

Centre for Policy Research on
Men and Boys

**International
Literature Review
on Boys' Education:**
Part II (Context)

Jordan Hill

Foreword by Nick Isles

Contents

Foreword	3
Executive Summary	5
Introduction	7
Original Analysis	8
Developmental Patterns	9
Energy and Movement	10
Executive Function	11
Pubertal Timing in Adolescence	11
Socio-economic Conditions	14
Country Wealth.....	15
Local Conditions.....	16
School Composition	18
Gender Norms and Expectations	21
Masculine Scripts	22
Working Class Masculinities.....	24
Academic Stereotypes	26
Education System Features	28
Student Stratification	29
Curriculum, Accountability and Assessment.....	33
Teaching Workforce.....	37
Conclusions	41
Recommendations	42
Research Gaps	44
References	45

Foreword

The issue of boys' aggregate lower achievement than girls has been longstanding. As this research project has found, this holds across jurisdictions, national borders and individual institutions. There are of course exceptions. There is also plenty of evidence of many boys doing well. Indeed, in mathematics boys hold a small advantage over girls. But what is driving this consistent picture of relative under performance by boys? Why are so many boys struggling in a variety of education systems? And importantly what can be done to change things? This last question the Centre for Policy Research on Men and Boys (CPRMB) is trying to address for England and Wales through its [Boys' Education Commission](#).

In this second report (in a series of three) the CPRMB has looked closely at the contexts that affect boys' educational outcomes across the high-income countries that are the focus of the research. The report looks at the reasons why a small disadvantage when boys start school relative to girls turns into a substantial gap in achievement and outcomes for so many at the end of school.

This clearly should matter. Lower school attainment often means less opportunity in the world of work. Employers often see school attainment and achievement (rightly or wrongly) as lenses through which to view aspects such as ability to learn, dedication to the task, orientation to work and potential to thrive in the workplace. This is, of course, not the only reason why doing well through school is important. Learning to learn enables individuals to enjoy a thriving internal life and to better navigate the challenges adult life may bring. It brings a sense of self-worth and self-respect. It allows those who achieve through school to feel fulfilled. The opposite can mean a life half lived; a life unfulfilled and higher and greater obstacles to overcome.

Boys under achievement also matters at a community and societal level. Failing boys can mean struggling communities. Boys not being able to maximise their talents can lead to an economy wide impact on areas such as levels of worklessness, levels of productivity, the costs of welfare and much more. And the report offers no single simple solution. The barriers to overcome are (at a minimum) structural, psychological and pedagogical. If this could be summed up in a phrase it might be that too many education systems are still stuck with models of schooling fit for the industrial age not the age of digital transformation.

The findings in the report, summarised in the Executive Summary below, offer strong insights into the contexts that drive boys' educational outcomes and then what might need to change to improve those outcomes. For change we must. All boys and every girl deserve the best opportunities through school to fulfil their potentials. This report points the way as to how.

Nick Isles

Director
Centre for Policy Research on Men and Boys

Centre for Policy Research on Men and Boys

The Centre for Policy Research on Men and Boys (CPRMB) is a research organisation and think tank dedicated to understanding and addressing issues that uniquely affect men and boys in the UK. Its vision is a world where men and boys of all backgrounds thrive in their families and communities, one where the sexes can rise together by supporting each other.

It will do this by supporting the development of new research, thinking and insight into policy affecting men and boys. This is by focusing on critical areas where they face unique challenges, promoting policy change, and fostering public awareness and understanding. These are in seven core areas: Economy, Employment and Skills; Education; Health; Fatherhood and Family; Criminal Justice; Male Identity; and Portrayal of Men in Media and Culture.

This report was authored by Jordan Hill. Jordan is a Research Fellow at the Centre, specialising in education policy. A former Education Policy Analyst at the OECD Directorate for Education and Skills, he has also held research and policy roles in Brussels and London.

Executive Summary

Part II of the International Literature Review on Boys' Education examines the contexts that shape boys' educational outcomes across high-income countries. Following on from part I, which described the gaps, part II asks what conditions contribute to them. It does so across four areas: developmental patterns, socio-economic conditions, gender norms and expectations, and the design of education systems.

One argument runs through the review. Boys and girls bring modest average differences into school and leave with substantial gaps in achievement and attainment. The differences at the start are real but small, and the two groups overlap far more than they differ. The gaps at the end are amplified along the way rather than given at the start.

Boys' outcomes are more tied to local socio-economic conditions than are girls'.

That holds most clearly for long-run trajectories into work and for progression into higher education. It does not hold for the gap in reading and language, which is large, present in almost every place studied, and mostly unexplained by local socio-economic conditions. Local conditions often determine the socio-economic composition of schools. This bears differently on boys and girls, with boys gaining more from an advantaged intake.

Developmental patterns matter mainly because schooling fixes consequences to age.

Girls develop self-regulation earlier, boys are on average more physically active, and pubertal timing differs. None of these is large enough on its own to account for the gaps Part I described. They become consequential because an education system decides at what age a child starts school, when it sorts students, and when it certifies attainment. Each of those decisions captures boys at a less developmentally mature point than girls. Boys' underachievement reflects the interaction between modest developmental differences, such as later brain maturation and pubertal timing, and the institutional, social and economic conditions in which boys grow up and learn. The conditions, rather than boys themselves, are where the room for action lies.

Gender norms shape what confers status among boys, and status runs against much of what school rewards.

Dominant masculine scripts, such as 'lad culture', often discourage visible effort, help-seeking and reading. Most boys do not hold those scripts in their strongest form, but departing from them carries a cost among peers. Norms of this kind do not shift as a society becomes more equal, and boys' stereotypes about what they are suited to are visible well before secondary school. Importantly, the evidence also demonstrates that masculinities are diverse, adaptable and shaped by institutional environments, suggesting that schools can play a positive role in fostering forms of masculinity that are compatible with academic engagement and aspiration.

The design of education systems shapes the size of the gap a system records.

Being held back a year or more (grade retention) is a gate that boys pass through more often than girls, even at the same level of measured performance. Whether a boy repeats a year depends substantially on which school he attends. Sorting decisions (e.g. tracking, streaming and setting) place boys downwards more often than their measured performance accounts for, and movement back up is rare. The balance between teacher-assigned grades and standardised examinations

changes the size and sometimes the direction of the recorded gap. How schools are held accountable is also crucial, as accountability systems determine which students a school directs its attention towards. Each of these is a design choice rather than a fact about children.

The review makes nine recommendations.

The strongest concern grade retention, which should be minimised, and the age of selection into different educational tracks and streams, where boys' underachievement is not a reason to select earlier. Others address the limits of recruiting more male teachers, curricular prescription in reading and language, the choice between threshold and growth school accountability measures, the balance between grades and examinations, the use of vocational routes as a default destination for struggling boys, teacher preparation for the classrooms where struggling boys are concentrated, and building movement into the school day.

The review does not endorse changes that improve one group's outcomes by worsening another's.

That applies between boys and girls, between boys and children from poorer homes, and between lower-attaining and higher-attaining students. Several of the levers examined would narrow the gender gap by lowering girls' attainment, and each is set aside on that basis. Redistribution of that kind moves who carries the cost and leaves the educational arrangements producing the gap in place. Improving boys' education should never be a zero-sum game.

Part III turns to the drivers closest to the child, among them family and home, engagement with school, special educational needs, the learning environment and wellbeing. The evidence in Part II establishes the conditions within which those drivers operate.

Introduction

Part I of this review described the gaps between boys' and girls' education outcomes across high-income countries, showing that boys are substantially overrepresented among students experiencing the greatest educational difficulties. Part II asks what contextual factors contribute to these outcomes.¹ It begins with the developmental patterns children bring into the classroom, then widens to the socio-economic conditions they grow up in, the gender norms they learn to perform, and the features of the education systems that sort and teach them. One argument runs through all four sections:

Boys and girls bring only modest psychological differences into school, real on average but small, and with the two groups overlapping far more than they differ. They leave it with gaps in achievement and attainment that are often considerable. Small differences at the start of schooling, in matters such as the timing of self-regulation, become large gaps at the end because of how schooling is arranged around them, through age-graded classes, fixed-age testing, the timing of selection into different tracks, and the cultures schools and neighbourhoods inhabit. The gaps at the end are amplified along the way, rather than given at the start.

Boys facing several forms of disadvantage at once have the poorest outcomes of any group. Part I established that at age fifteen more than four in ten boys in the poorest wealth quintile fall short of the minimum reading standard, against fewer than one in six boys in the wealthiest. However, the gap between boys and girls is not concentrated among disadvantaged students. Part I found boys' reading underperformance running across socio-economic groups, and on some measures it was wider among advantaged students than among disadvantaged ones. Although disadvantaged boys have the most to gain from improvement, they are not the group in which the difference between boys and girls is largest.

The evidence in this field often appears contradictory. Where one study finds an effect, another finds none, while a third finds the opposite. This can make it look as though nothing is settled and nothing can be acted on. That impression is mostly mistaken. When careful studies of a similar question reach different answers, it is usually because they were not measuring the same thing. Achievement, school grades, further education progression, confidence and career choice are different outcomes, and a single factor can move one while leaving another untouched. Evidence on similar themes often looks at different age groups and, because boys and girls develop at different speeds, a finding at ten and a finding at fifteen can both be correct despite being different. Studies also use different tests, and the way something is measured shapes the size of the gap it shows. They were carried out in different countries, where the gaps genuinely differ from place to place, so variation between studies is frequently the finding rather than a fault in it (Bianchi et al., 2025). And some studies can separate which students attend a school from what the school actually does, while others cannot, so an apparent advantage can shrink to nothing once the fairer comparison is made. None of this is a reason to abandon the evidence. One additional feature of the evidence base should be flagged early. The research on puberty and development, on stereotypes, and on curriculum reform has more often studied girls than boys, so the picture for boys is frequently thinner, or assembled by inference, or simply missing. The review tries to be transparent about this where it occurs.

² This second report does not try to explain boys' outcomes through individual motivation, behaviour or family circumstances, which are the subject of Part III. For an in-depth discussion of the overarching conceptual framework developed for all three parts of this international literature review, see the document "About This Review", available for download on the CPRMB website.

One axis of disadvantage deserves particular attention at the outset. In the United Kingdom, race and ethnicity are central to how educational disadvantage is understood, and for good reason. The underachievement of some minority ethnic boys is well documented. A reader coming to a review of boys' education will reasonably expect race and ethnicity to feature. They do, but not in the way a UK reader might expect, and the reason is the way the international evidence is structured. Race and ethnicity, as recorded in the United Kingdom, are national categories. Terms such as black Caribbean, Pakistani or Bangladeshi do not map onto other countries with different populations. Many European systems do not collect data on race or ethnicity at all, and France is legally prohibited from doing so. What is available instead is migration background, whether a child, or their parents, were born abroad. The disadvantage a British reader thinks of in racial or ethnic terms is therefore captured, elsewhere, as a migration category, and the two are not interchangeable. This review uses migration background when analysing across countries.

Part I showed this at work. Its analysis of international data could examine migration background and language, and on those measures boys from a migrant background record among the lowest reading scores of any group, with the penalty persisting across the socio-economic spectrum. Advantaged migrant boys still lag advantaged native boys, whereas for girls the migration gap narrows as socio-economic status rises (Varsik & Gorochevskij, 2023). The scale of that gap is established in Part I. Its mechanisms are mostly ones that act close to the child, in classrooms and families, and they are taken up in Part III. Part II carries the part that is structural. Where data is available by race/ ethnicity (for example in England and the United States), it is clearly signalled. The clearest evidence there is from the United States, where neighbourhood and opportunity structure bear on boys by race, and it is set out in the discussion of local conditions.

One constraint applies throughout. The review does not endorse changes that improve boys' outcomes by worsening girls'. That applies equally between boys and children from poorer homes, and between lower-attaining and higher-attaining students. Redistribution of that kind simply moves who carries the cost but it does not resolve educational inequalities.

Original Analysis

Alongside the published literature, the review presents a body of original analysis of international large-scale assessment data, conducted specifically for this review. The data are cross-sectional and the original analysis is positioned as a starting point, rather than a conclusion. It is used to establish what the broad comparative picture looks like. The questions it raises, above all questions of cause, mechanism and process, are then addressed by the longitudinal research, qualitative studies, controlled trials and theoretical literature. A full description of the original analysis, including the variables used, the construct it measures and a short definition, is listed in the "Original Analysis Note" downloadable from the CPRMB website. Readers who wish to see precisely what was analysed, and how each measure relates to the others, should look there.

Developmental Patterns

Developmental patterns refer to variability in physical, cognitive and psychosocial maturation that shapes how children interact with schooling at different ages. This section looks at three patterns: higher levels of energy and movement, later development of executive functions required for sustained engagement with learning and the educational implications of puberty. While these patterns are not exclusively gendered, their timing and expression often differ between boys and girls, with important implications for how students respond to school environments and expectations. These developmental patterns are average differences with substantial overlap rather than categorical divides, and their educational consequences arise not from the differences themselves but from the interaction between developmental timing and schooling demands that are age-graded and often ill-matched to the range of developmental trajectories in any given cohort.

Despite there being significant overlap between boys and girls in these developmental patterns, they tend to feature heavily in the attitudes and approaches parents, teachers, wider society, and boys themselves take to the male underperformance in education. This risks turning them into self-fulfilling prophecies (European Commission, 2021). Interventions to counter these stereotyped attitudes and approaches not only within schools but also within family and community contexts are important.

Energy and Movement

In their foundational meta-analysis of more than 90 studies covering birth-30 years of age, Eaton and Enns found substantial differences between boys and girls in terms of their activity levels and found evidence that social influences magnify existing differences, rather than create them (Eaton & Enns, 1986). A later meta-analysis finds that, on average, boys are slightly more active, less shy, and derive more pleasure than girls from high-intensity stimuli (Else-Quest et al., 2006). These differences show up in how boys and girls socialise. Both theory and empirical evidence suggest that by preschool, boys and girls have developed distinct patterns of social engagement (Maccoby, 2003). Boys tend to engage more often in rough-and-tumble play than girls, and play fighting appears to serve as an outlet for energy and a context for developing social skills rather than an expression of aggression (Maccoby, 2003; Pellegrini & Bohn-Gettler, 2005). Although it is more typically young boys who use physical play to understand their surroundings and relationships, not all boys show these behaviours, while a significant number of girls do (Hart & Tannock, 2018). It is therefore important not to overstate the role of gender in determining energy and movement levels. High-energy children who find extended sedentary school requirements challenging are more likely to be male, though many boys manage these demands comfortably and many girls do not.

There is some evidence that more opportunities for physical activity during the school day can positively influence children's cognitive performance. In the study of children in North America and the United Kingdom by Pellegrini and Bohn-Gettler (2005), children were more attentive after than before recess and boys' attention was especially sensitive to longer periods without recess. Short, structured activity breaks (sometimes called brain breaks) during the school day appear to improve on-task behaviour. For example, in evaluations for the whole-of-school programme TransformUs in Victoria, Australia, Salmon and colleagues record that more than 80% of teachers reported improved concentration among students and time-on-task following active breaks (Koorts et al., 2022; Salmon et al., 2023). This particular programme is designed to create supportive environments and integrate meaningful movement into daily teaching across all subjects for all students (ages 3-18).

Adults often misidentify high energy activities (for example rough play) as aggressive and try to prevent it. However, it serves important developmental functions such as learning emotional control and how to exercise restraint in potentially competitive or conflictual situations. Another outlet for children's energy and movement needs, while also fostering their sense of agency,

is independent mobility, such as walking or cycling to school alone. Children's opportunities to explore the environment on their own have been increasingly restricted by adults, impeding their ability to develop the risk-management skills that pay dividends in adulthood (Hill & Shewbridge, 2025). International analysis shows that more than two thirds of children growing up today do not engage in sufficient physical activity, putting them at increased risk of poorer health (Aubert et al., 2022). Integrating more physical movement and activity in the journey to and from school and throughout the school day offers significant benefits to all children both in terms of their education and wider life outcomes.

Executive Function

Research has shown that relative age within a class cohort matters for outcomes, with systematic evidence indicating that relatively younger children in same-class groups show significantly lower mean test scores, higher grade repetition rates, and lower social skills (Urruticoechea et al., 2021). Yet, little attention has been paid to gender as a source of maturational differences within cohorts (Torvik et al., 2021). Executive function refers to the deliberate, top-down neurocognitive processes involved in self-regulation. Self-regulation is a resource that allows for control of thought, feeling, and behaviour in the context of stressors and is important for life and education outcomes (Robson et al., 2020). On average, girls show earlier development of self-regulation and related executive function capacities, a difference that contributes to early gender gaps in academic performance because schooling rewards capacities such as sustained attention, task persistence and compliance, where girls hold a developmental advantage in the early years (Matthews et al., 2009).

Not all boys struggle in building self-regulatory competence, however Matthews et al (2009) identified a group of poorly regulating boys who may be at risk for school difficulties as they progress through the education system. The interaction between developmental timing and school demands is visible in research on age-of-school-entry. Children who start school young relative to their peers show persistent academic disadvantages, and these effects are typically larger for boys (Bedard & Dhuey, 2006; Dee & Sievertsen, 2018). In their meta-analysis, Else-Quest et al. (2006) found consistent evidence of female advantage in attention regulation and inhibitory control, capacities whose deficits are central to behavioural difficulties. Similarly, Duckworth and Seligman (2006) found that eighth-grade girls were more self-disciplined than boys across multiple measures (delay of gratification, self-report, teacher and parent ratings), and that self-discipline predicted academic performance more strongly than IQ. This leads the authors to suggest that gendered differences in self-regulation are consequential for academic outcomes, independently of cognitive ability.

