

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/201959>

Please be advised that this information was generated on 2020-10-30 and may be subject to change.

# Educational Systems and Gender Differences in Reading: A Comparative Multilevel Analysis

Margriet van Hek <sup>1,2,\*</sup>, Claudia Buchmann<sup>3</sup> and Gerbert Kraaykamp<sup>1</sup>

<sup>1</sup>Department of Sociology, Radboud University, PO Box 9104, 6500 HE Nijmegen, the Netherlands,

<sup>2</sup>Department of Sociology, Utrecht University, PO Box 80140, 3584 CH Utrecht, the Netherlands and

<sup>3</sup>Department of Sociology, The Ohio State University, 202 Townshend Hall, 1885 Neil Avenue Mall, Columbus, OH 43210, USA

\*Corresponding author. Email: m.vanhok@maw.ru.nl

Submitted January 2017; revised December 2018; accepted December 2018

## Abstract

Girls have a substantial advantage over boys in terms of reading performance throughout all OECD countries. This paper investigates whether the structure of a country's educational system is related to this gender inequality in reading performance. We assess whether standardization of educational curricula and the age at which students are selected into educational tracks affect boys' and girls' reading performance differently. To test our hypotheses, we employ data from all six Programme for International Student Achievement waves enriched with contextual information on countries' educational systems ( $N = 1,425,356$ ). Results show that in country-years with more standardized curricula overall reading performance is lower and the association between standardization and reading performance is more negative for boys than for girls. In countries with educational systems in which students are selected into educational tracks at later ages, gender differences in reading are larger because girls benefit more from late selection. These results indicate that educational policies at the country level are related not only to the reading performance of all students, but also to the underperformance of boys in reading.

## Introduction

Today, women obtain considerably more education than men in the vast majority of industrialized countries (DiPrete and Buchmann, 2013; Organisation for Economic Cooperation and Development (hereafter OECD), 2015; Van Hek, Kraaykamp, Wolbers, 2016). Educational researchers and policymakers therefore have become increasingly interested in understanding the lagging educational performance of boys (Buchmann, DiPrete and McDaniel, 2008; Garner, 2014; Van Hek, Kraaykamp and Pelzer, 2018). Insight into the factors related to boys' poorer reading

performance may help to understand boys' lower relative educational performance more generally, as reading is a fundamental skill for achieving educational success (Cheung and Andersen, 2003; Kraaykamp and Notten, 2016). After all, 'reading proficiency is the foundation upon which all other learning is built; when boys don't read well, their performance in other school subjects suffers too' (OECD 2015: p. 13).

Recent studies have found a sizable reading score gap favouring girls throughout all OECD countries (Stoet and Geary, 2013; OECD 2015). Figure 1 depicts this gender inequality in reading scores for the 37 countries we

include in our analyses (average scores across 2000–2015 Programme for International Student Achievement [PISA] waves), and it illustrates two important points. First, girls have higher average reading scores than boys in every country and, second, there is substantial variation across countries in the size of this gender gap. As Penner (2008: p. 140) stated, ‘there is no reason to believe that genetic factors involved in determining gender will vary across countries’. Instead, the cross-national variation in gender inequality in reading scores shown in Figure 1 is likely attributable to social or institutional factors that differ between countries.

Previous research has focused primarily on country-level measures related to gender inequality, such as female labour force participation or the prevalence of gender egalitarian attitudes, to explain cross-national variation in gender differences in educational performance (Penner, 2008; Else-Quest, Hyde and Linn, 2010; McDaniel, 2010; Stoet and Geary, 2013), but the results of these studies are inconclusive and contradictory. In contrast to much prior research, the current study focuses on differences in national educational systems as a possible explanation for the cross-national variation in reading scores between girls and boys (Ayalon and Livneh, 2013). The structure of educational systems is strongly related to students’ overall educational outcomes, but also to inequality therein (Hanushek and Wössmann, 2006; van de Werfhorst and Mijs, 2010). More specifically, educational systems structure students’ educational careers and their entry into the labour market (Kerckhoff, 2001), and by doing so, educational systems produce unequal opportunities for certain types of students (Hanushek and Wössmann, 2006; Montt, 2011; Bol *et al.*, 2014). Following prior work, van de Werfhorst and Mijs (2010) empirically classified educational systems along the features of *standardization* and *differentiation*. Although previous studies often have linked aspects of standardization and differentiation to educational inequality related to student’s family background, very few studies have investigated whether these structural features of educational systems are related to gender differences in educational performance across countries (Ayalon and Livneh, 2013; Scheeren, van de Werfhorst and Bol, 2018). Yet a decade ago, in a review article on gender inequality in education, Buchmann, DiPrete and McDaniel (2008) called for research on how structures, institutions, and practices of education affect gender inequality in educational outcomes. We heed this call and ask: *to what extent are the levels of standardization and differentiation of a country’s educational system related to girls’ and boys’ reading performance?*

We improve upon prior research in several ways. Our focus on reading performance enables us to investigate a domain with large and direct consequences for students’ educational careers. We examine how the structure of a country’s educational system relates to the reading performance of students generally, as well as to the lower performance of boys, which is a growing concern in industrialized societies. More specifically, we investigate whether the degree of standardization and differentiation of a country’s educational system is related to girls’ and boys’ reading performance. To do so, we pool all six waves of the PISA, conducted by OECD every 3 years between 2000 and 2015 and employ advanced four-level regression models to analyse information on 1,425,356 students from 59,001 schools in 37 countries. While our primary focus is on reading performance, we conduct parallel analyses for math performance and discuss these findings to provide broader and more robust insights into the relationships between dimensions of standardization and differentiation of educational systems and student performance.

## Theoretical Framework

### Features of Educational Systems: Standardization and Differentiation

Prior studies commonly distinguished between two features of educational systems: standardization and differentiation (Buchmann and Park, 2009; van de Werfhorst and Mijs, 2010; Montt, 2011). First, educational systems vary in their level of standardization (Bol and van de Werfhorst, 2013), which can be defined as either standardization of output, or standardization of input. Standardization of output requires that all students possess a similar level of knowledge by the end of an educational programme which is usually measured by central examinations. Standardization of input exists when governments control the organization of education by setting regulations for school policies and practices, for example, by prescribing the curricula schools should offer, such that individual schools and teachers have little room to deviate from these regulations (Montt, 2011). Herein, we examine level of standardization of educational curricula because we believe this aspect best captures processes of student learning. It specifically refers to the degree to which school leaders and teachers have the freedom to modify course offerings, course content, and textbooks; a similar definition was used by Montt (2011). As noted by Stevenson and Baker (1991), in the absence of state control over school curricula, teachers tend to modify learning processes according to

