Life Course Centre

WORKING PAPER SERIES

No. 2026-06

March 2026

Socioeconomic status and gender gaps in educational outcomes across the life course

New distributional evidence

Ha Nguyen

Bruce Chapman

Huong Le

Heather Royer

Lorraine Dearden

Francis Mitrou

The Australian Research Council Centre of Excellence for Children and Families over the Life Course

Phone +61 7 3346 7477 **Email** lcc@uq.edu.au

lifecoursecentre.org.au

Australian Government
Australian Research Council

The University Of Queensland
AUSTRALIA

The University of Sydney

The University of Western Australia

The University of Melbourne

Research Summary

Why was the research done?

This study is the first to leverage whole-of-population linked census-administrative data to examine gender gaps in educational outcomes from early primary school through early adulthood in Australia and to assess the contribution of socioeconomic factors to these gaps.

What were the key findings?

We find that females outperform males from as early as ages 5–6 across multiple developmental domains, and this advantage persists through university. The gender gap in favour of females is larger among lower-performing students. These findings are robust across population-wide analyses as well as sibling- and twin-based designs.

We also find that boys benefit more than girls from growing up in more advantaged families, particularly among academically lower-performing boys. However, this advantage is observed only for outcomes measured in the early years of primary school. By contrast, for outcomes measured at the tertiary level, most indicators of socioeconomic advantage confer stronger benefits to females, especially among individuals at the lower end of the educational attainment distribution. Finally, we identify gender differences across siblings in school sector choice and early childhood health conditions, both favouring females, as potential mechanisms underlying these patterns.

What does this mean for policy and practice?

This study demonstrates that gender differences in various socioeconomic factors help explain observed gaps in educational outcomes, including the reversal of their effects over an extended and critical developmental period. These findings improve our understanding of gender-based developmental disparities and provide evidence to guide interventions aimed at promoting better outcomes for both genders.

Citation

Nguyen, H., Chapman, B., Le, H., Royer, H., Dearden, L., & Mitrou, F. (2026). 'Socioeconomic status and gender gaps in educational outcomes across the life course: New distributional evidence', Life Course Centre Working Paper Series, 2026-06. Institute for Social Science Research, The University of Queensland.

The authors

Ha Nguyen
The Kids Research Institute Australia
Email: ha.nguyen@thekids.org.au

Bruce Chapman
The Australian National University

Huong Le
The Kids Research Institute Australia & The University of Western Australia
Email: huong.le@thekids.org.au

Heather Royer
University of California, Santa Barbara

Lorraine Dearden
University College London

Francis Mitrou
The Kids Research Institute Australia & The University of Western Australia
Email: francis.mitrou@thekids.org.au

Acknowledgements/Funding Sources

We appreciate the support of the Australian Bureau of Statistics, especially Robyn Berkovits, Caroline Blink, Lisa Commens, Brendan Harvey, Michael Rodriguez, Andrew Paull, Talei Parker, and Lei Yu, for their assistance with accessing the PLIDA data. We also thank the Australian Early Development Census (AEDC), particularly Alina Batoni, for reviewing the AEDC data. This research was partly funded by the Australian Research Council Centre of Excellence for Children and Families over the Life Course (CE140100027). Ha Nguyen's research is partly supported by the Australian Government Medical Research Future Fund (ID: 2024/MRF2036125).

Disclaimer: This study uses data from the Australian Early Development Census (AEDC). The AEDC is funded by the Australian Government Department of Education. The findings and/or views reported are those of the authors and should not be attributed to the Department or the Australian Government.

This paper uses unit record data from the Person Level Integrated Data Asset (PLIDA). These data are proprietary and researchers wishing to use them must seek approval from the relevant institutions. The views expressed here are those of the authors and do not represent the views of any data custodian or funder.

We acknowledge the Traditional Custodians of the lands on which we work and live across Australia. We pay our respects to Elders past and present and recognise their continued connections to land, sea and community.