These differences are not fixed traits. Executive function is shaped by experience and stays open to change, most of all in the preschool years and again at the transition to adolescence (Zelazo & Carlson, 2012). The periods of greatest malleability coincide with the points at which systems take their most consequential structural decisions, the age at which children start school and the age at which they are sorted into different classes, sets or programmes. Together, the evidence suggests that girls' earlier development of self-regulation and executive function, combined with schools' reliance on precisely these capacities for success in early years, contribute to the emergence of gender gaps in early academic performance. Although age-based schooling produces classes containing students at varying developmental stages regardless of gender, the consequences fall unevenly. Children who are young relative to their classmates show lower mean test scores, higher rates of grade repetition and weaker social skills, and these effects are typically larger for boys.

Pubertal Timing in Adolescence

The socio-emotional system that drives reward-seeking and sensation-seeking develops rapidly around puberty, while the cognitive control system that supports self-regulation

continues developing into the mid-twenties (Steinberg, 2008). Capability and control therefore mature on different schedules. Puberty and adolescence are separable both conceptually and neurobiologically. Puberty is the process leading to reproductive maturation. Adolescence is the wider set of behavioural, cognitive and socio-emotional changes. The evidence on both is weighted towards girls, whose pubertal development is easier to measure and has been studied more (Oakley, 2025), and the same imbalance runs through the literature on adolescent change (Steinberg, 2008). The picture for boys is correspondingly thinner.

Relevant evidence includes Torvik and colleagues (2021), using twin data from the British Twins Early Development Study, to examine the relationship between pubertal timing and academic achievement at age 16 among 13,477 twins. They found that earlier puberty was associated with higher academic achievement in both boys and girls, and that including measures of pubertal development as explanatory variables reduced the gender gap in achievement by 40%. Part of this association was rooted in prepubertal differences, indicating that the relationship predates the physical changes of puberty itself. The twin methodology used also allowed the authors to test whether the association reflected a direct influence of puberty or shared underlying factors.

The results of this analysis suggest that pubertal timing does not directly cause academic outcomes, but rather that both are shaped by shared genetic factors affecting broader developmental maturation of psychological traits relevant to schoolwork, with conscientiousness suggested as a candidate mechanism (Torvik et al., 2021). The authors suggest that grading students at a fixed age when developmental maturation varies substantially may not accurately measure academic potential, since age-based assessment systematically captures boys at a less developmentally mature point than girls. In other words, the achievement gap may reflect underlying developmental differences that shape both pubertal timing and traits like conscientiousness.

Research carried out on British populations also suggests that the effect of puberty on education outcomes is mediated by social context, rather than being purely physiological (Borra et al., 2023). Analysis of Norwegian data broadly supports this view, though the results explain a smaller share than in the British studies (Røed et al., 2025). The authors attribute this difference to institutional factors, particularly school tracking systems (which are discussed later in this report). Together, the findings point to an important moderating role for system design. In other words, the educational consequences of developmental timing are not fixed, but shaped in part by how school systems are structured.

Although some studies have shown that earlier pubertal timing is associated with better academic achievement, it is also associated with increased risk for poor psychosocial, behavioural, and physical health during adolescence for boys and girls, while later timing is generally protective (Hoyt et al., 2020). Despite the traditional view that early puberty is advantageous for boys, it has been robustly associated with negative psychological effects (Mendle & Ferrero, 2012). These effects are not uniform, however. In a longitudinal study of African American boys, researchers found that the effects of early puberty on boys' social competence and likelihood to commit crimes were contingent on important contextual factors, such as membership of deviant peer groups, poor school experience, harsh parenting, and neighbourhood disorganisation (Klopack et al., 2020). For some boys, earlier puberty was linked to higher crime and delinquency, while for others earlier puberty is associated with beneficial psychosocial development. These seemingly contradictory results represent two divergent pathways by which earlier pubertal timing can affect boys, depending on the wider context.

Taken together, the evidence suggests that average gender differences in pubertal timing and adolescent developmental maturation contribute substantively to the emergence and widening of gender gaps in educational attainment during lower secondary schooling. The mechanism is

not direct: pubertal timing and academic achievement appear to share a common developmental root rather than standing in a direct causal relationship. The educational consequences depend heavily on social context, institutional design, and the fit between developmental stage and school environment. These findings suggest that the gender gap in attainment at age 16 is partly a product of institutional arrangements such as age-graded schooling, fixed-age assessment, and the timing of critical school transitions.

Socio-economic Conditions

The wealth of the country, the conditions of the locality and region, and the social composition of the particular school are all important for education outcomes. Each works through somewhat different mechanisms. All children carry the effects of growing up in low-income households and communities, although researchers argue that boys are more exposed (Brozo, 2019). Boys growing up in poverty are at heightened risk of struggling to read, losing their way academically, leaving school early, and ending up underemployed, out of work, and socially and economically marginalised. This section takes the three levels in turn. Importantly, it excludes family-level socio-economic status (SES), which is addressed in Part III. This split acknowledges the closer proximity of family and home mechanisms on the education outcomes of boys.

Country Wealth

The wealth of a country shapes both how strongly other aspects of socio-economic background bears on achievement and how the gender gap presents. Analysis of PISA 2018 using data-mining techniques found that student achievement in low-SES schools was more strongly conditioned by the country's level of economic development than in high-SES schools (Gamazo & Martínez-Abad, 2020). Put simply, where a school served disadvantaged communities, the national economic context matters more for how its students performed.

As noted in Part I of this review, the gender gap itself looks different in richer and poorer countries, with boys doing comparatively better in poorer countries where the overall level of achievement tends to be much lower than richer countries. Two hypotheses have been advanced by Bossavie and Kanninen (2018) that might help explain this. The first, the mean hypothesis, holds that girls gain more from education than boys and so invest more in it, raising their average attainment. The second, the tail hypothesis, holds that boys and girls perform similarly on average but that boys' performance varies more, with more boys at the bottom of the distribution; because a minimum level of performance is needed to complete a given stage of schooling, fewer boys clear that bar. Gregory and Welmond (2021) argue that the mean hypothesis appears the more powerful account in high-income countries.

Focusing on only high-income countries, and mathematics in particular, reveals that country socio-economic status moderates the mathematics gap in a consistent direction. The gap tilts further towards boys where a country's average socio-economic status is higher, and it is smaller where average status is lower. The reading gap, which runs to boys' disadvantage, is wider in the lower-status, lower-performing countries instead (Bianchi et al., 2025). The mechanism the authors propose is occupational opportunity. Where occupational structures reward qualifications, boys may invest more in mathematics because it offers the clearest route to those returns. This research uses a measure of countries' average socio-economic status, strongly correlated with national income but not identical to it, drawn from a cross-sectional European sample of primary and lower-secondary students, so it should be read as an association rather than a wealth effect as such.

Two things keep this mathematics evidence in proportion. Firstly, there is an opposing country-level pattern. In the same analysis, the mathematics gap is narrower, not wider, where societal gender equality is higher, with girls' relative performance higher in those systems, so a country that is both prosperous and gender-equal sits under both patterns at once. That countervailing pattern, and the paradox it belongs to, are taken up under gender norms and expectations section of this review. Secondly, the tilt towards boys in mathematics for wealthier countries sits on a small and often invisible base. Despite having a slight lead at the international average, in most high-income systems boys are not reliably ahead in mathematics at age fifteen. In the 2018 PISA data analysed for this review, boys held a statistically significant advantage in only about four in ten high-income jurisdictions. In most there was no significant difference, and girls were significantly ahead in only one. The most recent 2022 PISA round widened the average male

advantage, but a clear female lead remained confined to a few, mostly Nordic, systems. Boys are, at the same time, over-represented at the bottom of the mathematics distribution in the PISA data. So, a higher average socio-economic status goes with a slightly stronger tilt towards boys in a gap that is itself modest and frequently non-significant. It does not make mathematics a subject that boys comfortably lead in.

Local Conditions

The locality in which a child grows up shapes both their prospects and their engagement with schooling. Individual educational outcomes are significantly associated with aspects such as levels of neighbourhood poverty, the proportion of ethnic/migrant groups, and social disorganisation, with stronger negative neighbourhood effects found in the United States compared to European studies (Nieuwenhuis & Hooimeijer, 2016). A recurring mechanism in this literature is occupational opportunity. Bianchi and colleagues (2025) find that local socio-economic context moderates the gender gap, lifting boys' performance in mathematics though not in reading, and interpret this as boys responding to the occupational opportunities they perceive around them, a perception shaped by many actors in a given social setting. The same mechanism runs in reverse where opportunity is scarce. Where the labour market offers educated young people little, where local social norms discount academic success, and where teaching and learning are weak, male underachievement is often high (Gregory & Welmond, 2021).

That places wider economic trends of high-income western countries in view. Shrinking low-skilled manual employment triggers contractions in the supply of economically secure young adult men (Autor et al., 2017). This, in turn, spurs a surge in males not in employment, education or training as well as premature mortality, a decline in marriage and fertility, an increase in the fraction of mothers who are unmarried and who are heads of single, non-cohabiting households, and a growth in the number of children raised in poverty (Autor et al., 2017). The long-run shift away from low-skilled manual occupations, the rising premium on soft and transferable skills, and the growing need to move between occupations and sectors all tend to concentrate opportunity in larger cities (European Commission, 2021). For those who do not move to larger cities, prospects narrow.

The United Kingdom illustrates the scale of the resulting disparities. The share of the working-age population holding a degree ranges from more than half in Brighton or London to around one in seven in a place such as Doncaster, a spread driven both by differences in the attainment of children growing up in each area and by the selective migration of graduates out of it (Overman & Xu, 2024). The spatial disparities in the United Kingdom are high by international standards, though not as extreme as some commentary suggests (Overman & Xu, 2024). They are further compounded by a governance structure that has struggled to lift the productivity of cities outside the prosperous core around London and parts of Scotland (Carrascal-Incera et al., 2020). This backdrop is not in itself gendered, but several strands of evidence show that local conditions bear on boys and girls differently, and not always in a coherent direction.

Some research suggests that place bears more heavily on boys than girls, and most heavily of all on disadvantaged boys. Analysis of United States population data finds that growing up in a poor family, a single-parent family, or a neighbourhood of concentrated poverty has particularly negative long-term effects on boys relative to girls, with boys who grow up in poor, highly segregated areas going on to much lower employment rates than girls raised in the same environment (Chetty, Hendren, Lin, et al., 2016). Most of this gap across places reflects the causal effect of childhood exposure rather than the sorting of families into areas, and it is concentrated among low-income families. Higher up the income distribution the gender difference in place sensitivity fades. Chetty, Hendren, Lin et al. (2016) read the employment gap as stemming from boys in concentrated poverty substituting away from formal work, a labour-market (not an educational) outcome, so it speaks

to boys' wider trajectories rather than to attainment as such. A separate randomised experiment, which moved families to lower-poverty areas, confirms that better neighbourhoods raise children's outcomes and improves them for young boys, though it does not itself detect a gender difference in the size of that effect (Chetty, Hendren, & Katz, 2016).

When it comes to education attainment, Norwegian administrative data covering cohorts born between 1965 and 1989 find that the gap in higher-education attainment between urban and rural areas is considerably wider among men than among women (Zahl-Thanem & Rye, 2024). Higher-education attainment among men from small towns and the countryside has not grown for cohorts born after the mid-1970s, even as it has continued to rise for urban men, so that the urban–rural divide has widened and now falls hardest on rural men (Zahl-Thanem & Rye, 2024). Norwegian population register data allow a more direct test of how local conditions bear on measured achievement. Borgen and Zachrisson (Borgen & Zachrisson, 2025) estimate how neighbourhood deprivation bears on standardised test scores at age thirteen. Boys are more affected than girls, and children of less educated and less affluent parents are more affected than their more advantaged peers. Both differences are secondary to the study's main result. How much a neighbourhood matters varies far more with a child's own academic standing than with either gender or family background. The students whose attainment is most sensitive to neighbourhood deprivation are those somewhat below average, not those at the very bottom, where the effect is smaller. Improving the conditions of a place would therefore be expected to do least for the lowest attainers, among whom boys are over-represented, and the authors suggest individualised instructional support for that group instead. The authors attribute this to features of the Norwegian system, where school quality varies little and is only weakly related to the socio-economic composition of neighbourhoods, and they note a recent United States study reaching the same conclusion (Wodtke et al., 2023). Whether the finding survives transfer to systems with wider variation in school quality is untested.

Not all the evidence points the same way. Longitudinal evidence from a 1970s United Kingdom cohort finds that growing up in a more deprived area depresses cognitive and non-cognitive development overall, as expected, but that the effect on cognitive skills, academic motivation and employability skills is more marked for girls than for boys, with no detectable gender difference for problem behaviour (Silva, 2023, 2024). On these measures, it is girls whose development is the more sensitive to neighbourhood disadvantage.

As set out in the introduction, most evidence on how race, ethnicity and migration background shape boys' education concerns what happens closest to the child, and it is taken up in Part III. One body of work is structural, and it suggests that how these factors interact with gender and local conditions warrants further international research.

A study following nearly every American child born between 1978 and 1983 into their thirties (Chetty et al., 2020) found that the black-white gap in individual earnings is almost entirely a gap between men. Conditional on the income of their parents, black men earn less than white men, while black women earn about the same as white women. The same asymmetry appears in schooling in weaker form, with wider gaps for boys in finishing school and entering college, and black women completing school and entering college at higher rates than white men at every level of parental income. It does not appear in primary school attainment. The black-white test score gap at age nine is close to identical for boys and girls. The divergence by gender emerges across schooling rather than being present at its start.

The study's evidence on place concerns adult economic outcomes rather than attainment. Boys who move earlier in childhood to areas with better outcomes for their own racial group earn more as adults, which establishes that childhood environment acts causally on the gap. Most of the gap nonetheless remains between boys whose families had similar incomes and who grew up on the same block, so the conditions that matter vary within neighbourhoods and not only between them.

Two are identified, both within low-poverty areas. The share of black fathers present in low-income families predicts better outcomes for black boys whether or not a boy's own father is at home, and predicts nothing for black girls or white boys. Lower racial prejudice among white residents also predicts better outcomes, though for black girls and white boys as well. Both are associations rather than established causes. The specificity of the first rules out anything acting on all children in a place, school quality included, and points instead at how boys are treated and by whom they are surrounded.

A disadvantage absent from primary school attainment, absent for girls in the same families and the same streets, and responsive to childhood environment is one produced by conditions rather than carried into school. It is not addressed by moving children between places. The authors are explicit that reducing residential segregation, or placing black and white children in the same schools without changing what happens inside them, would leave most of the gap standing (Chetty et al., 2020).

The same gradient appears in schooling. Estimates of gender gaps in test scores across nearly 10,000 United States school districts find the mathematics gap favouring boys widest in socio-economically advantaged districts, and in districts where disparities between adult men and women in income, education and occupation are larger (Reardon et al., 2019). Those two conditions together account for around a fifth of the differences between districts. No comparable association appears for English. There the gap favouring girls is present in virtually every district, is worth roughly two-thirds of a grade level, and is almost entirely unexplained by local socio-economic conditions. The authors read the variation between districts as evidence that local conditions shape gender gaps, while noting that most of that variation remains unaccounted for.

Taken together, the direction in which local disadvantage tilts the gender gap depends on the outcome in view. On measured skills and motivation, it is girls whose development appears the more sensitive to neighbourhood deprivation, in the British cohort evidence. On attainment and progression into higher education, it is boys, and rural boys in particular, who appear the more sensitive, in the Norwegian evidence. And on longer-run trajectories into work, the United States evidence points again to boys, with poorer employment prospects of disadvantaged boys suggesting that they are particularly exposed to the effects of growing up in concentrated poverty. The reach of these findings should not be overstated. They rest on different countries, cohorts and outcomes, the clearest causal evidence concerns employment rather than attainment. More evidence from more recent cohorts is needed. What recurs across them is that boys' outcomes are the more tied to the places in which they grow up. That holds most clearly for long-run trajectories into work and for progression into higher education and has direct support on measured achievement from Norwegian population data. It does not hold for language, where the gap favouring girls is large, near universal across places, and largely unexplained by local socio-economic conditions. The evidence that disadvantaged boys are the most exposed rests mostly on the employment findings, where the gender difference in place sensitivity is concentrated among low-income families and fades further up the income distribution.

School Composition

The extent to which socio-economically advantaged students are exposed to non-advantaged students varies considerably across high-income countries (OECD, 2023c). Meta-analytic evidence finds the association between school SES composition and achievement similar in language, mathematics and science, and stronger for older children (Van Ewijk & Slegers, 2010). How large it appears depends heavily on how composition is measured and which controls are applied, which is one reason individual studies in this field appear to disagree. This review now turns to the question of whether the SES compositional effect bears differently on boys and girls.

This has been tested in several systems, and the pattern is asymmetric. Analysis of six TIMSS cycles across 54 countries finds that both a student's own family socio-economic status and the socio-economic composition of their school are associated with achievement (Yigiter, 2025). However, in this study, family socio-economic status more strongly associated with girls' achievement, and the socio-economic composition of the school more strongly associated with boys'. Italian data show the same direction at secondary level, where high classroom and school SES composition is associated with a larger gain in boys' mathematics than in girls', although no gender difference was detected at low SES composition (Casella, 2020). The pattern is absent in the Italian primary level data, where a pupil's own family SES dominates. A quasi-experimental study using data from the United States also finds a mathematics effect (Legewie & DiPrete, 2011). Across 27 European systems the mathematics gap favouring boys is widest in schools with a predominantly middle-class intake and narrowest in those with a predominantly working-class one (Bianchi et al., 2025). The mechanism behind this is contested rather than established. Legewie and DiPrete propose a mechanism that operates through the masculine scripts discussed in the next section. Stated briefly, they argue that a classroom's social environment shapes which versions of masculinity carry standing in peer culture, and a learning-oriented environment raises the value placed on academic effort among boys. Girls' peer groups vary less with that environment. The authors identify two channels. Low-performing students benefit from advantaged peers whatever their gender, however the effect is larger for boys. The European analysis endorses Legewie and DiPrete's compositional finding but accounts for its direction differently, arguing that boys find more incentive to work at mathematics where the surrounding occupational structure rewards it (Bianchi et al., 2025).