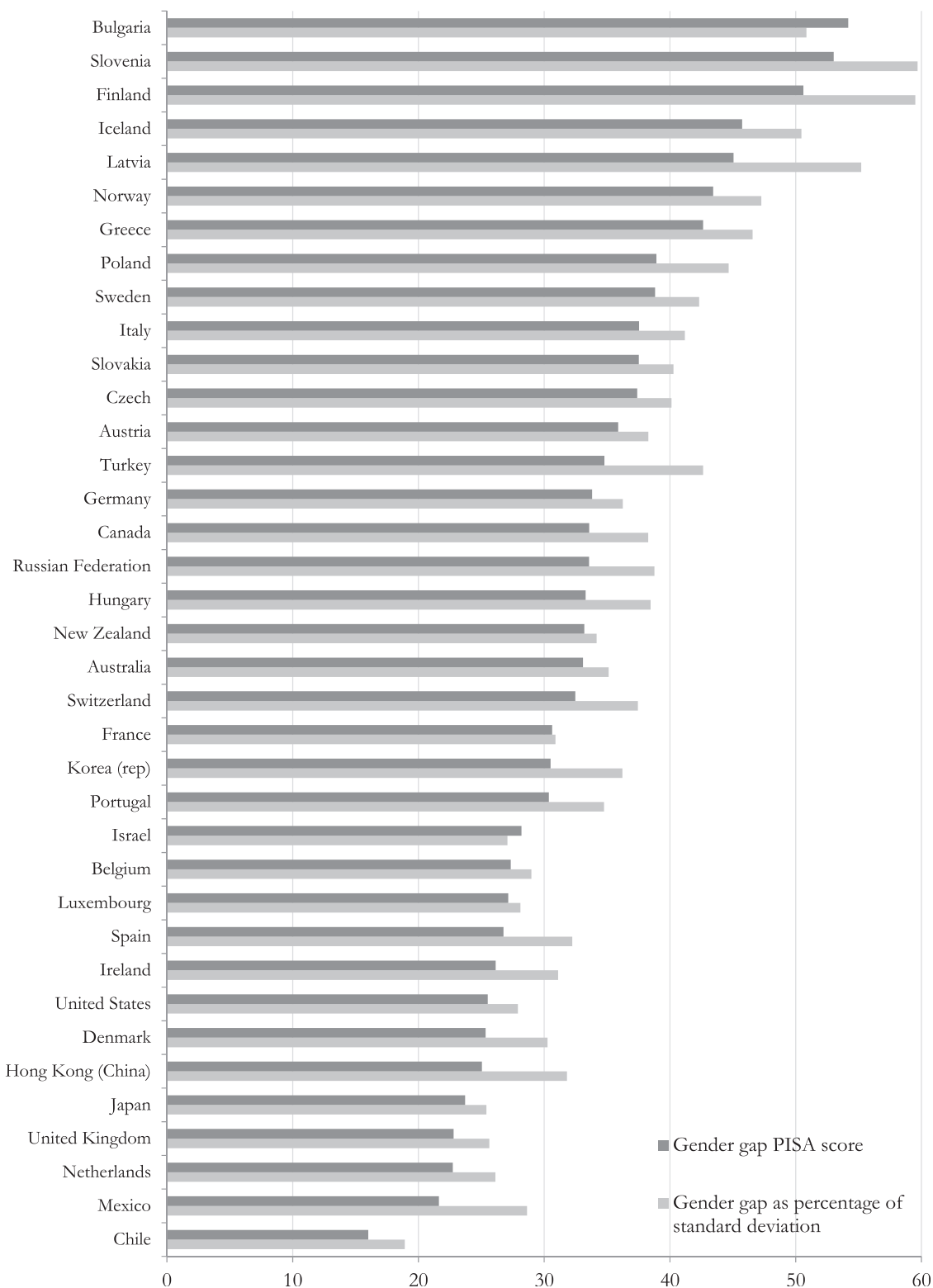


Figure 1. Gender gaps in reading scores in 37 countries

Source: 2000–2015 PISA waves; 1,425,356 students; 59,001 schools; 204 country-year combinations; 37 countries.

the needs of their students', thereby forging a link between student characteristics and their exposure to knowledge. As we explain below, teachers' inability to modify learning processes in highly standardized educational systems may be particularly relevant to boys' reading performance.

The level of differentiation within an educational system is indicated by the age at which students are selected for different educational tracks in secondary education in a country (Bol and Van de Werfhorst, 2013). This usually refers to the age at which students finish primary or lower secondary education and are allocated to specific educational tracks in (higher) secondary education, on the basis of their prior performance and/or teacher evaluations (Buchmann and Park, 2009). Such tracks in secondary education prepare students for post-secondary pathways including employment, vocational training, or university. Generally, tracks that prepare students for university enrolment are more prestigious and demanding than tracks that lead to lower levels of post-secondary education or direct entry into the labour market (Kerckhoff, 2001; van de Werfhorst and Mijs, 2010). Different educational tracks are often located in different secondary schools (or school buildings) and mobility between tracks is often difficult and thus rare.

In highly differentiated educational systems students are selected into different tracks and/or schools at an early age. Conversely, in undifferentiated systems there is no or late tracking; secondary school students with different ability levels then are situated within the same school and experience rather similar educational trajectories; not until tertiary education students attend distinct, more or less demanding, educational programmes.<sup>1</sup>

### Standardization of Educational Curricula and Gender Differences in Reading Performance

Standardization of educational curricula is an important feature of a country's educational system, as general regulations implemented by central or regional governments serve to constrain schools' and teachers' freedom in choosing course offering, course content and textbooks (Bol and Van de Werfhorst, 2013). Typically, these rules are implemented to exercise some quality control over learning processes in secondary schools. Research on the level of control that schools and teachers have over their educational curriculum is limited, however, as prior studies tended to focus more on school autonomy in terms of finances and teacher selection (Fuchs and Wössmann, 2004). It therefore is an open question as to how the degree of standardization in an

educational system is related to reading achievement for boys and girls.

Generally, it is expected that standardization of learning environments reduces social inequalities (van de Werfhorst and Mijs, 2010; Bol and Van de Werfhorst, 2013). Previous research has found that the differences in performance scores between students from lower and higher social class backgrounds are smaller in more standardized educational systems (Bol *et al.*, 2014; van de Werfhorst and Mijs, 2010). This could be due to the standardization of educational input between schools with lower and higher socio-economic student compositions (Montt, 2011). Looking specifically at standardization of educational curricula, Montt (2011) found no association with socio-economic achievement inequality. There are however good reasons to expect that far-reaching standardization of educational curricula and textbooks has different consequences for gender inequality in reading performance. In the only study linking standardization to gender gaps in education to date, Ayalon and Livneh (2013) showed that boys' math performance is harmed more than girls' math performance by a high level of standardization, as measured by between-teacher instructional variation. We argue that standardization is also more detrimental for boys' reading performance than girls', resulting in a larger female advantage in reading performance in countries with highly standardized educational systems.

As a starting point to understand why standardization may have negative consequences for boys' reading performance, it is well established that students' attitudes, motivations, and interests play an important role in their reading performance (Meece, Glienke and Brug, 2006). Prior research reports that boys experience less reading enjoyment and are less frequent readers in their free time than girls (Christin, 2012; Clark and Trafford, 1995). For example, the OECD (2015) found that boys read less often for enjoyment than girls in all but one OECD country (Korea). Additionally, research suggests that the relationship between reading interest and reading performance is stronger for boys than for girls. For instance, Logan and Medford (2011) showed that intrinsic reading motivation is more strongly associated with reading performance for boys than for girls. Also, Oakhill and Petrides (2007) found that boys have better reading comprehension when they find texts interesting, while this relationship is weaker for girls. As a result, especially boys' reading performance may be harmed in highly standardized educational systems (Ayalon and Livneh, 2013), because mandatory course content (texts) may not suit their interest, and teachers do not have the

flexibility to select reading materials that are tailored to students' interests.

Additionally, Wössmann (2003) argued that local schools and teachers are generally better able to assess students' needs than governmental institutions, as teachers and schools have personal knowledge about their students (see also Stevenson and Baker, 1991). As Montt (2011: p. 52) stated, teacher control over curricula may allow schools to 'meet the particular needs of low-achieving students in their local context, potentially reducing dispersion in achievement.' Within standardized curricula, however, teachers are required to use uniform course materials, methods, and textbooks. This uniformity may be least problematic for motivated students, who adapt well to standard educational methods, many of whom are girls. Conversely, the use of uniform teaching methods and courses likely is more detrimental to students who are less interested and experience less enjoyment in reading, many of whom are boys. To the degree that boys' reading ability is more dependent on motivation (Oakhill and Petrides, 2007), it may be extremely difficult for teachers within standardized educational systems to individualize boys' reading instruction in order to meet their needs. For these reasons, we hypothesize that *boys' reading performance is more negatively affected by standardized educational curricula than girls' reading performance (Hypothesis 1).*

### Differentiation in Educational Tracks and Gender Differences in Reading Performance

Research has yet to reach consensus about how the level of differentiation within a country's educational system affects students' educational performance. In an early meta-analysis, Slavin (1990) concluded that early selection into tracks had no effect on students' educational achievement. Using a difference-in-difference design, Hanushek and Wössmann (2006), however found that reading scores of students were somewhat lower in countries with highly differentiated systems. Some previous studies have linked the age at which students are selected into tracks to gender inequality in educational performance and attainment. Hadjar and Buchmann (2016) showed that the female favourable gender gap in educational attainment was larger in educational systems with later tracking than those with tracking at earlier ages. Jürges and Schneider (2011) and Pekkarinen (2008) claimed that selection into tracks at an early age disadvantages boys. Both studies derived insights from research that demonstrated that younger children in a class (i.e., children whose birthday is late in a school year) have a smaller chance of being allocated to

academic tracks (Schneeweis and Zweimüller, 2009). Crawford, Dearden and Meghir (2007) argued that this is due to their lower educational experience and lower maturity. If so, this may also have consequences for the inequality in reading performance of girls and boys. The finding that boys mature later than girls is well established (Tanner, 1978), and although social and cultural contexts may play a role in how developmental trajectories manifest themselves, biological causes of this phenomenon suggest that this male–female maturity gap is universal (Lim *et al.*, 2015). As a result, the maturity gap favouring girls may lead to more girls in the higher tracks of secondary education, and boys being over-represented in lower tracks in educational systems that track students at an early age. Indeed, studies from several European countries found that girls are more likely to be allocated to higher tracks in secondary education than boys (Ayalon and Shavit, 2004; Pekkarinen, 2008).

