Routledge
Taylor & Francis Group

ARTICLE

Do schools affect girls' and boys' reading performance differently? A multilevel study on the gendered effects of school resources and school practices

Margriet van Hek [a], Gerbert Kraaykamp[b] and Ben Pelzer[b]

[a]Department of Sociology, Utrecht University, Utrecht, The Netherlands; [b]Department of Sociology, Radboud University, Nijmegen, The Netherlands

ABSTRACT

Few studies on male–female inequalities in education have elaborated on whether school characteristics affect girls' and boys' educational performance differently. This study investigated how school resources, being schools' socioeconomic composition, proportion of girls, and proportion of highly educated teachers, and school practices, being schools' application of well-rounded assessment methods, influenced girls' and boys' reading performance differently. We hypothesised that positive effects of school resources would be greater for boys than for girls, and that more frequent use of well-rounded assessment methods would be associated with increased girls' and decreased boys' reading performance. Using advanced multilevel analyses of 2009 Programme for International Student Assessment (PISA) data, we found that boys profited more than girls from having a large proportion of girls in school. Contrary to our expectations, girls gained more than boys from a school's advantaged socioeconomic composition. These gendered effects of school resources were not explained by differences in school learning climate.

Introduction

This study focused on the differential effects of school characteristics on girls' and boys' reading performance. Schools are key socialising contexts in children's lives. Scientists and policymakers are therefore keenly interested in how schools can provide learning environments that draw out children's full potential. Educational research has long established inequality in educational performance by socioeconomic origin and race. Extensive research has since examined whether schools are equally effective for students from diverse familial backgrounds (Creemers & Kyriakides, 2008; Crul & Schneider, 2009; Dee, 2005; Hallinan, 1988). As schools do not seem to affect all students equally, an examination of whether and how schools affect girls' and boys' educational performance differently seems worthwhile, particularly considering the current female advantage in education (Eurostat, 2013; Stoet & Geary, 2013; Van Hek, Kraaykamp, & Wolbers, 2016).

An important focus in educational research nowadays is on why girls outperform boys on almost all indicators of educational achievement (DiPrete & Buchmann, 2013; Van Hek, Kraaykamp, & Wolbers, 2015, 2016; Van Houtte, 2004b). Yet, few studies on male–female differences in education have investigated the role of schools herein (Buchmann, DiPrete, & McDaniel, 2008). Concerning educational gender inequalities, Ma (2008) observed that "It is ironic that educational studies have given relatively little attention to the role of school experiences" (p. 441). Machin and McNally (2005) and Ma (2008) found that gender inequalities in educational performance differed between schools, but their studies did not go into great detail theoretically or analytically on how these differences in gender inequality come about. It thus remains unclear which school characteristics are beneficial or disadvantageous for girls and/or boys. Legewie and DiPrete (2012), in contrast, theorised elaborately on schools' differential effects on girls' and boys' reading performance. Their study, however, focused on Germany only and examined a single (though important) school characteristic. They found that boys were more affected than girls by their school's socioeconomic composition.

In the current study, we sought to better understand the established female advantage in education by investigating for a wide range of countries whether various school characteristics affected the reading performance of girls and boys differently. We focused on students' reading performance for two reasons. First, reading comprehends a core competency in a person's educational career (Organisation for Economic Co-operation and Development [OECD], 2010). Indeed, reading skills affect students' performance in other domains, including math and science (Martin & Mullis, 2013). Hence, gender inequalities in reading performance have consequences for educational gender inequalities overall. Second, Ma (2008) found that gender inequalities in reading skills varied more between schools (within countries) than gender inequalities in mathematics and science skills. This suggests that schools exert a relatively large influence on the gap between girls' and boys' reading scores.

In this research, we used a cross-national design to examine whether and which school features affect girls' and boys' reading test scores differently. Findings from this research could help schools build learning environments in which both boys and girls are encouraged to develop their full reading potential. We examined two main dimensions of the school context. First, we focused on a school's resources, defined as a school's socioeconomic composition, proportion of girls, and proportion of teachers with a college degree. We assessed whether these characteristics affected girls and boys differently. We theorised that school resources promote a positive learning climate in schools and in classrooms. In this, we built, among others, on the work of Legewie and DiPrete (2012), who proposed that boys in particular are affected by the learning environment in school. Second, we investigated whether frequent use of well-rounded assessment methods, like projects, homework, and student assignments, affected girls and boys differently. Prior research has repeatedly shown boys to lag behind girls in non-cognitive skills. Non-cognitive skills are relatively important, however, in such well-rounded assessment methods, meaning that their use may be detrimental to boys (Farkas, Grobe, Sheehan, & Shuan, 1990; Jacob, 2002). Our research question reads: *To what extent do school resources and school assessment methods affect girls' and boys' reading performance differently?*

This research improves on previous research in several ways. First, it introduces an examination of the role of schools into the study of male–female educational differentiation. Indeed, rising gender inequalities in educational performance are a key concern in Western societies, as they have consequences for family life, country demographics, and gender inequalities on the labour market (DiPrete & Buchmann, 2013). Second, a wide range of school characteristics is investigated, thereby broadening the understanding of school contexts provided by earlier work. Third, state-of-the-art multilevel methods are used on 2009 Programme for International Student Assessment (PISA) data, with students nested in schools nested in 33 OECD countries. For optimal analysis of these data, we applied weightings across schools to achieve robust estimates of schools' effects on girls' and on boys' reading test scores. Fourth, by examining reading performance as a core competence in education, advanced insights are pursued on how schools affect girls' and boys' educational careers differently (Cheung & Andersen, 2003; OECD, 2010).

Theory

Schools as learning institutions

Schools fulfil an indispensable role in students' learning and development (Scheerens & Bosker, 1997; Wentzel & Looney, 2007). Research on how exactly schools contribute to efficient learning is therefore of both scientific and policy relevance (Hanushek, 1997). First, policies affect educational practices directly through school financing, so the question of whether more funds actually leads to superior educational outcomes is utmost among policymakers. However, the direct effect of school funding levels remains a subject of debate. Hanushek (1997), for example, argued that funding bears only a very small relation to student performance after controlling for student familial background. Second, school composition has been studied frequently since Coleman et al. (1966) reported it as the most influential school characteristic. Many of these subsequent studies have confirmed that schools' socioeconomic and racial composition indeed affect the educational performance of students (Rumberger & Palardy, 2005). Third, the large body of mostly pedagogic research on instructional and teaching practices has produced mixed results. Few studies have focused explicitly on distinct assessment methods (Bishop, 1997; Driessen & Sleegers, 2000; Seidel & Shavelson, 2007). Most research on assessment has focused on whether the implementation of central or national examinations leads to gains in learning (Bishop, 1997; Maag Merki & Holmeier, 2015; Reardon, Arshan, Atteberry, & Kurlaender, 2010).

