Gender gaps in school grades:

An analysis on resistance behaviour and class compositions in England, Germany, the Netherlands, and Sweden

Isabel J. Raabe[*1] & Isa Steinmann[*2]

***Preprint*[3]**

Abstract

Research has highlighted boys' disadvantages in school grades. School resistance behaviour, i.e. non-adherence to school norms and rules, rejection of authority, and defiance of educational goals through lack of effort, is more frequently observed among boys. We argue that such behaviour partly explains gender differences in school grades and that class-level gender composition may also play a role. Drawing on large-scale data from four European countries (n = 18,234 students in 886 classes), we examine gender gaps in language and mathematics grades. Controlling for cognitive and language ability and demographic characteristics, we find that gender gaps do not consistently favour girls; boys hold an advantage in mathematics. School resistance behaviour partly explains the gender gap in language grades (where girls are advantaged) but not in mathematics. Moreover, resistance behaviour is more strongly penalised in girls. We find no evidence that class gender composition relates to individual grades.

[1] Department of Sociology, University of Zurich

[2] Department of Primary and Secondary Teacher Education, Oslo Metropolitan University

[*] joint first authorship

Corresponding author email address: isabel.raabe@uzh.ch

[3] Please cite as: Raabe, Isabel J., and Isa Steinmann. 2025. "Gender Gaps in School Grades: An Analysis on Resistance Behaviour and Class Compositions in England, Germany, the Netherlands, and Sweden." OSF. doi:10.17605/OSF.IO/78CY2.

Introduction

Comprehensive meta-analyses and large-scale studies have repeatedly shown that girls receive higher grades than boys, especially in language and social science subjects (O'Dea et al. 2018; OECD 2015; Voyer and Voyer 2014). This tendency of boys falling behind at school is mirrored in a variety of related outcomes. Boys are more likely to be delayed in the transition from preschool to primary school, as well as from primary to secondary school (Blossfeld et al. 2009; Hadjar et al. 2014). In stratified school systems with early tracking, boys are more often placed in lower school tracks (Caro et al. 2009; Klapproth et al. 2013), and in 2020, women were more likely than men to have completed a university degree in almost all countries (Encinas-Martín and Cherian 2023).

Notably, girls do not generally outperform boys on standardised achievement (OECD 2023; Rosén, Steinmann, and Wernersson 2021) or cognitive ability tests (Hyde 2014; Zell, Krizan, and Teeter 2015) across subject domains. While girls score higher than boys in reading in almost all countries, as shown by international large-scale assessment studies such as PISA, this picture does not apply to mathematics and science, where results are more mixed (OECD 2015; Rosén et al. 2021). Furthermore, grade advantages of girls typically decrease only slightly if differences in standardised achievement tests are accounted for (Angelo and Reis 2020; Gortazar, Lafuente, and Vega-Bayo 2022; Liberto, Casula, and Pau 2021).

Even though school grades are the most important instruments to assess student success over their educational careers, as they are the basis for class assignment, year repetition, career recommendation, and access to higher education (Hallinan 1992), they have been found to only modestly correspond to abilities (Brookhart et al. 2016; Liberto et al. 2021). From a social inequality perspective, it is alarming that half of the population seems to be disadvantaged, if boys indeed received lower grades than warranted. It is thus vital to examine in detail how grades come about, and to understand where and how gender differences appear. Several explanations have been brought forward to explain this discrepancy between school grades and standardised test scores, although this dynamic is not fully understood yet (Gortazar et al. 2022; Spinath, Eckert, and Steinmayr 2014). While there are several studies documenting teacher biases in grading (Hinnerich, Höglin, and Johannesson 2011; Lavy 2008; Terrier 2020), the most universally assumed mechanisms to explain the grade advantage for girls are differences in non-cognitive skills that manifest in school-related behaviour (Bowers 2011; Spinath et al. 2014). It has been found that girls are more likely to show behaviours that are beneficial to learning, such as more engagement and motivation to learn (Duckworth and Seligman 2006; OECD 2015; Steinmayr and Spinath 2008). It is plausible that skills fostered by such beneficial school behaviours are picked up by grades to a larger extent than ability tests, as grades are based on assessment of lesson-related content, whereas standardised ability tests are geared towards testing more general skills. Furthermore, showing more engagement and motivation to learn and more desirable classroom behaviour can be directly rewarded with better grades, since grades measure more than cognitive learning outcomes (Bowers 2011).

In this paper, we focus on resistance to schooling, a set of behaviours that show a lack of adherence to school norms and rules, of accepting authority, and of defying educational goals by not putting effort into schoolwork (McFarland 2001). Previous research has shown that girls are less likely to exhibit resistance to schooling (Hadjar and Gross 2016; Legewie and DiPrete 2012). These types of behaviours disrupt the individual learning process, as those who regularly disturb lessons or skip entire classes, for instance, are less able to pick up contents from lessons and might be punished with lower grades to discourage resistance to schooling. Consequently, resistance behaviour is closely related to undesirable educational outcomes, such as dropout, as they disrupt the individual learning process and are punished with lower grades (Downey 2008; Fredricks, Blumenfeld, and Paris 2004).

Our analysis adds to existing research in two ways. First, by investigating gender gaps in grades in mathematics and language in a large-scale dataset of n = 18,234 students in 886 classes in four European countries, we can compare patterns between countries and school subjects. We investigate the role of resistance to schooling behaviour for gender gaps in grades, assuming that part of the gender gap in grades goes back to boys displaying more resistance to schooling behaviour. Furthermore, we investigate if there is differential punishment by gender: Do girls who display such behaviour receive lower grades than boys who display such behaviour? Importantly, we are able to control for differences in standardised achievement tests and demographic characteristics in these analyses.

Second, we consider the effect of the gender composition in the classroom on gender gaps in grades. We argue that learning disruptions by resistance behaviour is likely to not only affect the individual, but to extend to the entire classroom. First, if there is a significant number of students that regularly disturb lessons or argue with the teacher, this will also hamper the learning process of everyone else in the class, for example because noise levels are higher (Shield and Dockrell 2003). Second, peer norms fostering such behaviour are more likely to emerge in classes with a higher share of resistant students (Allen et al. 2005, 2014). As girls are less likely to exhibit resistance to schooling (Hadjar and Gross 2016; Legewie and DiPrete 2012), we propose that these dynamics are less likely to occur in classes with a higher share of girls. We investigate whether students have higher grades in classrooms with larger shares of girls, and whether this applies differently for boys and girls. This part of our analysis addresses the literature on gender compositions and their effects on individual educational outcomes, by testing a concrete mechanism by which compositional differences can affect individual outcomes.

What's in a Grade?

Grades are assigned either to individual pieces of work or as composite measures on term cards. Teachers are usually expected to assess the students' learning outcomes relative to curriculum-based standards, rather than relative to their classmates (i.e., grading on a curve). Assessments may include a variety of standardised and non-standardised oral or written sources and, in some cases, behavioural aspects related to learning, such as completing homework or participating in class (Eurydice 2010). Therefore, grades have been found to reflect a combination of cognitive and non-cognitive skills and behaviours (Borghans et al. 2016; Bowers 2011).

A first potential explanation for gender gaps in grades is that girls perform better in school overall. Notably, we refer here not to performance on standardised achievement tests (such as PISA), but to performance in classroom tasks such as tests developed by teachers, homework, and oral participation. The content of these tasks is based on lesson material and does not aim to measure general knowledge or ability. Since these tasks are grounded in classroom instruction, it is likely that non-cognitive factors such as diligence, effort, and self-discipline play a significant role.

Girls tend to show higher levels of motivation, self-concept, and academic engagement (Borgonovi, Ferrara, and Maghnouj 2018; Spinath et al. 2014). They also show, on average, greater self-discipline (Duckworth and Seligman 2006; Silverman 2003), attentiveness, and organisational skills (Farkas et al. 1990; Jacob 2002). Additionally, girls are more likely to pay attention in class and complete assignments (Demanet and Houtte 2012). Given these differences, it seems plausible that girls are more likely to perform high on tests focusing on lesson content. Indeed, gender differences in effort have been shown to partly explain gender differences in language grades (Kessels and Heyder 2017). In fact, past research has shown that teachers reward students' engagement and effort, as they seek to promote a pro-learning culture (Kelly 2008). Since grades partly reflect class participation, girls' stronger performance on these non-cognitive skills also directly contributes to their higher average grades. While beneficial non-cognitive skills may also play a role in standardised ability tests, the expected return is likely higher in tests that are directly based on classroom material.

Moreover, grades are awarded in the process of social referencing (Calsamiglia and Loviglio 2019). Even if teachers aim to compare students' performance to national curriculum standards, they are likely influenced (consciously or unconsciously) by the performance of students in their immediate classroom context (i.e., grading on a curve). How exactly teachers engage in this social referencing may vary by context and individual and could contribute to gender differences in grades. A recent study examined this mechanism in Sweden, which shifted from a relative to an absolute grading scale in the 1990s. Since then, teachers are expected to base grades on national standards rather than on classmates' performance. The study found that this reform narrowed the gap between boys and girls, possibly due to reduced "frog-pond" effects (Hjorth-Trolle, Rosenqvist, and Hed 2021).

Gender Differences in School Resistance Behaviour

This paper focuses on another type of behaviour that is likely to affect gender gaps in school grades: School resistance behaviour, which is an umbrella term for behaviours such as disturbing lessons, defying school norms and rules, not accepting authority, and not putting effort into schoolwork; exact measurements vary between individual studies (e.g. Geven, Jonsson, and Tubergen 2017; McFarland 2001). This is also known as behavioural disengagement (e.g. Reichenberg 2018), disruptive behaviour (e.g. Lynch, Kistner, and Allan 2014), school misconduct (e.g. Heyder, Hek, and Houtte 2021), or lack of self-regulation (e.g. McGowan et al. 2021), depending on the strand of literature.