Gender inequality in track placement may reinforce gender inequality in reading performance, first because classroom homogeneity is shown to enhance the achievement of students in higher school tracks and hinder achievement in lower tracks. Huang (2009) explained this divergent effect by pointing to lower quality instruction, less qualified and experienced teachers and slower learning pace in lower tracks of highly differentiated educational systems. Moreover, in highly differentiated countries lower performing students have fewer opportunities to interact with high-performing students. So, if boys are more likely to be placed in a lower educational track, due to their developmental lag relative to girls, such placement may hinder their reading performance; conversely, if girls are more likely placed in higher tracks, their reading performance may benefit.

Another reason that differentiation at an early age may be detrimental to boys' reading performance is related to how norms regarding masculinity differ between school tracks. In early differentiating countries, students in different educational tracks are often situated in different schools with, as student background correlates with track placement, a lower or higher socio-economic composition. Legewie and DiPrete (2012) showed that in Germany, boys' disadvantage in reading is larger in schools with a large proportion of students from low socio-economic status backgrounds. They explicitly linked this phenomenon to the prevailing norms of masculinity arguing that, in these schools, boys gain status through sports, high-risk behaviours, and opposing authority, so non-academic norms are more common among boys. Contrastingly, in schools with a high socio-economic composition, school norms for boys are more directed at academic performance. For girls,



femininity norms tend to align more with academic performance, so variation between different school tracks for them is likely smaller. Boys' lower track placement, as a consequence of early tracking, may therefore hamper their reading performance because they are more likely to be exposed to non-academic masculine norms in lower educational tracks (van de Werfhorst and Mijs, 2010).

In sum, we expect that in countries with highly differentiated educational systems, where students are selected in tracks at a relatively young age, boys are more often placed in lower level tracks where general performance levels are lower and non-academic masculine norms are more prevalent. As gender differences in cognitive and motivational development are believed to weaken after the age of 15 (Halpern, 2013), we expect that the developmental penalty for boys is lower in countries with less differentiated educational systems where tracking occurs later (or not at all). This leads us to hypothesize that: *boys' reading performance is more negatively affected by early tracking than girls' reading performance (Hypothesis 2).*

## Data and Measurements

### Data

We employ data from all waves of PISA (2000, 2003, 2006, 2009, 2012 and 2015) containing information from 15-year-old students from 37 countries.<sup>2</sup> Through a two-stage selection procedure, in every country first schools are sampled, after which 15-year olds in those schools are randomly selected. PISA data are generally considered to be of very high quality (Else-Quest, Hyde and Linn, 2010). The pooled PISA data set provides us with information on 1,525,604 individual students across six waves and 37 countries.

### Measurements

Students' score on the PISA reading test indicates their *reading performance* (PISA, 2009). PISA provides measures of students' reading performance using a method based on Item Response Theory (Mislevy and Sheehan, 1987). Instead of a single measure, five 'plausible values' for a students' reading ability are provided. We estimate our models for each of the five plausible values of *student's reading performance* separately. Next, we merged results to arrive at correct estimates and standard errors (see OECD [2009] for details of this procedure).

*Male* is coded 1 for boys and 0 for girls. We further include four individual-level control variables. First, *parental educational level* (mean centred) is measured in

years of education associated with the ISCED level of the highest educated parent.<sup>3</sup> *Parental cultural resources* are indicated by the number of books present in the student's family home. Response options were 0–10 books (0), 11–100 books (1), 101–250 books (PISA 2000) or 101–200 books (other waves) (2), and more than 250 books (PISA 2000) or more than 200 books (other waves) (3). We also include a control for *students' age* (mean centred) indicated by year and month of birth (ranging from 15.167 to 16.420 years) as some studies showed that older students perform slightly better in school (Schneeweis and Zweimüller, 2009). Finally, since immigrant students tend to have a lower reading performance in most industrialized countries (OECD, 2012), we control for *students' immigrant status* measured with three dummy variables specifying native (born in the country to native-born parents), first generation immigrant (born outside the country to foreign born parents), and second generation immigrant (native born to immigrant parents). The 100,248 students with missing values on one or more of the individual variables were removed from the data set. The final data set consists of 1,425,356 students nested in 204 country-year combinations<sup>4</sup> and 37 countries.

At the country-year level, the level of *standardization* of the educational system is indicated by the degree to which school curricula are nationally or regionally standardized. We constructed this using the PISA school questionnaire. In all waves PISA asked school principals: 'Regarding your school, who has considerable responsibility for the following tasks': (i) choosing which textbooks are used, (ii) determining which courses are offered, and (iii) determining course content. For PISA 2000 and 2003, we determined for each country the percentage of school principals that reported these matters 'were not a school responsibility.' For PISA 2006, 2009, 2012 and 2015, we took the percentage of school principals who reported that local/regional or national education authorities were responsible. The average score of these three measurements indicates level of standardization in a country; a higher score refers to a higher level of standardization. Note that this measure of standardization varies over waves and over countries.

The level of differentiation of an educational system is derived from Bol and Van de Werfhorst (2013) and is measured at the country level; the younger students' *age of selection* into educational tracks in a country, the higher the level of differentiation. The countries with the highest level of differentiation are Austria and Germany that select students into educational tracks from age 10. Countries with no differentiation are assigned the age at which students leave secondary education, typically at

**Table 1.** Descriptive statistics

	Min.	Max.	Mean	SD
Plausible value reading 1	0	1083.506	493.368	96.883
Plausible value reading 2	0	1144.371	493.370	96.970
Plausible value reading 3	0	1002.878	493.377	96.954
Plausible value reading 4	0.122	1097.896	493.314	96.959
Plausible value reading 5	0	1065.790	493.358	96.936
Male	0	1	0.495	0.500
Parental education <sup>a</sup>	0	18	13.153	3.269
Books in household	0	3	1.494	0.984
Age of student <sup>a</sup>	15.167	16.420	15.779	0.290
Immigrant: native	0	1	0.900	0.300
Immigrant: second generation	0	1	0.051	0.219
Immigrant: first generation	0	1	0.049	0.217
Age of selection	0	6	4.191	1.977
Standardization	0.007	1	0.357	0.241
HDI <sup>a</sup>	0.675	0.949	0.859	0.056
Gender equality <sup>a</sup>	1.186	2.902	2.307	0.339

<sup>a</sup>These variables are (grand)mean centred in the analyses.

Source: 2000–2015 PISA waves; 1,425,356 students; 59,001 schools; 204 country-year combinations; 37 countries.

age 16. We subtract the minimum (10) from this variable so that it ranges between 0 and 6 (see Appendix A).

Following prior research, we control for level of prosperity and level of gender equality at the country-year level. Because country-level factors may also be related to gender gaps in reading scores (Ayalon and Livneh, 2013), we add interaction terms for both aspects with male. We used the Human Development Index (HDI) from the United Nations Development Programme (UNDP) as indicator for a country's level of prosperity. This HDI measure refers to general human development in three dimensions: health, knowledge, and standard of living (Malik, 2013); original values are multiplied by 10 for ease of interpretation. We used data from the World Value Survey (WVS) and European Social Survey (ESS) to determine a country's level of gender equality.<sup>5</sup> In both WVS and ESS respondents reported whether they agreed with the statement 'Men should have more right to a job than women when jobs are scarce' answer categories were: agree (0), neither agree nor disagree (1), or disagree (2). The aggregated country-year average indicates that the level of gender equality with a higher score reflecting more gender equality. Both control variables are mean-centred at the country level. Descriptive statistics for individual and contextual variables are presented in Table 1. Descriptive statistics of the contextual variables (not centred) per country are available in Appendix A.