Two things about that pattern matter for policy. The gendered difference sits towards the advantaged end of SES. It is a difference in how much boys gain from advantaged peers, not in how much they lose from disadvantaged ones. Stated the other way, girls' relative performance in mathematics is weakest in the most advantaged schools.

Reading is where the evidence divides. The same European analysis from Bianchi and colleagues (2025) finds no socio-economic compositional effect on the reading gap, which follows a school's average achievement instead. The strongest identification for reading comes from German data. Legewie and DiPrete (2012) use the near-random assignment of students to classrooms within Berlin schools to estimate the causal effect of peer SES composition on reading, and find boys more sensitive to it than girls. However, analysis of PISA 2009 across 33 OECD countries finds girls' reading more strongly associated with a school's socio-economic composition than boys', against the authors' own expectation, and additional analyses restricted to Germany find no gendered effect at all (van Hek et al., 2018). The study authors conclude that whether and how composition bears on boys and girls depends on the national context, and that the masculinity mechanism needs more rigorous testing with direct measures of school climate and of masculinity norms across many systems.

The reading studies differ in what they measure. The Berlin estimates identify the effect within schools, at the level of the classroom a child is actually in. The cross-national estimates measure composition at school level, cannot control for prior achievement, and cannot assume students are randomly distributed across schools. Meta-analytic evidence finds compositional effects estimated at class level larger than those estimated at school or cohort level (Van Ewijk & Sleegers, 2010). A compositional effect operating on the peer group immediately around a child, and attenuating as composition is measured over larger units, would fit both results. That is an inference across separate literatures rather than a tested claim, and it is the review's reconciliation of the findings.

A third line of evidence bears on the compositional interaction rather than on reading alone. Downey and colleagues (2022) compare how much children learn while school is in session with how much they learn over the summer, across more than 900,000 children in the United States. Across their

data as a whole, boys gain on girls during the school period. Furthermore, the gain to boys is much the same across the range of school SES composition, and in the later grades, where the difference does reach significance, it is slightly larger in the schools with the poorest intakes. The authors account for the divergence from earlier work by pointing to what each design compares. Studies of peer composition estimate the effect of a school's intake on achievement during the school period. This design contrasts learning in school with learning out of it, so the two are not measuring the same quantity and this result does not overturn the Berlin estimates. What the authors take from it is that the difficulties of boys in low socio-economic status schools originate more outside school than inside it.

The present review adds a distinct angle, using original analysis² to turn from achievement to the attitudes that underpin achievement. The international assessment data allows a specific question to be asked about school composition and gendered attitudes towards schooling. The measures come from PIRLS 2021 and TIMSS 2023 at Grade 4, and TIMSS 2023 at Grade 8. Overall gender differences in these attitudes (for example, boys reporting greater confidence in and liking for mathematics and girls reporting stronger reading engagement) are documented in the education engagement section of Part III. The current section asks whether the size of those attitudinal differences depend on the socio-economic composition of the school a student attends. In other words, whether a given gender gap is larger, smaller or unchanged in disadvantaged schools compared with affluent ones. For each measure, in each high-income system, the attitude was regressed on the pupil's gender, on the school's socio-economic tier, and on the interaction between the two.

According to these data, gender gaps in attitudes are real, but their size does not depend on the socio-economic context of the school. Across the measures available, in both studies and at both ages, the interaction between gender and school socio-economic tier is on average negligible. The gender difference in a disadvantaged school is, on average, close to that found in an affluent one. The gaps set out in Part III are each on average the same size in affluent, mixed and disadvantaged schools. School disadvantage shifts the overall level of some of these attitudes, which is a real and separate concern, but it does not change the size or direction of the gap between boys and girls.

That result bears on why school SES composition affects boys' achievement more than girls'. The mechanism proposed by Legewie and DiPrete runs through boys' orientation to school, which should vary more with school composition than girls' does. It does not do so in these data. The measures here are students' own reports rather than measures of peer culture or of masculinity norms, and composition is captured at school rather than classroom level, where compositional effects are weaker. The result is therefore a limited test rather than a decisive one. It is, however, consistent with the alternative account offered by Bianchi and colleagues (2025), in which boys respond to the occupational returns visible in more advantaged settings rather than to the academic culture of their peers.

School socio-economic composition does bear differently on boys and girls, but not in the way policy usually assumes. The difference sits at the advantaged end. Boys gain more than girls from an advantaged intake, in mathematics, while at the disadvantaged end there is no gender difference to speak of. Reducing school disadvantage would therefore not close the gender gap, though it matters a great deal for boys and girls together. Why the effect works this way is unsettled. The peer culture account and the occupational incentives account both remain live, and the evidence assembled here does not separate them.

² These results carry an important caution. The regressions were estimated separately within each country and then averaged across high-income systems, so the conclusion that the interaction is negligible describes that average rather than the outcome of a single cross-national test. In a small number of individual systems an interaction term does reach significance, but no more than chance would predict, and they point in no consistent direction. The Grade 8 results rest on twenty-five high-income systems, fewer than the thirty-two to thirty-four.

Gender Norms and Expectations

Studies suggest that measured psychological differences between boys and girls are generally small (Hyde, 2014). This section looks at how cultural amplification exaggerates small differences through gender norms and expectations, shaping what society expects of boys and what boys expect of themselves. Although boys are not a single category, their relative underperformance can be read against the particular social identities they inhabit (Jha et al., 2017). This section is about those norms, where they come from, how they bear on boys' engagement with school, and why they prove so resistant to progress made elsewhere.

That resistance is often framed through the gender equality paradox discussed in Part I, the observation that more prosperous and more gender-equal societies do not produce more equal educational outcomes, and in some respects show wider gaps in performance and subject choice (Balducci et al., 2024; Borgonovi et al., 2023). The paradox is contested, and for the present review, its purpose is only to establish that greater societal equality, on its own, does not close gender gaps. Dotti Sani and Quaranta (2017) suggest that part of the reason for the paradox may be that boys absorb egalitarian norms more slowly than girls, so that societal progress can mask a stalling in boys' own attitudes. This helps explain why laddish and anti-academic masculine performances persist even where the wider culture formally endorses equality. Boys are not simply unreached by egalitarian messaging but appear less likely to internalise it, which implies that passive exposure to it is unlikely to be enough.

This points to the asymmetry that frames the section. Gender norms are not held evenly by boys and girls. Across the OECD's Survey of Social and Emotional Skills, gender stereotypes are more prevalent among boys than girls in almost every site, yet most efforts to counter them are aimed at women and girls (OECD, 2024a). These norms take hold early, well before primary school (King et al., 2021). Bringing boys' own adherence to gender norms to the centre, rather than treating boys as the backdrop to girls' progress, is the thread that runs through what follows. That analysis moves from the norms boys are expected to perform, to the way those norms play out most sharply for working-class boys in places shaped by industrial decline, to the academic stereotypes that attach to what boys are thought capable of and suited to.

Masculine Scripts

Masculinity is constructed within institutional and cultural contexts that produce multiple forms, referred to in this review as 'scripts'. Scripts are commonly understood as a set of relational positions, hegemonic, subordinated, marginalised and complicit, defined against one another rather than as a fixed essence (Connell, 1995). A single, simple definition of masculinity is therefore neither possible nor desirable, and that complexity is part of the phenomenon rather than an obstacle to understanding it (Smiler, 2024). Schools are active players in the formation of masculinities, and their gender regimes tend to reinforce a gender dichotomy. Masculinising practices cluster at particular sites, such as curriculum divisions, discipline systems and sport, and students themselves help construct the gender hierarchies of their peer cultures. In doing so, they take up the offer of gender privilege in ways that range from protest masculinity, a compensatory, assertive form claimed where the usual routes to status are out of reach, to anti-sexism (Connell, 2023). This section examines the different forms of masculinity that may develop in different school contexts, and considers which are associated with greater engagement in education and which are associated with disengagement.

In its most familiar form, often labelled hegemonic masculinity, the script prizes toughness, physical assertiveness and self-reliance, often extending to a reluctance to ask for help or admit difficulty, since both jeopardise the projection of mastery the script is meant to display. It locates a boy's worth in the regard of his male peers rather than in the approval of the institution, and it treats visible effort, compliance and deference to teachers as marks of weakness (Connell,

1995; Younger & Warrington, 2005). It is suspicious of the abstract and the theoretical, valuing knowledge that is plainly useful over knowledge that is merely examined, and it casts reading and the verbal, expressive subjects as female territory (Brozo, 2019). Status, on these terms, is won by performance for an audience of peers, through banter, a degree of disruption and a reluctance to be seen to try.

The most extreme masculine scripts are contested by most boys. Rigid adherence to traditional masculinity is in fact a minority position among adolescent boys, with most boys in both UK and US samples resisting conventional masculine norms, and the boys who adhere most rigidly tending to do less well academically than those who resist the norms (Way et al., 2014; Yu et al., 2021). OECD evidence offers a partial profile of the minority of boys who do hold the most traditional gender stereotypes. They tend to have lower tolerance, in the sense of openness to different viewpoints and valuing diversity, and lower emotional control (OECD, 2024a). Yet the same boys are often more assertive, sociable and optimistic than their peers, which complicates any picture of the stereotype-holding boy as withdrawn or struggling, and which may itself offer a route to engaging them.

Despite being contested by most boys, the traditional masculine script is enforced by peers: boys who side with the institution, or who depart from the masculine template in other ways, are liable to be singled out (OECD, 2024a; UNESCO, 2022). Boys often face strong pressure to demonstrate masculinity by conforming to stereotypical roles, and departures are punished. The OECD Survey on Social and Emotional Skills notes that, across countries, 34% of students say that boys who are gender non-conforming are made fun of, against 24% who say the same of girls (OECD, 2024a). Other research has shown that gender non-conformity raises the risk of being bullied, while high conformity is associated with bullying others, and targeting non-conforming peers appears to be one way boys display their own conformity (Ioverno et al., 2021). Much of this is read through hegemonic masculinity, organised around heterosexuality, physical dominance and the normalisation of violence (Rosen & Nofziger, 2019), and boys are more likely than girls to face physical bullying, often on the basis of real or perceived sexual orientation or gender expression (UNESCO, 2022). The educational cost is direct in the sense that being bullied tends to lower achievement, and bullied boys tend to score lower than bullied girls in mathematics in particular (Gimenez et al., 2024).

The difficulty is that almost every one of these prescriptions runs against what schools reward. Schools ask for sustained and visible effort, for compliance and attention, for help-seeking when a pupil is stuck, and above all for engagement with reading and writing, and the hegemonic script codes each of these as unmasculine (Jha et al., 2017; Lundberg, 2020). The result is not only that some boys do less, but also that not trying becomes the safer course, since conspicuous effort that still ends in failure is the more humiliating outcome, whereas visible indifference protects a sense of self-worth whatever the result (Younger & Warrington, 2005). The cost falls heaviest on literacy, where the subject itself is gender-coded and a boy must first see reading as compatible with who he is before he will invest in it (Brozo, 2019). Most boys do not hold the hegemonic script in its purest form. Furthermore, some of the dominant masculine dispositions, for instance a sense of agency and of instrumentality, relate positively to achievement rather than negatively (Streck et al., 2022). But where the dominant script does take hold, it traditionally sets peer approval directly against the behaviours that contribute to good student performance and educational attainment.

That conflict has been tested directly. Heyder and colleagues (2021) followed more than 4,200 Flemish adolescents through the first two years of secondary school and measured how much pressure each felt to conform to gender stereotypes. Boys misbehaved in school more than girls at the outset and their misconduct rose faster over the period. The felt pressure predicted that rise in boys and not in girls. Girls reported both less pressure and a decline in it across the two years.

The authors read school misconduct as part of the enactment of a masculine identity rather than as a rejection of school authority in itself. The design follows the same students across three waves but does not identify a cause, so the relationship is associational. Further research in the Flemish context has shown that boys' gender role attitude development is associated with student culture and composition, while for girls it is parental socio-economic status that predicts development (Ullrich et al., 2024).

Because some forms of masculinity set academic effort and achievement at odds with a credible male identity, they can damage boys' prospects through the very norms they are expected to perform (Lundberg, 2020). Legewie and DiPrete (2011, 2012) argue that the school environment channels the local version of masculinity available to boys. The environment either fosters or restrains the anti-school attitudes and behaviour through which masculine status is sometimes claimed. Where a school sustains a learning-oriented peer culture, academic effort can be folded into a masculine identity rather than set against it. The authors suggest that sustaining such a culture requires the usual school resources, better teachers above all, alongside curriculum, facilities and support staff. These are all things that tend to raise the academic orientation of the peer group. Where those inputs are in place, academic competition can itself become a way of doing masculinity. Boys develop adaptive strategies, keeping up an outward coolness towards school while remaining instrumentally engaged in its work. Where resources are insufficient, engagement is more readily coded as unmanly.

As an object of empirical study, this learning-oriented culture is less examined than its opposite. Van Hek (2025) notes that the research concentrates on the anti-academic case, where status attaches to disengagement. The anti-academic behaviours are concrete, arguing with teachers, skipping lessons, lateness and punishments, and they are more common in boys' friendship groups than in girls'. On this basis, a learning-oriented culture can be seen as the setting in which those behaviours do not confer status, and effort does instead. How far school engagement aligns with the dominant gender script in a school environment seems to vary less in girls' peer groups than in boys'.

In short, the script is not fixed solely by being male. It is channelled by the local culture of school and classroom, which offers policy makers something with which to work. Two limits apply to this body of work. Differences between schools accounted for very little of the variation in misconduct among the Flemish adolescents in their first two years of secondary school, which limits the influence of the school environment at that age and sits alongside the null reported under the school composition sub-section of this review. And the quantitative evidence for the mechanism is drawn narrowly. Ullrich and colleagues (2024) and Heyder and colleagues (2021) both use Flemish adolescent panel data. The mechanism is plausible and consistently reported, but it has been tested in few systems, and whether it holds across high-income education systems more widely is not established.

Working Class Masculinities

Although this is an international literature review, there is a long and unusually rich tradition in British sociology of studying how working-class masculinity shapes boys' relationship with school. This work is worth highlighting, as it remains the deepest qualitative evidence base on the question. Its founding text is Paul Willis's *Learning to Labour* (1977), an ethnography of working-class boys in the West Midlands and among the most cited studies in the sociology of education. Willis described a group he called 'the lads', who rejected school not for lack of ability but because they judged it irrelevant to the manual work they expected to enter, and who earned status through an anti-school culture in which conformity was treated as weak and unmanly. Their disengagement was less a failure than a rehearsal for the manual 'breadwinner' role that anchored a particular kind of working-class manhood.

Willis's account is powerful but deterministic, and it has been criticised on two counts by Blower (2025). Firstly, for centring 'the lads' while overlooking their studious, upwardly mobile peers. Secondly, for lending weight to a stereotype of the working-class boy as inherently a problem in the classroom. Phillip Brown's *Schooling Ordinary Kids* (1987) softened that determinism by making the local economy the pivot. Where there were enough working-class jobs to walk into, school could broker the move from education to work; it was deindustrialisation, by severing that link, that unsettled the arrangement Willis had captured.

Later work follows what happened as deindustrialisation unfolded. The most recent account is Ward's study of young men in a former mining area of the South Wales Valleys (2015). Ward found not a single working-class masculinity but several coexisting versions, and traced how individual boys shifted between them, something that the earlier binary conceptualisations had missed. His central finding is the one most useful for this review: the symbolic tie between manhood and industrial work outlives the industries themselves and continues to make wholehearted engagement in education feel costly for boys in places shaped by that past.

A recent addition to this tradition comes from Gater's (2025) study of working-class boys in South Wales mining communities and their 'amalgamated masculinities'. This is a fusion of working-class and protest masculinity³ with softer, more emotionally open attributes drawn from popular culture, music and media. These young men are not simply inheriting the locally and historically produced masculinity of their place, but assimilating ideas of manhood from beyond their immediate community and reworking them (Gater, 2025). Gater suggests the continuing significance of place, since the industrial and social history of a locality still shapes how boys conceive of their masculinity, while at the same time acknowledging the pull of social and economic change and of cultural influences from well outside the immediate surroundings. It is a contemporary, empirical version of Connell's point, that masculine scripts are plural, assembled rather than simply inherited, and dynamic. At the same time, the openness that lets boys assemble such amalgamated identities also leaves them exposed to online influences pushing the other way. While some research suggests that the reach of certain specific 'manosphere' figures may already be waning among school-age boys, the lesson is that schools must equip boys to interrogate such content systematically, rather than assume exposure will fade (Milne & Baker, 2025).

