisonderwijs/evaluatie-instrumenten/gestandaardiseerde-proef-einde-bao-frans

Katholiek Onderwijs Vlaanderen. (2023d). *Gevalideerde toetsen einde bao: Mens en maatschappij*. Available at https://pro.katholiekonderwijs.vlaanderen/evaluatiebox-basisonderwijs/evaluatie-instrumenten/gevalideerde-toets-einde-bao-mens-en-maatschappij

Katholiek Onderwijs Vlaanderen. (2023e). *Gevalideerde toetsen einde bao: Wiskunde*. Available at https://pro.katholiekonderwijs.vlaanderen/evaluatiebox-basisonderwijs/evaluatie-instrumenten/gevalideerde-toets-einde-bao-wiskunde

Ministry of Education and Training. (2023). *Dataloep: Aan de slag met cijfers over onderwijs*. https://onderwijs.vlaanderen.be/nl/onderwijsstatistieken/dataloep-aan-de-slag-met-cijfers-over-onderwijs/download-je-dataset-uit-dataloep#bao-so-inschrijvingen

Mullis, l. V., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2019). *TIMSS 2019 International Results in Mathematics and Science*. TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/index.html

Mullis, l. V., von Davier, M., Foy, P., Fishbein, B., Reynolds, K. A., & Wry, E. (2023). *PIRLS 2021 International Results in Reading*. Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.tr2103.kb5342

OECD. (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. https://www.oecd.org/education/pisa-2015-results-volume-i-9789264266490-en.htm

OECD. (2020). *PISA 2018 Results*. OECD.

OECD. (2022). *Education at a Glance 2022: OECD Indicators*. OECD. https://doi.org/10.1787/3197152b-en

OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/education/pisa-2022-results-volume-i_53f23881-en

Schleicher, A. (2019). PISA 2018: Insights and Interpretations. In *OECD Publishing*. OECD Publishing.

Appendix A: Attrition

Table A1: Attrition of participating schools in the year 2023: test version 2019

	Attrited		**Participated**		*t*-Test
	N	Mean [SD]	*N*	Mean [SD]	*p*-Value
Number of Students	326	187.28	1077	191.83	0.42
Share of special needs students	326	0.09	1077	0.07	0.00
Teachers: Share with more than 20y experience	326	0.34	1077	0.36	0.08
Share of staff absences	326	0.17	1077	0.17	0.56
Teacher shortage: % of unfilled vacancies	326	2.01	1077	2.11	0.54
Year 6: Share of Girls	326	0.50	1077	0.50	0.56
Year 6: SES - Neighbourhood	325	0.23	1074	0.20	0.07
Year 6: SES - Mother's Education	325	0.17	1074	0.16	0.58
Year 6: SES - Home Language	325	0.20	1074	0.19	0.59
Year 6: Grade Repetition	326	0.00	1077	0.00	0.31
Number of Schools per District	324	16.50	1077	22.37	0.00
Population density (Inhab/Km2)	326	1337.64	1077	1154.79	0.17

Note: Standard deviations (SD) in parenthesis. 'Attrited' refers to schools in the school network that offer grade 6 and did not administer the 2019 test version in 2023. 'Participated' refers to the schools that offer grade 6 and administered the 2019 test version for at least one subject in 2023. The p-value is derived from the t-test comparing the attrited with the participating schools. Special needs schools (4 schools) were excluded from the attrition analysis.

Table A2: Attrition of participating schools in the year 2023: test version 2021

	Attrited		**Participated**		*t*-Test
	N	Mean [SD]	*N*	Mean [SD]	*p*-Value
Number of Students	446	184.99	957	193.47	0.09
Share of special needs students	446	0.08	957	0.07	0.00
Teachers: Share with more than 20y experience	446	0.35	957	0.36	0.21
Share of staff absences	446	0.17	957	0.16	0.19
Teacher shortage: % of unfilled vacancies	446	2.03	957	2.11	0.58
Year 6: Share of Girls	446	0.50	957	0.50	0.99
Year 6: SES - Neighbourhood	445	0.21	954	0.20	0.74
Year 6: SES - Mother's Education	445	0.16	954	0.17	0.55
Year 6: SES - Home Language	445	0.18	954	0.19	0.48
Year 6: Grade Repetition	446	0.00	957	0.00	0.67
Number of Schools per District	444	17.85	957	22.48	0.00
Population density (Inhab/Km2)	446	1147.44	957	1220.50	0.55

Note: Standard deviations (SD) in parenthesis. 'Attrited' refers to schools in the school network that offer grade 6 and did not administer the 2021 test version in 2023. 'Participated' refers to the schools that offer grade 6 and administered the 2021 test version for at least one subject in 2023. The p-value is derived from the t-test comparing the attrited with the participating schools. Special needs schools (4 schools) were excluded from the attrition analysis.

