SCHOOL EFFECTIVENESS and SCHOOL IMPROVEMENT
An International Journal of Research, Policy and Practice

Routledge
Taylor & Francis Group

School Effectiveness and School Improvement
An International Journal of Research, Policy and Practice

ISSN: 0924-3453 (Print) 1744-5124 (Online) Journal homepage: www.tandfonline.com/journals/nses20

Grade Retention Across Borders: A Cross-Country Analysis of the Role of Schools

Janneke Pepels, Barbara Belfi, Mieke Goos & Carla Haelermans

To cite this article: Janneke Pepels, Barbara Belfi, Mieke Goos & Carla Haelermans (25 Apr 2026): Grade Retention Across Borders: A Cross-Country Analysis of the Role of Schools, School Effectiveness and School Improvement, DOI: 10.1080/09243453.2026.2658545

To link to this article: https://doi.org/10.1080/09243453.2026.2658545

Published online: 25 Apr 2026.

SCHOOL EFFECTIVENESS AND SCHOOL IMPROVEMENT
https://doi.org/10.1080/09243453.2026.2658545

Routledge
Taylor & Francis Group

Grade Retention Across Borders: A Cross-Country Analysis of the Role of Schools

Janneke Pepels [a], Barbara Belfi [a], Mieke Goos [b] and Carla Haelermans [a]

Research Centre for Education and the Labour Market (ROA), School of Business and Economics, Maastricht University, Maastricht, the Netherlands; School of Educational Studies, Hasselt University, Wetenschapspark 24, 3590 Diepenbeek, Belgium

ABSTRACT

Grade retention – although widely used – remains controversial because of its implications for equity and long-term outcomes. While student-level predictors are well documented, less is known about how school and system characteristics shape grade retention practices across countries. Adopting a multilevel perspective, this study examines (1) which school-level characteristics are related to grade retention in lower-secondary education and (2) whether these relationships vary by national educational stratification. Using PISA 2018 data from 124,126 students in 5,693 schools across 24 systems, we estimate three-level regression models. School composition, within-class ability grouping, and shortages of educational materials are linked to grade retention probability, net of individual characteristics. Cross-level interaction analyses show that educational material shortages predict higher grade retention likelihoods only in low-stratification systems, while the association with within-class ability grouping is substantially lower in these contexts. These findings underscore that grade retention practices are shaped by both institutional structures and school organisation.

ARTICLE HISTORY
Received 9 September 2025
Accepted 7 April 2026

KEYWORDS
Grade Retention; Lower-Secondary Education; Educational Effectiveness; Cross-National Analysis; Multilevel Modelling; School Characteristics

Introduction

Grade retention – the practice of requiring students to repeat a school year when their performance is deemed insufficient – is a common but highly debated approach in education systems worldwide (Pipa et al., 2025). While it is often intended to give students an extra opportunity to master the curriculum, research consistently links grade retention to negative long-term outcomes, including lower academic achievement, reduced motivation, higher dropout rates, and poorer labour-market prospects (Goos et al., 2021; Valbuena et al., 2021). Moreover, grade retention is not applied equally: students from lower socio-economic backgrounds and students with a migration background are more likely to be retained than their peers (Choi et al., 2018; Klapproth & Schaltz, 2015;

CONTACT Janneke Pepels janneke.pepels@maastrichtuniversity.nl

Westphal et al., 2020). These patterns raise concerns about fairness and the reproduction of educational inequality, issues central to school effectiveness and school improvement (OECD, 2023).

Most prior research has focused on student-level predictors such as academic performance and background characteristics (Guèvremont et al., 2007; Valbuena et al., 2021). However, grade retention decisions are made within schools. School characteristics – including student composition, available resources, and internal policies – shape how schools respond to students who struggle academically. This implies that grade retention should not only be understood as a reaction to individual performance, but also as an organisational response to student heterogeneity. From this perspective, retention is embedded in school routines and decision-making structures rather than functioning solely as an individual corrective measure. Schools serving more disadvantaged populations may face greater instructional challenges, while schools operating with strict academic standards may be more likely to retain students who do not meet performance thresholds (Bulle, 2019; Hanushek & Woessmann, 2016). These findings suggest that grade retention is not solely an individual outcome but is also embedded in school-level organisational contexts. Multilevel research supports this view.

While a substantial portion of the variation in grade retention lies at the student level, considerable amounts are also found at the school and national levels (Goos et al., 2013). Cross-national differences are striking – grade retention rates range from below 1% in countries such as Ireland and Denmark to above 20% in Spain, Germany, and Belgium (OECD, 2019). These differences reflect national promotion rules, accountability systems, and degrees of school autonomy (Eurydice & Commission, 2011), as well as how schools implement these frameworks in practice (Dupriez et al., 2008). This suggests that the relationship between school characteristics and grade retention may depend on the broader institutional context.

One institutional dimension that may shape these patterns is educational stratification. In highly stratified systems, students are sorted into different tracks at an early stage, with very little mobility between tracks. This leads to larger differences between schools in student composition, academic expectations, and organisational practices (Pfeffer, 2008). As a result, school-level characteristics may show stronger associations with grade retention in such systems.

Against this background, two questions arise. First, which school characteristics are systematically associated with grade retention across countries (RQ1)? Second, to what extent do these associations vary depending on the degree of educational stratification within national systems (RQ2)?

This study addresses these questions in two steps. First, we examine which school-level characteristics are associated with grade retention across countries (RQ1). Second, we investigate whether the strength of these associations varies depending on the degree of educational stratification within national systems (RQ2). Using multilevel regression analysis of PISA 2018 data, we analyse school composition, school policy, and the learning environment as predictors of grade retention in lower-secondary education. By distinguishing between general school-level associations and context-dependent patterns, this study contributes to a more differentiated understanding of how institutional structures shape the relationship between schools and grade retention.

Determinants of Grade Retention Across Levels

Previous research has identified a wide range of factors associated with grade retention in secondary education, operating at multiple levels of the education system. These include individual characteristics, as well as classroom-level, school-level, and system-level features. This multilevel perspective reflects the idea that grade retention decisions are shaped by both student characteristics and the institutional contexts in which they are embedded.

Individual-Level Predictors

Grade retention has been most extensively studied at the individual level, focusing on various learning, psycho-social, and background characteristics as key predictors of grade retention. Learning characteristics such as lower academic performance, higher absenteeism, and weaker early childhood preparation have all been linked consistently to a higher likelihood of being retained (Choi et al., 2018; Cortázar et al., 2020; Klapproth & Schaltz, 2015; Westphal et al., 2020). In addition, lower mathematics self-concept, lower educational aspirations, and lower levels of conscientiousness have also been linked with a higher grade retention likelihood (Ferrão, 2015; Resino et al., 2019; Westphal et al., 2020). Background characteristics also play a consistent role. Students from low socioeconomic status (SES) families, those with a migration or minority background, and boys are more likely to repeat a grade (Chau et al., 2012; Goos et al., 2013; Kornfeld & Ochsen, 2017). This risk is also elevated for students who are relatively young, live in unfavourable neighbourhoods, or have younger mothers (Guèvremont et al., 2007; Klapproth & Schaltz, 2015; Kornfeld & Ochsen, 2017). Together, these findings indicate that grade retention is closely linked to pre-existing social and educational inequalities.

Classroom- and Teacher-Level Predictors

Evidence of grade retention predictors at the classroom and teacher level is rather limited. Two studies suggest that low peer acceptance increases the likelihood of being retained (Lubbers et al., 2006), as does being taught by non-certified teachers (Kornfeld & Ochsen, 2017). Overall, however, empirical evidence at this level remains scarce and fragmented. The limited research indicates a need for further study on how classroom dynamics and teacher characteristics influence grade retention, particularly given the broader evidence on the importance of the immediate learning environment for students' academic outcomes (Hattie, 2023a, 2023b).

School-Level Predictors

Grade retention predictors at the school level have been studied to a larger extent, as compared to the classroom and teacher level. School composition is the most frequently studied factor. Students in schools with a high share of low-SES or migration-background peers are generally at higher risk of being retained (Cordero Ferrera et al., 2014; Vinas-Forcade et al., 2021). In some cases, a disadvantaged school composition even weakens the protective effect of individual student achievement (Vinas-Forcade et al., 2021). However, results are not consistent: some studies have found no effect of migration background composition (Choi et al., 2018), and

Shi & Brown (2020) report lower grade retention likelihoods in schools with many economically vulnerable students. School location shows similarly mixed patterns. Grade retention is sometimes more frequently applied in urban schools (Choi et al., 2018; Ferrão, 2015), sometimes in rural areas (Vinas-Forcade et al., 2021). Counselling provision may also play a role: schools offering only non-targeted counselling report higher grade retention rates (Shi & Brown, 2020). Finally, some structural school features have been linked to a lower grade retention likelihood, such as larger school size, private status, stable student populations, and higher student-teacher ratios (Choi et al., 2018; Guèvremont et al., 2007). Yet, these findings have not been consistently replicated (Cordero Ferrera et al., 2014; Ferrão, 2015; Shi & Brown, 2020). Overall, while school-level characteristics appear relevant, empirical evidence remains fragmented and sometimes contradictory.