A general finding in this literature is that there are strong gender differences in the extent to which school resistance behaviour is exhibited. Boys are found to be more disruptive in lessons, less diligent in schoolwork (Demanet and Houtte 2012), whereas girls are less likely to show disruptive behaviour (Kenney-Benson et al. 2006). Some of the literature on non-cognitive skills reviewed above supports this trend as well, such as the finding that girls showing higher levels of self-discipline and effort. In addition to these studies in the school context, boys are also more often diagnosed with attention deficit hyperactivity disorder (ADHD) (Ramtekkar et al. 2010), which could point to a general tendency for boys to have more trouble following school rules and focusing on schoolwork. This should be treated with caution, however, as it is unclear whether this gender difference in ADHD diagnoses is due to actual differences in occurrence, or whether there are simply more undiagnosed girls than boys, for instance.

There are two main mechanisms by which resistance behaviour can affect grades: Direct punishment in grading by teachers, as well as lower performance due to compromised learning. First, it might be the case that teachers assign lower grades to students because of their resistance behaviour (Ferman and Fontes 2022). In fact, in many countries, grades are supposed to not only reflect test outcomes, but also effort, diligence, and rule-following (Brookhart et al. 2016). Second, resistance behaviour is likely to disrupt the learning experience, resulting in lower performance. Students who do not put effort into their schoolwork, who are late, and who are not paying attention are not going to learn as much as those who do not show these behaviours (Zimmermann et al. 2013). As girls show these behaviours to a lesser extent, these differences in resistance behaviour are likely to explain at least part of the gender gap in school grades, as found in an older study with U.S. data (Downey and Vogt Yuan 2005). Additionally, a recent study has provided evidence for differential punishment of resistance behaviour in boys and girls. A small-scale study on preschool children in the United States (n = 324) found that boys with high levels of resistance behaviour were punished more than girls with high levels of resistance behaviour (Horn et al. 2021); in this paper, we follow up on this finding in a different context. We focus on dynamics within the classroom and ask:

1) *Within classrooms, who gets better grades, boys or girls?*
2) *Within classrooms, does resistance behaviour explain the gender gap in grades?*

We expect to find gender differences in grades (in favour of girls) as well as a negative association between resistance behaviour and grades. We expect resistance behaviour to affect the gender gap, as we assume that boys show higher levels of resistance behaviour. Furthermore, we test for differential punishment by gender.

Classroom Composition Effects

The second part of our analysis is concerned with resistance behaviour and gender at the classroom level. Previous research has shown that girls as well as boys have higher levels of educational achievement in school classes with a higher share of girls (Hoxby 2000; Lavy and Schlosser 2011; Legewie and DiPrete 2012).

There are two ways in which the gender composition could affect individual school grades: through the disruption of the learning environment, and through the creation of unfavourable peer norms. First, general learning theories argue that student learning is affected by the environment in which it takes place (Vygotsky and Cole 1978). Instruction and hence learning can only function efficiently and fruitfully when it occurs in an orderly context, that is, in a classroom with low levels of noise and disruptions, a general adherence to school rules and acceptance of teacher authority, and a respectful behaviour among students (Sortkær and Reimer 2018). Individual studies (Figlio 2007; Marks 2010) as well as meta-analyses have shown the relationship between the classroom disciplinary climate and individual achievement (Hattie 2009; Wang, Haertel, and Walberg 1993). This pattern has also been identified in the 2009 (Ning et al. 2015) and the 2012 (Sortkær and Reimer 2018) PISA data. Boys show on average higher levels of school resistance behaviour, so we can plausibly assume that in classes with a higher share of girls, average levels of resistance behaviour are lower. Hence, learning environment and quality of instruction may be higher in classrooms with a higher share of girls, which is beneficial to all students in the class, regardless of individual gender or behaviour. It is likely that effects of gender composition on individual grades vary by gender. Given the traits that are beneficial to learning and school success which we reviewed above, we propose that girls are more likely to succeed in difficult learning environments due to their higher motivation, diligence, and effort.

Second, peer and group norms can establish a dynamic in which resistance behaviour is desirable from the students' perspective, as it contributes to their social standing in the peer group (Allen et al. 2005, 2014). Furthermore, peers' resistance to schooling has been found to increase students' own resistance to schooling (Geven et al. 2017). Peer dynamics are likely to differ between boys and girls. For example, it has been found that boys experience a peer culture that is less oriented towards studying and academic engagement (Houtte 2004). Similarly, perceived pressure to conform to gender stereotypes in resistance behaviour was found for male but not female students (Heyder et al. 2021). Qualitative research on so-called "lad cultures" in the U.K. has documented similar tendencies (Jackson 2002, 2003; Morris

2012). Furthermore, as we generally observe high levels of gender homophily in friendship networks (McPherson, Smith-Lovin, and Cook 2001) social influence processes as documented by Geven, Jonsson, and van Tubergen (2017) mainly take place within genders, which contributes to the cementation of the gender gap in resistance behaviour.

Our third research question asks:

3) *Between classrooms, do students receive higher grades in classrooms with a higher share of girls?*

We expect that we see higher grades in classrooms with a higher share of girls, and that average levels of resistance behaviour help explain this effect. We also test whether the gender composition of the school class matters differently for boys and girls.

Methods

Data

We used data from the Children of Immigrants Longitudinal Study in Four European Countries (*CILS4EU*), a nationally representative survey that comprises information on students, parents, and teachers from England, Germany, the Netherlands, and Sweden (Kalter et al. 2014, 2015). The data was collected longitudinally between the school years 2010/11 and 2012/13. A stratified three-stage sampling design was used with schools as first, classes as second, and students as third stages. Schools with a large proportion of students with migration backgrounds were oversampled. Per school, one or two classes were sampled from the grade level with mostly 14-year-old students in 2010/11 (i.e., grade 10 in England, grade 9 in Germany, secondary grade 3 in the Netherlands, and grade 8 in Sweden). Within the classes, all students who were able to participate in the survey were sampled (i.e., exclusion of students with severe physical, intellectual, or language barriers; see the Technical Report (Kruse and Jacob 2014) for further information). We removed a total of 14 students with missing gender information from the samples (9 from England and 5 from the Netherlands) to acknowledge that the binary gender categorisation might not apply to them. Further, we removed 66 classes that contained fewer than 10 students from the entire analytical sample to ensure consistency across models. Composition effects in very small groups are likely not comparable due to the strong potential for outliers and different social dynamics, a concern well-documented in research on group-level effects (e.g., social network analysis). Applying this exclusion criterion uniformly prevents comparability issues and ensures a stable analytical sample. Other missing values were handled using multiple imputation techniques (see below). As depicted in Tables 1 and 2, the four country samples included between 4201 and 4954 students from between 198 and 245 classes. Note that the dependent variable information was available for between 2963 and 4021 students (see overview of missing cases in Appendix A).

Variables

The *CILS4EU* survey included translated, standardised tests and questionnaires for all four countries (Kalter et al. 2014). In the present study, we used data from the student tests and questionnaires. Table 1 depicts the variables' descriptive statistics after multiple imputation of missing values. Appendix A lists the shares of missing values and descriptive statistics before missing value imputation.

< Table 1 about here >

School Grades

The teacher-awarded school grades in mathematics and language subjects served as outcome variables. The school grades were self-reported by the students in England, Germany, and the Netherlands in 2011/12. In Sweden, they were obtained from the official school records. The school grades had different metrics across countries (see Table 1). We reverse-coded the school grades in Germany, so that higher values imply more favourable ratings in all countries.

Female Gender

The students' gender was one of the main predictor variables. In the 2010/11 questionnaires, the students were asked if they were a boy or a girl. As depicted in Table 1, the share of girls ranged between 49% and 51% in the country samples.

Resistance to Schooling

Students' resistance to schooling was another main predictor variable (see Table 1). Based on four items that were assessed in the student questionnaires, we constructed a resistance to schooling scale following previous literature (Geven et al. 2017). The items were (1) "How often do you argue with a teacher?"; (2) "How often do you skip a lesson without permission?"; (3) "How often do you come late to school?"; and (4) "How often do you get a punishment in school?". Students responded to these items on a 5-point scale (1 = *never*, 2 = *less often*, 3 = *once to several times per month*, 4 = *once to several times per week*, 5 = *every day*).

To assess the factorial validity of the resistance behaviour scale, we conducted a confirmatory factor analysis (CFA) using maximum likelihood estimation with robust standard errors (MLR) in lavaan (R version 4.4.1). Details on validity and reliability of the extracted factor can be found in Appendix C. The resistance behaviour items were developed as part of the *CILS4EU* study, based on established theoretical considerations of student resistance to schooling (e.g. Harris 2006; McFarland 2001) . The survey underwent a standardized translation and adaptation process across participating countries to ensure linguistic and conceptual equivalence. While no separate face validity testing was conducted, the measures of resistance behaviour have been widely used in prior research, supporting its conceptual appropriateness (e.g. Bussemakers and Denessen 2024; Geven et al. 2017; Kalmijn 2023).