Although schools contribute to students' average educational performance, this does not mean that their influence is similar for all students. Nye, Konstantopoulos, and Hedges (2004), for example, found that qualified teachers were especially beneficial for students from disadvantaged families. Connor, Morrison, Fishman, Schatschneider, and Underwood (2007) showed that initially low-performing students benefited more from individualised instruction than high-performing students. Overall, most scholars agree that high-quality schools improve the achievement level of all students but particularly elevate the performance of disadvantaged students (Nye et al., 2004; Scheerens & Bosker, 1997). A high-quality school is typically defined as one with highly qualified teachers and optimal organisational conditions, but some previous studies

have proposed that a large proportion of students from advantaged socioeconomic backgrounds is another indicator of school quality (Clotfelter, Ladd, & Vigdor, 2010; Greenwald, Hedges, & Laine, 1996; Hallinger & Heck, 2011; Hopkins & Stern, 1996; Louis, Dretzke, & Wahlstrom, 2010; Rumberger & Palardy, 2005). Nonetheless, few studies have examined whether such school characteristics are relevant to the differential educational performances of girls and boys (Legewie & DiPrete, 2012; Ma, 2008; Machin & McNally, 2005). Those studies on gender differentiation that have addressed school characteristics focused mainly on sex-specific school features, such as the influence of single-sex schools or teachers' gender (Ehrenberg, Goldhaber, & Brewer, 1995; Salomone, 2003).

The current study combines theory on school effectiveness with research on gender differences in non-cognitive abilities and gender-related school norms to formulate and test hypotheses on how school resources and well-rounded assessment methods might affect girls' and boys' reading performance differently.

School resources

To address male–female differentiation in educational performance, we examined three aspects of a school's resources: school socioeconomic composition, proportion of girls in the student body, and proportion of teachers with a college degree. Regarding the first, students from advantageous socioeconomic backgrounds generally read better and misbehave less in class (Entwisle, Alexander, & Olson, 2007; Farkas et al., 1990). Regarding the second, girls are better and more frequent readers; they enjoy reading more, have more positive attitudes towards reading, and are intrinsically more motivated to perform well in school (Buchmann et al., 2008; Vantieghem & Van Houtte, 2015). Regarding the third, highly skilled teachers are known to make relatively large contributions to students' learning, due to their more effective teaching styles (Greenwald et al., 1996; Piopiunik, Hanushek, & Wiederhold, 2014; Rivkin, Hanushek, & Kain, 2005). We first elaborate on how these school characteristics affect girls' and boys' learning through what happens within schools. We then look specifically at what happens in classrooms.

Prior studies report that although students care about their educational performance, they often consider social status and popularity among peers to be even more important (Bishop et al., 2003; Van Houtte, 2004b). In peer cultures, success in sports, physical appearance, and attractiveness to the other sex are often valued more than educational performance. This, however, differs for girls and boys, because norms of femininity and masculinity set different boundaries for how girls and boys are expected to behave (Van Houtte, 2004b). Both qualitative and quantitative studies suggest that whereas femininity is relatively easily aligned with educational effort and performance, masculinity tends to be associated more with boldness and opposition to school authority (Francis, 2000; Jackson & Dempster, 2009). Male peer groups may therefore uphold stronger anti-academic norms than female peer groups (Van Houtte, 2004b). This was recently proposed as an explanation for boys' lower educational performance (DiPrete & Buchmann, 2013).

Legewie and DiPrete (2012) advocate that schools may advance academic competition in boys' perceptions of masculinity, thereby creating a stimulating learning environment for boys schoolwide:

> Such an environment promotes academic competition as an aspect of masculinity and encourages development of adaptive strategies that enable boys to maintain a show of emotional coolness toward school while being instrumentally engaged in the schooling process. In other words, academic competition as one of the "different ways of 'doing' masculinity". (p. 467)

Importantly for our purposes, Legewie and DiPrete claimed that establishing a successful learning environment (for boys) in a school requires resources, such as high-quality teachers and a talented and well-motivated student body. Although Legewie and DiPrete pointed out that their proposed mechanism applies to all sorts of school resources, their own empirical work considered only a school's socioeconomic composition. For future research, Legewie and DiPrete recommended to study the role of high-quality teachers. In addition, we believe that the proportion of girls in a school may be an important resource, considering girls' high educational performance and motivation. Van Houtte (2004a) found that boys performed better in schools with a large proportion of girls. These schools were thought to have a more study-oriented culture. The current study heeds Legewie and DiPrete's call and investigates, in addition to a school's socioeconomic composition, to what extent the proportion of teachers with a college degree and the proportion of female students affect girls' and boys' reading performance differently.

Next to the effects of school resources on the schoolwide learning climate, school resources may influence girls' and boys' reading performance differently through classroom experiences. Various scholars have shown that teachers in particular have a large and significant impact on student learning (Nye et al., 2004). According to Montt (2011), "Better qualified teachers are more able to adapt curricular material, subject knowledge, and pedagogical techniques to the needs of their students, thereby providing an enhanced schooling experience for all students and affecting student achievement" (p. 53). Besides teachers, students' learning experiences depend heavily on their fellow students in the classroom. Lazear (2001) observed that classroom learning is a public good because students who misbehave undermine learning by others. Disruptive students may hamper learning opportunities indirectly as well, by demotivating their teachers. As both students from lower socioeconomic families and boys more often exhibit social and behavioural problems, a larger share of low-socioeconomic background students and boys in a classroom would likely lead to more frequent class disruptions. In this respect, Betts and Shkolnik (1999) found that teachers spent more time on instruction and less time on discipline when classes were more female.

Several scholars have suggested that boys may be more sensitive to classroom interactions than girls. For instance, Wachs, Gurkas, and Kontos (2004) found that boys' already weaker self-regulating skills were more negatively affected by disruptive, chaotic, and disorganised classroom settings. Ponitz, Rimm-Kaufman, Brock, and Nathanson (2009) found relatively large learning gains among boys in well-organised classrooms. According to the OECD report *The ABC of Gender Equality in Education* (2015), "boys appear to be particularly sensitive to environmental factors, while girls are comparatively less affected by a lack of discipline, disorganization and chaos in the classroom" (p. 58). Possible explanations for boys' greater sensitivity to environmental factors may be found in their lower levels of intrinsic motivation compared to girls (Vantieghem & Van Houtte, 2015).

In sum, school resources may positively affect a student's learning, both in the classroom and at the schoolwide level. Although both girls and boys likely profit from a stimulating learning environment, we expect school resources to have a stronger effect on boys, as previous research suggests that boys are more sensitive to influences in the learning environment. This leads to the following hypotheses: *A higher proportion of students from advantageous socioeconomic backgrounds in a school positively affects the reading performance of girls and of boys, but this effect is stronger for boys* (H1); *a higher proportion of girls in a school (more than 60%) positively affects the reading performance of girls and of boys, but this effect is stronger for boys* (H2); and *a higher proportion of teachers with a college degree in a school positively affects the reading performance of girls and of boys, but this effect is stronger for boys* (H3).