Table A3: Comparison of Participating Schools in 2022 and 2023: test version 2019

	Participated in 2022		**Participated in 2023**		*t*-Test
	N	Mean [SD]	*N*	Mean [SD]	*p*-Value
Number of Students	573	194.22	1077	191.83	0.59
Share of special needs students	573	0.07	1077	0.07	0.68
Teachers: Share with more than 20y experience	573	0.35	1077	0.36	0.27
Share of staff absences	573	0.26	1077	0.17	0.00
Teacher shortage: % of unfilled vacancies	573	1.87	1077	2.11	0.04
Year 6: Share of Girls	573	0.51	1077	0.50	0.10
Year 6: SES - Neighbourhood	573	0.22	1074	0.20	0.18
Year 6: SES - Mother's Education	573	0.17	1074	0.16	0.21
Year 6: SES - Home Language	573	0.20	1074	0.19	0.26
Year 6: Grade Repetition	573	0.00	1077	0.00	0.55
Number of Schools per District	573	20.83	1077	22.37	0.13
Population density (Inhab/Km2)	573	1279.48	1077	1154.79	0.29

Note: Standard deviations (SD) in parenthesis. 'Participated in 2022' refers to schools in the school network that offer grade 6 and administered the 2019 test version in 2022. 'Participated in 2023' refers to schools in the school network that offer grade 6 and administered the 2019 test version in 2023. The p-value is derived from the t-test comparing the attrited with the participating schools. Special needs schools (1 school in 2022, and 4 schools in 2023) were excluded from the attrition analysis.

Table A4: Comparison of Participating Schools in 2022 and 2023: test version 2021

	Participated in 2022		**Participated in 2023**		*t*-Test
	N	Mean [SD]	*N*	Mean [SD]	*p*-Value
Number of Students	548	194.76	957	193.47	0.77
Share of special needs students	548	0.07	957	0.07	0.60
Teachers: Share with more than 20y experience	548	0.35	957	0.36	0.29
Share of staff absences	548	0.26	957	0.16	0.00
Teacher shortage: % of unfilled vacancies	548	1.83	957	2.11	0.02
Year 6: Share of Girls	548	0.50	957	0.50	0.36
Year 6: SES - Neighbourhood	548	0.22	954	0.20	0.34
Year 6: SES - Mother's Education	548	0.17	954	0.17	0.34
Year 6: SES - Home Language	548	0.20	954	0.19	0.42
Year 6: Grade Repetition	548	0.00	957	0.00	0.85
Number of Schools per District	548	20.74	957	22.48	0.11
Population density (Inhab/Km2)	548	1262.21	957	1220.50	0.73

Note: Standard deviations (SD) in parenthesis. 'Participated in 2022' refers to schools in the school network that offer grade 6 and administered the 2021 test version in 2022. 'Participated in 2023' refers to schools in the school network that offer grade 6 and administered the 2021 test version in 2023. The p-value is derived from the t-test comparing the attrited with the participating schools. Special needs schools (1 school in 2022, and 4 schools in 2023) were excluded from the attrition analysis.

Appendix B: Additional tables and figures

Table B1: Main results: Dutch score, domain "Reading comprehension" (2019 test version)

Dutch (Reading comprehension) Score, Sample 2019-2023				
Year-to-2019 change				
Change in 2020	-0.15***	-0.15***	-0.14***	-0.15***
	(0.02)	(0.02)	(0.02)	(0.02)
Change in 2021	-0.17***	-0.17***	-0.16***	-0.17***
	(0.02)	(0.02)	(0.02)	(0.02)
Change in 2022	-0.44***	-0.44***	-0.43***	-0.43***
	(0.03)	(0.03)	(0.03)	(0.03)
Change in 2023	-0.28***	-0.28***	-0.27***	-0.27***
	(0.02)	(0.02)	(0.02)	(0.02)
N	92423	92423	92420	92420
School fixed effects	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers & school district size	No	No	No	Yes

Note: Robust standard errors between parentheses. Sample 2019-2023, student-level data. 'Change in 2020', 'Change in 2021', 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2020, 2021, 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2020, 2021, 2022 and 2023, relative to pre-pandemic levels. The standardised test scores (mean 0, SD 1) come from the exact same test version 2019. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B2: Main results: Math score, domain "Geometry" (2019 test version)

Math (Geometry) Score, Sample 2019-2023				
Year-to-2019 change				
Change in 2020	-0.07***	-0.07***	-0.06**	-0.07***
	(0.03)	(0.03)	(0.03)	(0.03)
Change in 2021	-0.12***	-0.11***	-0.11***	-0.11***
	(0.02)	(0.02)	(0.02)	(0.02)
Change in 2022	-0.26***	-0.24***	-0.24***	-0.24***
	(0.02)	(0.02)	(0.02)	(0.02)
Change in 2023	-0.19***	-0.18***	-0.17***	-0.18***
	(0.02)	(0.02)	(0.02)	(0.02)
N	84215	84215	84177	84174
School fixed effects	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers & school district size	No	No	No	Yes