Control Variables

We included four control variables in the present study (Table 1). First, we included cognitive ability sum scores from a language-free cognitive ability test, which was assessed in all four countries in wave 1. Second, we included sum scores from four independent, country-specific language ability tests that measured the students' lexical abilities in the respective country languages in wave 1. Third, we constructed a dummy variable on the highest parental education degree (0 = *lower than tertiary education degree*, 1 = *tertiary education degree or higher*) based on the parental questionnaire data from wave 1 (retrieved from the student questionnaire in cases of missing data). Fourth, we included a dummy variable for the students' migration background (0 = *parents and students born in respective country*, 1 = *at least one parent or student born abroad*), based on the student questionnaire data from wave 1.

Class-Level Aggregates

Since our third research question is concerned with dynamics on the classroom level, we created simple class-level aggregates of the imputed, student-level variables female gender, resistance to schooling, as well as the control variables (see Table 2).

< Table 2 about here >

Analyses

To investigate our research questions, we first imputed missing data and then ran three sets of multilevel regression models. In all main analyses, we applied the total student sampling weights to account for the complex sampling design and non-response (Kalter et al. 2014), ensuring that the results are representative for the student populations. We used the statistics software R (R Core Team 2021), the R package *MplusAutomation* (Hallquist and Wiley 2018), and Mplus 8 (L. K. Muthén & Muthén, 2017).

Multiple Imputation

We imputed missing data at the student level five times by using the R package *mice* (Van Buuren and Groothuis-Oudshoorn 2011). In the imputation model, we included all student-level outcomes, predictors, and control variables (see Table 1). We accounted for the sampling weights and two-level hierarchical structure of the data. The imputations were conducted separately for the four country samples, but the same variables were included in all four cases. Note that cases with missing gender information were excluded from the samples (see data section above). All main analyses were replicated for the five imputed datasets and the estimates were combined using Rubin's (1987) rules in Mplus (Muthén and Muthén 2017). In all models, we z-standardised the non-binary variables grades, resistance to schooling, cognitive ability, and language ability within countries to facilitate the interpretability and comparability of the findings.

Models for First Research Question

The first two research questions on the association between students' characteristics and school grades are concerned with dynamics within school classes and consider student-level variables. We conducted multilevel regression analyses for of two reasons. First, we considered the hierarchical data structure with students being nested in classes (see Appendix D; Muthén and Satorra 1995). Second, we accounted for the fact that school grades are awarded in classroom contexts so that students might be evaluated relative to the performance of their classmates. Thus, we ran two-level regression models with random intercepts for every country.

To answer the research question on whether there is a gender gap in grades within classrooms, we regressed the z-standardised outcome $GRADE$ (in mathematics or language, respectively) of student i in class j on the dummy predictor variable $FEMALE$:

Within: $$GRADE_{ij} = \beta_{0j} + \beta_{1j} * FEMALE_{ij} + r_{ij} \quad (1)$$

Between: $$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} \quad (3)$$

In this model, the student-level intercept β_{0j} is decomposed into a class-level mean and variance of the school grade. The grade difference between boys and girls is captured in γ_{10}. The residual variance r_{ij} reflects that students' grades deviate from the predictions that are based on their class membership and gender. We ran separate models for the mathematics and language grade outcomes.

In this Model 1, we controlled for cognitive ability, language ability, tertiary parental education, and migration background. In case of the mathematics grade outcome, we included cognitive ability and in case of the language grade outcome language abilities as control variables. We did not include them simultaneously because of their collinearity (see Figure 1).

Models for Second Research Question

As for the first research question, we ran two-level regression models to answer the second research question on whether resistance behaviour explains gender gaps in grades. We expanded the model of the first research question and regressed the grades on female gender as well as resistance to schooling, and the control variables (cognitive ability or language ability, tertiary parental education, and migration background) in Model 2a. To furthermore test for a differential grade punishment by gender, we included the interaction term between female and resistance in a second step (Model 2b).

Models for Third Research Question

The third research question, which examines the association between grades and the share of females, is concerned with dynamics between school classes and focuses on class-level variables. Since school grades were measured at the student level, we ran two-level regression models with class-level predictor variables. Specifically, we regressed the outcome $GRADE$ (in mathematics or language, respectively) of student i in class j on the $SHARE\ OF\ FEMALES$ in class j:

Within: $$GRADE_{ij} = \beta_{0j} + \beta_{1j} * FEMALE_{ij} + r_{ij} \quad (4)$$

Between: $$\beta_{0j} = \gamma_{00} + \gamma_{01} * SHARE\ OF\ FEMALES_j + u_{0j} \quad (5)$$

$$\beta_{1j} = \gamma_{10} \quad (6)$$

The parameter γ_{00} reflects the class-level intercept of grade, γ_{01} the association between the share of girls and the class grade level, and u_{0j} the class-level residual variance. The parameter r_{ij} reflects the between-student variation in grades beyond class membership and the share of girls in the attended classes. In Model 3a, we included student-level (female gender, resistance to schooling, cognitive respectively language ability, tertiary parental education, and migration background) and class-level control variables (share of females, average resistance to schooling, average cognitive or language ability, share tertiary parental education, and share migration background). The class-level variables were aggregated from the imputed student-level variables. We z-standardised the continuous student-level as well as the class-level variables within countries to facilitate the interpretability and comparability of the findings.

We furthermore estimated Model 3b, in which we tested whether the gender gap differed by the share of females in the classroom:

Within: $$GRADE_{ij} = \beta_{0j} + \beta_{1j} * FEMALE_{ij} + r_{ij} \tag{7}$$

Between: $$\beta_{0j} = \gamma_{00} + \gamma_{01} * SHARE\ OF\ FEMALES_j + u_{0j} \tag{8}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} * SHARE\ OF\ FEMALES_j + u_{1j} \tag{9}$$

Results

Descriptive results

In addition to the univariate descriptive statistics that we provide in Tables 1 and 2, we furthermore depict some bivariate associations of interest. Table 3 displays descriptive statistics for the student-level outcome, predictor, and control variables by gender. We found that boys and girls differed in their mean grades in some cases, with boys receiving more favourable grades in mathematics and girls in language. Unreported analyses show that the gender difference in mathematics grades is significantly different from zero only in Germany, while the gender gaps in language grades are significant in all four countries. Furthermore, boys showed significantly more resistance to schooling in all four countries. Gender differences in cognitive ability, language ability, tertiary parental education, and migration background were inconsistent across country samples. Additionally, we report bivariate correlations of continuous variables in Appendix B.

< Table 3 about here >

Two-level Regression Results

Results Concerning Research Question 1

To investigate the first research question on gender gaps in grades within classrooms, we ran two-level regression models for mathematics and language grades as outcomes and gender as predictor variable (see Figure 1). We controlled for cognitive and language ability, respectively, tertiary parental education, and migration background. In mathematics, we found that boys received significantly better grades in all countries except Sweden, where the difference was not statistically significant. The effect sizes were small to medium-sized,[4] with boys receiving up to 0.18 standard deviations (SD) higher grades than girls in Germany (0.08 SD in the Netherlands and 0.09 SD in England). In language subjects, girls received significantly better grades in all four countries. The effect sizes were medium to large, with girls receiving up to 0.35 SD higher grades than boys in Germany (0.22 SD in England and Netherlands, 0.33 SD in Sweden). Students who scored higher on the cognitive and language ability tests had more favourable grades in comparison to their classmates. In most cases, students who had at least one parent with tertiary education also showed more favourable grades. After accounting for gender, the ability scores, as well as tertiary parental education, students with a migration background did not differ significantly from their classmates.

< Figure 1 about here >

Results Concerning Research Question 2

To investigate the second research question on the role of resistance to schooling for gender gaps in grades, we added resistance to schooling as another predictor variable (see Figure 2). In all cases, students with more resistance to schooling had less favourable grades than their classmates in these models (effect sizes varied between –0.07 SD in the Netherlands and –0.13 SD in Sweden for the mathematics grade and between –0.05 SD in the Netherlands and –0.09 in Sweden for the language grade), which controlled for ability test scores as well as demographic characteristics. We found that in comparison to the model without the resistance to schooling predictor (see Figure 1), the gender differences in grades shifted slightly to the advantage of boys in the models that included resistance to schooling as a predictor (see Figure 2). In case of mathematics grades as outcomes, the gender gaps were more negative and in case of language grades as outcomes, the gender gaps were less positive (-0.09 SD in Netherlands, -0.12 SD in England, -0.23 SD in Germany, not significantly different from zero in Sweden) and in case of language grades as outcomes, the gender gaps were less positive (0.20 SD in England, 0.21 SD in Netherlands, and 0.31 SD in Germany and Sweden). Our explanation for this finding is that negative school behaviour such as resistance to schooling has a negative effect on grades and that boys show more resistance to schooling than girls (see Table 3).

[4] We refer to effect sizes between 0.00-0.15 SD as small, 0.16-0.30 SD as medium-sized, and >0.30 SD as large effects.

< Figure 2 about here >

In separate models, we furthermore included an interaction term for female gender and resistance to schooling to test for a differential punishment of resistance to schooling by gender (see Figure 3). This interaction coefficient was significantly different from zero in three out of cases (-0.13 for Germany and –0.09 for Sweden regarding the mathematics grade, and –0.09 for Sweden regarding the language grade), in which it indicated that resistance to schooling had a more negative statistical effect on grades for girls than for boys. As an interesting result that does not pertain to any of our research questions, we want to highlight that in our analysis, migration background is not significantly related to higher or lower grades in either subject domain, in none of the four survey countries, after controlling for abilities and sociodemographic characteristics.