The above-formulated hypotheses reflect the assumption that school resources influence student reading performance partly via a school's learning climate. In our data, school principals actually reported on how often students in their school displayed improper behaviours (i.e., disruption of classes, skipping classes, and being disrespectful to teachers). This information allowed us to directly test the "schoolwide learning climate" mechanism. We extended our conceptual model to investigate whether possible gendered effects of schools' socioeconomic composition, proportion of girls, and proportion of teachers with a college degree on reading performance were interpreted (mediated) by the effects of this subjective measure of a schools' learning climate. We did not formulate an explicit hypothesis on this possible mediation, since it is unclear exactly what aspect of a school's learning climate this measure captures.

School practices

Several studies have shown that a student's educational performance is dependent on not only cognitive ability but also a variety of non-cognitive skills such as the ability to organise study materials, work together in groups, and stay focused in class (Downey & Vogt Yuan, 2005; Farkas et al., 1990; Jacob, 2002). DiPrete and Buchmann (2013) suggested that

> … the link between social and behavioral skills and academic outcomes (particularly teacher academic evaluations) is flowing largely through a direct connection between social and behavioural skills and learning, the production of homework, and other classroom exercises that factor into teachers' evaluations. (p. 163)

DiPrete and Buchmann distinguished three mechanisms by which non-cognitive skills may influence students' educational performance. First, when teachers use well-rounded evaluation methods, as opposed to narrow evaluation criteria, they intentionally evaluate non-cognitive skills. As such, higher grades are awarded to students who, for example, actively participate in groups and hand in assignments on time (Heckman, Stixrud, & Urzua, 2006). Second, so-called teacher bias implies that teachers give higher grades to students whose behaviour is more in line with preferred behaviour (Farkas et al., 1990). Third, in social learning environments, students with greater non-cognitive skills simply learn more, for instance, through active participation in group discussions (Entwisle et al., 2007; Farkas et al., 1990). Importantly, all these mechanisms may directly affect students' educational performance, but they may create negative feedback loops

as well; that is, students who receive poor teacher evaluations may become demoralised and consequently put less effort into schoolwork, resulting in lower school performance (Farkas et al., 1990).

The evaluation of students' performance thus relates to the assessment methods employed in school, in part because of the varying role played by non-cognitive qualities in assessment methods. Non-cognitive skills would likely have relatively little effect on standardised test scores, because their evaluation criteria are clear and given and leave little room for teacher bias. Such tests are furthermore administered in a well-defined classroom setting and timeslot, with scores determined by the answers that students write down. In contrast, the grading of homework assignments, group work, and projects is more rounded. Students' organisational skills and active participation come into greater play, as well as teacher bias.

Generally, adolescent girls have better non-cognitive skills than adolescent boys (Buchmann et al., 2008; Downey & Vogt Yuan, 2005; Jacob, 2002). Girls also have greater social and behavioural skills, such as self-control (DiPrete & Jennings, 2012). They are better organised (Farkas et al., 1990; Jacob, 2002) and find it easier to concentrate in class (Kenney-Benson, Pomerantz, Ryan, & Patrick, 2006). Moreover, boys receive more negative attention from teachers, and teachers' tolerance level for misbehaviour is lower for boys than for girls (Pickering & Lodge, 1998; Younger, Warrington, & Williams, 1999). Girls' greater non-cognitive skills likely affect their grades via direct assessments that include non-cognitive skills but also indirectly through teachers' subjective opinions (liking). The grading of adolescent boys' performances will probably be disadvantaged by their lower non-cognitive skill levels (Farkas et al., 1990). This initial mechanism of receiving higher or lower grades may easily accumulate into a self-fulfilling prophecy. Boys may become demotivated by their lower grades, negatively affecting their overall educational performance, whereas girls may be stimulated to greater achievement by their initial positive grading (Farkas et al., 1990; Voyer & Voyer, 2014).

In sum, school assessment methods, such as projects, homework, and group assignments depend in part on non-cognitive skills, so grading on these tasks is susceptive to teacher bias. We therefore hypothesise the following: *More frequent use of homework, group assignments, and projects in school evaluations positively affects the reading performance of girls and negatively affects the reading performance of boys* (H4).

Data and measurements

Data

We analysed data from the 2009 wave of PISA collected by the OECD. The PISA data are optimal for testing our hypotheses because they provide details on a great number of students and schools situated in a large number of countries. The OECD provides comprehensive elucidation on PISA sampling and survey methods in reports available online (OECD, 2012). PISA firstly samples schools and secondly selects students randomly within those schools. In the 2009 wave of PISA, a maximum of 35 students were sampled per school. An 80% response rate of sampled students in the participating schools was required. Note that students were thus nested in schools, and not in school classes.

PISA tests the reading, mathematics, and science performance of 15-year-olds, irrespective of the grade they were in, which may differ per country. PISA also asks students about a range of topics, including their study behaviour and family background. Information about the schools in which the students were nested was provided by the school principals, who filled in a PISA school questionnaire. For our analyses, we selected only OECD countries and removed single-sex schools. This resulted in a dataset consisting of 281,095 students from 10,425 schools in 33 countries.[1]

Measurements

Individual variables
Dependent variable. PISA provides measures of students' reading performance using a method based on item response theory (Mislevy & Sheehan, 1987). Instead of a single measure, five "plausible values" for a students' reading ability are provided. The plausible value method is especially useful if only part of a large battery of items is employed to measure ability, as done by PISA. It produces unbiased estimates of differences between subpopulations (like boys and girls) and their standard errors, as opposed to a single ability measure like the proportion of correct answers (von Davier & Hastedt, 2009; Wu, 2005).

Individual variables. Our primary interest was the variable *female*, which we coded 0 for boys and 1 for girls. We included several individual-level control variables. First, we controlled for students' *age*, as older students have been shown to perform better (Schneeweis & Zweimüller, 2014). We controlled for *parental education* by including the years of education of the highest educated parent. We subtracted 3 (the minimum) from these original values so that 0 represents the minimum years of education. We used the *number of books* in the family home as an indicator of cultural resources, with categories being 0–10 books (0), 11–25 books (1), 26–100 books (2), 101–200 books (3), and more than 200 books (4). We included this variable linearly in the analysis.[2] To control for students' immigrant background, we distinguished between *natives* (born in the country in which the PISA data were collected), *first-generation immigrants*, and *second-generation immigrants*. As family structure has proven to affect students' educational performance, especially that of boys, we controlled for whether students lived in a *two-parent family* (1) or had another family structure (0) (Amato, 2001).[3]

School variables
School resources and school practices. To determine the *socioeconomic composition of a school*, we aggregated the educational level of students' parents at the school level. School principals provided exact numbers for the student and teacher population. We used this information to construct the other two indicators of school resources, namely, the proportion of girls and the proportion of highly educated teachers. We calculated whether the *proportion of girls in a school* was more or less than 60% (0/1)[4], and we determined the *proportion of teachers with a university education* (Level 5a of the International Standard Classification of Education [ISCED]) in a school. With regard to a school's use of *well-rounded assessment methods*, school principals were asked how often in their school students were assessed using student assignments, projects, or homework (single item). All of these methods require a relatively high level of student

autonomy and were deemed more dependent on students' non-cognitive skills than tests.[5] Answer categories were never (0), 1–2 times a year (1), 3–5 times a year (2), monthly (3), and more than once a month (4). Additional analyses showed that we could include this variable linearly in our analyses. With respect to the overall *school climate*, principals were asked, "In your school, to what extent is the learning of students hindered by the following phenomenon?" Principals indicated whether the following happened a lot (0), to some extent (1), very little (2), or never (3): student absenteeism, poor student–teacher relations, disruption of classes by students, students skipping classes, students showing teachers disrespect, students using alcohol or illegal drugs, students intimidating or bullying other students, and students not being encouraged to achieve their full potential. We constructed the school climate variable by averaging these items. Finally, we grand-mean centred the socioeconomic composition of the school, the proportion of teachers with a university education, the use of well-rounded assessment methods, and the school climate for ease of interpretation.