Note: Robust standard errors between parentheses. Sample 2019-2023, student-level data. 'Change in 2020', 'Change in 2021', 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2020, 2021, 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2020, 2021, 2022 and 2023, relative to pre-pandemic levels. The standardised test scores (mean 0, SD 1) come from the exact same test version 2019. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B3: Main results: Science score, domain "Technology" (2019 test version)

Science Score (Technology), Sample 2019-2023				
Year-to-2019 change				
Change in 2020	-0.09**	-0.09**	-0.08**	-0.08**
	(0.04)	(0.04)	(0.04)	(0.04)
Change in 2021	-0.13***	-0.13***	-0.12***	-0.12***
	(0.02)	(0.03)	(0.02)	(0.03)
Change in 2022	-0.22***	-0.22***	-0.21***	-0.21***
	(0.03)	(0.03)	(0.03)	(0.03)
Change in 2023	-0.13***	-0.12***	-0.11***	-0.12***
	(0.02)	(0.02)	(0.02)	(0.03)
N	56740	56740	56720	56720
School fixed effects	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers & school district size	No	No	No	Yes

Note: Robust standard errors between parentheses. Sample 2019-2023, student-level data. 'Change in 2020', 'Change in 2021', 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2020, 2021, 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2020, 2021, 2022 and 2023, relative to pre-pandemic levels. The standardised test scores (mean 0, SD 1) come from the exact same test version 2019. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B4: Main results: Social Science score, domain "Time" (2019 test version)

Social Science (Time) Score, Sample 2019-2023				
Year-to-2019 change				
Change in 2020	-0.09***	-0.08***	-0.08***	-0.08***
	(0.02)	(0.02)	(0.02)	(0.02)
Change in 2021	-0.04	-0.03	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.03)
Change in 2022	-0.22***	-0.21***	-0.21***	-0.20***
	(0.02)	(0.03)	(0.03)	(0.03)
Change in 2023	-0.19***	-0.18***	-0.17***	-0.17***
	(0.02)	(0.02)	(0.02)	(0.03)
N	60961	60961	60958	60932
School fixed effects	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers & school district size	No	No	No	Yes

Note: Robust standard errors between parentheses. Sample 2019-2023, student-level data. 'Change in 2020', 'Change in 2021', 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2020, 2021, 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2020, 2021, 2022 and 2023, relative to pre-pandemic levels. The standardised test scores (mean 0, SD 1) come from the exact same test version 2019. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B5: Main results: Dutch score, domain "Listening" (2021 test version)

Dutch (Listening) Score, Sample 2021-2023				
Year-to-2021 change				
Change in 2022	-0.12***	-0.11***	-0.11***	-0.11***
	(0.02)	(0.02)	(0.02)	(0.02)
Change in 2023	-0.18***	-0.17***	-0.17***	-0.18***
	(0.02)	(0.02)	(0.02)	(0.03)
N	52374	52374	52357	52357
School fixed effects	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers & school district size	No	No	No	Yes

Note: Robust standard errors between parentheses. Sample 2021-2023, student-level data. 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2022 and 2023, relative to pre-pandemic levels. The standardised test scores (mean 0, SD 1) come from the exact same test version 2021. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B6: Main results: Math score, domain "Mental arithmetic & data visualisation" (2021 test version)

Math (Mental arithmetic & data visualisation) Score, Sample 2021-2023				
Year-to-2021 change				
Change in 2022	-0.11***	-0.10***	-0.10***	-0.10***
	(0.02)	(0.02)	(0.02)	(0.02)
Change in 2023	-0.01	0.00	0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.02)
N	68046	68046	67992	67991
School fixed effects	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers & school district size	No	No	No	Yes

Note: Robust standard errors between parentheses. Sample 2021-2023, student-level data. 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2022 and 2023, relative to pre-pandemic levels. The standardised test scores (mean 0, SD 1) come from the exact same test version 2021. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B7: Main results: Science score, domain "Nature" (2021 test version)

Science (Nature) Score, Sample 2021-2023				
Year-to-2021 change	-0.10***	-0.09***	-0.09***	-0.10***
Change in 2022	(0.03)	(0.03)	(0.03)	(0.03)
	-0.05**	-0.04**	-0.04*	-0.04
Change in 2023	(0.02)	(0.02)	(0.02)	(0.02)
N	41641	41641	41638	41638
School fixed effects	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers & school district size	No	No	No	Yes