System-Level Predictors

At the system level, educational policies and structural features have been shown to play a key role in shaping grade retention practices. A critical factor seems to be how the education system manages student heterogeneity. Systems that do not have flexible ways of responding to diverse learning needs – i.e. those that use early tracking or a uniform curriculum – show much higher grade retention rates than systems that have response mechanisms such as streaming or individualized curricula (Dupriez et al., 2008). Within tracked systems, students in vocational pathways are particularly at risk of being retained (Klapproth & Schaltz, 2015). Policy reforms can reduce the risk of grade retention: for example, the introduction of binding teacher track recommendations in Germany led to lower grade retention rates by reducing the number of students placed in a too demanding track (Kornfeld & Ochsen, 2017). System-level support measures also matter. Grade retention has been found to be less common in districts offering early childhood care and education (ECCE; Cortázar et al., 2020), and in systems that provide incentives for teaching students with special educational needs (SEN; (Goos et al., 2013)). These findings suggest that the application of grade retention is influenced by how education systems are designed to accommodate diverse learners and promote progression. Overall, these findings indicate that grade retention practices are embedded in national policy frameworks and institutional designs.

In sum, grade retention is shaped by factors at the student, classroom, school, and system levels. Research is most extensive at the individual level, while evidence on classroom and teacher characteristics is limited. School-level studies highlight the importance of school composition and school organisational features, yet findings are often inconsistent. System-level research further suggests that institutional structures influence how grade retention is applied. Despite this multilevel evidence, cross-national analyses that simultaneously examine school-level characteristics and their operation within different institutional contexts remain scarce. In particular, we lack systematic evidence on whether school-level predictors of grade retention operate differently in highly stratified versus more comprehensive systems. This study therefore examines which school characteristics are associated with grade retention and how these associations vary across countries.

Theoretical Framework

As the previous section has shown, grade retention is influenced by multiple factors across different levels of the education system. We use the Dynamic Model of Educational Effectiveness (DMEE; Creemers & Kyriakides, 2007) as our guiding framework and build on Carroll's Model of School Learning (Carroll, 1963, 1989) to understand how school characteristics are related to grade retention. We further examine whether the degree of educational stratification shapes the strength of these relationships across systems.

The DMEE offers a comprehensive framework for understanding educational outcomes by considering multiple hierarchical levels of influence. It assumes that students' learning is affected by a mixture of individual-, classroom-, school-, and system-level factors ("context" in the original model) and emphasizes how these levels interact to impact educational processes and outcomes. Although originally developed to explain variation in student achievement, the DMEE is relevant here, too, because grade retention decisions are typically based on students' academic progress within this multilevel context. For this study, we adapted the DMEE by focussing on grade retention (Figure 1). Our framework considers only three levels – student, school, and country – and centres on the school level. Within the school level, we distinguish three conceptual dimensions: school composition, school policy, and the learning environment. School composition refers to characteristics of the student body. School policy regards institutional practices and decisions that may influence the likelihood of grade retention. The school's learning environment

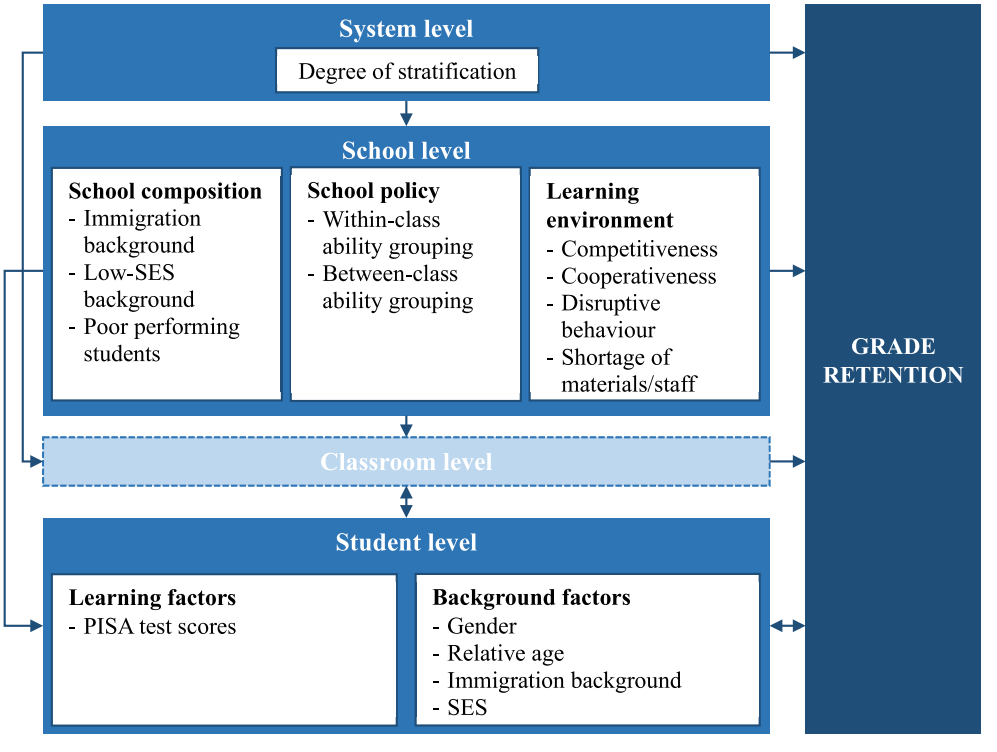

Figure 1. Conceptual model of school-level influences on grade retention.

contains behavioural and organisational aspects that reflect the school's day-to-day climate, shaped by school-specific policies and interactions between students and staff.

Since the single most important predictor of grade retention is low academic performance (Valbuena et al., 2021), understanding how school factors may be related to grade retention requires examining how schools shape student learning. To this end, we draw on Carroll's Model of School Learning (Carroll, 1963, 1989). This model suggests that school learning depends on five variables: (1) aptitude – the amount of time a student needs to learn under optimal conditions; (2) ability to understand instruction – influenced by prior knowledge or language barriers; (3) perseverance – the amount of time a student is willing to spend on learning; (4) quality of instruction – including clarity, organisation, and appropriateness; and (5) opportunity to learn – the time available for learning.

Grade retention decisions are typically based on students' academic achievement. Because achievement is shaped by learning conditions within schools, school factors may be indirectly related to grade retention particularly through their influence on perseverance, quality of instruction, and opportunity to learn. Pertaining to the school characteristics included in this paper, we hypothesize that school composition and school policy can influence students' perseverance and quality of instruction, while the learning environment may affect perseverance, quality of instruction, and opportunity to learn.

School compositional characteristics – i.e. SES, ethnicity, and achievement level - have been linked to teacher expectations, with expectations being higher in higher performing schools and lower in schools with more immigrant or low-SES students (Wang et al., 2018). In line with the Pygmalion effect (Rosenthal & Jacobson, 1968), teacher expectations have been found to impact students' learning behaviour, socio-psychological factors such as self-concept and academic motivation, and academic achievement (Wang et al., 2018). Furthermore, teacher expectations can also impact quality of instruction by modifying teachers' teaching behaviour (Aydin & Ok, 2022). Based on these mechanisms, we expect that schools with a higher proportion of low-SES students, students with a migration background, and low-achieving students will show higher odds of grade retention, net of individual student characteristics.

Similarly, we propose that a school's policy in the form of between- and within-class ability grouping will affect perseverance and quality of instruction. Between-class ability grouping has been linked to lower motivation and self-esteem among students in the lower groups (Slavin, 1990), and there are indications that lower quality or less experienced teachers are more likely to be assigned to lower ability groups (Francis et al., 2017). On the other hand, within-class ability has been associated with small positive effects on student achievement, attitudes and self-concept (Lou et al., 1996), suggesting that differentiated instruction within classrooms may improve learning conditions for some students. Based on these mechanisms, we expect that schools using between-class ability grouping will show higher odds of grade retention, whereas the association for within-class ability grouping may be reversed.

Finally, we argue that the learning environment can shape perseverance, quality of instruction, and opportunity to learn. Drawing on social constructivist perspectives (Amineh & Asl, 2015), learning and knowledge development take place in interaction with others. A cooperative school climate may therefore enhance students' engagement and willingness to invest effort in learning, thereby strengthening perseverance. In

contrast, a competitive environment may reduce perseverance, particularly among low-performing students who may experience pressure, anxiety, or fear of failure. In addition, disruptive behaviour can reduce effective instructional time and require teachers to devote attention to classroom management, thereby lowering instructional quality and opportunity to learn. Shortages of educational materials may constrain teachers' ability to provide appropriate instruction, while a lack of well-trained staff may diminish instructional quality. Taken together, we expect that schools characterised by higher levels of disruption, competitiveness, and resource shortages will show higher odds of grade retention, whereas more cooperative school environments will be associated with lower odds of retention.

As discussed in the literature review, studies on the association between school factors and grade retention show inconsistent findings across countries. One possible explanation for these differences lies in the institutional structure of educational systems. The degree of stratification of educational opportunity has been proposed to influence educational inequality (Marks, 2005; Pfeffer, 2008), and stratified systems are characterised by stronger segregation of disadvantaged students (Zapfe & Gross, 2021). Pfeffer conceptualises stratification in terms of the timing of selection, the rigidity of this selection, and its consequences for subsequent educational opportunities.