## Analyses

### Analytical Approach

We employ multilevel regression models in *R* to test our hypotheses. Based on Schmidt-Catran and Fairbrother (2016), we estimate four level models in which students (level-1 units) are nested in schools (level-2 units) that are nested in country-year combinations (level-3 units) that are nested in countries (level-4 units), and allow the effect of male to vary over all these levels.<sup>6</sup> Although we test hypotheses at the individual and (year-)country level, the structure of PISA data requires that we control for students being nested in schools; dealing with school level variation leads to more accurate estimates as effects of standardization and differentiation are possibly affected by processes in schools (Bol *et al.*, 2014). As PISA prescribes, we use the student weight provided by PISA in our models.

In Table 2, we first estimate a null-model that shows how much of the variation in students' reading scores is due to their nesting in schools, country-year combinations and countries. Model 1 shows the uncontrolled effect of male; the mean difference between girls' and boys' reading scores. Model 2 includes all individual and contextual variables. The main effects of the contextual characteristics indicate how they affect reading performance of all students. In model 3, we interact the characteristics of educational systems with male; these estimates show to what degree standardization and age of selection affect girls' and boys' reading scores differently.

### Results

In Table 2, the null-model shows an intraclass-correlation of 12.3 per cent for the country variance parameter, 4.7 per cent for country-year, and 22.5 per cent for the school variance parameter. Students' average reading scores are thus dependent on the year and country in which they live and the school they attend; this justifies multilevel modelling.

In model 1, we observe that boys overall have significantly lower reading scores than girls ( $b = -28.399$ ). The variances in the slope of male ( $\sigma^2 = 5.288$ ;  $\sigma^2 = 7.905$ ) illustrate the variation in the effect of male on reading performance between country-year combinations and countries; this confirms that the difference between girls' and boys' average reading performance varies across these two levels. In model 2, all individual and contextual characteristics are included as main effects. Individual variables behave as one would expect: students with highly educated parents and parents with cultural resources as well as older and native students have higher reading performance scores than their



**Table 2.** Linear multilevel models predicting reading scores

	Model 0	SE	Model 1	SE	Model 2	SE	Model 3	SE
Individual variables								
Intercept	491.114***	4.739	505.489***	4.886	453.958***	7.499	447.443***	7.919
Male 0/1			-28.399***	1.408	-28.153***	1.277	-16.931***	3.177
Parental education					1.887***	0.062	1.887***	0.062
Books in household					16.801***	0.133	16.801***	0.133
Age					9.301***	0.278	9.301***	0.278
Immigrant: first generation (ref)								
Second generation					16.059***	0.686	16.060***	0.686
Native					25.035***	0.569	25.034***	0.569
Country variables								
Standardization					-27.527***	6.255	-21.784**	6.711
Age of selection -10					3.126***	1.493	4.191**	1.580
HDI					25.854*	3.818	22.978***	4.056
Gender equality					-18.614***	5.407	-16.449**	5.795
Cross-level interactions								
Standardization × male							-8.128*	3.281
Age of selection × male							-1.983**	0.619
HDI × male							4.079*	1.893
Gender equality × male							-3.291	2.891
Variance statistics								
Individual variance	139.554		135.722		132.436		132.436	
School variance	51.984		50.300		43.930		43.932	
Country-year variance	10.783		11.190		11.387		11.258	
Country variance	28.365		29.228		16.608		17.176	
Slope male school			23.958		23.365		23.364	
Slope male country-year			5.288		5.427		5.228	
Slope male country			7.905		7.041		6.177	

Note: Linear multilevel regression models; LME in R.  $\sim P < 0.1$ .

\* $P < 0.05$ ;

\*\* $P < 0.01$ ;

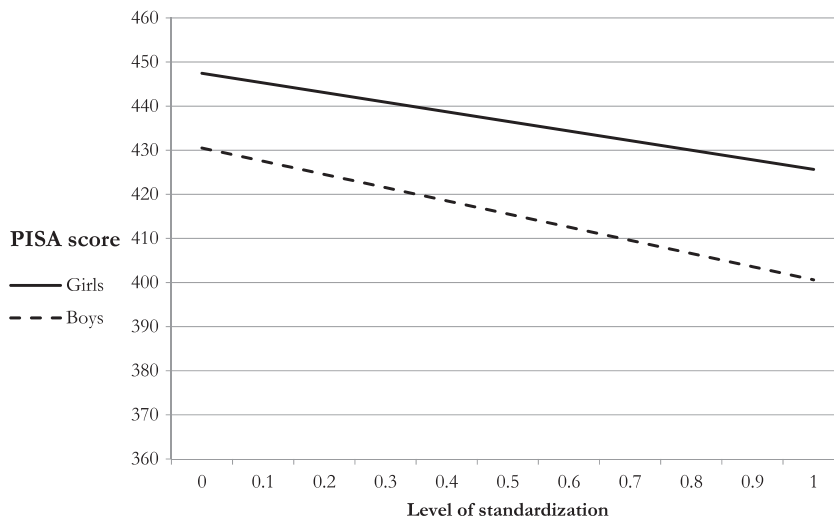
\*\*\* $P < 0.001$ .

Source: 2000–2015 PISA waves; 1,425,356 students; 59,001 schools; 204 country-year combinations; 37 countries.

counterparts. On the country-year level, standardization of educational curricula has a negative relationship with pupils' overall reading performance ( $b = -27.527$ ). In contrast, the age at which countries select students into tracks is positively related to reading scores ( $b = 3.126$ ), meaning that late (or no) differentiation is beneficial to students' overall reading performance. In their magnitude, the overall effect of standardization is somewhat larger than the effect of differentiation (given their range). In more prosperous countries, as measured by HDI, students have higher reading performance ( $b = 25.854$ ). A country's level of gender equality is negatively related to students' reading scores ( $b = -18.614$ ), but additional robustness checks show that this effect loses statistical significance when Japan,

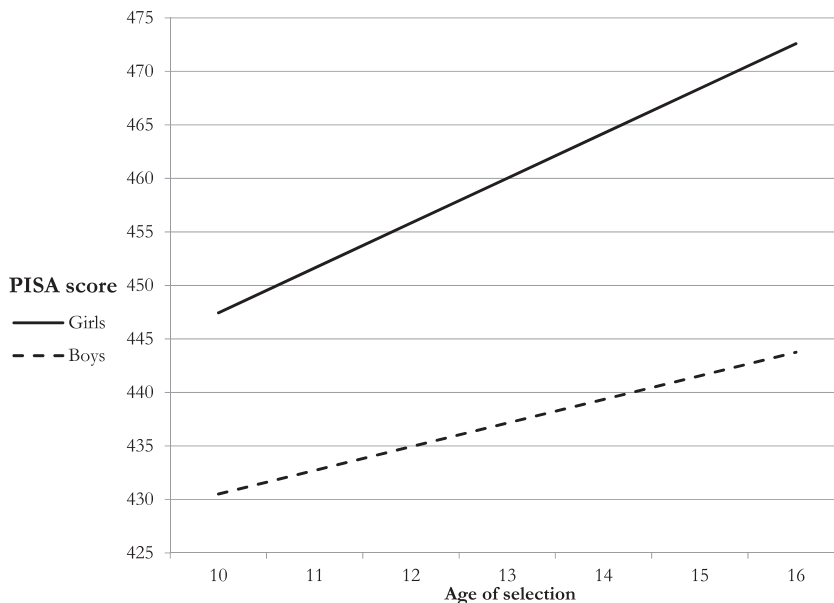
Turkey, Chile, Mexico, and Bulgaria are excluded from the analyses.

In model 3, we include cross-level interactions that together reduce the variance in the slope of male by 3.7 per cent on the country-year level and by 12.3 per cent on the country level.<sup>7</sup> Recall that standardization is measured on the country-year level and differentiation is measured on the country-level. The cross-level interactions represent the difference in the effect of standardization and differentiation of a country's educational system between girls and boys; main effects apply to girls (coded 0 on male). Figures 2 and 3 visualize these gendered effects, and Appendix B shows the associations between standardization and differentiation and the gender gap in reading performance.



**Figure 2.** Performance scores with level of standardization for girls and boys (Model 3: Table 2)

Source: 2000–2015 PISA waves; 1,425,356 students; 59,001 schools; 204 country-year combinations; 37 countries.



**Figure 3.** Performance scores with age of selection for girls and boys (Model 3: Table 2)

Source: 2000–2015 PISA waves; 1,425,356 students; 59,001 schools; 204 country-year combinations; 37 countries.