Note: Robust standard errors between parentheses. Sample 2021-2023, student-level data. 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2022 and 2023, relative to pre-pandemic levels. The standardised test scores (mean 0, SD 1) come from the exact same test version 2021. * p < 0.10, ** p < 0.05, *** p < 0.01

Table B8: Main results: Social Science score, domain "Space" (2021 test version)

Social Science (Space) Score, Sample 2021-2023				
Year-to-2021 change				
Change in 2022	-0.24***	-0.23***	-0.23***	-0.23***
	(0.03)	(0.03)	(0.03)	(0.03)
Change in 2023	-0.07***	-0.06***	-0.06***	-0.06**
	(0.02)	(0.02)	(0.02)	(0.03)
N	42894	42894	42876	42850
School fixed effects	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers & school district size	No	No	No	Yes

Note: Robust standard errors between parentheses. Sample 2021-2023, student-level data. 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2022 and 2023, relative to pre-pandemic levels. The standardised test scores (mean 0, SD 1) come from the exact same test version 2021. * p < 0.10, ** p < 0.05, *** p < 0.01

Table B8: Main results: French score, domain "Reading" (2021 test version)

French (Reading) Score, Sample 2021-2023				
Year-to-2021 change	-0.01	-0.00	-0.00	0.01
Change in 2022	(0.01)	(0.01)	(0.01)	(0.02)
	0.01	0.02	0.02	0.02
Change in 2023	(0.02)	(0.02)	(0.02)	(0.02)
N	76218	76218	76167	76095
School fixed effects	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers & school district size	No	No	No	Yes

Note: Robust standard errors between parentheses. Sample 2021-2023, student-level data. 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2022 and 2023, relative to pre-pandemic levels. The standardised test scores (mean 0, SD 1) come from the exact same test version 2021. * p < 0.10, ** p < 0.05, *** p < 0.01

Table B9: Main results: French score, domain "Listening" (2021 test version)

French (Listening) Score, Sample 2021-2023				
Year-to-2021 change				
Change in 2022	0.01	0.02	0.02	0.02
	(0.01)	(0.01)	(0.01)	(0.02)
Change in 2023	-0.02	-0.01	-0.01	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)
N	76394	76394	76343	76272
School fixed effects	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers & school district size	No	No	No	Yes

Note: Robust standard errors between parentheses. Sample 2021-2023, student-level data. 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2022 and 2023, relative to pre-pandemic levels. The standardised test scores (mean 0, SD 1) come from the exact same test version 2021. * p < 0.10, ** p < 0.05, *** p < 0.01

Table B10: External validity: participating schools in the test version 2019, year 2023

	Out-of-sample schools		In-sample schools		t-Test
	N	Mean [SD]	*N*	Mean [SD]	*p*-Value
% students living in neighbourhood with high 'grade retention rates'	1362	0.23	1049	0.18	0.00
% different home language	1426	0.27	1077	0.21	0.00
% students with low-educated mother	1433	0.22	1077	0.17	0.00
% students receiving financial support	1433	0.41	1077	0.35	0.00

Note: Mean values and p-values of each school characteristic are computed using a t-test. 'In-sample schools' are those that participated in the 2019 test version. 'Out-of-sample schools' include all remaining Flemish schools, affiliated with any school network, that offer grade 6 but did not participate in the test. Data source: (AGODI, 2023)

Table B11: External validity: participating schools in the test version 2021, year 2023

	Out-of-sample schools		In-sample schools		t-Test
	N	Mean [SD]	*N*	Mean [SD]	*p*-Value
% students living in neighbourhood with high 'grade retention rates'	1484	0.22	927	0.19	0.00
% different home language	1546	0.26	957	0.22	0.00
% students with low-educated mother	1553	0.22	957	0.18	0.00
% students receiving financial support	1553	0.40	957	0.35	0.00

Note: Mean values and p-values of each school characteristic are computed using a t-test. 'In-sample schools' are those that participated in the 2019 test version. 'Out-of-sample schools' include all remaining Flemish schools, affiliated with any school network, that offer grade 6 but did not participate in the test. Data source: (AGODI, 2023)

Table B12: Differences between 'Change' dummies

	Dutch	Math	Science	Social Science
Year-to-2019 change	*Reading comprehension*	*Geometry*	*Technology*	*Time*
Change in 2023 - Change in 2022	0.16***	0.06***	0.09***	0.03
Change in 2023 - Change in 2021	-0.09***	-0.06***	0.00	-0.14***
Year-to-2021 change	*Listening*	*Mental arithmetic*	*Nature*	*Space*
Change in 2023 - Change in 2022	-0.07***	0.11***	0.06**	0.17***

Note: The difference among coefficients is based on the student-level data. Year-to-2019 change is based on the sample 2019-2023. Year-to-2021 change is based on the sample 2021-2023. * p < 0.10, ** p < 0.05, *** p < 0.01