If school characteristics influence student learning through the mechanisms described above, the strength of these relationships may depend on the level of institutional stratification. In highly stratified systems, differences between schools reflect institutionalised sorting into distinct educational pathways, rather than minor organisational variation. As a result, schools may differ more substantially in composition, academic expectations, and performance thresholds. Under such conditions, the mechanisms linking school factors to student achievement – and ultimately to grade retention – are likely to operate more strongly. Consequently, we expect school factors to show stronger associations with grade retention in highly stratified countries.

Methods

Data

We used data from PISA 2018 to answer our research questions. PISA is a recurring international OECD survey that assesses the reading, mathematics, and science performance of 15-year-old students using a representative, multi-stage sampling design (OECD, 2019). Although PISA 2022 was available at the time of writing, we did not use this cycle because it was strongly affected by COVID-19 disruptions. Many countries experienced school closures, remote instruction, increased student absence, and temporary adjustments to grade promotion and retention policies (OECD, 2022). These exceptional circumstances likely influenced both learning processes and grade retention practices, thereby reducing cross-national comparability. Using pre-pandemic data improves the comparability and internal validity of our analyses. While PISA 2022 provides valuable insights into post-pandemic educational contexts, its exceptional conditions make it less suitable for examining structural cross-national differences in grade retention practices.

Our analytic sample was constructed in several steps. First, from the 81 participating countries, we excluded all countries that are not part of the OECD or Eurydice Network

Table 1. Overview of units of students, schools, and countries per level and descriptives of units within higher levels.

	N	*Min*	*M*	*Max*	*SD*
Students	124,126				
Schools	5,693				
Students per school		15	33.62	190	13.14
Retained students per school		0	1.14	25	2.30
Countries	24				
Students per country		2,281	5,172	25,190	4,563
Schools per country		47	237	1,036	203

($n = 33$), to ensure comparable educational contexts. OECD and Eurydice countries share broadly similar governance structures, grade progression systems, and accountability frameworks, which improves the comparability of grade retention practices across systems. Next, students who changed schools during lower-secondary education were excluded. Because grade retention decisions are typically made at the school level, including students who changed schools would introduce uncertainty about which school context was relevant for the grade retention decision. Since school change information was gathered in the Educational Career Questionnaire – an optional instrument (OECD, 2019) – we had to exclude 25 countries that did not measure these items (see the flowchart in Figure A1). This resulted in 24 educational systems across 23 countries. Belgium was split into Flanders and French-speaking Belgium, as PISA distinguishes these systems; other federal systems (e.g., Germany, Spain) could not be separated. The UK sample excluded Scotland due to missing school-change data. Then, students with missing grade retention data were excluded. Finally, to ensure robust school-level analysis, we excluded schools with less than 15 students[1] (5^{th} percentile) and excluded students lacking information regarding their grade retention status. In total, this school-size rule excluded 1,125 schools and 5,359 students.

Countries that were excluded from the final sample statistically differed from included countries on several student- and school-level characteristics. However, these differences were generally small among excluded OECD/EN countries. In contrast, non-OECD/EN countries showed more pronounced differences, particularly regarding ESCS, PISA performance, and indicators of school composition (Table A1). This suggests that restricting the sample mainly reduced institutional differences while keeping the selected systems comparable. Importantly, countries were not excluded based on their grade retention rates. Exclusions were driven by comparability criteria (OECD/Eurydice membership) and the availability of school-change information.

Our final sample included 124,126 students from 5,693 schools across 24 educational systems. Among them, 6,473 students (5.2%) reported having been retained in lower-secondary education. On average, schools had 1.14 retainees (SD = 2.30; Table 1), with considerable variation between national grade retention rates (Figure 2). The student population consisted of 50.8% female and 49.2% male students (Table 2). The large majority (88.9%) had no immigration background, respectively 3.8% and 1.2% had a second-generation or first-generation immigration background. On average, students

[1]Since several school-level variables were aggregated from the student level, schools with very few students would yield less reliable aggregate estimates and potentially unstable school-level effects in multilevel models.

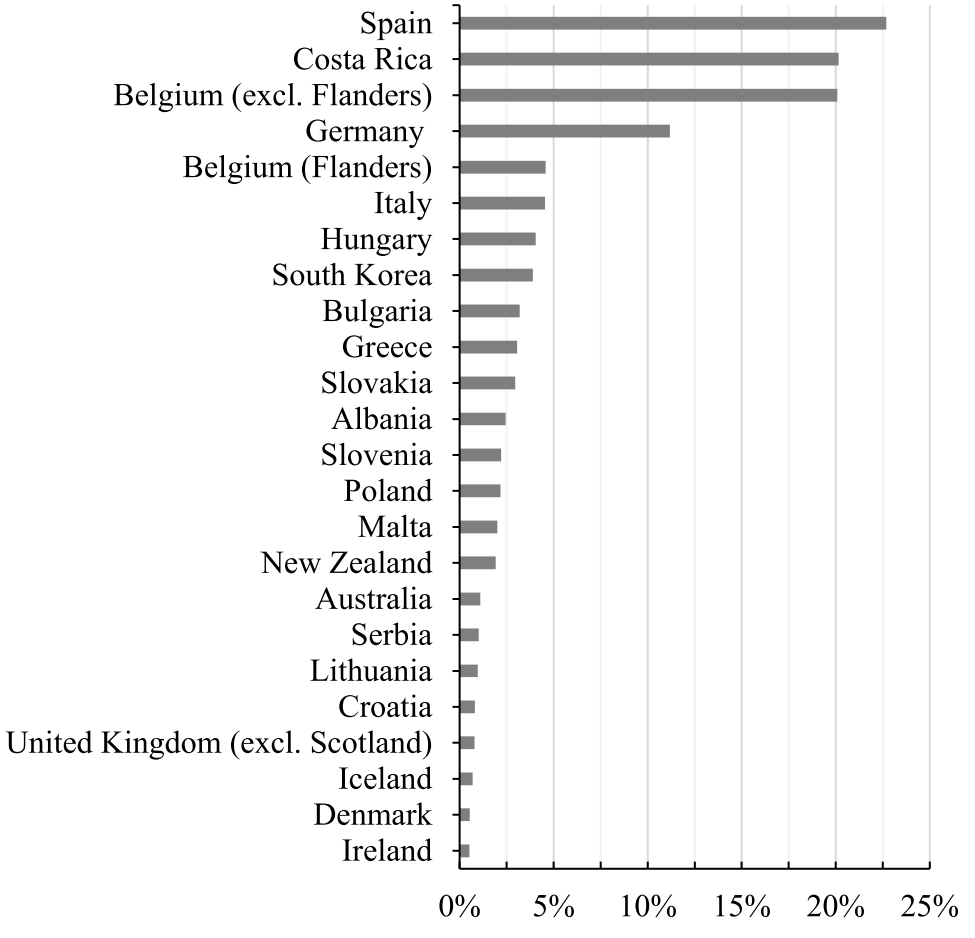

Figure 2. Lower-secondary grade retention prevalence per educational system.

Table 2. Descriptive statistics for student-level variables (*N*=124,126).

	n (%)	*M (SD)*	*Min;Max*	*Missing n (%)*
Gender				
Female	63,105 (50.8)			
Male	61,021 (49.2)			
Age		15.75 (0.29)	15.08;16.33	
ESCS		-0.01 (0.95)	-6.36;3.96	615 (0.5)
Immigration background				
Native	110,310 (88.9)			
Second-generation	7,516 (6.1)			
First-generation	4,759 (3.8)			
Missing	1,541 (1.2)			
PISA test average		495.10 (86.79)	168.20;832.00	

ESCS: economic, social, and cultural status

were 15.75 years old (*SD* = 0.29) and had a mean PISA test score of 495.10 (*SD* = 6.79; reflecting the analytic sample and scaling of plausible values). Their economic, social, and cultural status (ESCS) - a measure in PISA for SES – was comparable to that of the entire PISA 2018 population (*M* = -0.01, *SD* = 0.95).

Variables

Besides assessing students' academic performance, PISA also includes questionnaires for students and school administrators. The student questionnaire collects information on items such as students' background, perception of the learning environment, and school career, including information on grade retention. The school questionnaire covers several aspects of the school, such as its composition, resources, and policy. For an extensive description of the operationalization of these variables, we refer the reader to the PISA 2018 technical report (OECD, 2019).

Grade Retention

Data on students' grade retention history in lower secondary education was gathered from self-reported data from the student questionnaire item "*Have you ever repeated a grade at ISCED 2?*" with response options "*No, never*", "*Yes, once*", and "*Yes, twice or more.*"(*item* ST127Q01TA). For our analysis, we recoded this into a binary variable, where "0" indicated no grade retention and "1" indicated at least one grade repetition.

School-Level Variables

Data on school characteristics were gathered from the school questionnaire. Additionally, several variables were aggregated from the student level. Table 3 provides descriptive statistics of the school-level characteristics. Correlations among the school characteristics were generally small to moderate. The highest observed correlations were between the proportion of low-SES students and low-performing students ($r = .58$), and between disruptive student behaviour and disruptive teacher behaviour ($r = .50$) (see Table A2 for the correlation matrix). The school-level variables were categorised into three clusters, following our theoretical framework: school composition, school policy, and learning environment.