< Figure 3 about here >

Results Concerning Research Question 3

To investigate the role of the gender compositions of classrooms on grades, we ran two-level regression models with mathematics and language grades as outcomes, with share of females and resistance to schooling as main predictors, and student-level (female gender, resistance to schooling, cognitive respectively language ability, parental education, and migration background) and class-level control variables (average cognitive or language ability, share of students with tertiary parental education, and share of students with a migration background). These models showed no significant associations between grades and classrooms' share of female students, or average resistance to schooling (see Figure 4), after controlling for the students' individual gender and other classroom composition characteristics. We furthermore tested whether the gender gaps in grades differed by the share of females in the classroom and found no statistically significant differences (see Appendix G, Table G2).

< Figure about here >

Discussion

This study aimed to understand the dynamics of gender, grades, and resistance behaviour, considering individual and classroom-level factors. We first examined gender differences in mathematics and language grades. While prior research reports a general advantage for girls (O'Dea et al. 2018; OECD 2015; Voyer and Voyer 2014), our findings tell a more differentiated story. Girls outperformed boys in language, but boys received higher mathematics grades in Germany, Sweden, and the Netherlands, after controlling for ability test scores and demographic characteristics. This aligns with recent studies that question the idea of a consistent female advantage in grades (OECD 2015; Steinmann 2024). Although the magnitude of the gender gap varies, its direction is consistent across countries, indicating that gender differences in grades are subject-specific. These patterns point to the relevance of gender stereotypes. Boys are often considered better in STEM subjects, while girls are seen as stronger in language and the arts. If students internalise such beliefs, they may perceive their abilities differently despite having similar cognitive skills. Teachers' expectations and behaviours may also reflect these stereotypes, influencing grades either directly or indirectly (Gentrup et al. 2020; Riegle-Crumb and Humphries 2012).

Second, we examined the role of school resistance behaviour. Boys reported resistance more frequently, and this explained part of their grade disadvantage in language. In mathematics, however, boys' grade advantage increased when resistance was accounted for. These findings align with previous research in other contexts (Demanet and Houtte 2012; Downey and Yuan 2005). We also found tentative evidence for differential punishment: in three of eight cases (Germany and Sweden for mathematics, and Sweden for language), girls' grades declined more than boys' when they reported resistance behaviour. This may reflect teachers incorporating effort and diligence into grading, consciously or unconsciously. Alternatively, students who are frequently late or skip class may simply learn less, leading to poorer assessment performance (Downey 2008; Fredricks et al. 2004). The differential effects could also reflect gendered norms: some behaviours may be viewed as more acceptable for boys than for girls. Since the resistance measure was self-reported, it is also possible that boys exaggerate their behaviour due to masculine peer norms (Abraham 2008; Moret, Dümmler, and Dahinden 2017). If boys over-report resistance (Riegle-Crumb and Humphries 2012), this would explain the stronger grade penalties observed for girls. Still, the disadvantage for girls in mathematics remains even after controlling for ability, demographics, and resistance behaviour.

Third, we examined classroom gender composition. Prior work suggests that a higher proportion of girls in a class may foster a more academically supportive learning climate (Eisenkopf et al. 2015; Schneeweis and Zweimüller 2012). However, we did not find evidence of composition effects. Several factors may account for this. First, self-reported resistance may be influenced by the classroom context. Students may rate their own behaviour differently depending on what is typical among peers. Second, although our sample includes 886 school classes across four countries, there may not be sufficient variation in gender composition to detect effects. Third, since not all classmates are included in the dataset, measurement error at the class level is likely. Nonetheless, we found considerable variation in gender composition, especially in England, where single-sex classrooms are more common. These patterns suggest that gender imbalance in classrooms is not negligible, even if we did not detect effects. Unlike social or ethnic composition, which is shaped by neighbourhood segregation and school policies, gender composition likely reflects more random processes, such as school size or scheduling. This variation underscores the relevance of studying classroom gender composition, even if our findings on gender composition are null (see histograms in Figure H in Appendix H).

We also considered differences in national grouping and grading practices. England and Sweden follow comprehensive systems with delayed tracking, while Germany and the Netherlands track students into different educational paths at earlier ages (European Commission 2023; gov.uk 2025). Within-school grouping is still common in England. Germany begins tracking at age 10, based on teacher recommendations and parental input; the Netherlands does so at age 12, guided by test results and teacher advice. Sweden tracks only after age 16. Prior studies suggest that tracking increases segregation (Chmielewski 2014; Strello, Strietholt, and Steinmann 2022), though similar effects may emerge in comprehensive systems through internal grouping (Engzell and Raabe 2023). We accounted for these institutional differences by estimating models separately for each country.

Grading policies also differ across contexts. In Germany, grading formally includes classroom behaviour; in England and the Netherlands, this is at the discretion of individual schools; and in Sweden, grades must reflect academic attainment alone (Department for Education 2013; Inspectie van het Onderwijs 2024; Kultusministerkonferenz 2024; Skolverket 2024). Still, we found that resistance behaviour was negatively associated with grades in all countries, including Sweden. This suggests that the link may be due to learning disruption rather than grading bias. Notably, Swedish grades were drawn from official records, whereas in the other countries, grades were self-reported. The stronger effects observed in Sweden may therefore reflect greater data accuracy.

Limitations

Despite using large-scale representative data, our study has several limitations. First, better identification of class composition effects would require larger within-school samples to separate class-level from school-level influences. Second, we relied on available CILS4EU variables, which were not designed for our specific questions. The resistance measure, for instance, might conflate behaviour and teacher responses, and could be assessed more objectively. Similarly, grades were only objectively reported in Sweden. No curriculum-based performance tests were available, limiting our ability to capture subject-specific knowledge, which tends to correlate more strongly with grades (Alemán et al. 2024). Third, our design cannot account for reverse causality: students who feel unfairly graded may become resistant (Zimmermann et al. 2013).

More generally, our study is observational and cannot establish causality. Gender cannot be experimentally manipulated, and while classroom composition could theoretically be altered in an experiment, doing so across 886 classrooms in four countries would not be feasible. Our findings should therefore be interpreted as associations. Future work should also address intersecting dimensions of inequality such as ethnicity and socioeconomic status (Gortazar et al. 2022). Research shows, for instance, that Black boys in the U.S. face compounded disadvantages in how resistance is perceived by teachers (Owens 2022), and that gender and social class may interact in shaping reading outcomes (Becker and McElvany 2018). Finally, future studies could benefit from a broader set of control variables, such as personality traits (Mammadov 2022), which were not available in our data.

Implications and Conclusions

Our findings show that gender gaps in grades are subject-specific: girls are disadvantaged in mathematics and boys in language, even after adjusting for ability and other controls. These patterns may channel students into gendered educational paths, contributing to occupational segregation (Levanon and Grusky 2016).

The results also highlight the importance of teacher training. Grades shape future opportunities and should reflect actual learning rather than behaviours or assumptions. Yet classroom behaviour appears to influence grades, raising concerns about fairness. While rewarding effort may motivate students, penalising low engagement risks reinforcing bias. Teachers must understand what their grading is meant to assess and be aware of potential biases, including those shaped by gender norms. If girls are more heavily penalised for similar behaviours, this may reflect unconscious stereotyping, which can be addressed through awareness and reflection. Resistance behaviour should also be addressed proactively, as it hampers learning and contributes to lower achievement, regardless of how teachers assign grades.

In sum, this study offers insight into gender gaps in grades and resistance behaviour in four European countries. Boys more often report resistance, which partly explains their lower language grades but not their mathematics advantage. Girls may be more strongly penalised for such behaviour. Classroom gender composition showed no clear effect, but variation across contexts suggests this area warrants further attention.

References

Abraham, John. 2008. "Back to the Future on Gender and Anti-school Boys: A Response to Jeffrey Smith." *Gender and Education* 20(1):89–94. doi: 10.1080/09540250701749094.

Alemán, José Manuel Arencibia, Astrid Marie Jorde Sandsør, Henrik Daae Zachrisson, and Sigrid Blömeke. 2024. "Teacher-Assigned Grades and External Exams: Sources of Discrepancy." *Assessment in Education: Principles, Policy & Practice* 31(2):94–115. doi: 10.1080/0969594x.2024.2338764.

Allen, Joseph P., Maryfrances R. Porter, F. Christy McFarland, Penny Marsh, and Kathleen Boykin McElhaney. 2005. "The Two Faces of Adolescents' Success With Peers: Adolescent Popularity, Social Adaptation, and Deviant Behavior." *Child Development* 76(3):747–60. doi: 10.1111/j.1467-8624.2005.00875.x.

Allen, Joseph P., Megan M. Schad, Barbara Oudekerk, and Joanna Chango. 2014. "What Ever Happened to the 'Cool' Kids? Long-Term Sequelae of Early Adolescent Pseudomature Behavior." *Child Development* 85(5):1866–80. doi: 10.1111/cdev.12250.

Angelo, Catarina, and Ana Balcão Reis. 2020. "Gender Gaps in Different Grading Systems." *Education Economics* 29(1):1–15. doi: 10.1080/09645292.2020.1853681.

Becker, Michael, and Nele McElvany. 2018. "The Interplay of Gender and Social Background: A Longitudinal Study of Interaction Effects in Reading Attitudes and Behaviour." *British Journal of Educational Psychology* 88(4):529–49. doi: 10.1111/bjep.12199.

Blossfeld, Hans-Peter, Wilfried Bos, Bettina Hannover, Dieter Lenzen, Detlef Müller-Böling, Manfred Prenzel, and Ludger Wößmann. 2009. *Geschlechterdifferenzen Im Bildungssystem, Jahresgutachten 2009*. edited by vbw – V. der B. W. e.V. Wiesbaden: VS Verlag für Sozialwissenschaften.