Table B13: Robustness checks, schools participating in all years, sample 2019-2023

	Dutch (Reading Comprehension)	Math (Geometry)	Science (Technology)	Social Science (Time)
Year-to-2019 change				
Change in 2020	-0.13***	-0.06**	-0.06	-0.06***
	(0.02)	(0.03)	(0.04)	(0.02)
Change in 2021	-0.13***	-0.08**	-0.10***	0.03
	(0.03)	(0.03)	(0.04)	(0.04)
Change in 2022	-0.37***	-0.23***	-0.17***	-0.16***
	(0.04)	(0.03)	(0.05)	(0.04)
Change in 2023	-0.25***	-0.16***	-0.06*	-0.13***
	(0.04)	(0.03)	(0.04)	(0.05)
N	36855	30321	21225	24889

Note: Robust standard errors between parentheses. Sample 2019-2023, student-level data. The regressions only include schools that participated in the tests in each year. 'Change in 2020', 'Change in 2021', 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2020, 2021, 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2020, 2021, 2022 and 2023, relative to pre-pandemic levels.

Table B14: Robustness checks from coarsened exact matching (CEM), sample 2019-2023

	Dutch (Reading Comprehension)	Math (Geometry)	Science (Technology)	Social Science (Time)
Year-to-2019 change				
Change in 2020	-0.15***	-0.07**	-0.15**	-0.05*
	(0.03)	(0.04)	(0.07)	(0.03)
Change in 2021	-0.14***	-0.07**	-0.11***	0.02
	(0.03)	(0.03)	(0.04)	(0.04)
Change in 2022	-0.46***	-0.23***	-0.24***	-0.21***
	(0.04)	(0.03)	(0.04)	(0.04)
Change in 2023	-0.31***	-0.18***	-0.15***	-0.16***
	(0.03)	(0.03)	(0.04)	(0.04)
N	64130	60865	39536	42806

Note: Robust standard errors between parentheses. Sample 2019-2023, student-level data. . Schools were matched based on all school characteristics as coarsened variables of each year, compared to 2019. The 2019 cohort was retained in its entirety, while for the other cohorts, only matched observations were preserved to maximise both matching and statistical power. 'Change in 2020', 'Change in 2021', 'Change in 2022' and 'Change in 2023' are dummy variables equal to 1 in the year 2020, 2021, 2022 and 2023 (and 0 otherwise), respectively, measuring the change of standardised test scores in 2020, 2021, 2022 and 2023, relative to pre-pandemic levels