Table 3. Descriptives for school-level characteristics (*N*=5,693 schools).

	N (%)	*M (SD)*	*Min;Max*	*Missing N(%)*
School composition (%)				
Students with immigration background		12.48 (16.71)	0;100	-
Students with low-SES background		25.08 (17.62)	0;96.7	-
Poor-performing students		13.63 (16.06)	0;100	-
School policy				
Between-class ability grouping:				
Not for any subjects	2,985 (52.4)			
For some subjects	1,879 (33.0)			
For all subjects	355 (6.2)			
Missing	474 (8.3)			
Within-class ability grouping:				
Not for any subjects	2,334 (41.0)			
For some subjects	2,410 (42.3)			
For all subjects	549 (9.6)			
Missing	400 (6.0)			
Learning environment				
Perception of competitiveness		0.02 (0.38)	-1.99;2.04	13 (0.3)
Perception of cooperativeness		-0.04 (0.38)	-2.14;1.68	21 (0.4)
Disruptive student behaviour		-0.01 (1.06)	-4.09;3.44	297 (5.2)
Disruptive teacher behaviour		0.06 (0.99)	-2.09;3.81	300 (5.3)
Shortage of educational materials		-0.08 (1.01)	-1.86;4.04	315 (5.5)
Shortage of educational staff		0.09 (1.02)	-1.42;2.96	306 (5.4)

School composition – includes the proportion of students at school who (1) had an immigration background (i.e. first or second generation), (2) had a low-SES background ($\leq 25^{th}$ percentile in their country), and (3) had PISA test scores at the lowest level for all dimensions (reading, maths, and science). The low-performance measure refers to students scoring at or below Level 1 in each domain, capturing students with consistently weak academic performance. These variables were aggregated from student-level data and were based on items IMMIG, ESCS, and PV1MATH, PV1READ, and PV1SCIE, respectively. On average, the population of schools consisted for 12.48% of students with an immigration background ($SD = 16.71$), 25.08% of students had a low-SES background ($SD = 17.62$), and 13.63% ($SD = 16.06$) of students had PISA test results at the lowest levels (Table 3).

School policy – covers two forms of ability grouping, namely between-class and within-class grouping (never; sometimes; always), based on items SC042Q01TA and SC042Q02TA. Just over half of all schools did not use between-class ability grouping (52.4%), approximately a third used it for some subjects (33.0%), and 6.2% of schools used it for all subjects. For within-class ability grouping, 41.0% of schools reported not using it for any subjects, whereas 42.3% used it for some subjects, and 9.6% for all subjects (Table 3).

Learning environment – consists of measures for disruptive student and teacher behaviour, and shortages of educational materials and staff, as reported by the school questionnaire (items STUBEHA; TEACHBEHA; EDUSHORT; STAFFSHORT). It additionally includes the perception of competitiveness and cooperativeness among students at the school, which were aggregated from student level data (PERCOMP; PERCOOP). These variables were standardised by PISA and subsequently centred in our analyses; therefore, their means are close to zero. The standard deviation for perceived competitiveness and cooperativeness is smaller than 1, which reflects aggregation from student-level responses to the school level ($SD = 0.38$ for both, Table 3).

Stratification of Educational Systems

The degree of stratification in educational systems was operationalized based on (Pfeffer, 2008). Originally, this classification contained three levels of stratification (low, high, very high), which are based on a system's use of tracking and its between-track mobility and accessibility of post-secondary education. For this study, the high and very high levels were combined, since only Germany met the criteria for a very high degree of stratification. Additionally, we expanded the classification to educational systems that were not covered by Pfeffer. For this purpose, we used a conceptually similar classification proposed by Eurydice (European Commision; et al., 2020), which considers the same system characteristics. Countries within groups 1 and 2 of the Eurydice classification were reclassified as high in stratification (1), and countries within groups 3 to 5 were reclassified as low in stratification (0). Educational systems not included in the Eurydice or Pfeffer classifications (Australia, Costa Rica, and South Korea) were classified based on descriptions of their educational systems (OECD, 1998, 2025). In total, 14 educational systems were categorized as low in stratification and 10 as high in stratification (Table 4). In total, 79,596 students across 3,641 schools were in low-stratification systems, and 44,530 students in 2,052 schools were in high-stratification systems. Several school characteristics differed substantially between these systems, especially the proportion

Table 4. Degree of stratification in educational systems.

0 (Low stratification)	1 (High stratification)
Albania; Australia; Belgium (excl. Flanders); Costa Rica; Denmark; Greece; Ireland; Iceland; South Korea; Lithuania; Malta; New Zealand; Spain; United Kingdom	Belgium (Flanders); Bulgaria; Croatia; Germany; Hungary; Italy; Poland; Serbia; Slovakia; Slovenia

of students with an immigration background, the use of between-class ability grouping, the perception of competitiveness/cooperativeness, disruptive behaviour, and resource shortages (Table A3).

Covariates

We controlled for several student background characteristics, using data from the student questionnaire: gender (girl; boy; ST004D01 T), economic, social, and cultural status (ESCS), immigration background (native; first-generation immigrant; second generation immigrant; IMMIG), the PISA test average for reading, mathematics, and science (as a proxy for academic achievement, based on PV1MATH, PV1READ, and & PV1SCIE). Relative age was calculated based on students' age (AGE) and month of birth (ST003D02T).

Analyses

To account for the hierarchical structure of the data – students nested within schools, and schools nested within educational systems – we estimated three-level logistic regression models, with students at level 1, schools at level 2, and educational systems at level 3. Model coefficients are presented as odds ratios (ORs) to facilitate interpretation. An OR above 1 indicates a higher likelihood of grade retention, whereas an OR below 1 indicates a lower likelihood of being retained.

We followed a stepwise modelling strategy, aligned with our research questions. First, we estimated an empty model without any predictors, to partition the variance in grade retention across the three levels, using the latent variable approach (Browne et al., 2005). This provided intraclass correlations (ICCs) that indicate how much of the variation in grade retention is attributable to the different levels. Second, we added student-level covariates to control for students' individual background and academic performance, before introducing school characteristics. Third, to address Research Question 1, we estimated a series of school-level models. We first fitted three separate models, each including all variables from one of the conceptual dimensions in our theoretical framework: school composition, school policy, and learning environment. We then fitted a full school-level model that combined all variables, to identify which characteristics remained significant when considered jointly. Finally, to address Research Question 2, we included the degree of institutional stratification as an interaction term with all school-level variables. This allows us to test whether the associations between school-level variables and grade retention differ between high- and low-stratified systems. Significant interaction terms were probed by estimating marginal effects and predicted probabilities for each stratification group. All models controlled for the student-level covariates. All continuous predictors were grand-mean centred on their respective level. PISA test scores and the school composition variables were divided by 10 prior to centring. Missing values were handled

through multiple imputation for multilevel data ($m = 10$) using the R packages mitml v0.4-5, (Grund et al., 2019) and jomo v2.7-6 (Quartagno et al., 2019)). The imputation model included all variables used in the analyses to support the missing-at-random assumption. Due to convergence issues, the categorical variables measuring ability grouping could not be imputed and were therefore coded as "unknown", applying the missing-indicator approach (Enders, 2022), which allows partial inclusion of these cases in the multilevel models. Sensitivity analyses excluding these cases yielded substantively similar results. All models were weighted using senate weights. All analyses were conducted in RStudio v2025.05.0, using R v4.4.2, and the lme4 (v1.1.37) package (Bates et al., 2015).

Results

Variance Decomposition and Student-Level Characteristics associated with Grade Retention across Countries

The null model shows the distribution of the variance in grade retention across the different levels. Using the latent variable approach, we found that 53.8% of the variance in grade retention in our sample lies at the student level, 13.2% at the school level, and 33.1% at the country level. This indicates that a substantial share of the variation is located at the school and country levels, supporting the use of a multilevel approach.

When student-level covariates were included, several findings emerged. Boys have 79% higher odds of grade retention than girls (OR = 1.79, $SE = 0.02$, $p < .001$). First-generation migration students (OR = 1.59, $SE = 0.07$, $p < .001$) and second-generation migration students (OR = 1.19, $SE = 0.06$, $p = .005$) also have 59% resp. 19% higher odds of grade retention, compared to students with no immigration background. For ESCS, higher socioeconomic status is associated with lower odds of grade retention (OR = 0.69, SE = 0.02, p < .001). A one-unit increase in ESCS corresponds to 31% lower odds of grade retention; conversely, a one-unit decrease corresponds to about 45% higher odds ($1/0.69 \approx 1.45$). For academic achievement, a higher PISA test average is associated with lower odds of grade retention (OR = 0.86, SE = 0.002, p < .001). Because test scores were divided by 10, this OR reflects a 10-point difference: a 10-point increase corresponds to 14% lower odds; a 10-point decrease corresponds to about 16% higher odds ($1/0.86 \approx 1.16$), see also Table A4.

School Characteristics associated with Grade Retention across Countries

The results for the school-level models are presented in Table 5. The model including school composition variables (Table 5, Model 1) shows that a higher proportion of students with an immigration background is associated with higher odds of grade retention (OR = 1.21, $SE = 0.05$, $p < .001$). Similarly, a higher proportion of low-SES students is associated with higher odds of grade retention (OR = 1.29, $SE = 0.06$, $p < .001$). In contrast, a higher proportion of poor-performing students is associated with lower odds of grade retention (OR = 0.74, $SE = 0.06$, $p < .001$).