Control variables. We controlled for possible confounding variables at the school level. First, we considered whether a school was a *private school* (1) or a public school (0). We included the log function of the *school size*, as the original measures contained some very high values, which affected the results. We also controlled for the *availability of school materials*, measured by the question, "Is your school capacity to provide instruction hindered by any of the following issues?" Here, we considered the lack, shortage, or inadequacy of five items: instructional materials (e.g., textbooks), computers, internet, library staff, and library materials. Answer categories were a lot (0), to some extent (1), very little (2), and not at all (3). We took the average of the five items.[6] We grand-mean centred the variables indicating school size and availability of school materials. We listwise deleted students with missing values on our individual variables (14,788 students, 15 schools) and missing values on any of our school variables (50,190 students, 2,101 schools). We performed our analyses on a dataset containing 216,117 students, 8,306 schools, and 33 countries. Table 1 presents descriptive statistics.

Analyses and results

Analytical strategy

We employed the mixed package of the Julia programming language (Bezanson, Karpinski, Shah, & Edelman, 2012) to take the nested structure of our data into account. In the PISA data, students (Level 1 units) are nested in schools (Level 2 units) that are nested in countries (Level 3 units). We controlled for country-level variation in reading ability by including country-fixed effects in our models (these estimates are not presented, as our focus here is on the school level).[7] We first estimated our models for each of the five plausible values of reading performance. Next, we merged the results to arrive at point estimates and standard errors. The PISA manual (OECD, 2009) elaborately describes this procedure, so we do not repeat it here. In addition, PISA requires the use of one final student weight plus 80 replicate weight variables to account for PISA's two-stage sampling design, first selecting schools and next selecting students within schools. Also, student weight variables had to be used to adjust for overrepresentation or

Table 1. Descriptive statistics.

Variable	Min	Max	Mean	SD
Plausible value reading 1	12	871.120	492.919	92.255
Plausible value reading 2	60.990	855.070	492.972	92.449
Plausible value reading 3	19.600	844.750	492.847	92.438
Plausible value reading 4	33.640	877.240	492.956	92.383
Plausible value reading 5	64.910	904.500	492.956	92.401
Female	0	1	.502	.500
Age	0	1.160	.603	.290
Parental education	0	15	10.172	3.229
# of books at home	0	4	2.179	1.309
Two-parent family	0	1	.803	.398
Native (ref)	0	1	.909	.288
Second generation	0	1	.047	.211
First generation	0	1	.044	.206
SES composition	−10.052	4.448	0	2.025
>60% girls	0	1	.106	.308
Proportion of university teachers	−.754	.246	0	.365
Well-rounded assessment methods	−3.187	.813	0	1.116
School climate	−1.959	1.041	0	.515
Private school	0	1	.153	.360
School size (log)	−5.483	3.154	0	.907
Availability of teaching materials	−2.082	.918	0	.702

Source: PISA 2009. $N_{students}$: 216,117; $N_{schools}$: 8,306; 33 countries.

underrepresentation of students with certain characteristics. Consequently, for each of the five plausible values, we ran each regression model 81 times (one final student weight plus 80 replicate weights), yielding a total of 405 regression models. For this reason, we performed our analyses in the relatively fast programming language Julia.[8]

Table 2 presents the main effects of all individual and school variables. In Model 0, we estimated a null model that shows the variance at the individual level and at the school level. In Model 1, the uncontrolled effect of female represents the averaged (across schools) difference between girls' and boys' reading scores. Model 2 adds all individual control variables. In Model 3, we added all school variables. Table 3 shows in Model 4 the cross-level interactions in which female is interacted with the three indicators of school resources, and with schools' use of well-rounded assessment methods. The interaction of female with school climate was added in Model 5.[9] We tested our hypotheses with these cross-level interactions, which represent the extent to which the effects of school characteristics differed for girls and for boys.

Results

The null model in Table 2 shows that in addition to individual variation in reading scores, students' nesting in schools accounted for a considerable portion of the variance in students' reading performance. The intraclass correlation is 0.35. Model 1 shows that, on average, girls scored 29.4 points higher on the PISA reading test than boys. This effect was not interpreted by the individual control variables in Model 2. In Model 2, all individual control variables acted as expected: Native children from two-parent families with highly educated parents who had ample cultural resources in the family home were relatively good readers. Model 3 shows the effects of our independent school variables, controlling for possible confounding school characteristics. In line with our expectations,

Table 2. Main effects: individual and school variables.

	M0		M1		M2		M3	
	B	SE	B	SE	B	SE	B	SE
Individual variables								
Intercept	498.792***	.595	484.138***	.685	425.349***	1.293	410.181***	1.202
Female			29.437***	.306	29.115***	.293	28.545***	.281
Age					9.767***	.910	9.491***	.906
Parental education					1.754***	.087	1.372***	.086
# of books at home					15.310***	.256	15.092***	.256
Two-parent family (0/1)					10.685***	.548	10.612***	.546
Immigrant status								
Native (ref)								
Second generation					−14.113***	1.245	−14.319***	1.227
First generation					−25.580***	1.250	−25.355***	1.239
School variables								
School resources								
SES composition[a]							13.517***	.197
>60% girls (0/1)							13.384***	.467
Prop. university teachers[a]							22.255***	.731
School practices								
Well-rounded assessment methods[a]							−.118	.125
School control variables								
School climate[a]							19.203***	.313
Private school							−.294	.414
School size (log)[a]							11.879***	.280
Availability teaching materials[a]							−.160	.227
Variance parameters								
Student	5219.081		4948.241		4555.833		4554.207	
School	2800.541		2979.454		2176.510		1322.648	
Slope female			352.895		309.870		308.999	
−2 Log Likelihood	2,483,192.367		2,474,307.050		2,454,462.192		2,450,850.356	

Source: PISA 2009. $N_{students}$: 216,117; $N_{students}$: 8,306; 33 countries. *$p > 0.10$. **$p > 0.05$. ***$p > 0.001$.
[a]These variables are grand mean centred at the school level.
Notes: Controlled for country, parameters not shown. The output shown for each model is based on estimating that model for each of five plausible reading score values and each of 81 weight variables, amounting to $5 \times 81 = 405$ times estimating the model and combining the results as prescribed in the PISA data analysis manual 2009.

Table 3. Cross-level interactions with female and school characteristics.