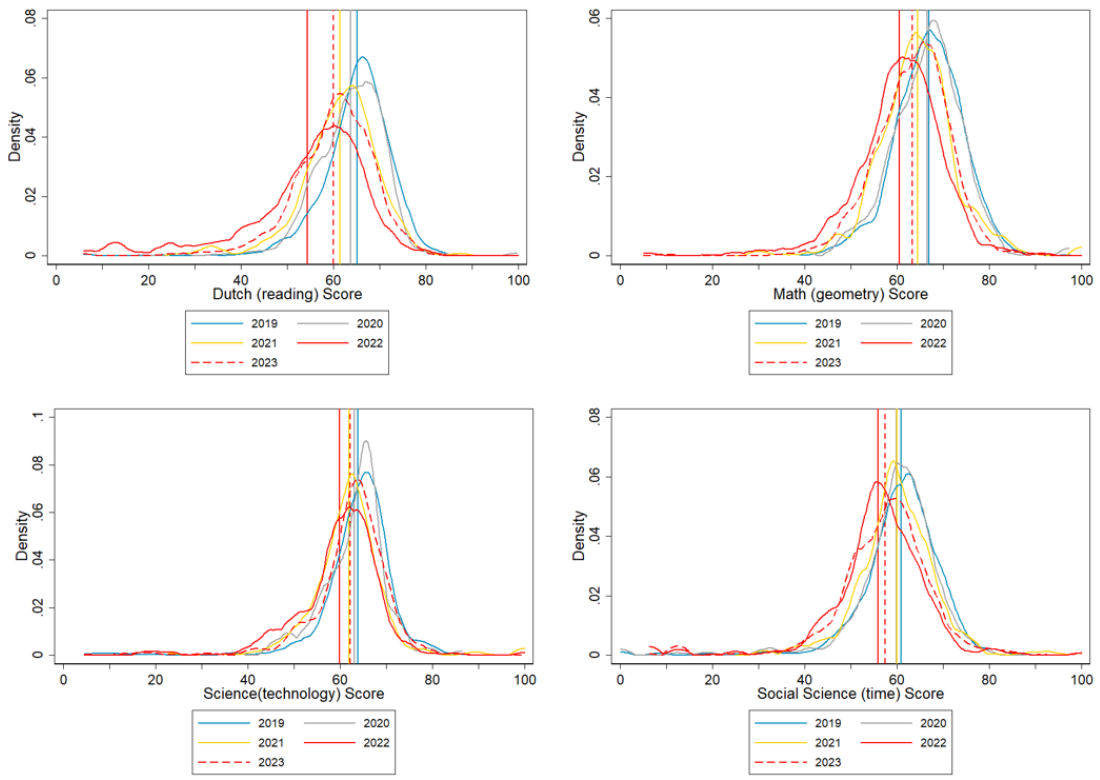

Figure B1: Distribution of scores from test version 2019, sample 2019-2023

Note: Comparison of the distribution of the scores for test version 2019 across the years 2019 to 2023. All distributions are based on raw test scores at student level.

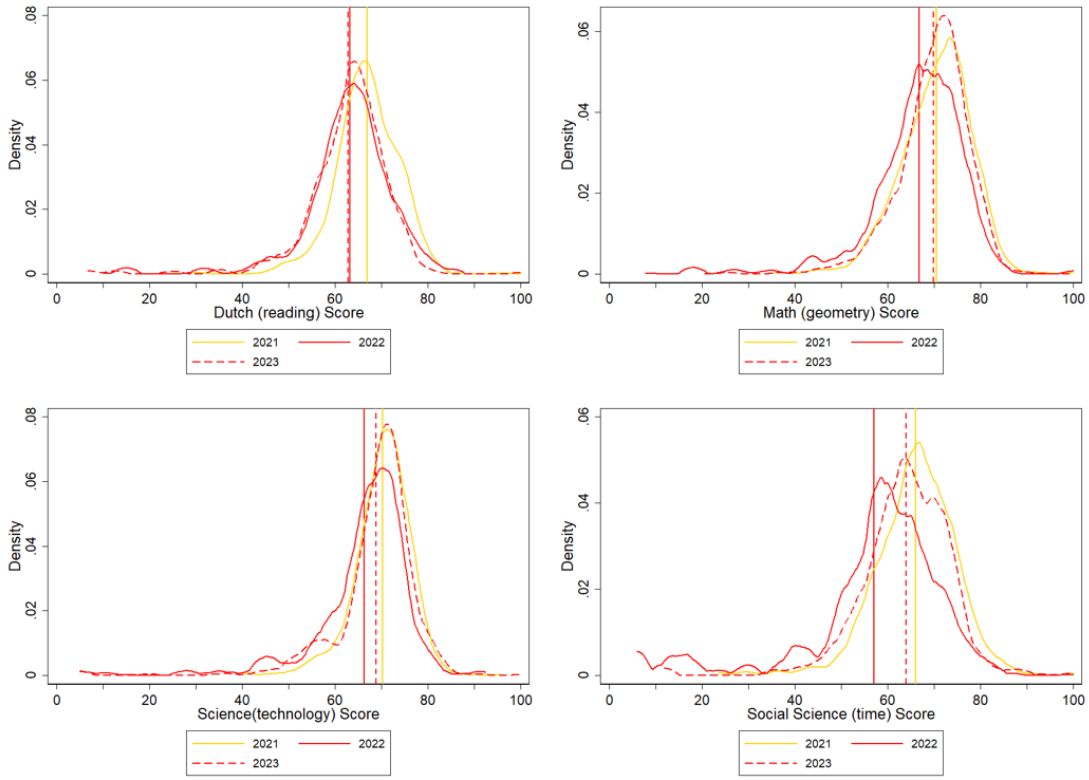

Figure B2: Distribution of scores from test version 2021, sample 2021-2023

Note: Comparison of the distribution of the scores for test version 2019 across the years 2019 to 2023. All distributions are based on raw test scores at student level.