The school policy model (Table 5, Model 2) shows that within-class ability grouping is associated with lower odds of grade retention. Compared to schools that do not use within-class grouping, schools that use this strategy for some subjects (OR = 0.78, $SE =$

Table 5. Associations between school- and student-level characteristics, and lower-secondary grade retention.

		Model 1		Model 2		Model 3		Model 4	
		OR	*SE*	OR	*SE*	OR	*SE*	OR	*SE*
Intercept		0.0027***	0.31	0.0031***	0.31	0.0026***	0.33	0.0032***	0.31
SCHOOL COMPOSITION									
Proportion of students:									
Immigration background[abc]		1.21***	0.05					1.20**	0.06
Low-SES background[abc]		1.29***	0.06					1.24**	0.06
Poor-performing[abc]		0.74***	0.06					0.72***	0.06
SCHOOL POLICY									
Between-class ability grouping:	For some subjects (ref: none)			0.96	0.06			0.95	0.06
	For all subjects			1.05	0.10			1.05	0.10
Within-class ability grouping:	For some subjects (ref: none)			0.78***	0.06			0.76***	0.06
	For all subjects			0.84*	0.09			0.84*	0.08
LEARNING ENVIRONMENT									
Perceived competitiveness[a]						0.93	0.09	0.95	0.10
Perceived cooperativeness[a]						0.83	0.10	0.81	0.10
Disruptive student behaviour[a]						1.02	0.03	1.03	0.03
Disruptive teacher behaviour[a]						1.07*	0.03	1.06	0.03
Shortage of educational staff[a]						0.95	0.03	0.96	0.03
Shortage of educational materials[a]						1.07*	0.03	1.08**	0.03
STUDENT LEVEL									
ESCS[a]		0.70***	0.02	0.69***	0.02	0.70***	0.02	0.70***	0.02
Gender:	Boy (ref: girl)	1.80***	0.03	1.79***	0.03	1.79***	0.03	1.79***	0.03
Immigration background:	Second generation (ref: native)	1.14*	0.06	1.20**	0.06	1.19**	0.06	1.14*	0.06
	First generation	1.52***	0.07	1.60***	0.07	1.58***	0.07	1.52***	0.07
Relative age[a]		1.00	0.004	1.00	0.004	1.00	0.004	1.00	0.004
PISA test average[ab]		0.86***	0.002	0.86***	0.002	0.86***	0.002	0.86***	0.002

ESCS: economic, social, and cultural status; ref: reference category; SE: standard error;[a]: grand mean centred;[b]: transformed by dividing by 10;[c]: log-transformed; * $p<.05$; ** $p<.01$; *** $p<.001$.

0.06, $p < .001$) or all subjects (OR = 0.84, $SE = 0.09$, $p = .04$) show lower odds of grade retention. Between-class ability grouping is not significantly associated with grade retention.

The learning environment model (Table 5, Model 3) indicates that higher levels of disruptive teacher behaviour (OR = 1.07, $SE = 0.03$, $p = .04$) and shortages of educational materials (OR = 1.07, $SE = 0.03$, $p = .01$) are associated with higher odds of grade retention. Other learning environment variables are not significantly related to grade retention.

When all school-level variables are included in the full fixed-effects model simultaneously, several associations remain significant. Higher proportions of students with an immigration background (OR = 1.20, $SE = 0.05$, $p = .001$) and low-SES background (OR = 1.24, $SE = 0.06$, $p < .001$), as well as lower proportions of poor-performing students (OR = 0.72, $SE = 0.06$, $p < .001$), remain associated with grade retention. Within-class ability grouping (for some subjects: OR = 0.76, $SE = 0.06$, $p < .001$; for all subjects: OR = 0.84, $SE = 0.08$, $p = .03$) continues to be associated with lower odds of grade retention. Shortage of educational materials (OR = 1.08, $SE = 0.03$, $p = .007$) also remains significantly associated with grade retention.

Cross-Level Interactions Between School Characteristics and the Degree of Stratification

In the interaction model (Table 6), the degree of institutional stratification itself is not significantly associated with grade retention while controlling for student- and school-level characteristics. The main school-level associations are largely comparable to those found in Model 4. Two cross-level interactions are statistically significant. First, the association between within-class ability grouping (for some subjects) and grade retention differs by stratification level. In low-stratification systems, within-class grouping (for some subjects) is associated with lower odds of grade retention (OR = 0.82, $SE = 0.07$, $p = .005$). In high-stratification systems, this association is stronger, with lower odds of grade retention in schools using within-class ability grouping compared to schools not using it (OR = 0.62). Second, the association between shortage of educational materials and grade retention also varies by stratification. In low-stratification systems, a higher shortage of educational materials is associated with higher odds of grade retention (OR = 1.11, $SE = 0.03$, $p = .001$). In high-stratification systems, this association is not statistically significant (OR = 0.97), as indicated by the significant interaction term ($OR_{interaction}$=0.87, $SE = 0.07$, $p = .04$). Other cross-level interactions are not statistically significant.

Discussion

The aim of this study was to identify school-level characteristics associated with students' probability of being retained in lower-secondary education, both across and between educational systems, using a cross-country multilevel approach. Drawing on PISA 2018 data from 23 countries, constituting 24 educational systems, we estimated three-level logistic regression models to examine how student, school, and country characteristics are associated with grade retention. By combining school-level fixed effects and a cross-level interaction between educational systems and school characteristics, we were able to distinguish between school characteristics with a broadly consistent

Table 6. Interaction model for the degree of stratification of the educational system and school-level characteristics.

		Model 5	
		OR	*SE*
Intercept		0.0032***	0.31
EDUCATION SYSTEM			
Stratification		1.09	0.63
SCHOOL COMPOSITION			
Proportion of students:			
Immigration background[abc]		1.19**	0.06
Low-SES background[abc]		1.31***	0.07
Poor-performing[abc]		0.74***	0.08
Immigration background*stratification		1.04	0.13
Low-SES background*stratification		0.76	0.15
Poor-performing*stratification		1.04	0.13
SCHOOL POLICY			
Between-class ability grouping:	For some subjects (ref: none)	0.95	0.07
	For all subjects	1.16	0.11
Within- class ability grouping:	For some subjects (ref: none)	0.82**	0.07
	For all subjects	0.87	0.09
Between-class*stratification:	For some subjects (ref: none)	0.98	0.16
	For all subjects	0.66	0.23
Within-class*stratification:	For some subjects (ref: none)	0.75*	0.14
	For all subjects	0.89	0.21
LEARNING ENVIRONMENT			
Perceived competitiveness[a]		1.03	0.10
Perceived cooperativeness[a]		0.85	0.12
Disruptive student behaviour[a]		1.04	0.04
Disruptive teacher behaviour[a]		1.03	0.04
Shortage of educational staff[a]		0.94	0.04
Shortage of educational materials[a]		1.11**	0.03
Perceived competitiveness*stratification		0.75	0.25
Perceived cooperativeness*stratification		0.88	0.20
Disruptive student behaviour*stratification		0.94	0.07
Disruptive teacher behaviour*stratification		1.11	0.08
Shortage of educational staff*stratification		1.06	0.08
Shortage of educational materials*stratification		0.87*	0.07
STUDENT LEVEL			
ESCS[a]		0.70***	0.02
Gender:	Boy (ref: girl)	1.80***	0.03
Immigration background:	Second generation (ref: native)	1.15*	0.06
	First generation	1.53***	0.07
Relative age[a]		1.00	0.004
PISA test average[ab]		0.86***	0.002

ESCS: economic, social, and cultural status; ref: reference category; SE: standard error;[a]: grand mean centred;[b]: transformed by dividing by 10;[c]: log-transformed;
* $p<.05$; ** $p<.01$; *** $p<.001$.

predictive value across countries and those that are associated with students' grade retention risk to a varying extent, depending on the degree of stratification in their educational system.

Main Findings and Interpretation

Our study has four central findings. First, grade retention seems to be shaped by more than just individual characteristics: 13.2% of the variance in grade retention lies at the school level, and 33.1% at the country level. This indicates that grade retention does not only reflect individual student performance but is also associated with broader

institutional and contextual environments (Dupriez et al., 2008; Goos et al., 2013). Understanding these contextual effects is important because they might point to levers for interventions beyond the student level.

Second, several school compositional characteristics seem consistently linked to students' likelihood of grade retention across countries, even after controlling for individual student-level risk factors. Schools with a higher proportion of migrant or low-SES students seem to use grade retention more often, confirming the relevance of school composition effects observed in earlier educational effectiveness research (Belfi et al., 2012; Coleman et al., 1966; Demanet & Van Houtte, 2013; Reynolds et al., 2014; Thrupp et al., 2002). In line with our theoretical framework, this may reflect how school composition shapes performance norms, expectations, and instructional conditions, which in turn relate to retention decisions.