Borghans, Lex, Bart H. H. Golsteyn, James J. Heckman, and John Eric Humphries. 2016. "What Grades and Achievement Tests Measure." *Proceedings of the National Academy of Sciences* 113(47):13354–59. doi: 10.1073/pnas.1601135113.

Borgonovi, Francesca, Alessandro Ferrara, and Soumaya Maghnouj. 2018. "The Gender Gap in Educational Outcomes in Norway." *OECD Education Working Papers*. doi: 10.1787/f8ef1489-en.

Bowers, Alex J. 2011. "What's in a Grade? The Multidimensional Nature of What Teacher-Assigned Grades Assess in High School." *Educational Research and Evaluation* 17(3):141–59. doi: 10.1080/13803611.2011.597112.

Brookhart, Susan M., Thomas R. Guskey, Alex J. Bowers, James H. McMillan, Jeffrey K. Smith, Lisa F. Smith, Michael T. Stevens, and Megan E. Welsh. 2016. “A Century of Grading Research.” *Review of Educational Research* 86(4):803–48. doi: 10.3102/0034654316672069.

Bussemakers, Carlijn, and Eddie Denessen. 2024. “Teacher Support as a Protective Factor? The Role of Teacher Support for Reducing Disproportionality in Problematic Behavior at School.” *The Journal of Early Adolescence* 44(1):5–40. doi: 10.1177/02724316231156835.

Calsamiglia, Caterina, and Annalisa Loviglio. 2019. “Grading on a Curve: When Having Good Peers Is Not Good.” *Economics of Education Review* 73:101916. doi: 10.1016/j.econedurev.2019.101916.

Caro, Daniel H., Jenny Lenkeit, Rainer Lehmann, and Knut Schwippert. 2009. “The Role of Academic Achievement Growth in School Track Recommendations.” *Studies in Educational Evaluation* 35(4):183–92. doi: 10.1016/j.stueduc.2009.12.002.

Chmielewski, Anna K. 2014. “An International Comparison of Achievement Inequality in Within- and Between-School Tracking Systems.” *American Journal of Education* 120(3):293–324. doi: 10.1086/675529.

Demanet, Jannick, and Mieke Van Houtte. 2012. “School Belonging and School Misconduct: The Differing Role of Teacher and Peer Attachment.” *Journal of Youth and Adolescence* 41(4):499–514. doi: 10.1007/s10964-011-9674-2.

Department for Education. 2013. “Behaviour and Discipline in Schools.” Retrieved March 6, 2025 (https://www.gov.uk/government/publications/behaviour-in-schools--2).

Downey, Douglas B. 2008. “Black/White Differences in School Performance: The Oppositional Culture Explanation.” *Annual Review of Sociology* 34(1):107–26. doi: 10.1146/annurev.soc.34.040507.134635.

Downey, Douglas B., and Anastasia S. Vogt Yuan. 2005. “Sex Differences in School Performance During High School: Puzzling Patterns and Possible Explanations.” *Sociological Quarterly* 46(2):299–321. doi: 10.1111/j.1533-8525.2005.00014.x.

Duckworth, Angela Lee, and Martin E. P. Seligman. 2006. “Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores.” *Journal of Educational Psychology* 98(1):198–208. doi: 10.1037/0022-0663.98.1.198.

Eisenkopf, Gerald, Zohal Hessami, Urs Fischbacher, and Heinrich W. Ursprung. 2015. “Academic Performance and Single-Sex Schooling: Evidence from a Natural Experiment in Switzerland.” *Journal of Economic Behavior & Organization* 115:123–43. doi: 10.1016/j.jebo.2014.08.004.

Encinas-Martín, Marta, and Michelle Cherian. 2023. “Gender, Education and Skills.” *OECD Skills Studies*. doi: 10.1787/34680dd5-en.

Engzell, Per, and Isabel J. Raabe. 2023. “Within-School Achievement Sorting in Comprehensive and Tracked Systems.” *Sociology of Education* 96(4):324–43. doi: 10.1177/00380407231183952.

Eurydice, European Education and Culture Executive Agency. 2010. *Gender Differences in Educational Outcomes: Study on the Measures Taken and the Current Situation in Europe*. Education, Audiovisual and Culture Executive Agency.

Farkas, George, Robert P. Grobe, Daniel Sheehan, and Yuan Shuan. 1990. “Cultural Resources and School Success: Gender, Ethnicity, and Poverty Groups within an Urban School District.” *American Sociological Review* 55(1):127. doi: 10.2307/2095708.

Ferman, Bruno, and Luiz Felipe Fontes. 2022. “Assessing Knowledge or Classroom Behavior? Evidence of Teachers’ Grading Bias.” *Journal of Public Economics* 216:104773. doi: 10.1016/j.jpubeco.2022.104773.

Figlio, David N. 2007. “Boys Named Sue: Disruptive Children and Their Peers.” *Education Finance and Policy* 2(4):376–94. doi: 10.1162/edfp.2007.2.4.376.

Fredricks, Jennifer A., Phyllis C. Blumenfeld, and Alison H. Paris. 2004. “School Engagement: Potential of the Concept, State of the Evidence.” *Review of Educational Research* 74(1):59–109. doi: 10.3102/00346543074001059.

Gentrup, Sarah, Georg Lorenz, Cornelia Kristen, and Irena Kogan. 2020. “Self-Fulfilling Prophecies in the Classroom: Teacher Expectations, Teacher Feedback and Student Achievement.” *Learning and Instruction* 66:101296. doi: 10.1016/j.learninstruc.2019.101296.

Geven, Sara, Jan O. Jonsson, and Frank van Tubergen. 2017. “Gender Differences in Resistance to Schooling: The Role of Dynamic Peer-Influence and Selection Processes.” *Journal of Youth and Adolescence* 46(12):2421–45. doi: 10.1007/s10964-017-0696-2.

Gortazar, Lucas, David Martinez de Lafuente, and Ainhoa Vega-Bayo. 2022. “Comparing Teacher and External Assessments: Are Boys, Immigrants, and Poorer Students Undergraded?” *Teaching and Teacher Education* 115:103725. doi: 10.1016/j.tate.2022.103725.

Hadjar, Andreas, and Christiane Gross. 2016. “Education Systems and Inequalities, International Comparisons.” doi: 10.1354/books/unregistered/9781447326113.

Hadjar, Andreas, Sabine Krolak-Schwerdt, Karin Priem, and Sabine Glock. 2014. “Gender and Educational Achievement.” *Educational Research* 56(2):117–25. doi: 10.1080/00131881.2014.898908.

Hallinan, Maureen T. 1992. “The Organization of Students for Instruction in the Middle School.” *Sociology of Education* 65(2):114. doi: 10.2307/2112678.

Hallquist, Michael N., and Joshua F. Wiley. 2018. “MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus.” *Structural Equation Modeling: A Multidisciplinary Journal* 25(4):621–38. doi: 10.1080/10705511.2017.1402334.

Harris, A. L. 2006. “I (Don’t) Hate School: Revisiting Oppositional Culture Theory of Blacks’ Resistance to Schooling.” *Social Forces* 85(2):797–834. doi: 10.1353/sof.2007.0006.

Hattie, John. 2009. *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*. London: Routledge.

Heyder, Anke, Margriet van Hek, and Mieke Van Houtte. 2021. “When Gender Stereotypes Get Male Adolescents into Trouble: A Longitudinal Study on Gender Conformity Pressure as a Predictor of School Misconduct.” *Sex Roles* 84(1–2):61–75. doi: 10.1007/s11199-020-01147-9.

Hinnerich, Björn Tyrefors, Erik Höglin, and Magnus Johannesson. 2011. “Are Boys Discriminated in Swedish High Schools?” *Economics of Education Review* 30(4):682–90. doi: 10.1016/j.econedurev.2011.02.007.

Hjorth-Trolle, Anders, Erik Rosenqvist, and Anders Hed. 2021. “Grading Practices and the Social Gradient in GPA: Quasi-Experimental Evidence from Sweden.” *European Sociological Review* 38(3):jcab053-. doi: 10.1093/esr/jcab053.

Horn, E. Parham, Meghan P. McCormick, Erin E. O’Connor, Sandee G. McClowry, and Frances C. Hogan. 2021. “Trajectories of Teacher–Child Relationships across Kindergarten and First Grade: The Influence of Gender and Disruptive Behavior.” *Early Childhood Research Quarterly* 55:107–18. doi: 10.1016/j.ecresq.2020.10.003.

Houtte, Mieke Van. 2004. “Gender Context of the School and Study Culture, or How the Presence of Girls Affects the Achievement of Boys.” *Educational Studies* 30(4):409–23. doi: 10.1080/0305569042000310336.

Hoxby, C. M. 2000. *Peer Effects in the Classroom: Learning from Gender and Race Variation*.

Hyde, Janet Shibley. 2014. “Gender Similarities and Differences.” *Psychology* 65(1):373–98. doi: 10.1146/annurev-psych-010213-115057.

Inspectie van het Onderwijs. 2024. “The Dutch Educational System.” Retrieved March 6, 2025 (https://english.onderwijsinspectie.nl/documents/publications/2024/06/05/dutch-education-system---background-information).

Jackson, Carolyn. 2002. “‘Laddishness’ as a Self-Worth Protection Strategy.” *Gender and Education* 14(1):37–50. doi: 10.1080/09540250120098870.

Jackson, Carolyn. 2003. "Motives for 'Laddishness' at School: Fear of Failure and Fear of the 'Feminine.'" *British Educational Research Journal* 29(4):583–98. doi: 10.1080/01411920301847.