	M4		M5	
	B	SE	B	SE
Individual variables				
Intercept	409.890***	1.210	409.896***	1.214
Female	29.040***	0.301	28.978***	0.305
School variables				
School resources				
SES composition[a]	13.174***	0.211	13.077***	0.220
>60% girls	17.700***	0.783	17.481***	0.787
Prop. qualified teachers[a]	22.123***	0.795	22.180***	0.792
School practices				
Well-rounded assessment methods[a]	0.010	0.185	−0.057	0.187
School control variables				
School climate[a]	19.191***	0.313	21.047***	0.513
Cross-level interactions				
SES composition*female	0.569***	0.150	0.723***	0.160
>60% girls*female	−5.714***	0.776	−5.407***	0.796
Prop. university teachers*female	0.157	0.870	0.033	0.865
Well-rounded assessment methods*female	−0.215	0.277	−0.116	0.282
School climate*female			−3.017***	0.639
Variance parameters				
Student	4554.177		4554.198	
School	1321.507		1321.506	
Slope female	306.016		303.977	
−2 Log Likelihood	2,450,822.920		2,450,807.262	

Source: PISA 2009. $N_{students}$: 216,117; $N_{schools}$: 8,306; 33 countries. *$p > 0.10$. **$p > 0.05$. ***$p > 0.001$.
Notes: Controlled for country, parameters not shown. The output shown for each model is based on estimating that model for each of five plausible reading score values and each of 81 weight variables, amounting to $5 \times 81 = 405$ times estimating the model and combining the results as prescribed in the PISA data analysis manual 2009.

students performed better in schools with a high socioeconomic composition ($b = 13.517$), where more than 60% of the students was female ($b = 13.384$), and where a large proportion of teachers had a university education ($b = 22.255$). The effect of a school's socioeconomic composition was especially large, considering the variance of this variable. The difference between schools with the lowest and the highest socio-economic composition was 196 points on the PISA reading test ($b = 13.517 *$ range $= 14.5$). The effect of schools' assessment methods was not significant. In addition, the presence of a positive school learning climate enhanced students' reading scores ($b = 19.203$), and students in larger schools appeared to be better readers ($b = 11.879$). Controlling for other school variables, it did not appear to matter whether students attended a private or a public school, or whether sufficient teaching materials were available.

Table 3 shows the extent to which girls and boys were affected differently by school resources and schools' use of well-rounded assessment methods. The cross-level inter-actions represent the differences in their effects on girls and boys. The main effects of school characteristics in these models apply to boys, as they scored 0 on female. The effect for girls was obtained by adding or subtracting the cross-level interaction term from the main effect. Figure 1 to 3 visualise all significant cross-level interactions. First, Model 4 shows that girls and boys were indeed affected differently by two indicators of school resources. In contrast to our hypothesis, girls seemed to be more positively affected by a school's socioeconomic composition than boys ($b = 0.569$). The benefit of being in a school with the highest socioeconomic composition versus being in a
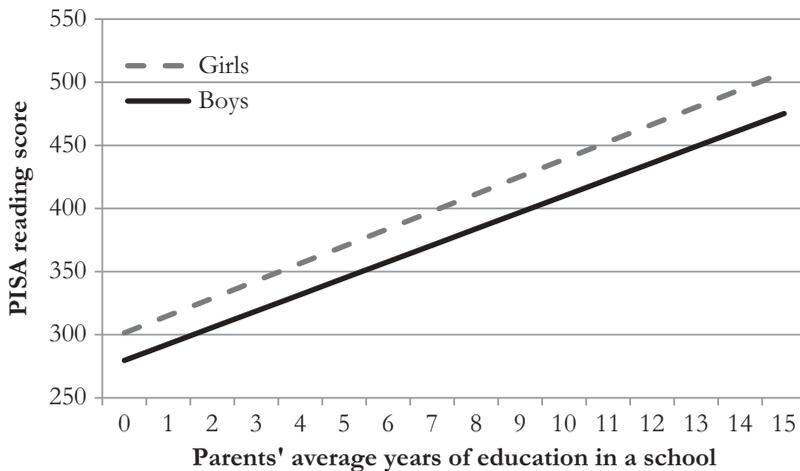
Figure 1. Female * socioeconomic composition.

Source: PISA 2009. $N_{students}$: 216,117; $N_{schools}$: 8,306; 33 countries.

school with the lowest socioeconomic composition was 8.25 points larger for girls than for boys ($b = 0.569$ * range = 14.5). We must therefore reject Hypothesis 1. Note, however, that this 8.25 gender difference must be considered in relation to the total effect of a school's socioeconomic composition. The difference between schools with the highest and the lowest socioeconomic composition was 196 points on PISA. As depicted in Figure 1, the divergence between the line for girls and the line for boys is hardly noticeable due to the strong main effect of a school's socioeconomic composition. Moreover, as this finding contradicts the results of Legewie and DiPrete (2012), we performed additional analyses in which we only selected German schools, like Legewie and DiPrete. These analyses showed no differential effect of a school's socioeconomic composition for girls and boys.[10]

In line with Hypothesis 2, in schools with more than 60% girls, boys on average perform better ($b = 17.700$), while for girls this effect is less strong ($b = -5.714$). Girls possibly set a more successful learning climate in the schools and classrooms, to which boys were more susceptible. Figure 2 depicts these differential effects on girls and boys. We see in Figure 2 that the gap in reading scores between students in a school with more than 60% girls versus students in a school with less than 60% girls was larger for boys than for girls. Contrary to Hypotheses 3 and 4, the proportion of teachers with a university education and schools' use of well-rounded assessment methods did not influence girls' and boys' reading performance differently. Lastly, we assessed the extent to which the differential effects of school resources were interpreted by the differential effect of a school's overall climate. Figure 3 visualises this cross-level interaction. We theorised that school resources affect girls and boys differently, in part through the school learning climate. Comparing Models 4 and 5, however, we observe that the effects of school resources were not substantially reduced by controlling for the cross-level interaction of female with school climate. Although school climate, in line with our theoretical model, exerted a stronger influence on boys than on girls, it did not seem to interpret the differential effects of school resources. Still, our results suggest that the advantage of attending a school with the
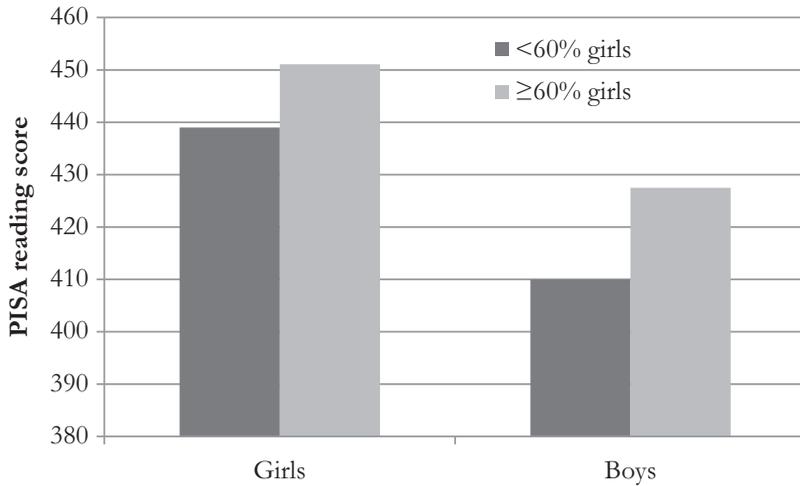
Figure 2. Female * 60% girls.