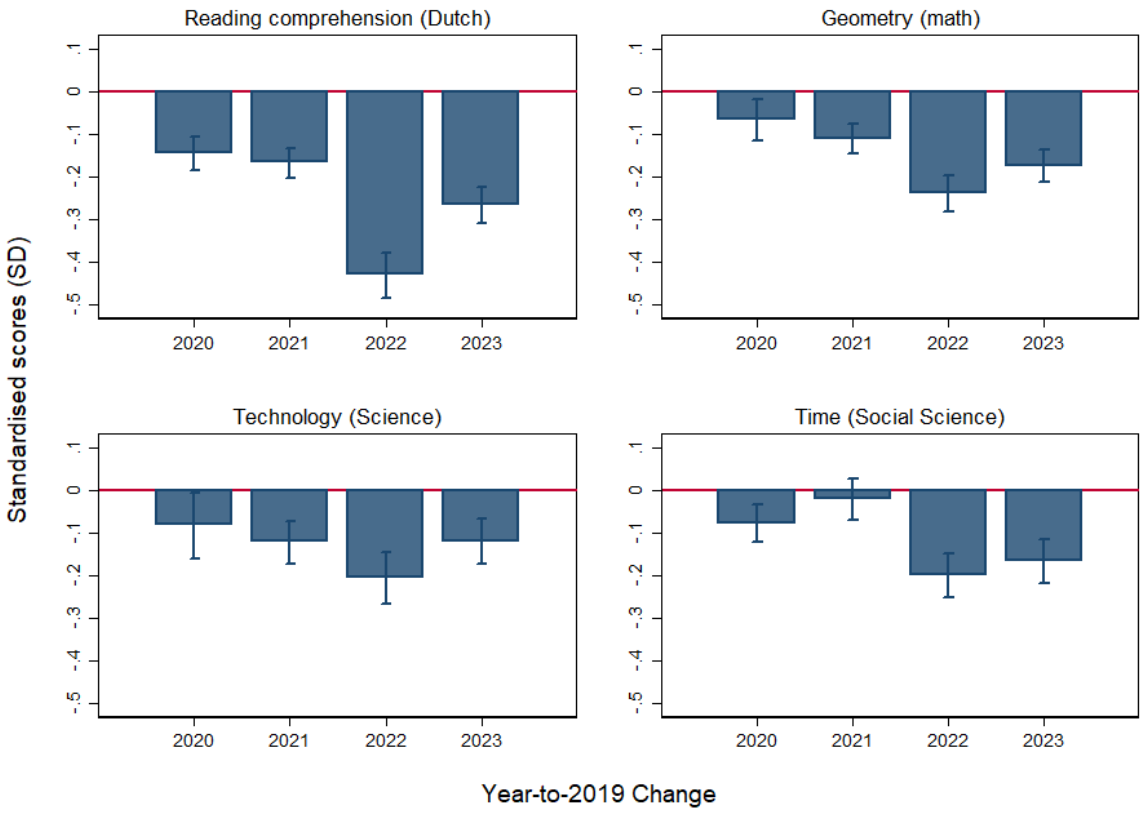

Figure B3: Visual exploration of the trend, sample 2019-2023

Note: These figures are based on the student-level data analysis from test version 2019.

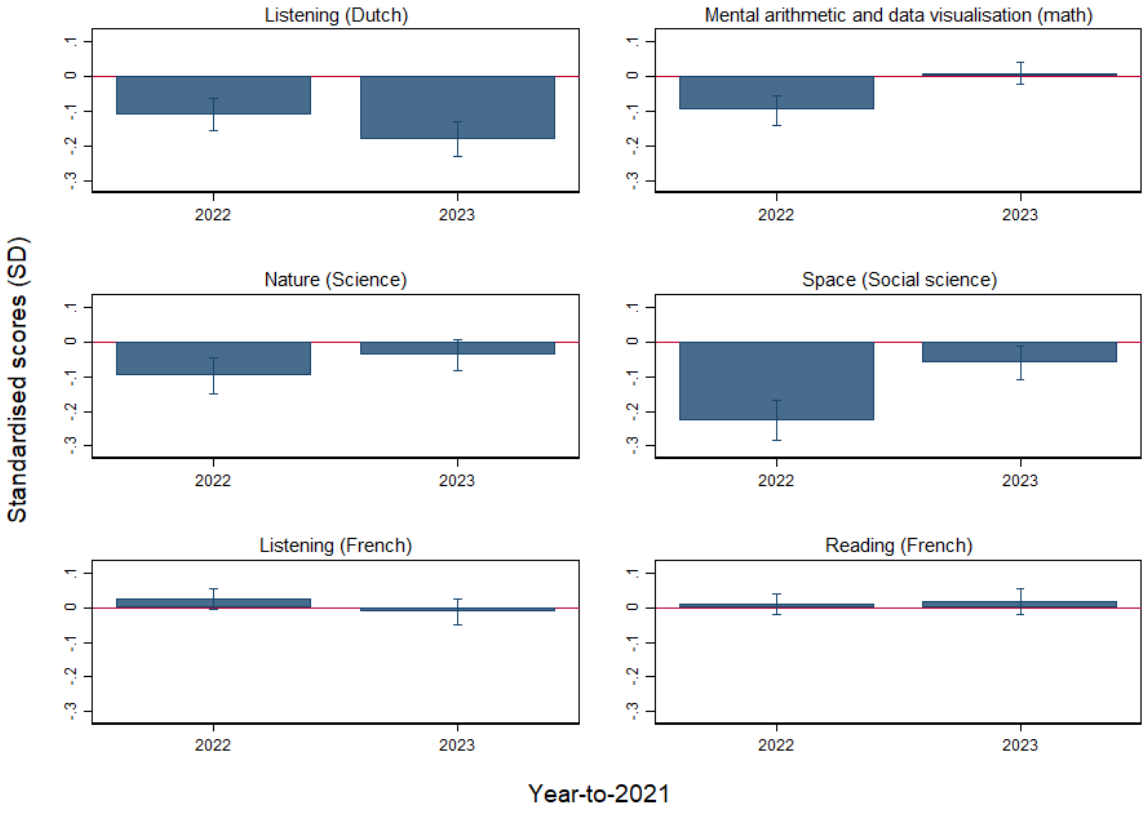

Figure B4: Visual exploration of the trend, sample 2021-2023

Note: These figures are based on the student-level data analysis from test version 2021.