Jacob, Brian A. 2002. "Where the Boys Aren't: Non-Cognitive Skills, Returns to School and the Gender Gap in Higher Education." *Economics of Education Review* 21(6):589–98. doi: 10.1016/s0272-7757(01)00051-6.

Kalmijn, Matthijs. 2023. "The Association between Parental Separation at the School Level and Student Outcomes in Four European Countries." *Research in Social Stratification and Mobility* 86:100813. doi: 10.1016/j.rssm.2023.100813.

Kalter, Frank, Anthony F. Heath, Miles Hewstone, Jan O. Jonsson, Matthijs Kalmijn, Irena Kogan, and Frank van Tubergen. 2014. "Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU)." doi: 10.4232/cils4eu.5353.1.1.0.

Kalter, Frank, Anthony F. Heath, Miles Hewstone, Jan O. Jonsson, Matthijs Kalmijn, Irena Kogan, and Frank van Tubergen. 2015. "Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU)." doi: 10.4232/cils4eu.5353.2.1.0.

Kelly, Sean. 2008. "What Types of Students' Effort Are Rewarded with High Marks?" *Sociology of Education* 81(1):32–52. doi: 10.1177/003804070808100102.

Kenney-Benson, Gwen A., Eva M. Pomerantz, Allison M. Ryan, and Helen Patrick. 2006. "Sex Differences in Math Performance: The Role of Children's Approach to Schoolwork." *Developmental Psychology* 42(1):11–26. doi: 10.1037/0012-1649.42.1.11.

Kessels, Ursula, and Anke Heyder. 2017. "Die Wertschätzung Schulischer Anstrengung Als Mediator von Geschlechtsunterschieden in Noten: Eine Fächerspezifische Analyse." *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie* 49(2):86–97. doi: 10.1026/0049-8637/a000171.

Klapproth, Florian, Sabine Glock, Sabine Krolak-Schwerdt, Romain Martin, and Matthias Böhmer. 2013. "Prädiktoren Der Sekundarschulempfehlung in Luxemburg." *Zeitschrift Für Erziehungswissenschaft* 16(2):355–79. doi: 10.1007/s11618-013-0340-1.

Kruse, Hanno, and Konstanze Jacob. 2014. *Children of Immigrants Longitudinal Survey in Four European Countries. Sociometric Fieldwork Report. Wave 1 – 2010/2011, v1.1.0.* Mannheim: Mannheim University.

Kultusministerkonferenz. 2024. "Schulgesetze Der Länder in Der Bundesrepublik Deutschland." Retrieved March 6, 2025 (https://www.kmk.org/dokumentation-statistik/rechtsvorschriften-lehrplaene/uebersicht-schulgesetze.html).

Lavy, Victor. 2008. "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment." *Journal of Public Economics* 92(10–11):2083–2105. doi: 10.1016/j.jpubeco.2008.02.009.

Lavy, Victor, and Analía Schlosser. 2011. “Mechanisms and Impacts of Gender Peer Effects at School.” *American Economic Journal: Applied Economics* 3(2):1–33. doi: 10.1257/app.3.2.1.

Legewie, Joscha, and Thomas A. DiPrete. 2012. “School Context and the Gender Gap in Educational Achievement.” *American Sociological Review* 77(3):463–85. doi: 10.1177/0003122412440802.

Levanon, Asaf, and David B. Grusky. 2016. “The Persistence of Extreme Gender Segregation in the Twenty-First Century.” *American Journal of Sociology* 122(2):573–619. doi: 10.1086/688628.

Liberto, Adriana Di, Laura Casula, and Sara Pau. 2021. “Grading Practices, Gender Bias and Educational Outcomes: Evidence from Italy.” *Education Economics* 1–28. doi: 10.1080/09645292.2021.2004999.

Lynch, Rebecca J., Janet A. Kistner, and Nicholas P. Allan. 2014. “Distinguishing among Disruptive Behaviors to Help Predict High School Graduation: Does Gender Matter?” *Journal of School Psychology* 52(4):407–18. doi: 10.1016/j.jsp.2014.05.001.

Marks, Gary N. 2010. “What Aspects of Schooling Are Important? School Effects on Tertiary Entrance Performance.” *School Effectiveness and School Improvement* 21(3):267–87. doi: 10.1080/09243451003694364.

McFarland, Daniel A. 2001. “Student Resistance: How the Formal and Informal Organization of Classrooms Facilitate Everyday Forms of Student Defiance.” *American Journal of Sociology* 107(3):612–78. doi: 10.1086/338779.

McGowan, Amanda L., Hope K. Gerde, Karin A. Pfeiffer, and Matthew B. Pontifex. 2021. “Physically Active Learning in Preschoolers: Improved Self-Regulation, Comparable Quantity Estimation.” *Trends in Neuroscience and Education* 22:100150. doi: 10.1016/j.tine.2021.100150.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology* 27:415–44. doi: 10.2307/2678628.

Moret, Joëlle, Kerstin Dümmler, and Janine Dahinden. 2017. “The Car, the Hammer and the Cables under the Tables.” *European Journal of Sociology* 58(2):265–93. doi: 10.1017/s0003975617000145.

Morris, Edward W. 2012. *Learning the Hard Way: Masculinity, Place, and the Gender Gap in Education. The Rutgers Series in Childhood Studies.* New Brunswick: Rutgers University Press.

Muthén, Bengt, and Linda Muthén. 2017. “Mplus.” Pp. 507–18 in *Handbook of item response theory*. Chapman and Hall/CRC.

Muthén, Bengt O., and Albert Satorra. 1995. “Complex Sample Data in Structural Equation Modeling.” *Sociological Methodology* 25:267. doi: 10.2307/271070.

Ning, Bo, Jan Van Damme, Wim Van Den Noortgate, Xiangdong Yang, and Sarah Gielen. 2015. “The Influence of Classroom Disciplinary Climate of Schools on Reading Achievement: A Cross-Country Comparative Study.” *School Effectiveness and School Improvement* 26(4):586–611. doi: 10.1080/09243453.2015.1025796.

O’Dea, R. E., M. Lagisz, M. D. Jennions, and S. Nakagawa. 2018. “Gender Differences in Individual Variation in Academic Grades Fail to Fit Expected Patterns for STEM.” *Nature Communications* 9(1):3777. doi: 10.1038/s41467-018-06292-0.

OECD. 2015. “The ABC of Gender Equality in Education.” doi: 10.1787/9789264229945-en.

OECD. 2023. “PISA 2022 Results (Volume I).” *PISA*. doi: 10.1787/8693309d-en.

Owens, Jayanti. 2022. “Double Jeopardy: Teacher Biases, Racialized Organizations, and the Production of Racial/Ethnic Disparities in School Discipline.” *American Sociological Review* 87(6):1007–48. doi: 10.1177/00031224221135810.

Ramtekkar, Ujjwal P., Angela M. Reiersen, Alexandre A. Todorov, and Richard D. Todd. 2010. “Sex and Age Differences in Attention-Deficit/Hyperactivity Disorder Symptoms and Diagnoses.” *Journal of the American Academy of Child & Adolescent Psychiatry* 49(3):217-228.e3. doi: 10.1097/00004583-201003000-00005.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reichenberg, Olof. 2018. “Student Behavioural Disengagement, Peer Encouragement and the School Curriculum: A Mechanism Approach.” *Educational Studies* 44(2):147–66. doi: 10.1080/03055698.2017.1347491.

Riegle-Crumb, Catherine, and Melissa Humphries. 2012. “Exploring Bias in Math Teachers’ Perceptions of Students’ Ability by Gender and Race/Ethnicity.” *Gender & Society* 26(2):290–322. doi: 10.1177/0891243211434614.

Rosén, Monica, Isa Steinmann, and Inga Wernersson. 2021. “Gender Differences in Achievement.” Pp. 1–48 in *International Handbook of Comparative Large-Scale Studies in Education: Perspectives and Findings*, *Springer International Handbooks of Education*, edited by T. Nilsen, A. Stancel-Piątak, and J.-E. Gustafsson. Springer.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Schneeweis, Nicole, and Martina Zweimüller. 2012. “Girls, Girls, Girls: Gender Composition and Female School Choice.” *Economics of Education Review* 31(4):482–500. doi: 10.1016/j.econedurev.2011.11.002.

Shield, Bridget M., and Julie E. Dockrell. 2003. "The Effects of Noise on Children at School: A Review." *Building Acoustics* 10(2):97–116. doi: 10.1260/135101003768965960.

Silverman, Irwin W. 2003. "Gender Differences in Delay of Gratification: A Meta-Analysis." *Sex Roles* 49(9–10):451–63. doi: 10.1023/a:1025872421115.

Skolverket. 2024. "Curriculum for Compulsory School, Preschool Class and School-Age Educare – Lgr22." Retrieved March 6, 2025 (https://www.skolverket.se/publikationsserier/styrdokument/2024/curriculum-for-compulsory-school-preschool-class-and-school-age-educare---lgr22).

Sortkær, Bent, and David Reimer. 2018. "Classroom Disciplinary Climate of Schools and Gender – Evidence from the Nordic Countries." *School Effectiveness and School Improvement* 29(4):1–18. doi: 10.1080/09243453.2018.1460382.

Spinath, Birgit, Christine Eckert, and Ricarda Steinmayr. 2014. "Gender Differences in School Success: What Are the Roles of Students' Intelligence, Personality and Motivation?" *Educational Research* 56(2):230–43. doi: 10.1080/00131881.2014.898917.

Steinmann, Isa. 2024. "Unexpected Gender Differences in Teacher Ratings of Academic Skills and School Track Recommendations." *Studies in Educational Evaluation* 83:101384. doi: 10.1016/j.stueduc.2024.101384.