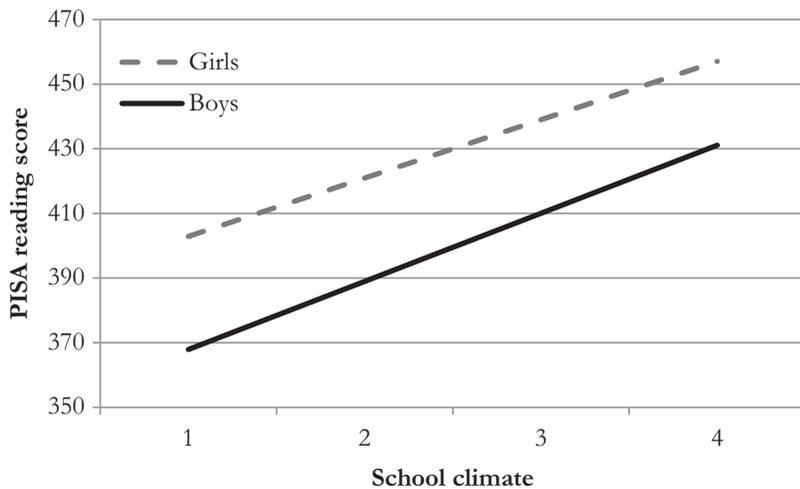Source: PISA 2009. $N_{students}$: 216,117; $N_{schools}$: 8,306; 33 countries.

Figure 3. Female * school climate.

Source: PISA 2009. $N_{students}$: 216,117; $N_{schools}$: 8,306; 33 countries.

best overall climate versus a school with the worst overall climate in terms of reading scores was 9 points higher for boys than for girls ($b = -3.017 *$ range = 3).

Conclusions

The aim of our study was to examine the extent to which school resources and school practices affect the reading performance of girls and boys differently. Such school characteristics have seldom been considered in prior research on the female advantage in education. We focused explicitly on reading scores because of the impact of reading

ability on overall educational attainment and later occupational careers. We tested hypotheses by performing state-of-the-art multilevel analyses in which we estimated cross-level interactions comparing effects on boys and on girls. Doing so produced robust results on whether school characteristics had differential effects on girls' and boys' reading performance in 33 OECD countries.

We found that schools with more than 60% girls, a large proportion of students with highly educated parents, and a large proportion of college-educated teachers had higher scoring students on the 2009 PISA reading test. Boys particularly seemed to be positively affected by a high proportion of female students in a school. We theorised that this was the result of classroom processes (learning opportunities) and also related to the schoolwide learning climate. Previously, Van Houtte (2004a) found that girls did not influence boys' learning directly but positively encouraged boys' educational performance through their contribution to a successful schoolwide learning climate. Although our direct measure of school learning climate (information provided by school principals) did not account for the gendered effect of the proportion of girls in a school, we feel this measure was probably too blunt to fully test this proposed indirect mechanism. In addition, we could not empirically distinguish between the working of these mechanisms in the classroom and at the school level. More research is needed to disentangle the mechanisms at these levels. This would not only shed more light on how a successful learning climate may be established in schools, how disorderly classrooms may hinder students' learning, and how different types of students are affected by these circumstances, but it would also have practical policy relevance. Redistributing students among classes within schools, for example, is a policy measure that could be implemented rather easily.

Our finding that boys in particular benefit from a large presence of girls in their school points to a possible negative side effect of vocationally partitioned educational systems. In contrast to comprehensive educational systems, vocationally partitioned educational systems separate students by field of study. As girls and boys consistently choose different fields of study, this leads to skewed gender distributions in schools (Charles & Bradley, 2002). The negative effects of horizontal gender segregation in education are well established for women, but our findings indicate that boys too may be disadvantaged by this phenomenon. These findings point to the relevance of policy measures that stimulate equal gender distributions in classes and schools.

Contrary to our expectations, the proportion of highly educated teachers and the socioeconomic composition of schools did not benefit boys more than girls. First, our results indicate that a large proportion of highly qualified teachers benefits both boys and girls equally. We recommend that future studies further explore the effects of teachers' pedagogical skills. This would shed light on the notion that more highly qualified teachers contribute more to boys' learning, in part because they are better able to maintain an orderly and stimulating classroom environment, to which boys are more susceptible. Second, an unexpected finding was that girls were more affected by a school's socioeconomic composition than boys. This contrasts with findings of Legewie and DiPrete (2012), though additional analyses on Germany showed no gendered effect. We therefore conclude that it depends on the country context whether and how schools' socioeconomic composition affects girls' and boys' reading scores. The mechanism proposed by Legewie and Diprete therefore needs to be tested more rigorously, using multiple indicators of school climate, measures of masculinity norms within schools, and analyses for various countries. Third, in our study, schools' use of

well-rounded assessment methods did not seem to affect girls' and boys' reading performance differently. In this regard, too, specific teacher or classroom data would provide a better test of the idea that boys are disadvantaged by more subjective evaluations by teachers. Considering the finding of Boonen, Van Damme, and Onghena (2014) that teachers' instructional practices are highly important for students' reading achievement and the robust finding of lower non-cognitive skill levels among adolescent boys (Jacob, 2002), we agree with DiPrete and Buchmann (2013) that greater understanding is needed of the extent to which non-cognitive abilities relate to educational performance and how this relation differs between contexts.

Our employment of information from a large dataset of schools and countries had some drawbacks. One downside of using PISA data to study school effects was an inability to control for students' prior achievement (Esser & Relikowski, 2015). Characteristics of students' primary and secondary school performance often correlate. However, we could not determine whether differences in students' reading scores were related to earlier performance due to primary school characteristics or to features of the current secondary school. In addition, we could not be sure that students were randomly distributed over the schools and classes, which implies the possibility of endogeneity problems. We dealt with this to the best of our ability by including relevant confounding school characteristics, such as public versus private, school size, and availability of school materials. Follow-up surveys should incorporate prior achievement in their models and strive for a random assignment of students to be included in the data collection.

Our study found, in line with previous research, that boys' lower reading performance in PISA was mitigated in an environment with predominantly female students (Van Houtte, 2004a). Future research might test whether these results hold when considering science or mathematics performance, school grades and other stages in educational careers (Downey & Vogt Yuan, 2005). Indeed, understanding of both differential educational effectiveness and the female advantage in education would be improved by application of the current study's hypotheses to other settings, such as subjects or stages in educational careers. Finally, future research might seek to deepen understanding of the theoretical mechanisms underlying the hypotheses explored here.