The young men in Gater's (2025) study display laddish, non-conformist behaviour but also selectively engage three or four subjects that meet the minimum entry requirement for the further-education course they intend to take (Gater, 2025). The result is a deliberate economy of effort, a departure from the anti-learning laddish culture of earlier accounts of masculinity cited in this review, driven by an awareness of the value of credentials. This cultural shift towards a more pragmatic approach is real but incomplete and risks leaving little margin for error. These young men have grasped that credentials matter faster than they have raised the effort they put in, because the pared-back approach to schoolwork was formed around a labour market that no longer exists. As deindustrialisation has continued to unfold, the move from education to work has become markedly harder for working-class men with few credentials than in the era of secure manual work (McDowell & Bonner-Thompson, 2020). Cultural changes can often lag changes to economic and social realities. In the aggregate, boys' underperformance does not yet show up as worse labour-market outcomes, which is part of why it attracts little urgency, but beneath the average sits a sub-group of low-educated men for whom the routes to secure work have largely closed (European Commission, 2021).

Dovetailing with, although slightly different from, Gater's pared back approach, Roberts (2012) describes a pattern of 'alienated instrumentalism' for the boys he describes as the 'missing middle'. He finds that many boys neither resist school, nor work to maximise attainment, treating school

³ See previous sub-section "masculine scripts" for discussion of this term.

effort as carrying a moral and social worth, even when it is not tied to a clear ambition (Roberts, 2012). Boys' behavioural responses to masculine scripts and academic stereotypes are best seen as a continuum, rather than a typology. On this continuum, laddish and swotty behaviour sit at the extremes but are not discrete categories, being defined locally by career prospects and culture (Roberts, 2012). At the sharper end, where school feels like an obstruction rather than an opportunity, some boys disengage completely from education, drop out of school, and instead favour faster (and often illegitimate) routes to money and masculine validation (Holt et al., 2019).

Morris (2012) offers a rare non-British test of this tradition, comparing boys in two American high schools, one rural and white, one urban and African American, and finding the same conflict between academic effort and a credible male identity in both. The comparison separates the mechanism from race, since the account in which opposition to schooling is treated as specific to low-income black boys does not survive the rural white case. Morris reads the resulting gaps not as a reversal of gender inequality but as a cost carried with masculine privilege, because the licence to expend less visible effort is itself a marker of standing.

Taken together, this work offers a genuine mechanism for one form of overlapping disadvantage. Boys are exposed to a masculine script that recodes effort and academic identity as out of place, operating most forcefully where local labour markets once offered a manual alternative route to adult status. That mechanism speaks directly to findings elsewhere in this review, among them boys' concentration in the skilled trades and their greater readiness to dismiss school as a waste of time. And in its more careful forms the literature is firmly non-deterministic. The same boys act differently across settings, and an apparent rejection of school is often a considered response to a particular history rather than a lack of ambition. The important caution across all of this work is that the evidence is emphatically Anglophone (and mostly British). It is rooted in the industrial geography of the specific parts of the UK and in the politics of post-1980s deindustrialisation that does not transfer cleanly to other high-income systems. Although the particular forms they take are local, the underlying processes, the interaction of class, masculinity and a shifting labour market, are general. The review treats this body of work as a deeply evidenced national case rather than a portable template.

Academic Stereotypes

If masculine scripts are about who boys should be, academic stereotypes are about what boys are assumed to be capable of and what they should aspire to. The clearest evidence concerns reading. Stereotypes that cast reading as a female domain depress boys' competence beliefs, motivation and achievement, and they do so at both the individual and the classroom level. Boys who hold a strong stereotype favouring girls in reading are less motivated to read and rate their own ability lower, while for girls the relationship runs the other way (Muntoni et al., 2021). The reading gap appears to be rooted, at least in part, in boys' fear of confirming a negative stereotype about their group, an effect already visible among nine-year-olds at the stage when learning to read is most critical (Pansu et al., 2016). It is reinforced at home, with research finding that sons of parents who hold a girls-favouring reading stereotype reporting lower competence beliefs and motivation, while daughters are unaffected (Muntoni & Retelsdorf, 2019). This points to parents as an early lever of academic stereotypes, something discussed further in Part III. A similar dynamic appears beyond reading, in second-language learning, where male learners are disadvantaged by female-language stereotypes (Li & McLellan, 2021), and in expressive domains such as dance, where stereotype threat undermines boys' motor performance and learning (Bastos et al., 2023).

These stereotypes are likely to influence the shape of gender gap across the distribution of student performance. When it comes to PISA mathematics and science, gaps favour girls at the bottom but flip to favour boys above roughly the median, widening towards the top. In reading, by contrast,

girls lead across the whole distribution, though by less at the top. The OECD suggests that boys thus face two stereotypes at once: a deficit stereotype in reading and the expressive subjects, and a brilliance stereotype in mathematics and science that works in their favour only at the very top (OECD, 2025a).

Stereotypes about what roles boys should aspire to are in place strikingly early. By age five, girls are likelier to name caring or creative roles and boys jobs in transport or construction (OECD, 2020). These aspirations prove resistant to societal progress in the sense that boys' career expectations are no less stereotypical in countries with more egalitarian beliefs, and gaps in aspiring to teach can even be wider where beliefs are more egalitarian (Han et al., 2020; OECD, 2024b). OECD data shows that, in most high-income-countries, boys are less likely to expect to work in health or education even when their skills would fit, so reducing stereotypes could widen the range of careers boys consider (OECD, 2024a).

Boys' aspirations sit within a wider attitudinal difference. Part I established that boys express weaker endorsement of gender equality than girls, as one of a broader set of civic attitudes. What matters here is narrower, whether those attitudes bear on achievement. The evidence suggests they do, though asymmetrically. More egalitarian gender-role attitudes are associated with better outcomes for all students, but the benefit may be larger for girls (Taraszow et al., 2024), and boys appear to concentrate effort where they perceive greater future returns, trying harder in mathematics than in reading (Bianchi et al., 2025).

These stereotypes carry a downstream cost. Norms of masculinity held by boys, teachers and parents can translate into low expectations of boys' ability and behaviour, feeding poor motivation, disengagement and eventual dropout (UNESCO, 2022), and in a number of countries boys at fifteen are already more likely than girls to say they do not expect to finish upper secondary school (Borgonovi et al., 2018).

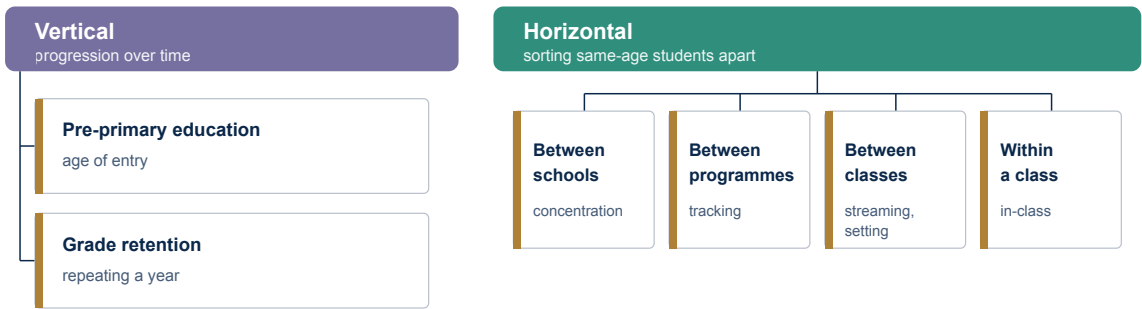
Education System Features

Few areas of education policy are as contested, or as laden with assumptions about boys, as the design of the school system itself. The vast majority of people in high-income countries have come into prolonged contact with the school system at some point in their lives, and these experiences leave lasting impressions, both positive and negative. Virtually every feature of an education system, and the way those features interact with the wider society, bears on the relative achievement, motivation and attainment of boys and girls (Borgonovi et al., 2018). Answering what matters is therefore relatively easy: everything matters. The harder task is deciding which features matter most for boys' outcomes. Unfortunately, the evidence is markedly uneven when it comes to boys specifically, though inferences can be drawn where a feature bears on groups in which boys are over-represented, among them the lowest attainers. This section takes three families of system features in turn: how and when systems sort students into tracks and groups; how schools are governed and how they assess, and who stands at the front of the classroom. For some of these the evidence is reasonably clear; for others it is thinner.

Student Stratification

Education systems stratify students by a range of different metrics, and at different points. Figure 1 shows a taxonomy which defines student stratification measures as either vertical or horizontal.

Figure 1 - How Students are Sorted



When it comes to vertical stratification, the extent of grade retention varies widely across education systems. Some systems, including Iceland, Japan, Korea, Norway and the United Kingdom promote students automatically to the next grade level in both primary and lower secondary school and so have very small numbers of repeating students (OECD, 2023c). Boys are more likely than girls to repeat a year, and the pattern holds in almost every country with data. What the headline pattern conceals is that the male share rises with each repetition. Across OECD countries, boys make up about half of students who have never repeated a year, 57% of those who have repeated once, and two thirds of those who have repeated more than once (OECD, 2023c). The concentration is not explained by attainment alone. Across 23 countries, and controlling for students' 2018 PISA scores, boys have 79% higher odds of being retained than girls (Pepels et al., 2026). There is substantial evidence, presented in Part I, that grade retention mostly results in poorer outcomes for the retained student. Furthermore, the costs associated to the policy can easily outweigh the weak benefits of retention (Valbuena et al., 2021).

Although retention explicitly aims to give students a chance to 'catch up', retention decisions seem to be based on more than a student's measured performance. Drawing on PISA 2018 data from 23 countries, Pepels et al. (2026) find that 13% of the variance in grade retention sits at school level and 33% at country level, so retention reflects institutional routine as much as individual performance. Schools with higher proportions of migrant or low socio-economic status students use retention more often, while schools with higher proportions of poor-performing students use it less. Across countries, national policy settings explain only a fifth to a quarter of the variation in retention, which suggests that educational traditions and beliefs carry much of the weight (Goos et al., 2013).

For boys the implication is structural rather than individual. Retention is a gate set by the education system and whose operation varies by school. Boys pass through it more often than girls, and more often even at the same level of measured performance. Whether a boy repeats a year depends substantially on where he goes to school, not only on how he performs. Although retention of struggling boys is likely not the answer to raising their attainment, there are promising alternatives (Figure 2).

Figure 2 - Alternatives to grade retention

Some practices have the potential for improving the academic achievement of low-performing students:

- Multi-age grouping, creating combined classes with students from different years;
- Looping, where a teacher or set of teachers remains with the same group for some years;
- Increasing instructional time for poor performers.
- Remedial tutoring, an expensive but effective alternative to grade retention that can take place within or outside school-time.
- Cooperative learning practices, which involves classmates tutoring children who are struggling.

None of the practices above were analysed with differential gender effects in mind, which suggests a research gap.

Source: (Valbuena et al., 2021) *Effects of Grade Retention Policies: A Literature Review of Empirical Studies Applying Causal Inference*.

Turning to horizontal stratification, streaming of students into classes based on perceived ability is common in high-income countries and one of the most researched topics in education. However, the literature generally shows that streaming has few advantageous effects. Negative effects include increased academic disadvantage for students placed in lower streams and segregation of students along the lines of race and class (Johnston & Wildy, 2016).

Placing students into academic or vocational tracks, used most heavily in systems such as Austria, Belgium and Germany, provides differentiated curricula depending on the track and thereby different academic pathways (OECD, 2023a). The average age of first selection into distinct tracks, classes or programmes across OECD countries is 14.3 years, but in some systems children are separated as early as ten (OECD, 2023a). The case for tracking by students' ability levels/attainment is that it allows more tailored teaching practices and a focus on skills aligned to the labour market.

However, this comes with the potential cost of increasing educational inequality (Brunello & Checchi, 2007). Early tracking in particular has been shown to increase inequality in educational achievement across international student assessments (Hanushek & Wößmann, 2006). Building directly on that work, a meta-analysis pooling 21 assessment cycles across 75 countries confirms that early tracking widens social achievement gaps, while doing little for overall performance (Strello et al., 2021). Early selection has one further consequence. It increases the extent to which boys and girls attend different schools, because tracks are chosen along gendered lines. However, this has not been found to translate into differences in achievement gaps (Steinmann et al., 2023).

Adding a further dimension to this, Montt (2011) reports that between-school tracking widens the total spread of achievement, independent of family background, and finds comprehensive systems producing a more even distribution. Bol and colleagues (2014) add additional texture to this landscape, showing that the extent to which tracking widens the socio-economic gap is reduced where central examinations exist. They are explicit that they do not argue central exams raise mean performance, only that the detrimental effect of tracking on the social background relationship is reduced. None of these studies disaggregates by gender; together they set the context for the tracking debate rather than speaking to boys' education directly.

Using internationally comparable data on educational test performance (mathematics, reading) and educational attainment, Scheeren and Bol (2022) show that differentiation between educational tracks benefits girls wherever it occurs, and the benefit is larger in systems that select late. The authors attribute this to the timing of selection relative to adolescence, when girls' advantage in the non-cognitive skills that shape track placement, attention, effort and goal-directed behaviour, is at its widest. The later a system selects, the more that advantage bears on where students are placed. The association is modest for test scores and larger for final qualifications. The same authors caution that late-tracking systems tend to be more gender-egalitarian in other respects, which their data could not separate from the effect of tracking age itself.

The gender gap therefore widens under later tracking because girls gain, not because boys lose. That distinction carries the policy weight. Evidence from tracking-age reforms across 21 European countries finds later tracking raising women's completed years of education relative to men's (Scheeren et al., 2018). The authors describe this as a trade-off, since a higher tracking age has repeatedly been found to reduce inequality by socio-economic background. Selecting earlier would lower girls' attainment and widen socio-economic gaps while returning nothing to boys, and nothing in this literature supports it as a response to boys' underachievement. Those reforms span most of the twentieth century, so the finding rests on cohorts educated some decades ago.

One study points the other way. Longitudinal analysis of the Finnish comprehensive reform of the 1970s, which postponed selection from age ten or eleven to fifteen or sixteen, found a positive effect on girls' probability of entering the academic track and a slightly negative one for boys, concentrated among boys whose fathers had no academic education (Pekkarinen, 2008). It concerns one reform in one system and the wider evidence does not reproduce it.

The question that matters more for this review is whether later tracking reaches disadvantaged boys, and the evidence suggests it does not. Later tracking reduces inequalities by social origin in test scores among girls. The same effect is not detected among boys (Scheeren, 2022). The main equity lever in the comparative literature appears to work through girls. Why additional time in untracked classes delivers for low socio-economic status girls and not for boys from the same homes is unresolved, and it is a more useful question than the tracking age debate itself.

A second mechanism concerns direction of movement rather than timing. Boys, and disadvantaged boys in particular, more often underperform early, so they are more often placed in lower tracks and less academic schools, and movement back up is rare, sustained by peer influence, falling motivation and often less stimulating teaching (European Commission, 2021; UNESCO, 2022). For example, research in Toronto, Canada, shows that not only are boys more likely to be placed in less academic tracks, they are disproportionately disadvantaged by the streaming system that keeps them there (Parekh et al., 2021). Which track a student is recommended for also depends in part on the instrument used. Dutch primary schools choose between several end-of-primary tests, and a comparison of the three most used, across cohorts of around 160,000 students each, finds that they produce different recommendations and that those recommendations relate differently to the level a student reaches three years later (Douma et al., 2025).

A student's recommendation therefore depends partly on which test their school has adopted. The study does not examine whether the accuracy of these tests differs for boys and girls, and excludes students who repeated a year, among whom boys are over-represented.

Many education systems do not provide different academic tracks, instead relying on heavier use of setting students (i.e. top set, middle set, bottom set). Evidence from England shows that boys had higher odds of bottom-stream placement than girls even when they performed similarly and the attainment gap widened across ages 5 to 7 (Hallam & Parsons, 2013; Parsons & Hallam, 2014). International reviews have taken this as grounds for abandoning the routine grouping of low-attaining boys by attainment, arguing that it isolates and demotivates them (Gregory & Welmond, 2021; Jha et al., 2017).

The general form of that argument has weakened in the English context. A matched quasi-experimental evaluation comparing English secondary schools that set by attainment in mathematics with schools teaching mixed attainment classes found no difference in progress for low-prior attainers, and none for students eligible for free school meals, which the authors note questions earlier evidence that setting disadvantages students from disadvantaged backgrounds (Hodgen et al., 2026). Mixed attainment teaching was associated with lower mathematics self-confidence among low-prior attainers. Neither outcome was reported by gender, so the boys-specific version of the argument is untested rather than refuted. What has weakened is the general premise beneath it, that grouping by attainment holds low attainers back.

The practices also need separating. Between-school tracking, whole-curriculum streaming and subject-specific setting are different arrangements, and the evaluation above notes that most international evidence concerns tracking rather than the setting common in several systems. This review draws no conclusion on which arrangement serves low-attaining boys better, because no study has tested it.

Channelling underachieving boys into trades and technical programmes is a recurring response to their underachievement. The composition of these routes is already gendered and socially skewed, with students in vocational schools more often older boys with low school performance from less educated and poorer families (Brunello & Checchi, 2007). Boys entering them frequently lack the basic skills the courses require, and the programmes are widely regarded as less prestigious than a degree and are often under-resourced (Gregory & Welmond, 2021).

Vocational partitioning has a second effect that is less often noticed. Because boys and girls choose different fields, systems separating students by field of study produce schools with skewed gender distributions. This matters because boys' reading performance is higher in schools with a higher proportion of girls, and the authors read their finding as pointing to a negative side effect of vocationally partitioned systems (van Hek et al., 2018). The objection is not to vocational provision, which has value when it is well resourced and freely chosen. It is that treating it as the default destination for struggling boys concentrates them in settings that are under-resourced, low in prestige, and short of the peer conditions associated with better outcomes for boys.