Steinmayr, Ricarda, and Birgit Spinath. 2008. "Sex Differences in School Achievement: What Are the Roles of Personality and Achievement Motivation?" *European Journal of Personality* 22(3):185–209. doi: 10.1002/per.676.

Strello, Andrés, Rolf Strietholt, and Isa Steinmann. 2022. "Does Tracking Increase Segregation? International Evidence on the Effects of between-School Tracking on Social Segregation across Schools." *Research in Social Stratification and Mobility* 78:100689. doi: 10.1016/j.rssm.2022.100689.

Terrier, Camille. 2020. "Boys Lag behind: How Teachers' Gender Biases Affect Student Achievement." *Economics of Education Review* 77:101981. doi: 10.1016/j.econedurev.2020.101981.

Van Buuren, Stef, and Karin Groothuis-Oudshoorn. 2011. "Mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software* 45:1–67.

Voyer, Daniel, and Susan D. Voyer. 2014. "Gender Differences in Scholastic Achievement: A Meta-Analysis." *Psychological Bulletin* 140(4):1174–1204. doi: 10.1037/a0036620.

Vygotsky, Lev Semenovich, and Michael Cole. 1978. *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press.

Wang, Margaret C., Geneva D. Haertel, and Herbert J. Walberg. 1993. "Toward a Knowledge Base for School Learning." *Review of Educational Research* 63(3):249–94. doi: 10.3102/00346543063003249.

Zell, Ethan, Zlatan Krizan, and Sabrina R. Teeter. 2015. "Evaluating Gender Similarities and Differences Using Metasynthesis." *American Psychologist* 70(1):10–20. doi: 10.1037/a0038208.

Zimmermann, Friederike, Kerstin Schütte, Päivi Taskinen, and Olaf Köller. 2013. "Reciprocal Effects Between Adolescent Externalizing Problems and Measures of Achievement." *Journal of Educational Psychology* 105(3):747–61. doi: 10.1037/a0032793.

Table 1. Overview of Student-Level Variables and their Distributions in the four Samples.

		England	Germany	Netherlands	Sweden
Number of students		4231	4848	4201	4954
Mathematics grade					
	Range	[1.00, 4.00]	[1.00, 6.00]	[0.00,10.00]	[1.00, 4.00]
	M(SD)	2.99(0.72)	4.07(1.00)	6.59(1.39)	2.54(0.80)
Language grade					
	Range	[1.00, 4.00]	[1.00, 6.00]	[0.00,10.00]	[1.00, 4.00]
	M(SD)	3.07(0.66)	4.27(0.81)	6.69(0.96)	2.68(0.75)
Female gender					
	Range	[0.00,1.00]	[0.00,1.00]	[0.00,1.00]	[0.00,1.00]
	Percent	51 %	49 %	49 %	50 %
Resistance to schooling					
	Range	[-1.05,3.90]	[-1.05,3.63]	[-1.05,3.90]	[-1.05,3.90]
	M(SD)	0.31(0.99)	-0.14(0.72)	0.10(0.84)	-0.30(0.74)
Cognitive ability					
	Range	[0.00,26.00]	[1.00,27.00]	[1.00,27.00]	[0.00,27.00]
	M(SD)	18.73(3.99)	19.85(3.77)	19.85(3.55)	18.01(4.75)
Language ability					
	Range	[0.00,24.00]	[0.00,25.00]	[0.00,30.00]	[0.00,30.00]
	M(SD)	16.74(3.40)	12.91(4.31)	17.23(4.19)	18.93(4.96)
Tertiary parental education					
	Range	[0.00,1.00]	[0.00,1.00]	[0.00,1.00]	[0.00,1.00]
	Percent	38 %	25 %	13 %	52 %
Migration background					
	Range	[0.00,1.00]	[0.00,1.00]	[0.00,1.00]	[0.00,1.00]
	Percent	15 %	18 %	8 %	19 %

Note. The means, standard deviations, and percentages reflect weighted distributions after multiple imputation. See Appendix A for the original distributions and shares of missing values. Mathematics grade and language grade served as outcomes, female gender and resistance to schooling served as main predictors, and cognitive ability, language ability, tertiary parental education, and migration background served as control variables in the present study.

Table 2. Overview of Class-Level Variables and their Distributions in the four Samples

	England	Germany	Netherlands	Sweden
Number of classes	198	245	201	242
Average mathematics grade				
Range	[2.00,3.96]	[2.83,5.00]	[1.00,7.86]	[1.60,3.74]
M(SD)	2.99(0.38)	4.07(0.35)	6.59(0.42)	2.54(0.29)
Average language grade				
Range	[2.00,4.00]	[2.94,5.25]	[2.62,8.00]	[1.64,3.47]
M(SD)	3.07(0.32)	4.27(0.35)	6.69(0.34)	2.68(0.26)
Share of females				
Range	[0.00,1.00]	[0.00,1.00]	[0.00,1.00]	[0.17,0.90]
M(SD)	0.51(0.27)	0.49(0.15)	0.49(0.24)	0.50(0.13)
Average resist. to school.				
Range	[-0.50,1.38]	[-0.62,0.86]	[-0.58,1.07]	[-0.84,0.44]
M(SD)	0.31(0.35)	-0.14(0.24)	0.10(0.29)	-0.30(0.25)
Average cognitive ability				
Range	[12.70,23.75]	[10.60,23.29]	[12.00,23.61]	[9.69,21.95]
M(SD)	18.73(1.74)	19.85(2.04)	19.85(1.59)	18.01(1.79)
Average language ability				
Range	[8.10,20.64]	[4.80,20.29]	[8.75,24.41]	[9.00,23.16]
M(SD)	16.74(1.54)	12.91(2.59)	17.23(2.25)	18.93(1.87)
Share tertiary parental ed.				
Range	[0.00,0.95]	[0.00,0.83]	[0.00,0.71]	[0.00,1.00]
M(SD)	0.38(0.20)	0.25(0.20)	0.13(0.14)	0.52(0.16)
Share migration background				
Range	[0.00,0.91]	[0.00,1.00]	[0.00,1.00]	[0.00,1.00]
M(SD)	0.15(0.18)	0.18(0.18)	0.08(0.13)	0.19(0.20)

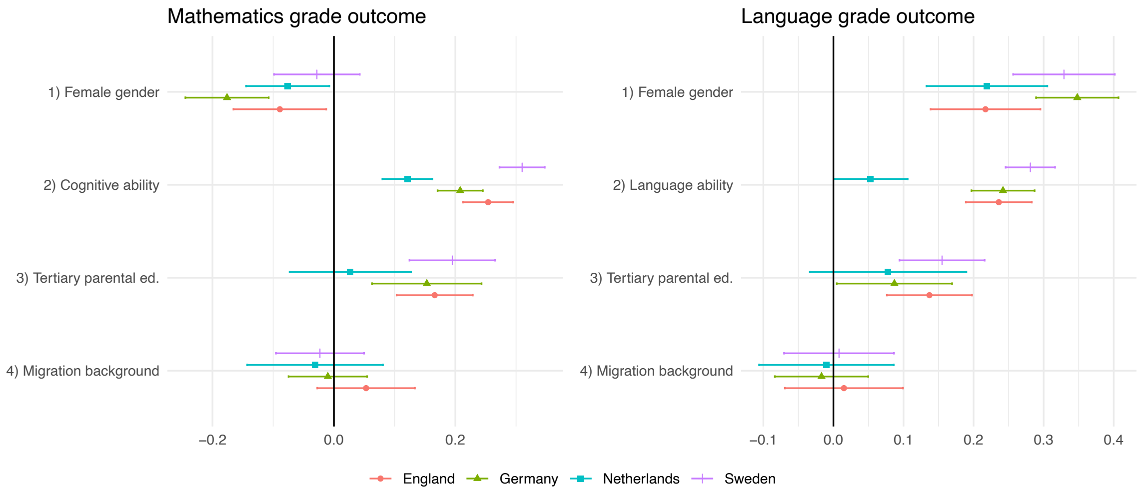
Note. The class-level variables are aggregated, imputed student-level variables (see Table 1). Average mathematics grade and average language grade served as outcomes, share of females and average resistance to schooling served as main predictors, and average cognitive ability, average language ability, share tertiary parental education, and share migration background served as control variables in the present study.