First, as hypothesized, in model 3 we see that a highly standardized curriculum in a country-year affects boys' reading scores ( $b = [-21.784 - 8.128] -29.912$ ) more negatively than girls' ( $b = -21.784$ ). Also in

Figure 2, we observe a steeper negative slope for boys' reading scores than for girls' with rising standardization of curricula. This implies that the female-favourable gender gap in reading performance increases when

educational curricula are more standardized and teachers have less freedom to individualize reading instruction in order to meet students' needs. Gender differences in reading performance range from a 16.931 advantage for girls in countries with the least standardized educational systems, to 25.002 in countries with highly standardized educational curricula; this more than eight point difference approximates the gap in PISA reading scores between Latvia and the Switzerland. So, in line with hypothesis 1, boys are more than girls harmed by a lack of autonomy among teachers and schools.

Second, we hypothesized that boys would be more negatively affected by early differentiation than girls. Our results in Table 2 show the opposite: early differentiation is actually more detrimental for girls' reading scores than boys'. Model 3 and Figure 3 show that girls' reading scores increase ( $b=4.191$ ) more than boys' reading scores ( $b=[4.191-1.983] 2.208$ ) as the age of selection in a country increases. In fact, additional analyses show that the main effect of country's age of selection is not statistically significant for boys. So, in countries that select students at a younger age (more differentiated countries), the gender gap in reading is smaller due to girls' lower reading performance. We therefore reject Hypothesis 2. Finally, living in a prosperous country-year affects boys' reading scores more positively than girls' ( $b=4.079$ ), and a country's level of gender equality does not significantly affect the difference between girls' and boys' reading scores.

### Robustness Analyses

We performed several robustness analyses to broaden the scope of our paper and shed light on some of the underlying theoretical mechanisms for male–female differences in reading performance. In this section, we discuss additional analyses we did for math scores, presented in Appendix C, and results when students' grades are accounted for, presented in Supplementary Appendix D, where we also present and discuss the results of six other robustness analyses.

We first test whether our results are specific to reading, or also apply to students' math performance. Models 2 and 3 in Table C1 of Appendix C show that standardization of educational curricula is related to overall lower math performance scores. Standardization, however, does not affect the gender gap in math performance; the interaction is not statistically significant ( $b=0.924$ ). This is also visualized in Figure C1. So, whereas we find negative effects of standardization on both students' overall reading and math performance, we find that standardization enlarges the

female favourable gender gap in reading, but is unrelated to the male favourable gender gap in math. It could well be that standardization asserts a less negative effect on boys' math performance because math is considered a more masculine field in which boys (are socially allowed to) have more interest. This difference in findings more generally supports the idea that different mechanisms underlie gender gaps in reading and gender gaps in math. For example, whereas social norms about gender atypical behaviour are more often brought forward to explain boys' lack of interest in reading, insecurity about math abilities has mainly been related to girls' disadvantage in math (DiPrete and Buchmann, 2013).

Second, model 3 of Table C1 shows no significant main effect of the age of selection in a country on students' overall math performance. This result contradicts earlier research on this topic (Montt, 2011). Since we employed all six available PISA waves this result reflects the most comprehensive evidence currently available on this issue. In concordance with the results for reading performance, we find that girls' math performance profits somewhat more from an later age of selection (i.e., less differentiation) ( $b=3.646$ ,  $P=0.103$ ) than boys' ( $b=[3.646-1.572] 2.074$ ). Early differentiation in a country's educational system is thus more negatively related to girls' reading and math performance, than to boys' reading and math performance.

Lastly, in Table D3 in Supplementary Appendix D controls are included for whether students are below, at or above the national modal grade (ranging from  $-3$  to  $3$ ). Model G shows that the effect of male is partly interpreted by students' grade position, which is consistent with the fact that boys more often repeat a grade and girls more often skip a grade. As shown in model H, the main effects of standardization and differentiation, as well as their interaction with male, are in the same direction and remain (marginally) significant when accounting for students' deviation from the national modal grade. This seems plausible since student's grade position likely to a large extent reflects earlier educational performance.

### Conclusion and Discussion

Employing information on girls' and boys' reading performance from all six waves of PISA, this study is the first to assess how two important features of a country's educational system, standardization and differentiation, are related to gender differences in students' reading performance scores. In light of the fact that men lag behind women in educational achievement and attainment in

large parts of the world today, this explicit focus on boys' disadvantage in reading, and how countries' educational policies are related to this lower reading performance, is of great importance.

First, it is important to acknowledge that standardization of educational curricula and early differentiation are related to lower overall reading performance scores in OECD countries. Our results clearly underscore the general importance of educational structures in the schooling of adolescents across the world. In countries where curricula and textbooks are more tailored to individual students, students earn higher reading scores than in countries where a one-size-fits-all ideology is prevalent. Additionally, our results support the idea that the moment of selection into different educational tracks should not be too early in a students' life; early tracking generally leads to lower reading (and math) scores.

Second, our study indicates that differences between girls and boys in reading performance in secondary school are substantially related to standardization and differentiation of a country's educational system. We find that standardization of educational curricula is more negatively related to boys' reading performance than girls' reading performance. In countries where governmental regulations largely determine school curricula and learning materials, gender gaps in reading scores are even more to the advantage of girls. These results are robust in that they hold when we control for standardization of output (central examinations) in a country (see [Supplementary Appendix D1](#)). They are consistent with theoretical notions that imply that restrictions placed upon schools and teachers to act upon students' individual needs are especially detrimental for boys. This may be because 15-year-old boys are often poorer readers who need more personalized attention, or because boys are less motivated readers who not develop their reading competencies even when they are obliged to spend time reading in class. More in depth research might be directed at the further implications of standardization for girls' and boys' motivation and learning opportunities in reading (and math), to fully probe the mechanisms behind this gender gap in learning.

We also find that in countries in which students are selected into educational tracks at later ages, gender differences in reading are larger because girls benefit more from late selection. While this finding does not confirm our hypothesis, it does align with other research that finds a larger female advantage in educational attainment in later tracking relative to earlier tracking educational systems ([Hadjar and Buchmann, 2016](#); [Scheeren, van de Werfhorst and Bol, 2018](#)). Perhaps it is the case that in early differentiating countries, selection on

performance restricts students' exposure to, and inspiration from high-ability students ([Slavin, 1990](#)), and this may be especially detrimental for studious girls ([Jackson and Dempster, 2009](#)). Future investigations of why girls and boys are not equally affected by early selection into educational tracks would be valuable and may dig deeper into aspects of class composition and school-related norms in various tracks. Because of the cross-sectional nature of the PISA data, we were not able to test the claim that low reading performance could be both a source and an outcome of low track placement. Longitudinal investigations that assess girls' and boys' achievement during their school career in different educational systems would provide a more direct test of how girls' and boys' reading achievement differs as a result of track placement. Ideally, such investigations would also include information on students' primary schools as the skills of students entering (different tracks of) secondary education may already be shaped by structures in primary education. This however will be difficult in a cross-national design that allows for variation in educational systems characteristics.

More generally, features of educational systems not only seem to affect inequality between students from different social backgrounds, but also between girls and boys. A focus on how various types of students are affected by the structures of educational systems therefore should be a central focus for future educational research. We find that in countries with more standardized educational curricula, gender inequalities are larger, whereas others found that standardization is associated with lower inequality between students from lower and higher socio-economic backgrounds ([Montt, 2011](#); [Bol et al., 2014](#)). In addition, we conclude that early tracking is linked to less gender inequality in reading performance, whereas earlier studies found that in differentiated school systems differences between students from low and high socio-economic backgrounds are larger ([Hanushek and Wössmann, 2006](#)). Taken together, these findings suggest that structural features of educational systems do not simply enlarge or reduce inequality between student subgroups. A promising direction for future research therefore could be to address the reading performance of other vulnerable groups, such as children from immigrant backgrounds or single-parent families.

Dealing with information from a wide variety of countries is challenging, and like all cross-national research, this study has limitations. First, our measurement of differentiation of a country's educational system refers to data from 2009 ([van de Werfhorst and Mijs, 2010](#); [Bol and van de Werfhorst, 2013](#)). Although early/

late tracking is likely a relatively stable characteristic of a country's educational system, it is possible that changes within countries occurred over time. Additionally, in constructing a differentiation measure researchers were understandably limited in dealing with all countries of PISA (Bol and Van de Werfhorst, 2013). So, in our analyses we could not include several less economically developed OECD-partner countries from the PISA samples. We recommend that future studies update the information on country's educational differentiation and extend this classification to more non-OECD countries. Moreover, working with large-scale comparable international information clearly poses limitations with respect to a thorough investigation of mechanisms. While we made an effort to consider mechanisms in several robustness analyses, our ability to do so was limited. At any rate this study is one of the first that links features of educational systems to gender inequality in reading performance, and by doing so, it provides future research a relevant foundation from which to further develop theory on this issue and establish more precisely the processes through which girls and boys may be differently affected by standardization of educational curricula and differentiation in secondary education.