Contrary to our hypothesis, schools with higher proportions of poor-performing students show lower odds of grade retention compared to schools with fewer low-performing students. This finding does not align with our initial expectation that a concentration of low achievement would be associated with higher retention risk through lower expectations or instructional quality. One possible explanation is that retention decisions are shaped by internal performance norms within schools (Alexander et al., 2003; Dupriez et al., 2008). In higher-achieving schools, lower-performing students may stand out more clearly relative to their peers, which may increase the likelihood that teachers judge their performance as insufficient. In contrast, in schools with many low-performing students, promotion standards may be calibrated differently, and large-scale retention may be less feasible or less aligned with school practice (Rumberger, 1995). This suggests that school composition may influence how similar levels of achievement are evaluated, rather than how achievement itself develops.

Third, school policy appears to matter. Students attending schools that use within-class ability grouping show lower odds of grade retention. This finding is consistent with research suggesting that differentiated instruction within classrooms can support student learning outcomes (Hattie, 2023a), which may reduce the likelihood that retention is considered necessary. Moreover, this association is stronger in highly stratified systems. In highly stratified systems, where schools differ more strongly in their student intake (Marks, 2005; Pfeffer, 2008), within-class differentiation may further refine instruction within these more homogeneous school contexts. This may strengthen the association between instructional practices and retention decisions. Importantly, stratification does not uniformly amplify all school-level associations; rather, it shapes which organisational mechanisms become more or less relevant in different institutional contexts.

Last, regarding the learning environment, shortages of educational materials are associated with higher odds of grade retention, but this association is only observed in low-stratification systems. This finding is consistent with research suggesting that insufficient instructional resources can hinder learning processes (Hanushek & Woessmann, 2016), which may increase the likelihood that retention is considered. In more comprehensive systems, where formal tracking mechanisms are limited, schools may rely more strongly on retention when instructional support is constrained. In contrast, in highly stratified systems, other structural mechanisms may partly absorb these pressures.

Together, these findings may help explain why earlier studies of school-level predictors of grade retention have produced mixed results (Choi et al., 2018; Ferrão, 2015;

Guèvremont et al., 2007). Without considering the broader institutional context, school-level associations may appear inconsistent across countries. These findings suggest that grade retention operates not merely as an individual sanction, but as an organisational mechanism shaped by institutional design

Implications for Educational Practitioners and Policy-makers

For school leaders aiming to reduce grade retention, our findings suggest that school organisation and instructional practices are relevant areas of attention. In particular, within-class differentiation is consistently associated with lower odds of grade retention. From a school effectiveness perspective, this suggests that retention is not merely an outcome of student performance, but also a reflection of organisational routines and decision-making processes within schools.

For policymakers, our study highlights the importance of the institutional context. Efforts to reduce grade retention should take into account national promotion rules, tracking structures, and the availability of support mechanisms. Our findings illustrate the risks of policy borrowing without adaptation: interventions that are associated with lower retention in one type of system may not operate in the same way in another. International comparisons can support more targeted interventions by identifying which school-level characteristics show broadly consistent associations and which are context-dependent. In particular, the moderating role of stratification suggests that school-level practices are embedded in system-level structures, and that effective school improvement strategies require alignment between classroom practices and institutional design. In addition, our study once more highlights the significance of school composition for educational opportunities.

Strengths and limitations

This study's main strength lies in its integration of individual- and school-level predictors of grade retention across multiple countries, addressing an important gap in grade retention research. By using three-level logistic regression models with cross-level interactions between school and system characteristics, we were able to distinguish between associations that appear broadly consistent and those that vary across institutional contexts.

However, several limitations should be acknowledged. First, PISA 2018 is based on a cross-sectional, observational design. We cannot examine how changes in school composition, school policy, or learning environment affect grade retention over time. As a result, the findings should be interpreted as associations rather than causal effects. Longitudinal or quasi-experimental designs would allow stronger causal inference. Second, grade retention and several school-level variables are based on self-reported questionnaire data, which may introduce measurement error or differences in interpretation across contexts. Third, our operationalisation of stratification relies on a dichotomous classification, which may not fully capture variation in institutional complexity across systems. More fine-grained measures of stratification could provide additional insight. Fourth, our focus on student-, school-, and country-level factors does not include classroom- and teacher-level processes, even though such processes likely play a central role in retention

decisions (Creemers & Kyriakides, 2007). Future research could examine whether school-level characteristics influence grade retention indirectly through student achievement or motivational processes, for example using mediation models or linked longitudinal administrative data. Such approaches would help to better understand the mechanisms underlying retention decisions.

Conclusion

Grade retention in lower-secondary education is not only a reflection of individual student disadvantage. It is also shaped by school composition, organisational practices, school culture, and national educational contexts. Certain school characteristics – in particular student composition –are consistently linked to grade retention, across all countries. Other school characteristics – such as resource shortages or within-class ability grouping – operate differently depending on the stratification of the national system.

By combining cross-country evidence with school-level insights, this study provides a more nuanced understanding of where and how interventions aimed at reducing grade retention rates can be most effective. Reducing unnecessary grade retention could improve educational equity over time. This requires strategies that are both school-informed and system-aware, ensuring that local practices align with the broader educational framework in which they are embedded.

Acknowledgements

This work was supported by the Erasmus + KA2 2001 programme under Grant 2021-1-BE02-KA220-HED-000022966; European Association for Researching on Learning and Instruction (EARLI vzw) to the EARLI centre for excellence in grade retention research.

Declaration of Interest

The authors have nothing to declare.

ORCID

Janneke Pepels http://orcid.org/0000-0001-7671-0939
Barbara Belfi http://orcid.org/0000-0003-2145-4311
Mieke Goos http://orcid.org/0000-0003-0966-7410
Carla Haelermans http://orcid.org/0000-0002-9202-8427

Data availability statement

The data that support the findings of this study are openly available on the OECD website, at https://www.oecd.org/en/data/datasets/pisa-2018-database.html#data.

Biographical notes

Janneke Pepels (MScRes) is a PhD candidate at the Research Centre for Education and the Labour Market (ROA) at Maastricht University. Her research focuses on educational effectiveness, classroom heterogeneity, and educational inequality.

Dr. Barbara Belfi is a senior researcher at the Research Centre for Education and the Labour Market (ROA) at Maastricht University. Her research focuses on socioemotional development and the influence of school contexts.

Prof. Dr. Mieke Goos is assistant professor at the School of Educational Studies at Hasselt University. Her research focuses on educational effectiveness, especially regarding inequality topics in primary education.

Prof. Dr. Carla Haelermans is a professor of Education Economics at the Research Centre for Education and the Labour Market (ROA) at Maastricht University. Her research focuses on educational effectiveness, digitalisation in education, and inequality.

References

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (2003). *On the success of failure: A reassessment of the effects of retention in the primary school grades*. Cambridge University Press.

Amineh, R. J., & Asl, H. D. (2015). Review of constructivism and social constructivism. *Journal of Social Sciences, Literature and Languages*, *1*(1), 9–16.

Aydin, Ö., & Ok, A. (2022). A Systematic Review on Teacher's Expectations and Classroom Behaviors. *International Journal of Curriculum and Instructional Studies*, *12*(1), 247–274.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. Ben. (2015). Package 'lme4.' *Convergence*, *12*(1), 2.

Belfi, B., Goos, M., De Fraine, B., & Van Damme, J. (2012). The effect of class composition by gender and ability on secondary school students' school well-being and academic self-concept: A literature review. *Educational Research Review*, *7*(1), 62–74.

Browne, W. J., Subramanian, S. V, Jones, K., & Goldstein, H. (2005). Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *168*(3), 599–613.

Bulle, N. (2019). Democratization of educational systems, inequality, opportunity, and selection process: a re-examination of the case of France. *School Effectiveness and School Improvement*, *30*(4), 432–454.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, *64*(8), 1–9.

Carroll, J. B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher*, *18*(1), 26–31.

Chau, K., Baumann, M., Kabuth, B., & Chau, N. (2012). School difficulties in immigrant adolescent students and roles of socioeconomic factors, unhealthy behaviours, and physical and mental health. *BMC Public Health*, *12*(1). https://doi.org/10.1186/1471-2458-12-453

Choi, Á., Gil, M., Mediavilla, M., & Valbuena, J. (2018). Predictores y Efectos de la Repeticion de Curso [Predictors and effects of grade repetition]. *Revista de Economia Mundial*, 2018(48), 21–42.

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). The coleman report. *Equality of Educational Opportunity*, *3*(18), 1–32.

Cordero Ferrera, J. M., Manchón López, C., & Simancas Rodríguez, R. (2014). La repetición de curso y sus factores condicionantes en España [Repetition and explanatory factors in Spain]. *Revita de Educación*, *365*(Julio-septiembre), 12–37. https://doi.org/DOI:10.44381988-592X-RE-2014-365-263

Cortázar, A., Molina, M. de los Á., Sélman, J., & Manosalva, A. (2020). Early childhood education effects on school outcomes: Academic achievement, grade retention and school drop Out. *Early Education and Development*, *31*(3), 376–394.

Creemers, B., & Kyriakides, L. (2007). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. Routledge.

Demanet, J., & Van Houtte, M. (2013). Grade retention and its association with school misconduct in adolescence: A multilevel approach. *School Effectiveness and School Improvement*, *24*(4), 417–434.