Table 3. Student-Level Descriptive Statistics by Gender in the four Samples

	England M(SD) or %		Germany M(SD) or %		Netherlands M(SD) or %		Sweden M(SD) or %	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Math. grade	3.01(0.72)	2.97(0.72)	4.15(0.99)	3.99(1.00)	6.64(1.36)	6.54(1.42)	2.54(0.80)	2.54(0.81)
Language grade	2.98(0.65)	3.15(0.65)	4.15(0.83)	4.40(0.78)	6.55(0.96)	6.83(0.94)	2.53(0.74)	2.84(0.73)
Resist. to school.	0.47(0.99)	0.17(0.96)	0.05(0.77)	-0.33(0.60)	0.18(0.83)	0.02(0.85)	-0.22(0.80)	-0.38(0.67)
Cognitive ability	18.78(4.07)	18.68(3.92)	19.70(3.86)	20.01(3.67)	19.58(3.64)	20.13(3.43)	17.86(4.94)	18.16(4.54)
Language ability	16.72(3.50)	16.76(3.30)	13.24(4.32)	12.56(4.27)	17.48(4.21)	16.98(4.16)	18.59(5.03)	19.27(4.87)
Tertiary par. ed.	37 %	39 %	24 %	26 %	12 %	14 %	52 %	52 %
Migration backgr.	16 %	14 %	18 %	19 %	7 %	10 %	18 %	19 %

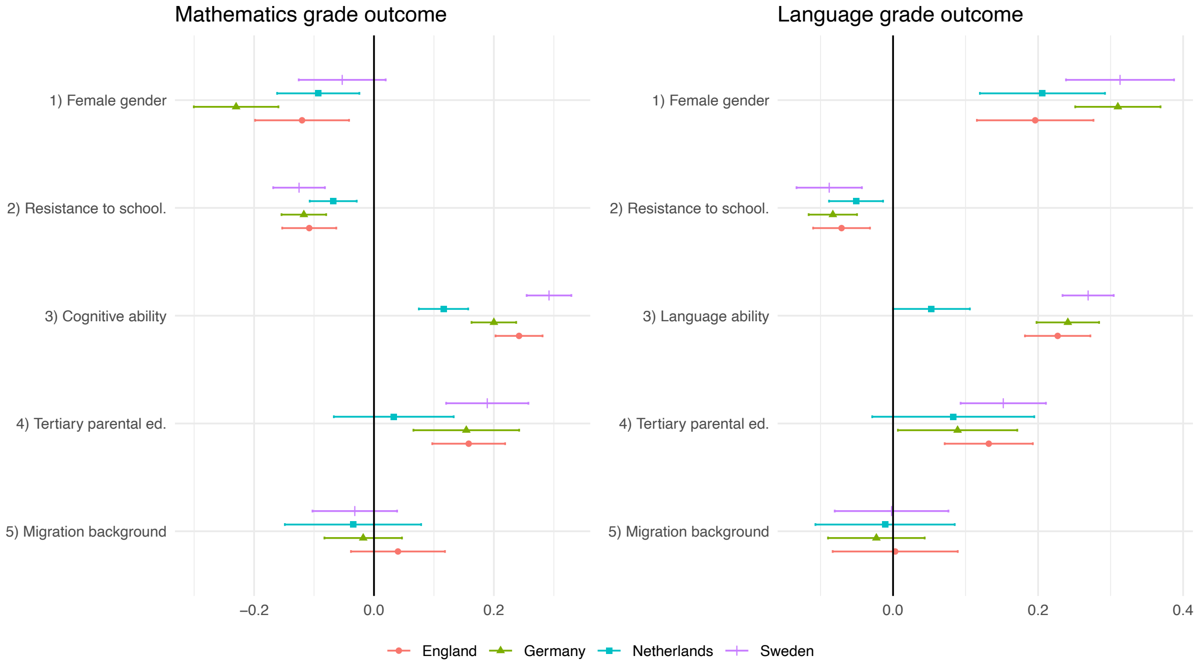
Note. The means, standard deviations, and percentages reflect weighted distributions after multiple imputation.

Figure 1. Coefficient Plot for Models 1 (Research Question 1) in the four Countries



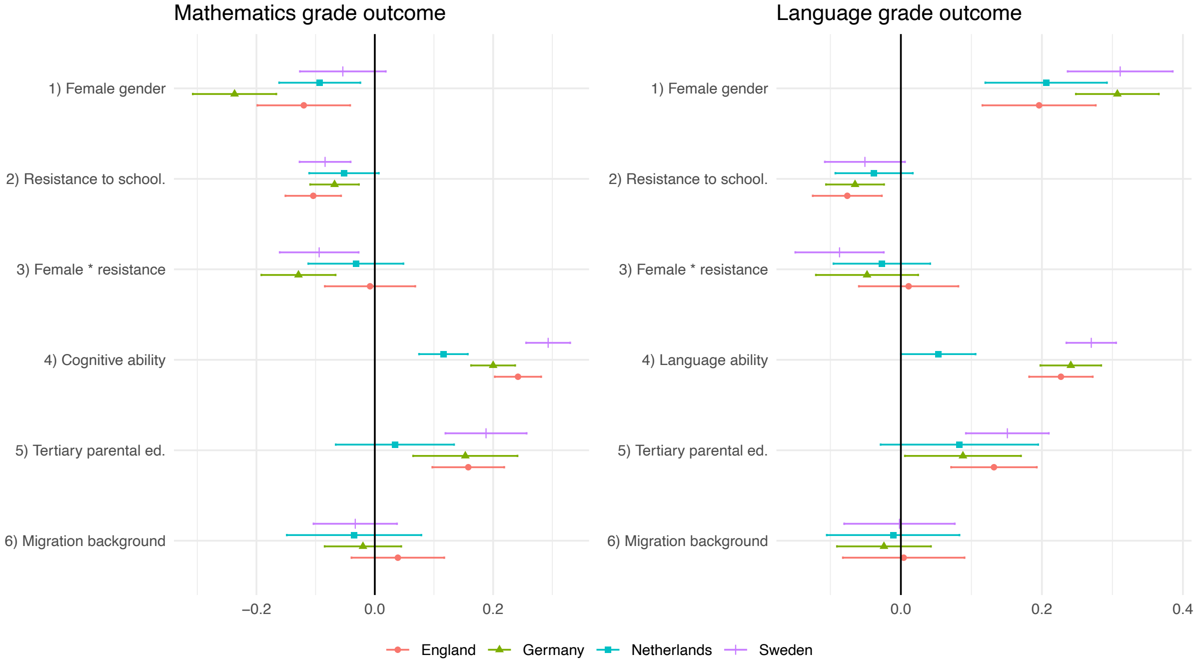
Note. Displayed are weighted two-level regression results after multiple imputation. Separate by outcome (left and right graph), the regression coefficients are displayed per predictor variable (rows). Regression coefficients are indicated by different colours and shapes per country. Horizontal lines indicate confidence intervals around the regression coefficients (+/- 1.96 * SE). Full results can be found in Table E in Appendix E.

Figure 2. Coefficient Plot for Models 2a (Research Question 2) in the four Countries



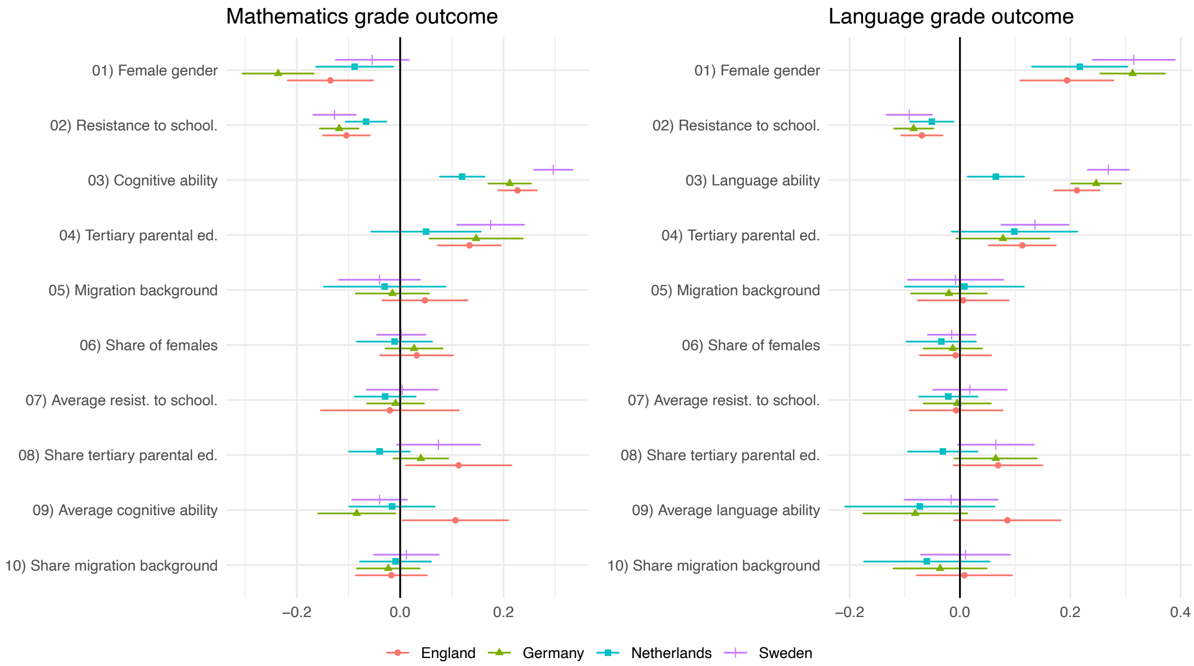
Note. Displayed are weighted two-level regression results after multiple imputation. Separate by outcome (left and right graph), the regression coefficients are displayed per predictor variable (rows). Regression coefficients are indicated by different colours and shapes per country. Horizontal lines indicate confidence intervals around the regression coefficients (+/- 1.96 * SE). Full results can be found in Table F1 in Appendix F.

Figure 3. Coefficient Plot for Models 2b (Research Question 2) in the four Countries



Note. Displayed are weighted two-level regression results after multiple imputation. Separate by outcome (left and right graph), the regression coefficients are displayed per predictor variable (rows). Regression coefficients are indicated by different colours and shapes per country. Horizontal lines indicate confidence intervals around the regression coefficients (+/- 1.96 * SE). Full results can be found in Table F2 Appendix F.

Figure 4. Coefficient Plot for Models 3a (Research Question 3) in the four Countries



Note. Displayed are weighted two-level regression results after multiple imputation. Separate by outcome (left and right graph), the regression coefficients are displayed per predictor variable (rows). Regression coefficients are indicated by different colours and shapes per country. Horizontal lines indicate confidence intervals around the regression coefficients (+/- 1.96 * SE). Full results can be found in Table G1, in Appendix G.