Appendix C: Education system in Flanders

The primary education cycle in the Flemish community of Belgium spans six years (from age 6 to 12), after which students transition to secondary education (from age 12 to 18). Dutch serves as the medium of instruction, while French is taught as the first foreign language. Flanders[10] affords schools a significant degree of autonomy, and both children and their parents possess the freedom to select any primary or secondary school without the imposition of catchment areas. Moreover, the organization of education falls under two primary school networks: public education, primarily administered by the Flemish government (known as GO! Gemeenschapsonderwijs), municipal, city, or provincial schools (referred to as OVSG for urban and municipal education, and POV for provincial education), and privately-run yet publicly funded education, predominantly represented by Catholic schools. The network of Catholic schools, known as Katholiek Onderwijs Vlaanderen, stands as the largest education provider in Flanders, attracting a slightly more advantaged student population (Cherchye et al., 2010). As there is no standardised national assessment in Belgium, this paper draws upon data on standardised curriculum-based tests administered annually during the last year (6th grade) of primary school the network of Catholic schools. These tests enable schools to internally assess their students' performance, forming a component of the network's internal quality evaluation. The results of the tests are not disclosed to the public or the central government. Since 2019 two different test versions were administered either as the main official test of the specific year, or as an additional test taken along with the official test version: a test version administered between 2019 and 2023 (with a focus on the content subdomains: reading comprehension, geometry, technology, and time), and a test version introduced in 2021 and then administered every year till 2023 (with a focus on the content subdomains: listening, mental arithmetic and data visualisation, nature, and space). Table C1 provides additional information about the content subdomains tested in each test version.

While the Flemish education system is widely recognized for its high standards[11], there has been a downward trend in students' performances across all domains tested by international education assessments in the past decade (e.g. PIRLS, PISA, TIMSS) (Dockx et al., 2019; Mullis et al., 2019, 2023; OECD, 2023). Gambi & De Witte (2023) further confirmed this decline, reporting a similar downward trend in standardised test scores since 2018, based on data from annual curriculum-based tests at the end of primary school in Flanders.

[10] In our sample, schools situated in the Brussels Region where Dutch serves as the language of instruction are also included. This inclusion is warranted by the identical funding mechanisms applied to these schools, mirroring those employed in the Flemish region.

[11] Flemish students perform above average in science, mathematics, and reading according to 2022 PISA data (OECD, 2023), and in TIMMS (Dockx et al., 2019; Mullis et al., 2019) and PIRLS tests (Mullis et al., 2023). Furthermore, investment in education is one of the highest among the OECD countries (OECD, 2022).

Table C1: Content subdomains for each subject by test version

Test Version	Content domain	Details
Dutch (Katholiek Onderwijs Vlaanderen, 2023a)		
2019	Reading comprehension	Reading comprehension
2021	Listening	Listening
Mathematics (Katholiek Onderwijs Vlaanderen, 2023e)		
2019	Geometry	Rounding, estimation and length, circumference, and area
2021	Mental arithmetic & data visualization	Mental arithmetic, measuring and calculating with time and money and interpreting graphs
Science (Katholiek Onderwijs Vlaanderen, 2023b)		
2019	Technology	Components of technical systems, technical principles, and natural phenomena, evaluating existing technical systems, selecting technical solutions, recognizing different energy sources and their importance, impact of technical systems on the environment
2021	Nature	Organisms in their habitats, dependency of human activities on nature, growth and reproduction of living organisms, impact of human activity on nature and environment
Social Science (Katholiek Onderwijs Vlaanderen, 2023d)		
2019	Time	Calendars and planning, historical perspective, evolution of reality and knowledge of time, difference between historical facts and opinions about them
2021	Space	Delimitation of space and borders, landscape geography, maps and orientation, traffic safety, itinerary planning
French (Katholiek Onderwijs Vlaanderen, 2023c)		
2021	Listening, Reading	Reading comprehension and listening, each tested using 10 questions.

Note: 'Test version' refers to the first year in which the test was administrated. The test versions 2019 and 2021 have been administered between 2019 and 2023, and between 2021 and 2023, respectively.