Dupriez, V., Dumay, X., & Vause, A. (2008). How do school systems manage pupils' heterogeneity? *Comparative Education Review*, *52*(2), 245–273. https://doi.org/10.1086/528764

Enders, C. K. (2022). *Applied missing data analysis*. Guilford Publications.

European Commision;, EACEA;, & Eurydice. (2020). *Equity in school education in Europe: Structures, policies and student performance*. Publications Office of the European Union.

European Commission: European Education and Culture Executive Agency. (2011). *Grade Retention during Compulsory Education in Europe*. Publications Office. https://data.europa.eu/doi/10.2797/50570

Ferrão, M. E. (2015). Tópicos sobre retenção escolar em Portugal através do PISA: qualidade e equidade [Topics on school retention in Portugal through PISA: quality and equity]. *Education Policy Analysis Archives*, *23*, 114.

Francis, B., Archer, L., Hodgen, J., Pepper, D., Taylor, B., & Travers, M.-C. (2017). Exploring the relative lack of impact of research on 'ability grouping'in England: A discourse analytic account. *Cambridge Journal of Education*, *47*(1), 1–17.

Goos, M., Pipa, J., & Peixoto, F. (2021). Effectiveness of grade retention: A systematic review and meta-analysis. *Educational Research Review*, *34*(March 2020), 100401. https://doi.org/10.1016/j.edurev.2021.100401

Goos, M., Schreier, B. M., Knipprath, H. M. E., De Fraine, B., Van Damme, J., & Trautwein, U. (2013). How can cross-country differences in the practice of grade retention be explained? A closer look at national educational policy factors. *Comparative Education Review*, *57*(1), 54–84. https://doi.org/10.1086/667655

Grund, S., Robitzsch, A., Luedtke, O., & Grund, M. S. (2019). *Package 'mitml*.' R Studio.

Guèvremont, A., Roos, N. P., & Brownell, M. (2007). Predictors and Consequences of Grade Retention: Examining Data From Manitoba, Canada. *Canadian Journal of School Psychology*, *22*(1), 50–67. https://doi.org/10.1177/0829573507301038

Hanushek, E. A., & Woessmann, L. (2016). School resources and student achievement: A review of cross-country economic research. *Cognitive Abilities and Educational Outcomes: A Festschrift in Honour of Jan-Eric Gustafsson*, 149–171.

Hattie, J. (2023a). Classroom effects. In *Visible learning: The sequel: A synthesis of over 2,100 meta-analyses relating to achievement* (pp. 182–212). Routledge.

Hattie, J. (2023b). The teacher. In *Visible learning: The sequel: A synthesis of over 2,100 meta-analyses relating to achievement* (pp. 213–246). Routledge.

Klapproth, F., & Schaltz, P. (2015). Who is retained in school, and when? Survival analysis of predictors of grade retention in Luxembourgish secondary school. *European Journal of Psychology of Education*, *30*(1), 119–136. https://doi.org/10.1007/s10212-014-0232-7

Kornfeld, M., & Ochsen, C. (2017). Student assessment and grade retention: evidence from a natural experiment. *Education Economics*, *25*(2), 127–141. https://doi.org/10.1080/09645292.2016.1199660

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, *66*(4), 423–458.

Lubbers, M. J., Van Der Werf, M. P. C., Snijders, T. A. B., Creemers, B. P. M., & Kuyper, H. (2006). The impact of peer relations on academic progress in junior high. *Journal of School Psychology*, *44*(6), 491–512.

Marks, G. N. (2005). Cross-national differences and accounting for social class inequalities in education. *International Sociology*, *20*(4), 483–505.

OECD. (1998). *Reviews of National Policies for Education: Korea*. OECD Publishing.

OECD. (2019). *PISA 2018 Technical Report*. OECD Publishing.

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

OECD. (2025). *Education GPS*. http://gpseducation.oecd.org/

Pfeffer, F. T. (2008). Persistent inequality in educational attainment and its institutional context. *European Sociological Review*, *24*(5), 543–565.

Pipa, J., Pepels, J., Gaitas, S., Goos, M., Peixoto, F., Belfi, B., Meissner, F., & Klapproth, F. (2025). *Summary of predictors and effects of grade retention in Europe*. Maastricht University Press. https://doi.org/10.26481/mup.rep.alter.2502.en

Quartagno, M., Grund, S., & Carpenter, J. (2019). Jomo: a flexible package for two-level joint modelling multiple imputation. *R Journal*, *9*(1).

Resino, D. A., Amores, I. A. C., & Muñoz, I. A. (2019). La repetición de curso a debate: un estudio empírico a partir de PISA 2015 [Debate on grade repetition: An empirical study from pisa 2015]. *Educacion XX1*, *22*(2), 69–92. https://doi.org/10.5944/educxx1.22479

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement*, *25*(2), 197–230.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, *3*(1), 16–20.

Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal*, *32*(3), 583–625.

Shi, Q., & Brown, M. H. (2020). School Counselors' Impact on School-Level Academic Outcomes: Caseload and Use of Time. *Professional School Counseling*, *23*(1_part_3), 2156759X2090448. https://doi.org/10.1177/2156759×20904489

Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research*, *60*(3), 471–499.

Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research*, *37*(5), 483–504.

Valbuena, J., Mediavilla, M., Choi, Á., & Gil, M. (2021). Effects of Grade Retention Policies: a Literature Review of Empirical Studies Applying Causal Inference. *Journal of Economic Surveys*, *35*(2), 408–451. https://doi.org/10.1111/joes.12406

Vinas-Forcade, J., Mels, C., Van Houtte, M., Valcke, M., & Derluyn, I. (2021). Can failure be prevented? Using longitudinal data to identify at-risk students upon entering secondary school. *British Educational Research Journal*, *47*(1), 205–225. https://doi.org/10.1002/berj.3683

Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, *24*(3–5), 124–179.

Westphal, A., Vock, M., & Lazarides, R. (2020). Are more conscientious seventh- and ninth-graders less likely to be retained? Effects of Big Five personality traits on grade retention in two different age cohorts. *Journal of Applied Developmental Psychology*, *66*(October 2019), 101088. https://doi.org/10.1016/j.appdev.2019.101088

Zapfe, L., & Gross, C. (2021). How do characteristics of educational systems shape educational inequalities? Results from a systematic review. *International Journal of Educational Research*, *109*, 101837.

Appendix

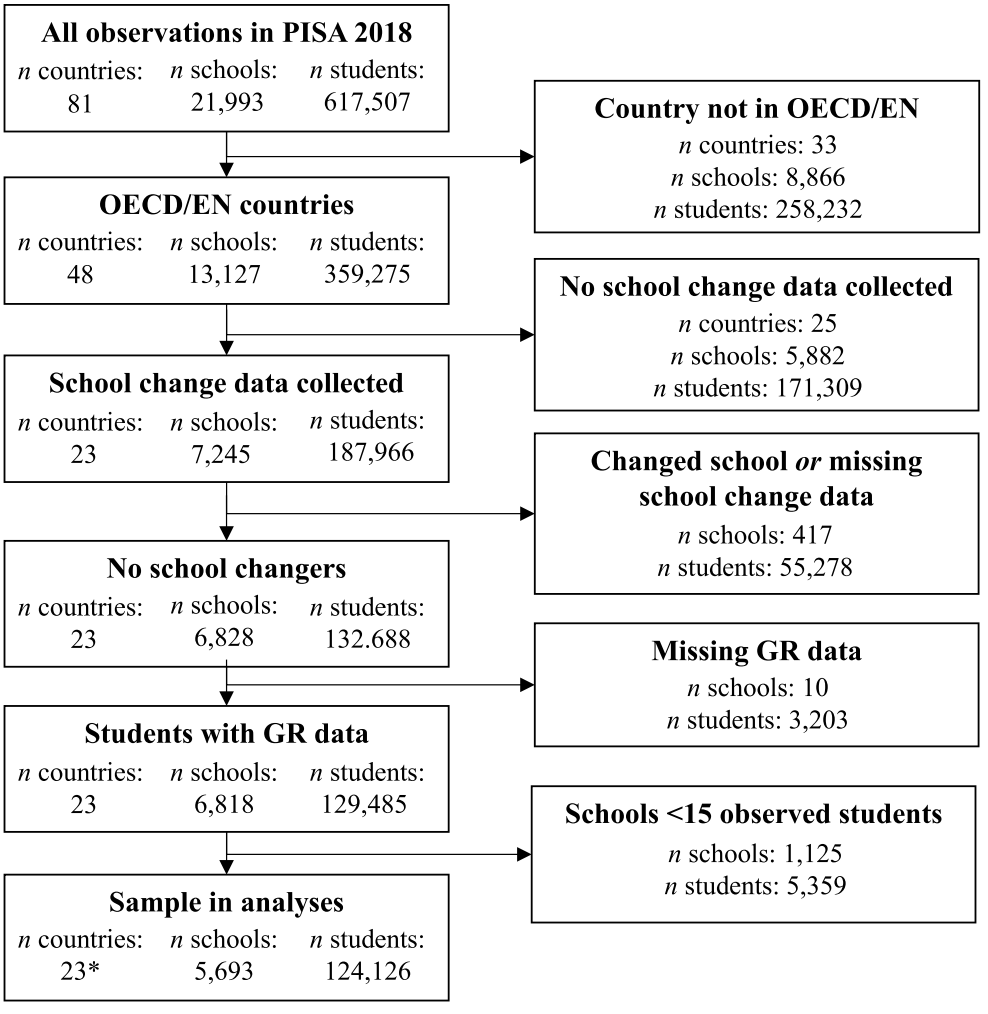

Figure A1. Flowchart of the study population.