Notes

1. France was not included because the school questionnaire was not administered there. We included only OECD countries because we were more confident about assuming similar mechanisms when studying countries with more similar contexts. All OECD countries have highly developed economies. The 2009 PISA included 448 schools in which 100% of the student population consisted of girls or boys.
2. Additional analyses showed that the Eta R square did not differ whether including this variable linearly or in categories.
3. There were considerable missing values on the variable indicating students' family structure. Students with missing values on this variable were most similar to students who lived in families that were not two-parent or one-parent families (e.g., they lived with their grandparents or with brothers or sisters). As this concerned a control variable, and we wanted to keep our models as parsimonious as possible, we added students with missing values on this variable to the 0 category.
4. We distinguished between schools with more and less than 60% girls because our proposed mechanisms were considered most applicable when a clear majority of the students

was female instead of expecting a linear effect for every added percentage. Results did not differ when we took other cut-off points (55%, 65%) or added the proportion of girls to the models linearly (results available upon request). Robustness checks employing the jackknife procedure showed that the (cross-level) effects of gender composition remained virtually unchanged when subsequently leaving out one country at a time.

5. The 2009 PISA data offered additional measures of well-rounded assessment methods. We chose this variable because the interpretation of others was less clear (i.e., teachers' ratings, student portfolios). In Canada, Germany, Ireland, and the United States, there was little between-school variance on this item.

6. School materials can be considered an indicator of school resources. Since our theoretical framework focused on resources in terms of teacher and student population, we did not formulate a hypothesis on the availability of school materials but included it as control variable. Additional analyses showed no effect of school materials on gender differences in reading performance.

7. We controlled for country fixed effects because we did not focus on the country level, the country means of our sample were not normally distributed, and the number of countries in our dataset was limited.

8. The online appendix contains the R syntax for normalising weights and the Julia syntax in which we determined our models.

9. Adding all cross-level interactions separately yielded virtually the same estimates. Robustness checks in which controls for ability tracking of students were included produced practically identical results.

10. Outcomes of these analyses are available upon request.

Acknowledgement

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Margriet van Hek is a postdoctoral researcher in the Department of Sociology/ICS, Utrecht University, The Netherlands. Her postdoctoral research focuses on the impact of female leadership on social inequalities in public and private organizations and in public administration. Her main research interests comprise gender issues, social stratification, educational sociology, and cultural tastes and lifestyles.

Gerbert Kraaykamp is a full Professor in Empirical Sociology in the Department of Sociology/ICS, Radboud University, Nijmegen, The Netherlands. His major research interests include educational inequality, partner effects, and health inequality. He has published on these subjects widely in international journals.

Ben Pelzer is an assistant professor at the Department of Sociology at the Radboud University, Nijmegen, The Netherlands. His main interests and recent publications are in the field of multilevel analysis and categorical data analysis.

ORCID

Margriet van Hek   http://orcid.org/0000-0003-3989-9933

References

Amato, P. R. (2001). Children of divorce in the 1990s: An update of the Amato and Keith (1991) meta-analysis. *Journal of Family Psychology*, 15, 335–370. doi:10.1037//0893-3200.15.3.355

Betts, J. R., & Shkolnik, J. L. (1999). The behavioral effects of variations in class size: The case of math teachers. *Educational Evaluation and Policy Analysis*, 21, 193–213. doi:10.3102/01623737021002193

Bezanson, J., Karpinski, S., Shah, V. B., & Edelman, A. (2012). *Julia: A fast dynamic language for technical computing*. Retrieved from https://arxiv.org/pdf/1209.5145.pdf

Bishop, J. H. (1997). The effect of national standards and curriculum-based exams on achievement. *The American Economic Review*, 87, 260–264. doi:10.2307/1183407

Bishop, J. H., Bishop, M., Gelbwasser, L., Green, S., Zucherman, A., Schwartz, A. E., & Labaree, D. F. (2003). Nerds and freaks: A theory of student culture and norms. *Brookings Papers on Education Policy*, 6, 141–213. doi:10.1353/pep.2003.0002

Boonen, T., Van Damme, J., & Onghena, P. (2014). Teacher effects on student achievement in first grade: Which aspects matter most? *School Effectiveness and School Improvement*, 25, 126–152. doi:10.1080/09243453.2013.778297

Buchmann, C., DiPrete, T. A., & McDaniel, A. (2008). Gender inequalities in education. *Annual Review of Sociology*, 34, 319–337. doi:10.1146/annurev.soc.34.040507.134719

Charles, M., & Bradley, K. (2002). Equal but separate? A cross-national study of sex segregation in higher education. *American Sociological Review*, 67, 573–599. doi:10.2307/3088946

Cheung, S. Y., & Andersen, R. (2003). Time to read: Family resources and educational outcomes in Britain. *Journal of Comparative Family Studies*, 34, 413–433. doi:10.1108/s1479-353920150000019010

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school. A cross-subject analysis with student fixed effects. *The Journal of Human Resources*, 45, 655–681. doi:10.3386/w13617

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). Algorithm-guided individualized reading instruction. *Science*, 315, 464–465. doi:10.1126/science.1134513

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. Abingdon, UK: Routledge.

Crul, M., & Schneider, J. (2009). Children of Turkish immigrants in Germany and the Netherlands: The impact of differences in vocational and academic tracking systems. *Teachers College Record*, 111, 1508–1527. doi:10.1057/9781137308634_3

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *The American Economic Review*, 95, 158–165. doi:10.1257/000282805774670446

DiPrete, T. A., & Buchmann, C. (2013). *The rise of women: The growing gender gap in education and what it means for American schools*. New York, NY: Russell Sage Foundation.

DiPrete, T. A., & Jennings, J. L. (2012). Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research*, 41, 1–15. doi:10.1016/j.ssresearch.2011.09.001

Downey, D. B., & Vogt Yuan, A. S. (2005). Sex differences in school performance during high school: Puzzling patterns and possible explanations. *The Sociological Quarterly*, 46, 299–321. doi:10.1111/j.1533-8525.2005.00014

Driessen, G., & Sleegers, P. (2000). Consistency of teaching approach and student achievement: An empirical test. *School Effectiveness and School Improvement*, 11, 57–79. doi:10.1076/0924-3453(200003)11:1;1-#;ft057

Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do teachers race, gender, and ethnicity matter? Evidence from the national educational longitudinal study of 1988. *Industrial and Labor Relations Review*, *48*, 547–561. doi:10.2307/2524781

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2007). Early schooling: The handicap of being poor and male. *Sociology of Education*, *80*, 114–138. doi:10.1177/003804070708000202

Esser, H., & Relikowski, I. (2015). *Is ability tracking (really) responsible for educational inequalities in achievement? A comparison between country states Bavaria and Hesse in Germany* (Discussion Paper No. 9082). Bonn, Germany: IZA.

Eurostat. (2013). *Persons with tertiary education attainment by age and sex (%)*. Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=edat_lfse_03&lang=en

Farkas, G., Grobe, R. P., Sheehan, D., & Shuan, Y. (1990). Cultural resources and school success: Gender, ethnicity, and poverty groups within an urban school district. *American Sociological Review*, *55*, 127–142. doi:10.2307/2095708

Francis, B. (2000). *Boys, girls, and achievement: Addressing the classroom issues*. London, UK: Routledge Falmer.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, *66*, 361–396. doi:10.3102/00346543066003361

Hallinan, M. T. (1988). Equality of educational opportunity. *Annual Review of Sociology*, *14*, 249–268. doi:10.1146/annurev.soc.14.1.249

Hallinger, P., & Heck, R. H. (2011). Exploring the journey of school improvement: Classifying and analyzing patterns of change in school improvement processes and learning outcomes. *School Effectiveness and School Improvement*, *22*, 1–27. doi:10.1080/09243453.2010.536322

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, *19*, 141–164. doi:10.2307/1164207

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). *The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior* (Working Paper No. 12006). Cambridge, MA: NBER.