The case of single-gender schooling is more ambiguous in terms of positive differential effects on boys. The effect of single-gender schooling on academic outcomes is hard to isolate, because in most countries single-gender schools are selective and few, leaving small samples with which to work. Some evidence from randomised studies broadly supports positive effects of single-gender schooling for both boys and girls. One of the landmark studies in this area exploited the unique random assignment of students into single-gender versus coeducational high schools in Korea (Park et al., 2013). While controlling for socio-economic factors, prior academic achievement, as well as school-level variables, attending all-boys schools or all-girls schools was significantly associated with higher average scores on Korean and English test scores, as well as a higher

percentage of university entrants. Using 2018 PISA data for Ireland, Clavel and Flannery (2023) examined the relationship between single-gender schooling and performance in mathematics, reading and science for boys and girls. The results were more ambiguous compared to the findings of Park et al. (2013) in Korea. After controlling for SES effects and other variables, the analysis found no association between attending single-gender schools and performance in mathematics, reading or science scores for boys or girls. A key meta-analysis on the topic comes from Pahlke et al. (2014), who analysed data from 184 studies, representing the testing of 1.6 million students from 21 countries, for multiple outcomes. They found that uncontrolled studies showed some modest advantages for single-gender schooling, for both girls and boys, while controlled studies showed only trivial differences between students. Findings are therefore mixed as to whether there are meaningful effects on the size of gender gaps in education outcomes (Delaney & Devereux, 2021).

Curriculum, Accountability and Assessment

Education systems differ in where authority over curriculum, assessment, resources and staffing sits, and many high-income countries have shifted that balance since 2000. Greater school autonomy over academic content, personnel and budgets is associated with higher student achievement in high-income countries, and most strongly for decisions about academic content (Hanushek et al., 2013). The distributional picture differs. Across 69 countries and six cycles of PISA, autonomy shows no overall association with socio-economic inequality in mathematics, but autonomy over academic content is associated with wider inequality in high- and upper-middle-income countries specifically, while autonomy over budgets and personnel shows no such association (Hossain, 2023).

None of that work reports results by gender. A synthesis of 195 studies of system-level features finds those concerning decentralisation, accountability and stratification inconsistently related to student outcomes, with small effect sizes where significant, and stronger and more consistent findings for the wealth and inequality of the wider society (Mejía-Rodríguez & Kyriakides, 2022). The features with gender-differentiated evidence lie a level down, in how assessment is designed, how schools are held externally accountable, and how far they can shape what and how they teach.

School curriculum frameworks are extensively studied in comparative education but rarely with gender-differentiated effects in view. What evidence there is suggests modest and specific effects. Some studies argue that curriculum standardisation appears to widen the gender gap, and the degree to which its curriculum is centralised or decentralised has been reported as having some bearing on boys' attainment (European Commission, 2021). The primary evidence supports that for reading and reverses it for mathematics, and the difference turns on what the term is taken to cover.

Three arrangements are often grouped under the term standardisation: How far central government prescribes curricula and learning materials; whether attainment is certified through national examinations; and how uniformly teachers cover the same content in practice. The studies that separate them find them working in different directions.

Prescription of curricula and materials is the arrangement associated with a wider gap. Across all six waves of PISA, reading scores are lower where governmental regulation largely determines curricula and learning materials, and the association is stronger for boys than for girls (Van Hek et al., 2019). The authors suggest that boys are more often the weaker readers who need individual attention, or that requiring reading time in class does not produce reading development among less motivated readers. Neither mechanism is tested. This is one country-level study of an association, and no comparable study of curricular prescription has been found for mathematics.

National examinations point the other way. Across 32 countries, the use of national examinations and greater uniformity between teachers in time given to major mathematics topics are each

associated with a smaller gender gap in mathematics, in scores and in the odds of excelling (Ayalon & Livneh, 2013). The association runs against boys' advantage rather than in favour, and it comes with no association with lower average achievement.

Formal policy and classroom practice diverge. Curriculum standardisation as recorded in national policy shows no association with the overall spread of achievement, which has been attributed to implemented curricula varying less between schools than formal frameworks suggest (Montt, 2011). Uniformity between teachers, which is what Ayalon and Livneh measure, is closer to what children experience than what a framework specifies. A systematic review of thirteen studies reaches the same subject-differentiated position and draws a policy implication from it, that standardisation might be applied differently by subject, held lower in reading and language and retained in mathematics (Zapfe & Gross, 2021). That is a proposal rather than a tested arrangement.

This review takes no position on mathematics standardisation. The association there runs in girls' favour, it carries no cost to average achievement, and reversing it would improve boys' relative standing by worsening girls'. What the evidence supports is narrower. Where a system prescribes closely what and how reading and language are taught, boys' reading appears the more affected. The equity arguments usually made for standardisation rest on examinations rather than on curricular prescription, and the two can be held apart.

Devolving decisions about content and materials is associated with higher average achievement and a wider socio-economic gap. Prescribing them centrally is associated with lower reading achievement and a wider gender gap. Those come from different studies using different outcomes and subjects, so they describe a pattern rather than a tested trade-off. The literature does not establish strong, systematic mechanisms linking accountability to boys' outcomes specifically. There is no systematic cross-national work on how accountability regimes affect boys' engagement or attainment. On curriculum, this review finds one country-level associational study on curricular prescription and boys' reading, and one on national examinations and the mathematics gap. These are genuine gaps, and they feed directly into the review's account of where the evidence base needs building.

Student assessment features throughout this review, since it is a cross-level phenomenon designed at the system level but enacted by teachers in classrooms. The system-level design aspects are set out here, while the teacher-level mechanisms, including marking behaviour, are taken up in Part III. At the system level, the design of assessments measurably shapes the size and even the direction of measured gender gaps.

The stakes for the student who is taking the assessment can be set by policy, and there is evidence that gender gaps shrink in high-stakes testing. Following several cohorts through six years of secondary school at one high-performing school in Spain, Azmat et al. (2016) exploited variation in how much individual tests counted towards the final grade, from 5% to 27%. Girls outperformed boys in every test, but by a wider margin where the stakes were low. In the tests taken at the end of secondary school, which counted for half of the university entry grade, the gender gap disappeared. Boys' disadvantage on school-based assessments may reflect not only differences in cognitive skill but differences in engagement and motivation with the test. Across eighty countries in PISA 2018, a low stakes assessment, boys were significantly more likely than girls to leave items unanswered, a pattern reversed in only one country and more pronounced on reading-related material (Marcq & Braeken, 2024). This implies that changing how proficiency is assessed, or helping boys to display the proficiency they have, could itself narrow measured gaps (Borgonovi et al., 2018). The practical implication for schools and systems is to interpret gender gaps in literacy with an awareness of how far assessment design is shaping them, rather than reading them as purely differences in ability (Miyamoto et al., 2024). Standardised external assessments, used more heavily in secondary

education, can signal clearly what is expected, but where high stakes attach to them they can also distort teaching towards the test and, for students who judge the bar out of reach, depress motivation and raise drop-out (OECD, 2013).

Although assessment design changes the size, and sometimes the direction, of the gender gap, this does not mean that a system should pick the assessment that shows the smallest gap without knowing why. For instance, Delaney and Devereux (2021) report that removing negative marking from multiple-choice tests, as in the redesigned US SAT, improves girls' relative performance, but the literature has not separated whether girls underperform under competitive pressure or boys raise their effort when stakes are higher.

The importance given to coursework and teacher-assigned grades should be considered. Comparing components within GCSE subjects in England, Bramley and colleagues (2015) found girls' more often lead in coursework components compared to written examination components. Guez and colleagues (2020), analysing French students across teacher evaluations, national examinations and standardised tests, found the gap shifts with evaluation type, girls doing relatively better in teacher evaluations and relatively worse in standardised tests. Part I recorded the same divergence between grades and standardised tests in Sweden and the United States. The Covid years supplied a natural experiment. Carroll (2023) and Devereux and Kapur (2024) both record gaps favouring girls widening, and gaps favouring boys narrowing, when systems increased teacher assessment.

The classroom-level reasons for these variations are discussed in further detail in Part III, the task for the current sub-section, education system features, is understanding where the balance should be struck according to the evidence. Neither of the assessment modes (teacher assessed grades/standardised tests) can be judged against actual student proficiency, which cannot be observed directly. Instead, this report uses how well each predicts later educational outcomes as a proxy. On that criterion, teacher assessed grades do better. Brookhart and colleagues (2016), reviewing a century of grading research, report that standardised scores are not as good at predicting a student's school completion and entry to higher education, while grades consistently predict persistence, completion and the transition from school to further study. They also find that roughly a quarter of the variation in teacher-assigned grades is attributable to what tests measure. The rest reflects mostly the achievement of classroom learning intentions, and to a lesser degree students' effort and engagement. The authors are explicit that expecting the two to correlate highly is not a sound validity argument, because school achievement is not fully defined by what tests capture. Graetz and Karimi (2022) find similar results across a Swedish national population, where grade point average predicts graduation better than the national admissions test. Part I recorded the gender pattern in that study. Girls outperform boys on grade point average by around a third of a standard deviation, while boys score slightly higher on the test. On this evidence, the mode that predicts outcomes better is the one in which boys do worse, on average.

How far the two modes of assessment diverge is a feature of a national system rather than a fixed property of the assessment instruments. Two Nordic cases show the range. Arencibia Alemán and colleagues (2024) examined nearly a decade of Norwegian data in mathematics, English and Norwegian, in a system where grading criteria and written examinations are built on the same construct by design. The correlation between grades and examinations was higher than the prior literature predicts. The authors conclude that grades in Norway measure academic achievement better than previous research has shown, and note that weaknesses in the reliability of the examinations themselves mean the true figure is probably higher still.

Finland sits at the other end. Lehti and Laaninen (2024) matched national school register grades to test scores for the same cohorts and found the gender gap in literacy grades significantly

exceeded the gap in test scores. What accounts for that difference matters more than its size. Lehti and Laaninen find that boys' lower interest in reading, lower effort in schoolwork and lower conscientiousness about homework account for the whole of the excess. Had boys matched girls on those three, the gap recorded in grades and the gap recorded in tests would have been the same. In other words, the part of grades that diverges from tests is the part that carries the gender difference, which is why the mode that predicts better is also the mode on which boys do worse. Whether that divergence reflects something real about how boys approach schoolwork or something about how their work is marked is a classroom question, and Part III takes it up.

The criterion for a defensible balance between grades and tests has two parts. Each mode should measure proficiency as accurately for boys as for girls, and the balance overall should retain the predictive strength that makes assessment worth doing. On present evidence the second is met better by grades than by examinations, with one caveat. Studies of general and non-selective settings find school grades the stronger predictor of college completion and later grades, while studies of highly selective admission find the opposite, with standardised scores predicting first-year performance at the most selective American universities several times more strongly than school grades (Friedman et al., 2025). German data on the selection of medical students show the same asymmetry, with aptitude tests adding substantially to the prediction of later performance beyond school-leaving grades (Jaehn et al., 2025). Friedman and colleagues read the pattern across the literature as a matter of selectivity and suggest that grade inflation may be eroding the information in grades most at the top, where high-scoring students press against the ceiling of the scale. That explanation is offered as one possibility among multiple, rather than a tested finding. All this evidence concerns admission to higher education and none of it tests whether the difference between the two modes as predictors is the same for boys and girls.

No study found in this review has tested whether school grades and school examinations predict later outcomes equally well for boys and girls, which is the test needed. Overall, it is very difficult to make student assessment wholly inclusive and neutral since all format, content, construct, and method may be biased in some direction (OECD, 2013). With this in mind, high-stakes decisions about students should not be based on the results of a single test and collecting multiple data, using a range of assessment tasks involving a variety of contexts, response formats and styles is the fairest way.

In addition to measuring student performance, education systems regularly use student assessments for school accountability purposes. Building a dataset of over two million students in 59 countries across six PISA waves (2000 to 2015), Bergbauer et al. (2018) find that the expansion of standardised external comparisons, both school-based and student-based, is associated with improvements in student achievement. By contrast, the introduction of solely internal testing and monitoring teacher practices by principals and external inspectorates does not affect student achievement. One qualification runs through this literature. Improvements in student performance are measured on the high-stakes test the accountability system itself uses, which leaves open how far they reflect broader learning rather than improved performance on that particular instrument (Ladd & Lauen, 2010). An average gain also says nothing about which students within a school are improving.

Accountability regimes have the ability to change where schools direct their effort, and one way a school can move its measured performance is by reallocating attention across its students. They can also have some negative side effects, such as teaching to the test, curriculum narrowing, altering the pool of tested students for better school average attainment, erosion of teacher solidarity and diminishing student trust in educators (Levatino et al., 2024). Most accountability regimes fall broadly into two designs: A threshold measure, which this literature also calls a status measure, judges a school on how many of its students reach a fixed standard in a given year and;

a growth measure, which judges it on how much its students improve over the year, relative to where they started. Comparing cohorts who sat the same exams before and after Chicago introduced threshold accountability, Neal and Schanzenbach (2010) find that the changes raised reading and mathematics scores among students near the proficiency bar but not among the lowest attainers. Schools gave extra help to the students near the bar while providing no special attention either to those already comfortably above it or to those judged unlikely to reach it. There is some evidence that a threshold approach produced negative effects for students above the proficiency standard, while a growth approach does not (Ladd & Lauen, 2010).

England provides a worked example of moving from one rule to the other. Until 2016, secondary schools were judged principally on the proportion of students gaining five or more A*-C grades including English and mathematics, a threshold measure. It was replaced by Progress 8, a growth measure, which assesses how far students move from their own starting point at the end of primary school. Using seven years of secondary attainment data, Burgess and Thomson (2023) examine how schools responded. The results are consistent with schools having previously focused on the group whose results counted most under the old threshold.

Once that incentive was removed, that group made less relative progress while other students progressed faster. The gender element is that, under Progress 8, girls make slightly more progress per subject than boys, and because girls already lead at the end of primary school, the gender gap widens across secondary schooling (Leckie & Goldstein, 2019). The research reveals a plausible mechanism acting on boys, who are over-represented among the lowest attainers. Where an accountability rule concentrates attention on students near a threshold, it concentrates it away from the lowest attainers. Where a rule rewards progress from a starting point, it removes that particular penalty without redistributing towards the bottom.

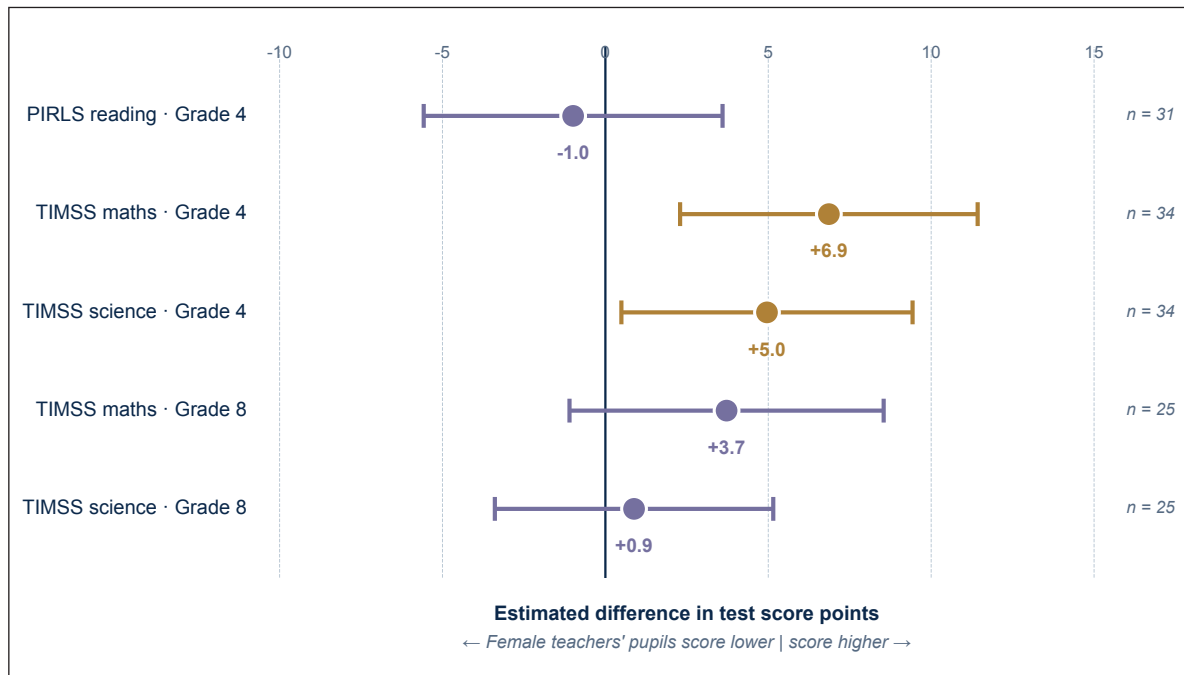
Teaching workforce

The gender balance of the teaching workforce is set by who enters teaching and who stays, and both respond to policy on recruitment and retention. The composition of the teaching workforce is one of the most widely discussed policy levers in debates about boys' education. On average, 84% of primary teachers in OECD countries were women in 2019, compared to 64% at secondary level and 44% at tertiary level (OECD, 2022). There are significant differences across countries, the share of female primary teachers ranges from 64% in Japan and to 97% in Lithuania.