Our study comprehensively assessed the relationship between the structure of a country's educational system and girls' and boys' reading performance. As girls' advantage in reading performance and women's lead in educational attainment continues (DiPrete and Buchmann, 2013; Van Hek, Kraaykamp and Wolbers, 2016), it is important to gain insight in whether and how countries' educational institutional arrangements contribute to opportunities and outcomes for both girls and boys. Our conclusion that girls and boys are differentially affected by features of educational systems implies that some countries do a better job in providing environments conducive for learning by all students. The central finding that standardization of course offerings, curricula, and reading materials is detrimental to the average reading scores of all students, but especially those of boys, is a meaningful starting point for future research.

### Notes

1 In undifferentiated educational systems within-school tracking may exist. Our main focus is not on this form of tracking, as internal differentiation 'is hard to capture in cross-national research' (van de Werfhorst and Mijts, 2010: p. 408). In [Supplementary Appendix D](#), we report on a robustness check in which we consider ability tracking within schools.

- 2 There are 52 countries available in both PISA and the Bol and Van de Werfhorst (2013) data set of which 15 countries lacked sufficient information on one or more of the country-level control variables. Not all countries are present in every PISA wave.
- 3 PISA 2000 only provided information on parents' ISCED level, not the years of schooling associated with parents' ISCED level. For students from the 2000 wave, we therefore took the average years of education per ISCED level based on information from the other waves.
- 4 We omitted Korea in 2000, Japan in 2000, and the United States in 2006 because all students in that country-year combination had missing values on at least one individual-level control variable.
- 5 Data were derived from WVS 1999–2004, 2005–2009, and 2010–2014 and ESS 2004, 2008, and 2010. If countries were present in both surveys, we preferred WVS. The years that were available differed per country; we inter- and extrapolated missing years. For Luxembourg, Latvia, Iceland, and Italy, we only had one value and assigned this to every year. For our set of countries and survey years, only one item was available to indicate countries' egalitarian gender norms. We prefer using this attitudinal item for gender equality since we consider it is a direct indicator of gender norms. Other measures of country-level gender equality, such as women's labour market participation, are possibly affected by other country variables (such as the economic necessity for women to work).
- 6 A replication package can be found in the [Supplementary Material](#).
- 7 The results are not driven by outliers.

### Supplementary Data

Supplementary data are available at *ESR* online.

### References

- Ayalon, H. and Livneh, I. (2013). Educational standardization and gender differences in mathematics achievement: a comparative study. *Social Science Research*, 43, 432–445.
- Ayalon, H. and Shavit, S. (2004). Educational reforms and inequalities in Israel: the MMI hypothesis revisited. *Sociology of Education*, 77, 103–120.
- Bol, T. and Van de Werfhorst, H. (2013). *The Measurement of Tracking, Vocational Orientation, and Standardization of Educational Systems: A Comparative Approach*. AIAS: GINI Discussion Paper 81.
- Bol, T. et al. (2014). Curricular tracking and central examinations: counterbalancing the impact of social background on student achievement in 36 countries. *Social Forces*, 92, 1545–1572.

- Buchmann, C., DiPrete, T. A. and McDaniel, A. (2008). Gender inequalities in education. *Annual Review of Sociology*, **34**, 319–337.
- Buchmann, C. and Park, H. (2009). Stratification and the formation of expectations in highly differentiated educational systems. *Research in Social Stratification and Mobility*, **27**, 245–267.
- Cheung, S. Y. and Andersen, R. (2003). Time to read: family resources and educational outcomes in Britain. *Journal of Comparative Family Studies*, **34**, 413–433.
- Christin, A. (2012). Gender and highbrow cultural participation in the United States. *Poetics*, **40**, 423–443.
- Clark, A. and Trafford, J. (1995). Boys into modern languages: an investigation of the discrepancy in attitudes and performance between boys in modern languages. *Gender & Education*, **7**, 315–325.
- Crawford, C., Dearden, L. and Meghir, C. (2007). *When You Are Born Matters: The Impact of Date of Birth on Child Cognitive Outcomes in England*. Centre for the Economics of Education (CEE). Discussion Paper 0093.
- DiPrete, T. A. and Buchmann, C. (2013). *The Rise of Women. The Growing Gender Gap in Education and What It Means for American Schools*. New York: Russell Sage Foundation.
- Else-Quest, N. M., Hyde, J. S. and Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin*, **136**, 103–127.
- Fuchs, T. and Wössmann, L. (2004). *What Accounts for International Differences in Student Performance?* CESifo Working Paper No. 1235, Center for Economic Studies and Ifo Institute (CESifo), Munich.
- Garner, R. (2014). Women are now a third more likely than men to opt to go to university, according to UCAS application figures. *The Independent*, available from: <<http://www.independent.co.uk>> [accessed November 2014].
- Hadjar, A. and Buchmann, C. (2016). Education systems and gender inequalities in educational attainment. In Hadjar, A. and Gross, C. (Eds.), *Education Systems and Inequalities: International Comparisons*. Bristol: Policy Press, pp. 159–184.
- Halpern, D. F. (2013). *Sex Differences in Cognitive Abilities*. New York: Psychology press.
- Hanushek, E. A. and Wössmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal*, **116**, 63–76.
- Huang, M. (2009). Classroom homogeneity and the distribution of student math performance: a country-level fixed-effects analysis. *Social Science Research*, **38**, 781–791.
- Jackson, C. and Dempster, S. (2009). ‘I sat back on my computer ... with a bottle of whisky next to me’: constructing ‘cool’ masculinity through ‘effortless’ achievement in secondary and higher education. *Journal of Gender Studies*, **18**, 341–356.
- Jürges, H. and Schneider, K. (2011). Why young boys stumble: early tracking, age and gender bias in the German school system. *German Economic Review*, **12**, 371–394.
- Kerckhoff, A. C. (2001). Education and social stratification processes in comparative perspective. *Sociology of Education*, **74**, 3–18.
- Kraaykamp, G. and Notten, N. (2016). Parental cultural socialization and educational attainment. Trend effects of traditional cultural capital and media involvement. *Research in Social Stratification and Mobility*, **45**, 63–71.
- Legewie, J. and DiPrete, T. A. (2012). School context and the gender gap in educational achievement. *American Sociological Review*, **77**, 463–485.
- Lim, S. *et al.* (2015). Preferential detachment during human brain development: age- and sex-specific structural connectivity in diffusion tensor imaging (DTI) data. *Cerebral Cortex*, **25**, 1477–1489.
- Logan, S. and Medford, E. (2011). Gender differences in the strength of association between motivation, competency beliefs and reading skill. *Educational Research*, **53**, 85–94.
- Malik, K. (2013). *Human Development Report 2013. The Rise of the South: Human Progress in a Diverse World*. United Nations Development Programme: UNDP-HDRO Human Development Reports.
- McDaniel, A. (2010). Cross-national gender gaps in educational expectations: the influence of national-level gender ideology and educational systems. *Comparative Education Review*, **54**, 27–50.
- Meece, J. L., Glienke, B. B. and Brug, S. (2006). Gender and motivation. *Journal of School Psychology*, **44**, 351–373.
- Mislevy, J. and Sheehan, K. M. (1987). Marginal estimation procedures. In Beaton, A. E. (Ed.), *Implementing the New Design: The NAEP 1983-84 Technical Report*. Princeton: Educational Testing Service, National, Assessment of Educational Progress, pp. 293–360.
- Montt, G. (2011). Cross-national differences in educational achievement inequality. *Sociology of Education*, **84**, 49–68.
- Oakhill, J. V. and Petrides, A. (2007). Sex differences in the effects of interest on boys’ and girls’ reading comprehension. *British Journal of Psychology*, **98**, 223–235.
- OECD. (2009). *PISA Data Analysis Manual*, SPSS 2nd edn. Paris, France: OECD Publishing.
- OECD. (2012). *Untapped Skills. Realising the Potential of Immigrant Students*. Paris, France: OECD Publishing.
- OECD (2015). *The ABC of Gender Equality in Education. Aptitude, Behavior and Confidence*. Paris, France: OECD Publishing.
- Pekkarinen, T. (2008). Gender differences in educational attainment: evidence on the role of tracking from a Finnish quasi-experiment. *The Scandinavian Journal of Economics*, **110**, 807–825.
- Penner, A. M. (2008). Gender differences in extreme mathematical achievement. An international perspective on biological and social factors. *American Journal of Sociology*, **114**, 138–170.
- Scheeren, L., van de Werfhorst, H. and Bol, T. (2018). The gender revolution in context: how later tracking in education benefits girls. *Social Forces*, **97**, 193–220.
- Schmidt-Catran, A. W. and Fairbrother, M. (2016). The random effects in multilevel models: getting them wrong and getting them right. *European Sociological Review*, **32**, 23–38.