Hopkins, D., & Stern, D. (1996). Quality teachers, quality schools: International perspectives and policy implications. *Teaching and Teacher Education*, *12*, 501–517. doi:10.1016/0742-051x(95)00055-o

Jackson, C., & 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. doi:10.1080/09589230903260019

Jacob, B. A. (2002). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics Education Review*, *21*, 589–598. doi:10.1016/s0272-7757(01)00051-6

Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., & Patrick, H. (2006). Sex differences in math performance: The role of children's approach to schoolwork. *Developmental Psychology*, *42*, 11–26. doi:10.1037/0012-1649.42.1.11

Lazear, E. P. (2001). Educational production. *The Quarterly Journal of Economics*, *116*, 777–803. doi:10.1037/0012-1649.42.1.11

Legewie, J., & DiPrete, T. A. (2012). School context and the gender gap in educational achievement. *American Sociological Review*, *77*, 463–485. doi:10.1177/0003122412440802

Louis, K. S., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national US survey. *School Effectiveness and School Improvement*, *21*, 315–336. doi:10.1080/09243453.2010.486586

Ma, X. (2008). Within-school gender gaps in reading, mathematics and science literacy. *Comparative Education Review*, *52*, 437–460. doi:10.2307/30218831

Maag Merki, K., & Holmeier, M. (2015). Comparability of semester and exit exam grades: Long-term effect of the implementation of state-wide exit exams. *School Effectiveness and School Improvement*, *26*, 57–74. doi:10.1080/09243453.2013.861353

Machin, S., & McNally, S. (2005). Gender and student achievement in English schools. *Oxford Review of Economic Policy*, *21*, 357–372. doi:10.1093/oxrep/gri021

Martin, M. O., & Mullis, I. V. S. (Eds). (2013). *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade – Implications for early learning*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Mislevy, J., & Sheehan, K. M. (1987). Marginal estimation procedures. In A. E. Beaton (Ed), *Implementing the new design: The NAEP 1983–84 technical report* (Report No. 15-TR-20) (pp. 293–360). Princeton, NJ: National Assessment of Educational Progress, Educational Testing Service.

Montt, G. (2011). Cross-national differences in educational achievement inequality. *Sociology of Education*, *84*, 49–68. doi:10.1177/0038040710392717

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*, 237–257. doi:10.3102/01623737026003237

Organisation for Economic Co-operation and Development. (2009). *PISA data analysis manual: SPSS, second edition*. Paris, France: Author.

Organisation for Economic Co-operation and Development. (2010). *Pathways to success – How knowledge and skills at age 15 shape future lives in Canada*. Paris, France: Author.

Organisation for Economic Co-operation and Development. (2012). *PISA 2009 technical report*. Paris, France: Author.

Organisation for Economic Co-operation and Development. (2015). *The ABC of gender equality in education: Aptitude, behaviour, confidence*. Paris, France: Author.

Pickering, J., & Lodge, C. (1998). Boys' underachievement – Challenging some assumptions about boys. *Improving Schools*, *1*(1), 54–60. doi:10.1177/136548029803010120

Piopiunik, M., Hanushek, E. A., & Wiederhold, S. (2014). *The impact of teacher skills on student performance across countries* (Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik; Session: Education 1, No. A18-V3). Retrieved from http://hdl.handle.net/10419/100356

Ponitz, C. C., Rimm-Kaufman, S. E., Brock, L. L., & Nathanson, L. (2009). Early adjustment, gender differences, and classroom organizational climate in first grade. *The Elementary School Journal*, *110*, 142–162. doi:10.1086/605470

Reardon, S. F., Arshan, N., Atteberry, A., & Kurlaender, M. (2010). Effects of failing a high school exit exam on course taking, achievement, persistence, and graduation. *Educational Evaluation and Policy Analysis*, *32*, 498–520. doi:10.3102/0162373710382655

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*, 417–458. doi:10.1111/j.1468-0262.2005.00584.x

Rumberger, R. W., & Palardy, G. J. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. *Teachers College Record*, *107*, 1999–2045. doi:10.1111/j.1467-9620.2005.00604.x

Salomone, R. C. (2003). *Same, different, equal: Rethinking single-sex schooling*. New Haven, CT: Yale University Press.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford, UK: Pergamon.

Schneeweis, N., & Zweimüller, M. (2014). Early tracking and the misfortune of being young. *The Scandinavian Journal of Economics*, *116*, 394–428. doi:10.1111/sjoe.12046

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*, 454–499. doi:10.3102/0034654307310317

Stoet, G., & 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*(3), e57988. doi:10.1371/journal.pone.0057988

Van Hek, M., Kraaykamp, G., & Wolbers, M. H. J. (2015). Family resources and male–female educational attainment: Sex specific trends for Dutch cohorts (1930–1984). *Research in Social Stratification and Mobility*, *40*, 29–38. doi:10.1016/j.rssm.2015.02.001

Van Hek, M., Kraaykamp, G., & Wolbers, M. H. J. (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. doi:10.1080/13803611.2016.1256222

Van Houtte, M. (2004a). Gender context of the school and study culture, or how the presence of girls affects the achievement of boys. *Educational Studies*, *30*, 409–423. doi:10.1080/0305569042000310336

Van Houtte, M. (2004b). Why boys achieve less at school than girls: The difference between boys' and girls' academic culture. *Educational Studies*, *30*, 159–173. doi:10.1080/0305569032000159804

Vantieghem, W., & Van Houtte, M. (2015). Differences in study motivation within and between genders: An examination by gender typicality among early adolescents. *Youth and Society*. Advance online publication. doi:10.1177/0044118X15602268

von Davier, M., & Hastedt, D. (Eds). (2009). *IERI monograph series: Issues and methodologies in large-scale assessments: Volume 2*. Hamburg, Germany: IEA-ETS Research Institute (IERI).

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, *140*, 1174–1204. doi:10.1037/a0036620

Wachs, T. D., Gurkas, P., & Kontos, S. (2004). Predictors of preschool children's compliance behavior in early childhood classroom settings. *Journal of Applied Developmental Psychology*, *25*, 439–457. doi:10.1016/j.appdev.2004.06.003

Wentzel, K. R., & Looney, L. (2007). Socialization in school settings. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization: Theory and research* (pp. 382–403). New York, NY: The Guilford Press.

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, *31*, 114–128. doi:10.1016/j.stueduc.2005.05.005

Younger, M., Warrington, M., & Williams, J. (1999). The gender gap and classroom interactions: Reality and rhetoric? *British Journal of Sociology of Education*, *20*, 325–341. doi:10.1080/01425699995290