The specific proposition that recruiting more male teachers would raise boys' achievement recurs in policy commentary, in public debate and in the literature on gendered educational outcomes. The supposed mechanisms for this are diverse and include the provision of role models, catalysing changes in the gendered character of classroom interaction and by altering boys' sense of 'who school is for'. The international assessment data permits a direct, if narrow, test of this proposition (Figure 3). What it shows is that, across high-income mixed-gender education systems, the gender of a child's teacher does not predict their measured achievement to any consistent degree.⁴

⁴ A few methodological cautions apply. The models were estimated separately within each country and then averaged across high-income systems, so the conclusion is a description of the typical pattern rather than a single combined statistical test. Country coverage varies across cycles: thirty-one high-income systems for PIRLS Grade 4 reading, thirty-four for TIMSS Grade 4 mathematics and science, and twenty-five for TIMSS Grade 8 mathematics and science. The Grade 8 figures rest on the fewest systems and are the least precise of the five. Teacher gender, as captured in the IEA studies, refers to the teacher administratively linked to each pupil in the data; at primary level this is typically the main class teacher, but at Grade 8 it is the relevant subject teacher, so the meaning of "the pupil's teacher" differs slightly between cycles.

Figure 3 - Estimated difference in boys' test scores between those taught by female and male teachers, across five regression models



Notes: Positive values indicate boys taught by women scored higher on average than boys taught by men. Two associations reach statistical significance (in gold); both point in the same direction. High-income systems, descriptive cross-sectional pooling. N = number of high-income systems with data.

Source: PIRLS 2021 database, TIMSS 2023 database.

The analysis covers reading, mathematics and science achievement, across PIRLS 2021 at Grade 4 and TIMSS 2023 at both Grade 4 and Grade 8. In each case, students' achievement was regressed on the gender of their linked teacher, on the pupil's own gender, and on the interaction between the two. Having a female rather than a male teacher makes almost no difference to boys' achievement. The estimated effect ranges from roughly one and a half score points lower to six score points higher across the five models, differences negligible against assessment scales on which the spread of pupil scores runs into the hundreds of points. The two effects large enough to reach statistical significance, on Grade 4 mathematics and science, point opposite to the popular thesis: boys taught by women score slightly higher, not lower. None of this is to say that gender gaps in achievement are themselves small. It simply shows that these achievement gaps are not appreciably changed by whether a child's teacher is a man or a woman. They are present in classrooms led by men and in classrooms led by women alike.

These findings are supported by other analyses. For example, using data on upper secondary school students and their teachers from the municipality of Stockholm, Sweden, Holmlund and Sund (2008) found no strong support for their initial hypothesis that a same gender teacher improves student outcomes. This is in line with other studies which seek to test role-model effects of teacher gender. For instance, de Gendre et al. (2024) carried out a study of same-gender role model effects of teachers with a meta-analysis and their own study of three million students in 90 countries. They established that role models have a negligible effect on student test performance. However, they found larger average effects and more variation for non-test score outcomes, such as job preferences, which are more pronounced in high-income and gender-equal countries. They conclude by suggesting that hiring more male teachers in primary school will not close gender gaps in student performance.

These findings sit in some tension with some parts of the literature, which has argued that male teachers, role-model effects and the gendered character of classroom interaction matter for boys' engagement with school. That tension is real and is the central question this section addresses, but it is not necessarily a contradiction. For example, Dee (2007) exploits a unique "matched pairs" feature of a major longitudinal survey in the United States and finds that gender interactions between teachers and students had statistically significant effects on test scores, teacher perceptions of student performance and student engagement with academic subjects. These effects were also quite large relative to the subject-specific gender gaps. For example, they implied that just one year with a male English teacher would eliminate nearly a third of the gender gap in reading. However, and crucially, it would do so by improving the performance of boys and simultaneously harming the performance of girls.

Doornkamp et al. (2024) adds to these findings with data from the Netherlands. They found that positive effects of student–teacher gender matching emerge in different subjects for male and female students. Female students benefited from female teachers in mathematics and physics, while male students benefited from male teachers in Dutch language and French language. Other analyses conclude that same-gender teaching has mixed effects, with some evidence of benefits for girls in mathematics, and analyses of linked teacher and student data find that a student's performance tends to rise modestly as the share of teachers of their own gender in the school increases (Delaney & Devereux, 2021; OECD, 2021). Moving up to the country level, gender achievement gaps covary with how far women are overrepresented in teaching relative to their share of the wider labour force, with girls' advantages larger where that overrepresentation is more pronounced, a pattern that points to broad societal dynamics rather than to anything happening in the individual classroom (Luyten, 2026).

The cross-national correlational data can test whether teacher gender predicts students' achievement on standardised assessments, averaged across high-income systems. It is not well placed to detect within-classroom processes and differences in pedagogy or relational dynamics between male and female teachers. Nor can it speak to effects on engagement and motivation rather than on attainment, effects that accumulate over a longer school career, or effects concentrated in particular subgroups of boys for whom the question may matter most. A null finding at the cross-national, cross-subject, cross-grade level does not rule out that any of these mechanisms operate; it shows only that, in the aggregate, no clear achievement effect emerges from a teacher-gender variable on its own. What the original analysis conducted for this review therefore establishes is a useful starting point for policy, rather than a complete answer to the question.

There is also a self-reinforcing dimension that bears directly on boys. The teaching workforce is heavily female, and the more so it is, the wider the gender gap in students' own expectations of becoming teachers, which suggests that the overrepresentation of women may itself discourage boys from entering the profession (Han et al., 2020). The same imbalance is sharpest in the subject where boys are most disadvantaged. Most reading teachers are women, so boys are less likely to encounter a male reading mentor in school (Eck et al., 2024). It follows that the case for a more gender-balanced teaching workforce is better made on grounds other than a direct effect on boys' test scores. A better balance could benefit all students by helping them develop positive gender identities and challenge biased assumptions, and it would help address teacher labour shortages, both reasons to want balance independent of any achievement effect (OECD, 2023b).

If the gender of the teacher is largely besides the point for boys' achievement, the preparation and practice of teachers is not, and here the evidence is more actionable. Teaching methods that promote student autonomy and individualisation, giving students supported choice in tasks suited to their interests and abilities, are associated with smaller gender differences in motivation (Gregory &

Welmond, 2021). Teacher professional development directed specifically at personalised instruction for disadvantaged students is part of a credible strategy for using teacher quality to lift attainment (Agasisti et al., 2021). The difficulty is that teachers often do not feel equipped for exactly the situations in which struggling boys are concentrated. Among recent teaching graduates, only a third feel prepared for multicultural or multilingual classrooms and fewer than half for supporting students' social and emotional development, with reported preparedness on several fronts having declined according to OECD assessments (OECD, 2019b, 2021, 2025b). Here is where the room for action lies at the level of the education system. Comprehensive initial teacher education curriculum covers both content and pedagogical knowledge, while developing practical skills. A recurring diagnosis is the disconnect between teacher-education institutions and the realities of schools (OECD, 2019b). Policy makers have a role in promoting a shared vision of teacher learning as a continuous process, providing strategic governance of knowledge around teacher preparation, and capacity building at all levels (OECD, 2019a). Since struggling students are disproportionately boys, and struggling boys from minority backgrounds are among the most at risk for poorer completion, progression and attainment, these preparation gaps fall unevenly on the group this review is concerned with.

Conclusion

Part I described the gaps between boys' and girls' education outcomes. Part II asked what conditions contribute to them. Four conclusions can be drawn.

Boys and girls arrive at school with modest average differences and leave with substantial ones. Girls develop self-regulation earlier. Boys are on average more physically active. Pubertal timing differs, and age-based assessment captures boys at a less developmentally mature point than girls. These differences are real, they are small, and the two groups overlap far more than they differ. None of them is large enough on its own to account for the gaps Part I described. Morris (2012) cautions that biological and socialisation accounts alike tend towards over-determinism, leaving too little room for variation among boys, for individual agency, and for the part institutions play.

Institutional arrangements determine how much those differences matter. Age-graded classes and fixed-age assessment convert a difference in developmental timing into a difference in measured attainment. Grade retention has been shown to operate as a gate that boys pass through more often than girls at the same level of measured performance, and whether a boy repeats a year depends substantially on which school he attends. The balance a system strikes between teacher-assigned grades and standardised examinations changes the size and sometimes the direction of the gap it records. Accountability rules determine which students a school directs its attention towards. Curricular prescription in reading and language bears more heavily on boys than on girls. Each of these is a design choice rather than a fact about children.

Boys' outcomes are more tied to their surroundings than girls', but the claim has limits. It holds most clearly for long-run trajectories into work and for progression into higher education. It holds for school socio-economic composition in mathematics, and there it operates at the advantaged end, where boys gain more from an advantaged intake rather than losing more from a disadvantaged one. It does not hold for the language gap, which is large, present in almost every place studied, and largely unexplained by local socio-economic conditions. And the mechanism most often proposed for it, that school peer culture shapes which versions of masculinity carry standing and thereby boys' orientation to school, is not supported by this review's own analysis of boys' reported attitudes, which finds no difference between advantaged and disadvantaged schools in terms of the size of the gender gap.

Where class, place and masculine norms run together, boys face compounding disadvantage. The evidence for this concerns long-run employment trajectories, where the gender difference in sensitivity to place is concentrated among low-income families and fades further up the income distribution. That does not extend to the size of the gender gap in achievement. The gender gap is not concentrated among disadvantaged students. Part I found boys' reading underperformance spread across the socio-economic range, with boys in the wealthiest quintile more than twice as likely as girls in the same quintile to fall short of the minimum standard, and the gap in reaching high proficiency wider among advantaged students than among poor ones. Disadvantaged boys have the poorest outcomes. The difference between boys' and girls' outcomes is not a specific feature of disadvantage. Both are true and they answer different questions. The conditions around boys, rather than boys themselves, are where the room for action lies.

Recommendations

This review does not endorse changes that improve one group's outcomes by worsening another's. That applies between boys and girls. It applies equally between boys and children from poorer homes, and between lower-attaining and higher-attaining students. Redistribution of that kind leaves the arrangement producing the gap in place and moves who carries its cost.

This is a necessary constraint to the recommendations, because some levers examined would technically shrink the gender gap. For example, selecting into tracks earlier would narrow the gender gap in attainment, by lowering girls and widening socio-economic gaps.

Similarly, reducing standardisation in mathematics would improve boys' relative position, by worsening girls'. Each is set aside on principle rather than on evidence. The recommendations that follow below are organised in terms of strength of their evidence base.

1. Minimise grade retention.

Retention mostly produces poorer outcomes for the student retained, and its costs outweigh its benefits. Boys are retained more often than girls in almost every system with data, and more often at the same level of measured performance. Retention practice varies more between schools than pupil performance does. Alternatives with better evidence exist, among them multi-age grouping, looping, additional instructional time and remedial tutoring.

2. The evidence on tracking age is not an argument for selecting earlier.

Later selection benefits girls but has not been found to harm boys. Earlier selection would lower girls' attainment and widen socio-economic gaps while returning nothing to boys. Where a system is considering the age of first selection, boys' underachievement is not a reason to make the change.

3. Recruiting more male teachers should not be expected to raise boys' achievement.

Across high-income systems, the gender of a child's teacher does not predict measured achievement to any consistent degree, and the largest international study of the question reaches the same conclusion. The case for a more gender-balanced teaching workforce rests on other grounds, among them addressing shortages, supporting caring male role models, and widening the range of careers boys consider.

4. Where a system prescribes closely what and how reading and language are taught, expect that to bear more heavily on boys.

Curricular prescription in reading and language bears more heavily on boys than on girls. This review takes no position on standardisation in mathematics, where the association runs in girls' favour at no cost to average achievement. The equity arguments usually made for standardisation rest on national examinations rather than on curricular prescription, and the two can be held apart.

5. Choose accountability rules knowing which students they direct attention towards.

Threshold measures concentrate schools' attention on students near the standard and away from the lowest attainers, among whom boys are over-represented. Growth measures remove that particular penalty without redistributing attention towards the bottom. Neither design is clearly favourable to boys, and the choice between them should be made with that in view rather than as a technical matter.

6. Do not choose an assessment regime by the size of the gap it displays.

Teacher-assigned grades and standardised examinations record different gaps between boys and girls, and each carries information the other does not. Boys as a group do worse on grades. However, grades predict later educational outcomes better than examinations for most students, though not at the most selective margins of entry to higher education. A system that shifts weight from examinations towards grades will widen the gap it records in girls' favour, and will raise the bar boys face where selection turns on that mode. That is a reason to establish why the two diverge before moving between them, not a reason to prefer whichever displays the smaller gap. No study located by this review has tested whether grades and examinations predict later outcomes equally well for boys and for girls, which is the question a system choosing a balance needs answered and cannot currently get.

7. Vocational and technical routes should be a positive choice, not a default destination for struggling boys.

These routes already draw disproportionately on boys with low prior attainment from poorer homes. Directing more struggling boys into them concentrates that group in provision that is often under-resourced and lower in prestige, and in schools with skewed gender composition, which is itself associated with weaker reading outcomes for boys.

8. Build teacher preparation for the classrooms in which struggling boys are concentrated.

Among recent teaching graduates, only a third report feeling prepared for multicultural or multilingual classrooms and fewer than half for supporting social and emotional development. Teaching that gives students supported choice over tasks is associated with smaller gender differences in motivation.

9. Build movement into the school day.

Boys' attention is more sensitive than girls' to long periods without a break, and short structured activity breaks are associated with improved on-task behaviour for boys and girls. The evidence is modest and the cost is low.

Research Gaps

No study located by this review has tested whether teacher-assigned grades and standardised examinations predict later outcomes equally well for boys and for girls. That is the test a system choosing a balance between them needs.

The largest recent evaluation of attainment grouping in England reported outcomes by free school meal status and by prior attainment, and not by gender. The boys-specific version of the argument against setting is untested rather than refuted.

No large-scale study of the tests used for placement into tracks has examined whether their accuracy differs for boys and girls. The most recent comparison of the Dutch instruments states this as an open question.

Later tracking reduces socio-economic inequality in test scores among girls, and the same effect is not detected among boys. Why additional time in untracked classes delivers for low socio-economic status girls and not for boys from comparable homes is unresolved, and it is an important question.

Curricular prescription and boys' reading rests on a single country-level associational study. No comparable study of curricular prescription exists for mathematics.

There is no systematic cross-national work on how accountability regimes bear on boys' engagement or attainment.

Research on pubertal development, adolescent change, stereotypes and curriculum reform has studied girls more than boys, so the picture for boys is thinner, assembled by inference, or absent.

Across research on educational systems and inequality, most work addresses ethnic inequality, and research on system characteristics and gender inequality is comparatively scarce.

The mechanism through which school composition bears on boys' achievement is contested. The peer culture account and the occupational incentives account both remain live, and no study has tested them against one another with direct measures of school climate and of masculinity norms across multiple systems.

References

- Agasisti, T., Avvisati, F., Borgonovi, F., & Longobardi, S. (2021). What School Factors are Associated with the Success of Socio-Economically Disadvantaged Students? An Empirical Investigation Using PISA Data. *Social Indicators Research*, 157(2), 749–781. <https://doi.org/10.1007/s11205-021-02668-w>
- Arencibia Alemán, J. M., Sandsør, A. M. J., Zachrisson, H. D., & Blömeke, S. (2024). Teacher-assigned grades and external exams: Sources of discrepancy. *Assessment in Education: Principles, Policy & Practice*, 31(2), 94–115. <https://doi.org/10.1080/0969594X.2024.2338764>
- Aubert, S., Barnes, J. D., Demchenko, I., Hawthorne, M., Abdeta, C., Abi Nader, P., Adsuar Sala, J. C., Aguilar-Farias, N., Aznar, S., Bakalár, P., Bhawra, J., Brazo-Sayavera, J., Bringas, M., Cagas, J. Y., Carlin, A., Chang, C.-K., Chen, B., Christiansen, L. B., Christie, C. J.-A., Tremblay, M. S. (2022). Global Matrix 4.0 Physical Activity Report Card Grades for Children and Adolescents: Results and Analyses From 57 Countries. *Journal of Physical Activity and Health*, 19(11), 700–728. <https://doi.org/10.1123/jpah.2022-0456>
- Autor, D., Dorn, D., & Hanson, G. (2017). *When Work Disappears: Manufacturing Decline and the Falling Marriage-Market Value of Young Men* (No. W23173; p. w23173). National Bureau of Economic Research. <https://doi.org/10.3386/w23173>
- Ayalon, H., & Livneh, I. (2013). Educational standardization and gender differences in mathematics achievement: A comparative study. *Social Science Research*, 42(2), 432–445. <https://doi.org/10.1016/j.ssresearch.2012.10.001>
- Azmat, G., Calsamiglia, C., & Iriberry, N. (2016). GENDER DIFFERENCES IN RESPONSE TO BIG STAKES: Gender Differences in Response to Big Stakes. *Journal of the European Economic Association*, 14(6), 1372–1400. <https://doi.org/10.1111/jeea.12180>
- Balducci, M., Larose, M.-P., Stoet, G., & Geary, D. C. (2024). The Gender-Equality Paradox in Intraindividual Academic Strengths: A Cross-Temporal Analysis. *Psychological Science*, 35(11), 1246–1259. <https://doi.org/10.1177/09567976241271330>
- Bastos, B. de P., Chiviawsky, S., Drews, R., & Cardozo, P. (2023). Gender Stereotype Threat Undermines Dance Performance and Learning in Boys. *Journal of Motor Behavior*, 55(3), 256–261. <https://doi.org/10.1080/00222895.2023.2166454>
- Bedard, K., & Dhuey, E. (2006). The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *The Quarterly Journal of Economics*, 121(4), 1437–1472. JSTOR.
- Bergbauer, A., Hanushek, E., & Woessmann, L. (2018). *Testing* (No. W24836; p. w24836). National Bureau of Economic Research. <https://doi.org/10.3386/w24836>
- Bianchi, D., Cabrera, L., & Sicilia, G. (2025). Contextualising Gender Gaps: The Moderating Role of Social Context on Gender Educational Inequality in European Primary Schools. *Gender Issues*, 42(3), 16. <https://doi.org/10.1007/s12147-025-09360-0>
- Blower, A. (2025). *Lost boys: How education is failing young working-class men*. Policy Press.