*The 24 countries reported in the final sample includes the two distinct educational systems in Belgium, therefore the 23 countries mentioned here are 24 educational systems.

Table A1. Descriptive statistics and comparison of student and school characteristics for the countries in the final sample, and OECD/EN and non-OECD/EN countries that were excluded.

	Countries in final sample	Excluded OECD/EN countries		Non-OECD/EN countries	
	N (%) / mean(*SD*)	*n*(%) / mean(*SD*)	p-value	*N* (%) / mean(*SD*)	p-value
STUDENT LEVEL	***n*=190,964**	***n*=168,311**		***n*=258,232**	
Lower-secondary grade retention	11,716 (6.1)	7,258 (4.3)	<.001	17.908 (6.9)	<.001
Gender[a]					
Female	94,622 (49.5)	83,679 (49.7)	.31	129,343 (50.1)	<.001
Male	96,342 (50.5)	84,630 (50.3)		128,889 (49.9)	
Missing	-	2 (0.0)		-	
Age[b]	15.78 (0.29)	15.80 (0.29)	<.001	15.80 (0.29)	<.001
ESCS[b]	-0.04 (0.98)	-0.10 (1.06)	<.001	-0.55 (1.18)	<.001
Immigration background[a]					
Native	162,382 (85.0)	133,587 (79.4)	<.001	216,418 (83.8)	<.001
Second-generation	11,306 (5.9)	11,701 (7.0)		13,346 (5.2)	
First-generation	9,639 (5.0)	8,826 (5.2)		17,580 (6.8)	
Missing	7,637 (4.0)	14,197 (8.4)		10,888 (4.2)	
PISA test average[b]	481.94 (90.54)	474.30 (95.89)	<.001	431.84 (102.09)	<.001
SCHOOL LEVEL	***n*=7,355**	***n*=5,772**		***n*=8,866**	
Students with immigration background[b]	12.26 (17.98)	12.60 (19.95)	.32	10.47 (22.28)	<.001
Students with low-SES background[b]	28.29 (21.74)	27.76 (21.33)	.17	28.62 (23.17)	0.36
Poor-performing students[b]	16.43 (20.48)	18.08 (22.20)	<.001	37.78 (31.18)	<.001
Between-class ability grouping:[a]					
Not for any subjects	430 (5.8)	548 (9.5)	<.001	1,310 (14.8)	<.001
For some subjects	2,418 (32.9)	1,668 (28.9)		2,205 (24.9)	
For all subjects	3,811 (51.8)	3,164 (54.8)		4,414 (49.8)	
Missing	696 (9.5)	392 (6.8)		937 (10.6)	
Within-class ability grouping:[a]					
Not for any subjects	708 (9.6)	355 (6.2)	<.001	1,538 (17.3)	<.001
For some subjects	3,119 (42.4)	2,373 (41.1)		3,124 (35.2)	
For all subjects	2,952 (40.1)	2,738 (47.4)		3,435 (38.7)	
Missing	576 (7.8)	306 (5.3)		769 (8.7)	
Perception of competitiveness[b]	0.03 (0.44)	-0.01 (0.43)	<.001	0.09 (0.43)	<.001
Perception of cooperativeness[b]	-0.04 (0.44)	-0.01 (0.44)	.001	0.03 (0.49)	<.001
Disruptive student behaviour[b]	-0.05 (1.09)	0.19 (1.00)	<.001	0.02 (1.46)	.001
Disruptive teacher behaviour[b]	-0.01 (1.01)	0.11 (0.99)	<.001	0.20 (1.35)	<.001
Shortage of educational staff[b]	-0.10 (1.01)	0.02 (1.02)	<.001	0.03 (1.20)	<.001
Shortage of educational materials[b]	0.09 (1.02)	0.01 (1.03)	<.001	0.22 (1.18)	<.001

ESCS: economic, social, and cultural status;[a] *n*(%);[b] mean(*SD*); differences between the countries in final sample and the excluded OECD/EN and non-OECD/EN countries were tested with Mann-Whitney tests for categorical data and t-tests for continuous data.

Table A2. Correlation matrix of school-level variables.

		1	2	3	4	5	6	7	8	9	10
1	Proportion of immigration background	-									
2	Proportion of low-SES background	0.15	-								
3	Proportion of poor-performing students	0.09	0.58	-							
4	Perceived competitiveness	-0.07	-0.10	-0.01	-						
5	Perceived cooperativeness	-0.11	-0.28	-0.28	0.24	-					
6	Disruptive student behaviour	0.16	0.37	0.34	-0.10	-0.27	-				
7	Disruptive teacher behaviour	0.19	0.11	0.04	-0.07	-0.13	0.50	-			
8	Shortage of educational materials	-0.04	0.17	0.15	-0.12	-0.13	0.25	0.19	-		
9	Shortage of educational staff	0.15	0.15	0.09	-0.15	-0.17	0.28	0.36	0.45	-	
10	Ability grouping between classes	-0.12	0.02	-0.01	-0.23	-0.03	-0.03	-0.03	0.08	0.06	-
11	Ability grouping within classes	0.02	-0.03	-0.05	-0.14	-0.05	-0.03	-0.02	-0.02	0.00	0.27

Table A3. Descriptive statistics and comparison of student- and school-level characteristics for low- and high-stratification educational systems.

	Low stratification countries (n=14)			High stratification countries (n=10)			
	n (%)	mean (*SD*)	min;max	*n* (%)	mean (*SD*)	min;max	*p*-value
STUDENT LEVEL	**n=79,596**			**n=44,530**			
Lower-secondary grade retention	5,694 (7.2)			779 (1.7)			<.001
Gender							
Female	40,276 (50.6)			22,829 (51.3)			.02
Male	39,320 (49.4)			21,701 (48.7)			
Age		15.78 (0.29)	15.08;16.33		15.79 (0.29)	15.25;16.33	<.001
ESCS		0.04 (1.00)	-6.16;3.96		-0.08 (0.86)	-6.36;3.75	<.001
Immigration background							
Native	69,228 (87.0)			41,082 (92.3)			<.001
Second-generation	5,329 (6.7)			2,187 (4.9)			
First-generation	3,882 (4.9)			877 (2.0)			
Missing	1,157 (1.5)			384 (0.9)			
PISA test average		494.24 (87.06)	168.24; 832.02		496.52 (86.29)	168.15; 798.39	<.001
SCHOOL LEVEL		n=3,641			n=2,052		
Students with immigration background		14.94 (18.44)	0;100		8.12 (11.91)	0;87.10	<.001
Students with low-SES background		25.29 (17.60)	0;94.44		24.70 (17.65)	0;99.67	0.23
Poor-performing students		13.31 (14.18)	0;100		14.20 (19.00)	0;100	0.07
Between-class ability grouping:							
Not for any subjects	187 (5.1)			168 (8.2)			<.001
For some subjects	1,461 (40.1)			418 (20.4)			
For all subjects	1,690 (46.4)			1,295 (63.1)			
Missing	303 (8.3)			171 (8.3)			
Within-class ability grouping:							
Not for any subjects	368 (10.1)			181 (8.8)			0.68
For some subjects	1,508 (41.4)			902 (44.0)			
For all subjects	1,486 (40.8)			848 (41.3)			
Missing	279 (7.7)			121 (5.9)			
Perception of competitiveness		0.11 (0.38)	-1.99;2.04		-0.15 (0.32)	-1.99;2.04	<.001
Perception of cooperativeness		0.01 (0.37)	-1.72;1.68		-0.12 (0.38)	-2.14;1.68	<.001
Disruptive student behaviour		-0.07 (1.06)	-3.38;3.44		0.10 (1.07)	-4.09;3.44	<.001
Disruptive teacher behaviour		0.03 (0.98)	-2.09;3.81		0.11 (1.00)	-2.04;3.79	.004
Shortage of educational staff		-0.02 (1.02)	-1.86;4.04		-0.18 (1.00)	-1.4731;3.10	<.001
Shortage of educational materials		0.04 (1.05)	-1.42;2.96		0.18 (0.95)	-1.42;2.96	<.001

ESCS: economic, social, and cultural status; differences between low- vs. high-stratification systems were tested with Mann-Whitney tests for categorical data and t-tests for continuous data.

Table A4. Fixed effects estimates for the null model and student-level model for lower-secondary grade retention.

		Null model		Student model	
		OR	*SE*	OR	*SE*
Intercept		0.0087***	0.30	0.0026***	0.32
STUDENT LEVEL					
ESCS [a]				0.69***	0.02
Gender	Boy (ref: girl)			1.79***	0.03
Immigration background	Second generation (ref: native)			1.19**	0.06
	First generation			1.59***	0.07
Relative age [a]				1.00	0.004
PISA test average[ab]				0.86***	0.002

ESCS: economic, social, and cultural status; *SE*: standard error; [a]: grand mean centred;
[b]: transformed by dividing by 10. * $p<.05$; ** $p<.01$; *** $p<.001$.