- Schneeweis, N. and Zweimüller, M. (2009). *Early Tracking and the Misfortune of Being Young*. Working Paper No. 0911. Department of Economics, Johannes Kepler University of Linz, Auhof, Austria.
- Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: a best-evidence synthesis. *Review of Educational Research*, 60, 471–499.
- Stevenson, D. L. and Baker, D. P. (1991). State control of the curriculum and classroom instruction. *Sociology of Education*, 64, 1–10.
- Stoet, G. and Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: within- and across-nation assessment of 10 years of PISA data. *PLoS One*, 8, e57988.
- Tanner, J. M. 1978. *Foetus into Man: Physical Growth from Conception to Maturity*. Cambridge, MA: Harvard University Press.
- van de Werfhorst, H. G. and Mijs, J. J. B. (2010). Achievement inequality and the institutional structure of educational systems: a comparative perspective. *Annual Review of Sociology*, 36, 407–428.
- Van Hek, M., Kraaykamp, G. and Pelzer, B. (2018). Do schools affect girls' and boys' reading performance differently? A multilevel study on the gendered effects of school resources and school practices. *School Effectiveness and School Improvement*, 29, 1–21.
- Van Hek, M., Kraaykamp, G. and Wolbers, M. H. (2016). Comparing the gender gap in educational attainment: the impact of emancipatory contexts in 33 cohorts across 33 countries. *Educational Research and Evaluation*, 22, 260–282.
- Wössmann, L. (2003). *Central Exams as the "Currency" of School Systems: International Evidence on the Complementarity of School Autonomy and Central Exams*. CESifo DICE Report, ISSN 1613-6373, ifo Institut für Wirtschaftsforschung an der Universität München, München, Vol. 1, Iss. 4, pp. 46–56.
- Margriet van Hek** is a postdoctoral researcher in the Department of Sociology/ICS, Radboud University in Nijmegen, the Netherlands. Her main research interests comprise social stratification, gender issues, educational sociology, and cultural tastes and lifestyles.
- Claudia Buchmann** is a Professor and chair in the Department of Sociology at The Ohio State University, Columbus, OH, USA. She is the coauthor of *The Rise of Women: The Growing Gender Gap in Education and What it Means for American Schools* (2013, Russell Sage Foundation). Her research on gender, race, and class inequalities in higher education has been published in many journals. Buchmann has served as deputy editor of the *American Sociological Review* and as chair of the Sociology of Education Section of the American Sociological Association.
- Gerbert Kraaykamp** is a full Professor of Empirical Sociology in the Department of Sociology and director of the Radboud Institute of Social Cultural Research at Radboud University Nijmegen. His main research interests lie with intergenerational transmission of inequality, parental socialization, educational careers, and cultural capital effects.

## Appendix A: Descriptive Statistics

Table A1. Descriptive statistics per country: average reading scores, female reading advantage and unstandardized country variables

Country	Average reading score <sup>a</sup>	Average reading score girl-boy <sup>a</sup>	Standardization <sup>a</sup>	Age of selection	Human Development Index <sup>a,b</sup>	Egalitarian gender norms <sup>a,b</sup>
Australia	512.61	33.09	0.25	16	0.92	2.39
Austria	491.56	35.89	0.35	10	0.87	2.17
Belgium	516.52	27.34	0.37	12	0.88	2.64
Bulgaria	435.18	54.18	0.59	14	0.77	2.38
Canada	518.66	33.58	0.58	16	0.90	2.06
Chile	458.50	16.01	0.32	13	0.82	2.79
Czech	507.60	37.40	0.15	11	0.86	2.61
Denmark	492.12	25.35	0.28	16	0.91	2.16
Finland	532.93	50.61	0.27	16	0.88	2.47
France	512.80	30.63	0.38	15	0.88	1.86
Germany	512.98	33.81	0.36	10	0.90	2.12
Greece	477.02	42.64	0.89	15	0.85	1.59
Hong Kong (China)	533.74	25.05	0.08	15	0.88	2.58
Hungary	488.72	33.29	0.29	11	0.81	2.47
Iceland	493.75	45.74	0.26	16	0.89	2.24
Ireland	520.63	26.15	0.27	15	0.90	1.84
Israel	482.47	28.20	0.44	15	0.88	1.84
Italy	493.26	37.55	0.28	14	0.87	1.95
Japan	518.29	23.71	0.08	15	0.89	2.38
Korea (rep)	536.19	30.51	0.08	14	0.88	2.27
Latvia	488.72	45.06	0.44	16	0.80	2.45
Luxembourg	485.09	27.14	0.86	13	0.88	2.56
Mexico	431.31	21.63	0.58	12	0.74	2.68
Netherlands	517.87	22.73	0.07	12	0.91	2.73
New Zealand	527.73	33.19	0.09	16	0.90	2.16
Norway	506.31	43.44	0.47	16	0.94	2.13
Poland	510.37	38.92	0.22	15	0.82	2.09
Portugal	485.12	30.37	0.48	15	0.81	2.10
Russian Federation	466.59	33.57	0.33	15	0.77	2.58
Slovakia	469.96	37.53	0.36	11	0.81	2.60
Slovenia	471.77	53.02	0.61	15	0.88	2.76
Spain	494.50	26.76	0.44	16	0.87	2.35
Sweden	507.90	38.82	0.28	16	0.90	1.65
Switzerland	498.16	32.47	0.72	15	0.92	2.35
Turkey	454.15	34.79	0.71	11	0.72	2.66
United Kingdom	507.31	22.80	0.14	16	0.89	2.39
United States	502.80	25.51	0.42	16	0.91	2.17
Average	496.249	33.473	0.373	14,190	0.862	2.318

Note: readings scores weighed for student weight.

<sup>a</sup>Average over waves.

<sup>b</sup>These variables are grand mean centred in the analyses.

Source: 2000–2015 PISA waves; 1,425,356 students; 59,001 schools; 204 country-year combinations; 37 countries.

## Appendix B

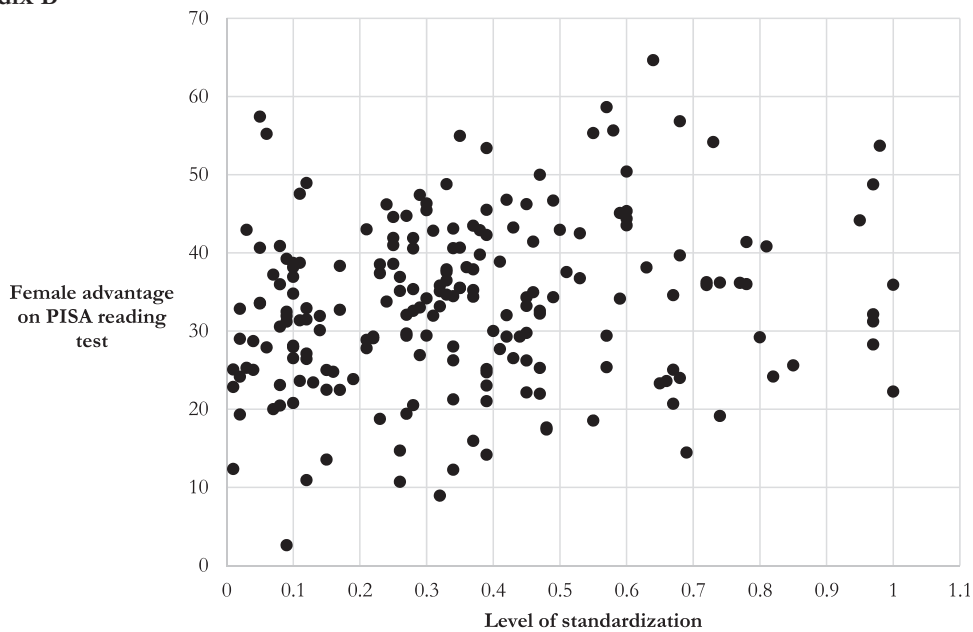


Figure B1. Association between the level of standardization and the gender gap in reading performance