- Bol, T., Witschge, J., Van De Werfhorst, H. G., & Dronkers, J. (2014). Curricular Tracking and Central Examinations: Counterbalancing the Impact of Social Background on Student Achievement in 36 Countries. *Social Forces*, 92(4), 1545–1572. <https://doi.org/10.1093/sf/sou003>
- Borgen, N. T., & Zachrisson, H. D. (2025). How neighbourhood effects vary by achievement level. *European Sociological Review*, 41(2), 215–231. <https://doi.org/10.1093/esr/jcae034>
- Borgonovi, F., Ferrara, A., & Maghnouj, S. (2018). *The gender gap in educational outcomes in Norway* (OECD Education Working Papers No. 183; OECD Education Working Papers, Vol. 183). <https://doi.org/10.1787/f8ef1489-en>
- Borgonovi, F., Han, S. W., & Greiff, S. (2023). Gender differences in collaborative problem-solving skills in a cross-country perspective. *Journal of Educational Psychology*, 115(5), 747–766. <https://doi.org/10.1037/edu0000788>
- Borra, C., Iacovou, M., & Sevilla, A. (2023). Adolescent development and the math gender gap. *European Economic Review*, 158, 104542. <https://doi.org/10.1016/j.euroecorev.2023.104542>
- Bossavie, L., & Kanninen, O. (2018). *What Explains the Gender Gap Reversal in Education? Theory and Evidence* (Policy Research Working Paper No. 8303). World Bank Group. <https://openknowledge.worldbank.org/server/api/core/bitstreams/ec486986-2b4f-5fd2-938d-a9b466a4d93b/content>
- Bramley, T., Vidal Rodeiro, C. L., & Vitello, S. (2015). *Gender differences in GCSE* [Cambridge Assessment Research Report]. Cambridge, UK: Cambridge Assessment. <https://www.cambridgeassessment.org.uk/Images/gender-differences-in-gcse.pdf>
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A Century of Grading Research: Meaning and Value in the Most Common Educational Measure. *Review of Educational Research*, 86(4), 803–848. <https://doi.org/10.3102/0034654316672069>
- Brown, P. (1987). *Schooling Ordinary Kids (1987): Inequality, Unemployment, and the New Vocationalism*. Routledge. <https://doi.org/10.4324/9781351009720>
- Brozo, W. G. (2019). *Engaging boys in active literacy: Evidence and practice*. Cambridge University Press. <https://doi.org/10.1017/9781108654111>
- Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, 22(52), 782–861. <https://doi.org/10.1111/j.1468-0327.2007.00189.x>
- Burgess, S., & Thomson, D. (2023). Beyond the threshold: The implications for pupil achievement of reforming school performance metrics. *Economics of Education Review*, 97, 102477. <https://doi.org/10.1016/j.econedurev.2023.102477>
- Carrascal-Incera, A., McCann, P., Ortega-Argilés, R., & Rodríguez-Pose, A. (2020). UK INTERREGIONAL INEQUALITY IN A HISTORICAL AND INTERNATIONAL COMPARATIVE CONTEXT. *National Institute Economic Review*, 253, R4–R17. <https://doi.org/10.1017/nie.2020.26>

Carroll, M. (2023). *Sex gaps in education in England*. Cambridge University Press & Assessment. <https://www.cambridgeassessment.org.uk/Images/698454-sex-gaps-in-education-in-england.pdf>

Cascella, C. (2020). Intersectional effects of Socioeconomic status, phase and gender on Mathematics achievement. *Educational Studies*, 46(4), 476–496. <https://doi.org/10.1080/03055698.2019.1614432>

Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2020). Race and Economic Opportunity in the United States: An Intergenerational Perspective*. *The Quarterly Journal of Economics*, 135(2), 711–783. <https://doi.org/10.1093/qje/qjz042>

Chetty, R., Hendren, N., & Katz, L. F. (2016). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review*, 106(4), 855–902. <https://doi.org/10.1257/aer.20150572>

Chetty, R., Hendren, N., Lin, F., Majerovitz, J., & Scuderi, B. (2016). *Childhood Environment and Gender Gaps in Adulthood* (No. W21936; p. w21936). National Bureau of Economic Research. <https://doi.org/10.3386/w21936>

Clavel, J. G., & Flannery, D. (2023). Single-sex schooling, gender and educational performance: Evidence using PISA data. *British Educational Research Journal*, 49(2), 248–265. <https://doi.org/10.1002/berj.3841>

Connell, R. (1995). *Masculinities* (2. ed., [Nachdr.]). Univ. of California Pr.

Connell, R. (2023). Teaching the Boys: New Research on Masculinity and Gender Strategies for Schools. In A. Darder, K. Hernandez, K. D. Lam, & M. Baltodano, *The Critical Pedagogy Reader* (4th edn, pp. 332–354). Routledge. <https://doi.org/10.4324/9781003286080-25>

de Gendre, A., Feld, J., Salamanca, N., & Zöhlitz, U. (2024). *Same-sex teacher effects*. <https://doi.org/10.5167/UZH-234072>

Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources*, XLII(3), 528–554. <https://doi.org/10.3368/jhr.XLII.3.528>

Dee, T. S., & Sievertsen, H. H. (2018). The gift of time? School starting age and mental health. *Health Economics*, 27(5), 781–802. <https://doi.org/10.1002/hec.3638>

Delaney, J. M., & Devereux, P. J. (2021). The Economics of Gender and Educational Achievement: Stylized Facts and Causal Evidence. In J. M. Delaney & P. J. Devereux, *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190625979.013.663>

Devereux, P., & Kapur, K. (2024). Teacher Bias, Grade Inflation, and Gender Gaps in Achievement: Evidence from Ireland. *The Economic and Social Review*, 55(3, Autumn), 281–310.

Doornkamp, L., Doornkamp, F., Van Der Pol, L. D., Groeneveld, S., Mesman, J., & Groeneveld, M. G. (2024). Student–teacher gender congruence and student performance: The role of context. *Social Psychology of Education*, 27(6), 3299–3328. <https://doi.org/10.1007/s11218-024-09922-2>

- Dotti Sani, G. M., & Quaranta, M. (2017). The Best Is Yet to Come? Attitudes Toward Gender Roles Among Adolescents in 36 Countries. *Sex Roles*, 77(1–2), 30–45. <https://doi.org/10.1007/s11199-016-0698-7>
- Douma, N. H., Warrens, M. J., Fleur, E., Dijks, M. A., & Korpershoek, H. (2025). The predictive power of Dutch end of primary school tests for educational attainment in ninth grade. *British Educational Research Journal*, 51(1), 280–298. <https://doi.org/10.1002/berj.4074>
- Downey, D. B., Kuhfeld, M., & Van Hek, M. (2022). Schools as a Relatively Standardizing Institution: The Case of Gender Gaps in Cognitive Skills. *Sociology of Education*, 95(2), 89–109. <https://doi.org/10.1177/00380407211070319>
- Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, 98(1), 198–208. <https://doi.org/10.1037/0022-0663.98.1.198>
- Eaton, W. O., & Enns, L. R. (1986). Sex differences in human motor activity level. *Psychological Bulletin*, 100(1), 19–28. <https://doi.org/10.1037/0033-2909.100.1.19>
- Eck, M., Hencke, J., Kennedy, A., Meinck, S., & Sass, J. (2024). *Why boys may need more support in building their reading, motivation, confidence, and engagement* (IEA Compass: Briefs in Education No. 25).
- Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., & Van Hulle, C. A. (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin*. <https://doi.org/10.1037/0033-2909.132.1.33>
- European Commission. (2021). *Study on gender behaviour and its impact on education outcomes (with a special focus on the performance of boys and young men in education): Final report*. Publications Office. <https://data.europa.eu/doi/10.2766/509505>
- Friedman, J., Sacerdote, B., Staiger, D., & Tine, M. (2025). *Standardized Test Scores and Academic Performance at Ivy-Plus Colleges* (No. W33570; p. w33570). National Bureau of Economic Research. <https://doi.org/10.3386/w33570>
- Gamazo, A., & Martínez-Abad, F. (2020). An Exploration of Factors Linked to Academic Performance in PISA 2018 Through Data Mining Techniques. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.575167>
- Gater, R. (2025). *The 21st Century Ladz: Continuity and Changes among Marginalised Young Men from the South Wales Valleys*. Emerald Publishing Limited. <https://doi.org/10.1108/9781837976317>
- Gimenez, G., Mediavilla, M., Giuliodori, D., & Rusteholz, G. C. (2024). Bullying at School and Students' Learning Outcomes: International Perspective and Gender Analysis. *Journal of Interpersonal Violence*, 39(11–12), 2733–2760. <https://doi.org/10.1177/08862605231222457>
- Goos, M., Schreier, B. M., Knipprath, H. M. E., De Fraine, B., Van Damme, J., & Trautwein, U. (2013). How Can Cross-Country Differences in the Practice of Grade Retention Be Explained? A Closer Look at National Educational Policy Factors. *Comparative Education Review*, 57(1), 54–84. <https://doi.org/10.1086/667655>

- Graetz, G., & Karimi, A. (2022). Gender gap variation across assessment types: Explanations and implications. *Economics of Education Review*, 91, 102313. <https://doi.org/10.1016/j.econedurev.2022.102313>
- Gregory, L., & Welmond, M. J. (2021). *Educational Underachievement Among Boys and Men (English)*. World Bank Group. <http://documents.worldbank.org/curated/en/111041644611110155>
- Guez, A., Peyre, H., & Ramus, F. (2020). Sex differences in academic achievement are modulated by evaluation type. *Learning and Individual Differences*, 83–84, 101935. <https://doi.org/10.1016/j.lindif.2020.101935>
- Hallam, S., & Parsons, S. (2013). Prevalence of streaming in UK primary schools: Evidence from the Millennium Cohort Study. *British Educational Research Journal*, 39(3), 514–544. <https://doi.org/10.1080/01411926.2012.659721>
- Han, S. W., Borgonovi, F., & Guerriero, S. (2020). Why don't more boys want to become teachers? The effect of a gendered profession on students' career expectations. *International Journal of Educational Research*, 103, 101645. <https://doi.org/10.1016/j.ijer.2020.101645>
- Hanushek, E. A., Link, S., & Woessmann, L. (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics*, 104, 212–232. <https://doi.org/10.1016/j.jdeveco.2012.08.002>
- Hanushek, E. A., & Wößmann, L. (2006). Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries. *The Economic Journal*, 116(510), C63–C76. <https://doi.org/10.1111/j.1468-0297.2006.01076.x>
- Hart, J. L., & Tannock, M. T. (2018). Rough Play: Past, Present, and Potential. In P. K. Smith & J. L. Roopnarine (Eds), *The Cambridge Handbook of Play* (1st edn, pp. 200–221). Cambridge University Press. <https://doi.org/10.1017/9781108131384.012>
- Heyder, A., Van Hek, M., & Van Houtte, M. (2021). When Gender Stereotypes Get Male Adolescents into Trouble: A Longitudinal Study on Gender Conformity Pressure as a Predictor of School Misconduct. *Sex Roles*, 84(1–2), 61–75. <https://doi.org/10.1007/s11199-020-01147-9>
- Hill, J., & Shewbridge, C. (2025). *How can OECD countries empower children to be more physically active?* (339th edn, OECD Education Working Papers) [OECD Education Working Papers]. <https://doi.org/10.1787/f534b2d0-en>
- Hodgen, J., Taylor, B., Tereshchenko, A., Anders, J., Jacques, L., Cockerill, M., Kwok, R., Sirl, D., Burn, H., & Bretscher, N. (2026). *The Student Grouping Study: Evaluation Report*. Education Endowment Foundation. https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/student_grouping_study_-_evaluation_report_-_v1.3.0.pdf?v=1785063183
- Holmlund, H., & Sund, K. (2008). Is the gender gap in school performance affected by the sex of the teacher? *Labour Economics*, 15(1), 37–53. <https://doi.org/10.1016/j.labeco.2006.12.002>

- Holt, C., Zlotowitz, S., Moloney, O., & Chentite, M. (2019). Improving the Mental Health and Well-Being of Excluded Young Men. In J. A. Barry, R. Kinglerlee, M. Seager, & L. Sullivan (Eds), *The Palgrave Handbook of Male Psychology and Mental Health* (pp. 369–396). Springer International Publishing. https://doi.org/10.1007/978-3-030-04384-1_19
- Hossain, M. (2023). The Linkage between School Autonomy and Inequality in Achievement in 69 Countries: Does Development Level Matter? *The Journal of Development Studies*, 59(10), 1491–1507. <https://doi.org/10.1080/00220388.2023.2232916>
- Hoyt, L. T., Niu, L., Pachucki, M. C., & Chaku, N. (2020). Timing of puberty in boys and girls: Implications for population health. *SSM - Population Health*, 10, 100549. <https://doi.org/10.1016/j.ssmph.2020.100549>
- Hyde, J. S. (2014). Gender Similarities and Differences. *Annual Review of Psychology*, 65(1), 373–398. <https://doi.org/10.1146/annurev-psych-010213-115057>
- Ioverno, S., DeLay, D., Martin, C. L., & Hanish, L. D. (2021). Who Engages in Gender Bullying? The Role of Homophobic Name-Calling, Gender Pressure, and Gender Conformity. *Educational Researcher*, 50(4), 215–224. <https://doi.org/10.3102/0013189X20968067>
- Jaehn, M., Hissbach, J., Frickhoeffer, M., Weppert, D., Zimmerhofer, A., Hampe, W., Kadmon, M., & Becker, N. (2025). Predictive validity of admission tests and educational attainment on preclinical academic performance – a multisite study. *BMC Medical Education*, 25(1), 1255. <https://doi.org/10.1186/s12909-025-07974-2>
- Jha, J., Menon, N., & Chatterjee, D. (2017). *Boys' Underperformance in the Commonwealth: Revisiting the Issue*. Commonwealth of Learning. <https://oasis.col.org/server/api/core/bitstreams/0654d268-ee6f-42dc-a279-c888643018fa/content>
- Johnston, O., & Wildy, H. (2016). The effects of streaming in the secondary school on learning outcomes for Australian students – A review of the international literature. *Australian Journal of Education*, 60(1), 42–59. <https://doi.org/10.1177/0004944115626522>
- King, T. L., Scovelle, A. J., Meehl, A., Milner, A. J., & Priest, N. (2021). Gender stereotypes and biases in early childhood: A systematic review. *Australasian Journal of Early Childhood*, 46(2), 112–125. <https://doi.org/10.1177/1836939121999849>
- Klopack, E. T., Sutton, T. E., Simons, R. L., & Simons, L. G. (2020). Disentangling the Effects of Boys' Pubertal Timing: The Importance of Social Context. *Journal of Youth and Adolescence*, 49(7), 1393–1405. <https://doi.org/10.1007/s10964-019-01141-9>
- Koorts, H., Timperio, A., Abbott, G., Arundell, L., Ridgers, N. D., Cerin, E., Brown, H., Daly, R. M., Dunstan, D. W., Hume, C., Chinapaw, M. J. M., Moodie, M., Hesketh, K. D., & Salmon, J. (2022). Is level of implementation linked with intervention outcomes? Process evaluation of the TransformUs intervention to increase children's physical activity and reduce sedentary behaviour. *International Journal of Behavioral Nutrition and Physical Activity*, 19(1), 122. <https://doi.org/10.1186/s12966-022-01354-5>

- Ladd, H. F., & Lauen, D. L. (2010). Status versus growth: The distributional effects of school accountability policies. *Journal of Policy Analysis and Management*, 29(3), 426–450.
<https://doi.org/10.1002/pam.20504>
- Leckie, G., & Goldstein, H. (2019). The importance of adjusting for pupil background in school value-added models: A study of Progress 8 and school accountability in England. *British Educational Research Journal*, 45(3), 518–537. <https://doi.org/10.1002/berj.3511>
- Legewie, J., & DiPrete, T. A. (2011). *Gender Differences in the Effect of Peer SES: Evidence from a Second Quasi-Experimental Case Study*. <https://api.semanticscholar.org/CorpusID:146658933>
- Legewie, J., & DiPrete, T. A. (2012). School Context and the Gender Gap in Educational Achievement. *American Sociological Review*, 77(3), 463–485.
<https://doi.org/10.1177/0003122412440802>
- Lehti, H., & Laaninen, M. (2024). The gender achievement gap in grades and standardised tests – What accounts for gender inequality? *Frontiers in Sociology*, 9, 1448488.
<https://doi.org/10.3389/fsoc.2024.1448488>
- Levatino, A., Verger, A., Camphuijsen, M., Termes, A., & Parcerisa, L. (2024). School Governance Through Performance-Based Accountability: A Comparative Analysis of Its Side Effects Across Different Regulatory Regimes. *Review of Research in Education*, 48(1), 248–286.
<https://doi.org/10.3102/0091732X241270672>
- Li, J., & McLellan, R. (2021). Is language learning a feminine domain? Examining the content and stereotype threat effect of female-language stereotypes among EFL learners in China. *Contemporary Educational Psychology*, 66, 101991.
<https://doi.org/10.1016/j.cedpsych.2021.101991>
- Lundberg, S. (2020). Educational gender gaps. *Southern Economic Journal*, 87(2), 416–439.
<https://doi.org/10.1002/soej.12460>
- Luyten, H. (2026). Overrepresentation of female teachers in secondary education and gender achievement gaps in PISA 2022. *Studies in Educational Evaluation*, 88, 101544.
<https://doi.org/10.1016/j.stueduc.2025.101544>
- Maccoby, E. E. (2003). *The two sexes: Growing up apart, coming together* (5. printing). The Belknap Press of Harvard University Press.
- Marcq, K., & Braeken, J. (2024). Gender differences in item nonresponse in the PISA 2018 student questionnaire. *Educational Assessment, Evaluation and Accountability*, 36(3), 301–336.
<https://doi.org/10.1007/s11092-023-09412-7>
- Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology*, 101(3), 689–704.
<https://doi.org/10.1037/a0014240>
- McDowell, L., & Bonner-Thompson, C. (2020). The other side of coastal towns: Young men's precarious lives on the margins of England. *Environment and Planning A: Economy and Space*, 52(5), 916–932. <https://doi.org/10.1177/0308518X19887968>

- Mejía-Rodríguez, A. M., & Kyriakides, L. (2022). What matters for student learning outcomes? A systematic review of studies exploring system-level factors of educational effectiveness. *Review of Education*, 10(3), e3374. <https://doi.org/10.1002/rev3.3374>
- Mendle, J., & Ferrero, J. (2012). Detrimental psychological outcomes associated with pubertal timing in adolescent boys. *Developmental Review*, 32(1), 49–66. <https://doi.org/10.1016/j.dr.2011.11.001>
- Milne, B., & Baker, C. R. (2025). From ‘villains’ to ‘idols’: Exploring teenage boys’ conflicting attachments to manospheric masculinities. *Gender and Education*, 1–22. <https://doi.org/10.1080/09540253.2025.2568407>
- Miyamoto, A., Gauly, B., & Zabal, A. (2024). Gender differences in literacy in PIAAC: Do assessment features matter? *Large-Scale Assessments in Education*, 12(1), 21. <https://doi.org/10.1186/s40536-024-00208-9>
- Montt, G. (2011). Cross-national Differences in Educational Achievement Inequality. *Sociology of Education*, 84(1), 49–68. <https://doi.org/10.1177/0038040710392717>
- Morris, E. W. (2012). *Learning the hard way: Masculinity, place, and the gender gap in education*. Rutgers University Press.
- Muntoni, F., & Retelsdorf, J. (2019). At their children’s expense: How parents’ gender stereotypes affect their children’s reading outcomes. *Learning and Instruction*, 60, 95–103. <https://doi.org/10.1016/j.learninstruc.2018.12.002>
- Muntoni, F., Wagner, J., & Retelsdorf, J. (2021). Beware of Stereotypes: Are Classmates’ Stereotypes Associated With Students’ Reading Outcomes? *Child Development*, 92(1), 189–204. <https://doi.org/10.1111/cdev.13359>
- Neal, D., & Schanzenbach, D. W. (2010). Left Behind by Design: Proficiency Counts and Test-Based Accountability. *Review of Economics and Statistics*, 92(2), 263–283. <https://doi.org/10.1162/rest.2010.12318>
- Nieuwenhuis, J., & Hooimeijer, P. (2016). The association between neighbourhoods and educational achievement, a systematic review and meta-analysis. *Journal of Housing and the Built Environment*, 31(2), 321–347. <https://doi.org/10.1007/s10901-015-9460-7>
- Oakley, C. M. (2025). *Longitudinal study of the influence of puberty on school engagement and widening achievement gaps during secondary education*. [A thesis submitted for the degree of PhD in Psychology]. Department of Psychology University of Essex. <https://repository.essex.ac.uk/41181/1/Claire%20Oakley%20June%202025.pdf>
- OECD. (2013). *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*. OECD. <https://doi.org/10.1787/9789264190658-en>
- OECD. (2019a). *A Flying Start: Improving Initial Teacher Preparation Systems*. OECD Publishing. <https://doi.org/10.1787/cf74e549-en>

- OECD. (2019b). *TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners*. OECD Publishing. <https://doi.org/10.1787/1d0bc92a-en>
- OECD. (2020). *Early Learning and Child Well-being: A Study of Five-year Olds in England, Estonia, and the United States*. OECD Publishing. <https://doi.org/10.1787/3990407f-en>
- OECD. (2021). *Positive, High-achieving Students?: What Schools and Teachers Can Do*. OECD Publishing. <https://doi.org/10.1787/3b9551db-en>
- OECD. (2022). *Why is the gender ratio of teachers imbalanced?* (81st edn, Education Indicators in Focus) [Education Indicators in Focus]. <https://doi.org/10.1787/8fea2729-en>
- OECD. (2023a). *Education at a Glance 2023: OECD Indicators*. OECD Publishing. <https://doi.org/10.1787/e13bef63-en>
- OECD. (2023b). *Joining Forces for Gender Equality: What is Holding us Back?* OECD Publishing. <https://doi.org/10.1787/67d48024-en>
- OECD. (2023c). *PISA 2022 Results (Volume II): Learning During – and From – Disruption*. OECD Publishing. <https://doi.org/10.1787/a97db61c-en>
- OECD. (2024a). *Nurturing Social and Emotional Learning Across the Globe: Findings from the OECD Survey on Social and Emotional Skills 2023*. OECD Publishing. <https://doi.org/10.1787/32b647d0-en>
- OECD. (2024b). *Social and Emotional Skills for Better Lives: Findings from the OECD Survey on Social and Emotional Skills 2023*. OECD Publishing. <https://doi.org/10.1787/35ca7b7c-en>
- OECD. (2025a). *Education at a Glance 2025: OECD Indicators*. OECD Publishing. <https://doi.org/10.1787/1c0d9c79-en>
- OECD. (2025b). *Results from TALIS 2024: The State of Teaching*. OECD Publishing. <https://doi.org/10.1787/90df6235-en>
- Overman, H. G., & Xu, X. (2024). Spatial disparities across labour markets. *Oxford Open Economics*, 3(Supplement_1), i585–i610. <https://doi.org/10.1093/ooec/odae005>
- Pahlke, E., Hyde, J. S., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychological Bulletin*, 140(4), 1042–1072. <https://doi.org/10.1037/a0035740>
- Pansu, P., Régner, I., Max, S., Colé, P., Nezlek, J. B., & Huguet, P. (2016). A burden for the boys: Evidence of stereotype threat in boys' reading performance. *Journal of Experimental Social Psychology*, 65, 26–30. <https://doi.org/10.1016/j.jesp.2016.02.008>
- Parekh, G., Brown, R., & Abdulkarim, F. (2021). *Streaming in Education: Thinking beyond Grade 9*. Technical Report. York University – Toronto District School Board.

- Park, H., Behrman, J. R., & Choi, J. (2013). Causal Effects of Single-Sex Schools on College Entrance Exams and College Attendance: Random Assignment in Seoul High Schools. *Demography*, 50(2), 447–469. <https://doi.org/10.1007/s13524-012-0157-1>
- Parsons, S., & Hallam, S. (2014). The impact of streaming on attainment at age seven: Evidence from the Millennium Cohort Study. *Oxford Review of Education*, 40(5), 567–589. <https://doi.org/10.1080/03054985.2014.959911>
- 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(4), 807–825. <https://doi.org/10.1111/j.1467-9442.2008.00562.x>
- Pellegrini, A., & Bohn-Gettler, C. (2005). The Role of Recess in Children's Cognitive Performance and School Adjustment. *Educational Researcher*, 34, 13–19. <https://doi.org/10.3102/0013189X034001013>
- Pepels, J., Belfi, B., Goos, M., & Haelermans, C. (2026). Grade Retention Across Borders: A Cross-Country Analysis of the Role of Schools. *School Effectiveness and School Improvement*, 37(2), 196–221. <https://doi.org/10.1080/09243453.2026.2658545>
- Reardon, S. F., Fahle, E. M., Kalogrides, D., Podolsky, A., & Zárate, R. C. (2019). Gender Achievement Gaps in U.S. School Districts. *American Educational Research Journal*, 56(6), 2474–2508. <https://doi.org/10.3102/0002831219843824>
- Roberts, S. (2012). 'I just got on with it': The educational experiences of ordinary, yet overlooked, boys. *British Journal of Sociology of Education*, 33(2), 203–221. <https://doi.org/10.1080/01425692.2011.649832>
- Robson, D. A., Allen, M. S., & Howard, S. J. (2020). Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review. *Psychological Bulletin*, 146(4), 324–354. <https://doi.org/10.1037/bul0000227>
- Røed, M., Schøne, P., & Strøm, M. (2025). How important is girls' 'Biological Head Start' in explaining gender differences in education and the labour market? *Economics & Human Biology*, 56, 101466. <https://doi.org/10.1016/j.ehb.2024.101466>
- Rosen, N. L., & Nofziger, S. (2019). Boys, Bullying, and Gender Roles: How Hegemonic Masculinity Shapes Bullying Behavior. *Gender Issues*, 36(3), 295–318. <https://doi.org/10.1007/s12147-018-9226-0>
- Salmon, J., Arundell, L., Cerin, E., Ridgers, N. D., Hesketh, K. D., Daly, R. M., Dunstan, D., Brown, H., Della Gatta, J., Della Gatta, P., Chinapaw, M. J. M., Sheppard, L., Moodie, M., Hume, C., Brown, V., Ball, K., & Crawford, D. (2023). Transform-Us! cluster RCT: 18-month and 30-month effects on children's physical activity, sedentary time and cardiometabolic risk markers. *British Journal of Sports Medicine*, 57(5), 311–319. <https://doi.org/10.1136/bjsports-2022-105825>
- Scheeren, L. (2022). The Differential Impact of Educational Tracking on SES Gaps in Educational Achievement for Boys and Girls. *European Sociological Review*, 38(6), 942–958. <https://doi.org/10.1093/esr/jcac012>

- Scheeren, L., & Bol, T. (2022). Gender inequality in educational performance over the school career: The role of tracking. *Research in Social Stratification and Mobility*, 77, 100661. <https://doi.org/10.1016/j.rssm.2021.100661>
- Scheeren, L., Van De Werfhorst, H. G., & Bol, T. (2018). The Gender Revolution in Context: How Later Tracking in Education Benefits Girls. *Social Forces*, 97(1), 193–220. <https://doi.org/10.1093/sf/soy025>
- Silva, L. (2023). *Beyond “Do neighbourhoods matter?”: Investigating heterogeneous neighbourhood effects on youth development* (Issue 2023IEPP0035) [Theses, Institut d'études politiques de paris - Sciences Po]. <https://theses.hal.science/tel-04465516>
- Silva, L. (2024). Gendered Contexts? The Effect of Neighbourhood Socio-Economic Deprivation on Girls' and Boys' Cognitive and Non-Cognitive Development. *Sociology Compass*, 18(10), e70010. <https://doi.org/10.1111/soc4.70010>
- Smiler, A. P. (2024). Understanding masculinity: A multifaceted organizing framework. *Psychology of Men Masculinities*, 25(4), 357–374. <https://doi.org/10.1037/men0000495>
- Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review*, 28(1), 78–106. <https://doi.org/10.1016/j.dr.2007.08.002>
- Steinmann, I., Strello, A., & Strietholt, R. (2023). The effects of early between-school tracking on gender segregation and gender gaps in achievement: A differences-in-differences study. *School Effectiveness and School Improvement*, 34(2), 189–208. <https://doi.org/10.1080/09243453.2023.2165510>
- Streck, H., Nishen, A. K., & Kessels, U. (2022). Instrumentality Gives Girls the Edge: Gender-Differential Relations Between Instrumentality, Achievement Motivation, and Self-Esteem. *Sex Roles*, 86(5–6), 379–394. <https://doi.org/10.1007/s11199-021-01270-1>
- Strello, A., Strietholt, R., Steinmann, I., & Siepmann, C. (2021). Early tracking and different types of inequalities in achievement: Difference-in-differences evidence from 20 years of large-scale assessments. *Educational Assessment, Evaluation and Accountability*, 33(1), 139–167. <https://doi.org/10.1007/s11092-020-09346-4>
- Taraszw, T., Gentrup, S., & Heppt, B. (2024). Egalitarian gender role attitudes give girls the edge: Exploring the role of students' gender role attitudes in reading and math. *Social Psychology of Education*, 27(6), 3425–3452. <https://doi.org/10.1007/s11218-024-09913-3>
- Torvik, F. A., Flatø, M., McAdams, T. A., Colman, I., Silventoinen, K., & Stoltenberg, C. (2021). Early Puberty Is Associated With Higher Academic Achievement in Boys and Girls and Partially Explains Academic Sex Differences. *Journal of Adolescent Health*, 69(3), 503–510. <https://doi.org/10.1016/j.jadohealth.2021.02.001>
- Ullrich, R., Van Houtte, M., & Becker, M. (2024). Student and Teacher Culture and Composition and the Development of Gender Role Attitudes among Young Adolescents. *Journal of Youth and Adolescence*, 53(3), 563–580. <https://doi.org/10.1007/s10964-023-01897-1>

UNESCO. (2022). *Leave no child behind: Global report on boys' disengagement from education*. UNESCO. <https://doi.org/10.54675/BDLL3314>

Urruticoechea, A., Oliveri, A., Vernazza, E., Giménez-Dasí, M., Martínez-Arias, R., & Martín-Babarro, J. (2021). The Relative Age Effects in Educational Development: A Systematic Review. *International Journal of Environmental Research and Public Health*, 18(17), 8966. <https://doi.org/10.3390/ijerph18178966>

Valbuena, J., Mediavilla, M., Choi, Á., & Gil, M. (2021). EFFECTS OF GRADE RETENTION POLICIES: A LITERATURE REVIEW OF EMPIRICAL STUDIES APPLYING CAUSAL INFERENCE. *Journal of Economic Surveys*, 35(2), 408–451. <https://doi.org/10.1111/joes.12406>

Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review*, 5(2), 134–150. <https://doi.org/10.1016/j.edurev.2010.02.001>

Van Hek, M. (2025). Exploring pathways: How friends' anti-academic behavior contributes to the gender gap in language and math grades. *Research in Social Stratification and Mobility*, 97, 101042. <https://doi.org/10.1016/j.rssm.2025.101042>

Van Hek, M., Buchmann, C., & Kraaykamp, G. (2019). Educational Systems and Gender Differences in Reading: A Comparative Multilevel Analysis. *European Sociological Review*, 35(2), 169–186. <https://doi.org/10.1093/esr/jcy054>

van Hek, M., Kraaykamp, G., & 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), 1–21. <https://doi.org/10.1080/09243453.2017.1382540>

Varsik, S., & Gorochovskij, J. (2023). *Intersectionality in education: Rationale and practices to address the needs of students' intersecting identities* (OECD Education Working Papers No. 302; OECD Education Working Papers, Vol. 302). <https://doi.org/10.1787/dbb1e821-en>

Ward, M. R. M. (2015). *From Labouring to Learning: Working-Class Masculinities, Education and De-Industrialization*. Palgrave Macmillan UK. <https://doi.org/10.1057/9781137441751>

Way, N., Cressen, J., Bodian, S., Preston, J., Nelson, J., & Hughes, D. (2014). "It might be nice to be a girl... Then you wouldn't have to be emotionless": Boys' resistance to norms of masculinity during adolescence. *Psychology of Men & Masculinity*, 15(3), 241–252. <https://doi.org/10.1037/a0037262>

Willis, P. (1977). *Learning to Labour* (0 edn). Routledge. <https://doi.org/10.4324/9781351218788>

Wodtke, G. T., Yildirim, U., Harding, D. J., & Elwert, F. (2023). Are Neighborhood Effects Explained by Differences in School Quality? *American Journal of Sociology*, 128(5), 1472–1528. <https://doi.org/10.1086/724279>

- Yigiter, M. S. (2025). The effect of socioeconomic status on academic achievement: A big data study across countries and time with integrative data analysis. *PLOS ONE*, 20(10), e0335485. <https://doi.org/10.1371/journal.pone.0335485>
- Younger, M., & Warrington, M. (2005). *Raising boys' achievement* [University of Cambridge Faculty of Education]. DFES Publications.
- Yu, J., McLellan, R., & Winter, L. (2021). Which Boys and Which Girls Are Falling Behind? Linking Adolescents' Gender Role Profiles to Motivation, Engagement, and Achievement. *Journal of Youth and Adolescence*, 50(2), 336–352. <https://doi.org/10.1007/s10964-020-01293-z>
- Zahl-Thanem, A., & Rye, J. F. (2024). Spatial inequality in higher education: A growing urban–rural educational gap? *European Sociological Review*, 40(6), 1067–1081. <https://doi.org/10.1093/esr/jcae015>
- Zapfe, L., & Gross, C. (2021). How do characteristics of educational systems shape educational inequalities? Results from a systematic review. *International Journal of Educational Research*, 109, 101837. <https://doi.org/10.1016/j.ijer.2021.101837>
- Zelazo, P. D., & Carlson, S. M. (2012). Hot and Cool Executive Function in Childhood and Adolescence: Development and Plasticity. *Child Development Perspectives*, 6(4), 354–360. <https://doi.org/10.1111/j.1750-8606.2012.00246.x>

Published by the Centre for Policy Research on Men and Boys (CPRMB),
September 2026

E: Contact@menandboys.org.uk

W: <https://www.menandboys.org.uk>

LinkedIn: [centre-policy-research-menboys](#)

Instagram: [cprmenandboys](#)

X: [CPRMenBoys](#)

The logo consists of a white rectangular box with a green border. Inside the box, the text "Centre for Policy Research on" is in a smaller, dark green font, and "Men and Boys" is in a larger, bold, dark green font.

Centre for Policy Research on
Men and Boys

The contents and opinions expressed in this paper are those of the authors only.

Address: Centre for Policy Research on Men and Boys
2 Dean Trench Street, Westminster, London, SW1P 3HE

© CPRMB 2026

Publication Number: 05/26