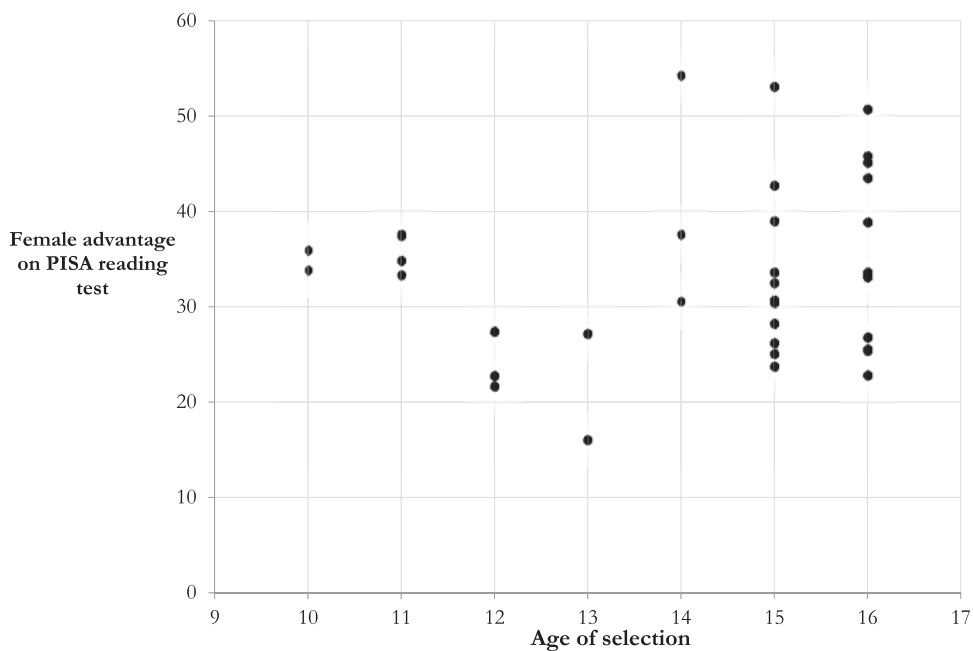


Figure B2. Association between the age of selection and the gender gap in reading performance

## Appendix C

Table C1. Linear multilevel models predicting math scores

	Model 0	SE	Model 1	SE	Model 2	SE	Model 3	SE
Individual variables								
Intercept	492.876***	5.658	485.429***	5.774	440.522***	10.792	433.780***	10.862
Male 0/1			15.487***	1.052	16.014***	0.905	22.282***	2.063
Parental education					1.833***	0.037	1.832***	0.037
Books in household					16.964***	0.123	16.963***	0.123
Age					7.566***	0.422	7.566***	0.422
Immigrant: first generation (ref)								
Second generation					7.231***	0.800	7.231***	0.800
Native					17.898***	0.554	17.899***	0.554
Country variables								
Standardization					-16.466*	6.954	-16.280*	6.971
Age of selection -10					2.055	2.212	3.646	2.235
HDI					17.485***	4.390	17.000***	4.402
Gender equality					-16.562**	5.854	-15.902**	5.876
Cross-level Interactions								
Standardization × male							0.924	2.241
Age of selection × male							-1.572***	0.401
HDI × male							-1.217	1.315
Gender equality × male							-0.926	1.945
Variance Statistics								
Individual variance	133.584		123.479		127.130		127.130	
School variance	49.487		50.061		43.456		43.460	
Country-year variance	11.097		10.111		10.525		10.517	
Country variance	34.015		34.766		25.476		25.369	
Slope male school			22.014		22.443		22.443	
Slope male country-year			3.003		3.087		3.010	
Slope male country			5.884		4.904		3.708	

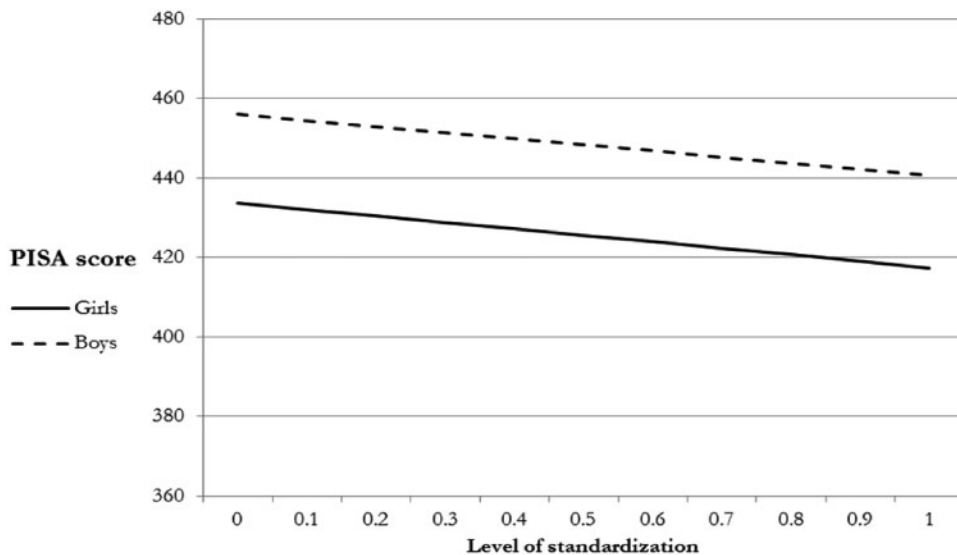
Note: Linear multilevel regression models; LME in R.  $\sim P < 0.1$ .

\* $P < 0.05$ ;

\*\* $P < 0.01$ ;

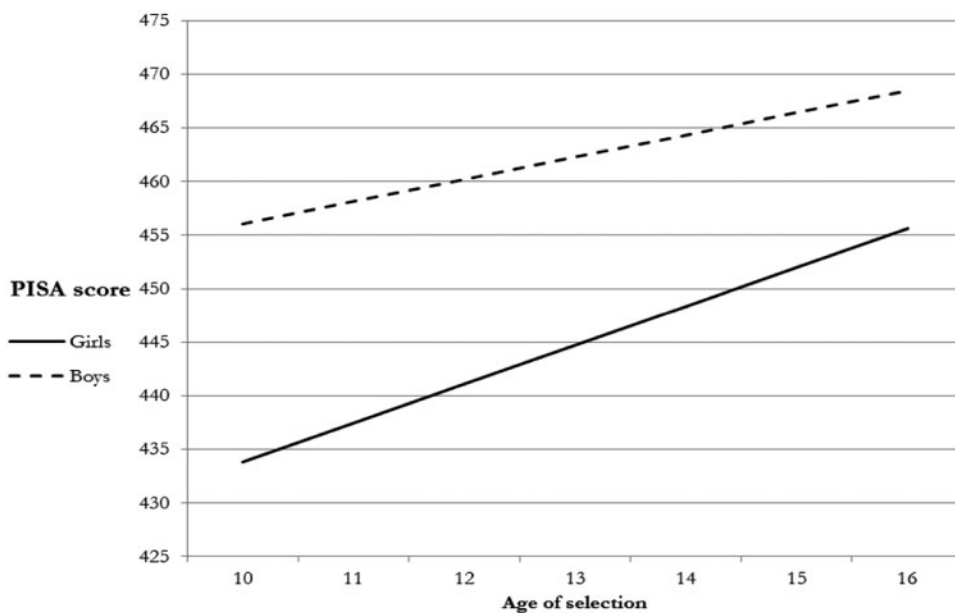
\*\*\* $P < 0.001$ .

Source: 2000–2015 PISA waves; 1,425,356 students; 59,001 schools; 204 country-year combinations; 37 countries.



**Figure C1.** Performance scores with level of standardization for girls and boys (Model 3: Table C1)

Source: 2000–2015 PISA waves; 1,425,356 students; 59,001 schools; 204 country-year combinations; 37 countries.



**Figure C2.** Performance scores with age of selection for girls and boys (Model 3: Table C1)

Source: 2000–2015 PISA waves; 1,425,356 students; 59,001 schools; 204 country-year combinations; 37 countries.