Socioeconomic status and gender gaps in educational outcomes across the life course: New distributional evidence

Ha Nguyen[*, †] Bruce Chapman[‡] Huong Le[†] Heather Royer[§]
Lorraine Dearden[⁂] Francis Mitrou[†]

This study leverages whole-of-population linked census-administrative data to examine gender gaps in educational outcomes from early primary school through early adulthood in Australia and to assess the contribution of socioeconomic factors to these gaps. We find that females outperform males from as early as ages 5–6 across multiple developmental domains, and this advantage persists through university. The gender gap in favour of females is larger among lower-performing students. These findings are robust across population-wide analyses as well as sibling- and twin-based designs. We also find that boys benefit more than girls from growing up in more advantaged families, particularly among academically lower-performing boys. However, this advantage is observed only for outcomes measured in the early years of primary school. By contrast, for outcomes measured at the tertiary level, most indicators of socioeconomic advantage confer stronger benefits to females, especially among individuals at the lower end of the educational attainment distribution. Finally, we identify gender differences across siblings in school sector choice and early childhood health conditions, both favouring females, as potential mechanisms underlying these patterns.

Keywords: Education; Gender Gap; Socioeconomic Status; Administrative data; Census; Australia.

JEL classifications: I24; J13; J15; J16; J62; R23

[*] **Corresponding author**: The Kids Research Institute Australia | Postal: GPO Box 855, Perth WA 6872, Australia | Email: ha.nguyen@thekids.org.au.
[†] The Kids Research Institute Australia & The University of Western Australia. [‡] The Australian National University. [§] University of California, Santa Barbara. [⁂] University College London.

Acknowledgements: We appreciate the support of the Australian Bureau of Statistics, especially Robyn Berkovits, Caroline Blink, Lisa Commens, Brendan Harvey, Michael Rodriguez, Andrew Paull, Talei Parker, and Lei Yu, for their assistance with accessing the PLIDA data. We also thank the Australian Early Development Census (AEDC), particularly Alina Batoni, for reviewing the AEDC data. This research was partly funded by the Australian Research Council Centre of Excellence for Children and Families over the Life Course (CE140100027). Ha Nguyen's research is partly supported by the Australian Government Medical Research Future Fund (ID: 2024/MRF2036125).

Disclaimer: This study uses data from the Australian Early Development Census (AEDC). The AEDC is funded by the Australian Government Department of Education. The findings and/or views reported are those of the authors and should not be attributed to the Department or the Australian Government.
This paper uses unit record data from the Person Level Integrated Data Asset (PLIDA). These data are proprietary and researchers wishing to use them must seek approval from the relevant institutions. The views expressed here are those of the authors and do not represent the views of any data custodian or funder.

1. Introduction

An emerging global phenomenon is the rise in women's educational attainment relative to men, with females now achieving higher levels of educational achievement and attainment than males in many countries (Delaney & Devereux 2021; OECD 2025). Given the well-documented and wide-ranging benefits of education for life outcomes (Blundell *et al.* 1999; Clark & Royer 2013; Galama *et al.* 2018; Gunderson & Oreopolous 2020), it is unsurprising that a large body of research has examined the factors contributing to gender gaps in education (see, for example, Delaney and Devereux (2021), for a review, and Section 2 for a brief overview of the literature).

Within this literature, a growing number of studies examines the role of family socioeconomic status in shaping gender gaps in educational outcomes (Bertrand & Pan 2013; Brenøe & Lundberg 2018; David *et al.* 2019; Autor *et al.* 2023; Paterson *et al.* 2024). However, the empirical evidence remains mixed, and it is still unclear how family socioeconomic status influences gender gaps beyond the school years or how this relationship varies across children with different levels of academic performance. This inconclusive evidence, together with the growing policy and academic emphasis on gender equality across a range of outcomes (Blau & Kahn 2017; Blundell *et al.* 2025), underscores the need for further investigation using newly available data and more rigorous empirical approaches (Delaney & Devereux 2021).

This study contributes to the literature by, for the first time, using recently available Australian population-wide linked census–administrative data from the Person Level Integrated Data Asset (PLIDA) to examine gender gaps in educational outcomes and the role of socioeconomic factors from early primary school through early adulthood in Australia. These unique data, together with our empirical approach, enable three important contributions. First, the richness and scale of the data enable us to implement the most rigorous empirical approaches used in this literature, including sibling comparisons with family fixed effects (FE), thereby addressing concerns regarding the potential endogeneity of gender and socioeconomic variables in the educational outcome equation (Brenøe & Lundberg 2018; David *et al.* 2019; Autor *et al.* 2023). Moreover, where relevant, we advance the literature by exploiting twin data and applying twin fixed effects—an approach not previously used in this context—to examine gender gaps in educational outcomes and their underlying mechanisms.

Second, this study examines educational outcomes over a longer span of the life course—from the first year of primary school through early adulthood—and incorporates a more

comprehensive set of socioeconomic factors than most prior studies. Whereas much of the existing literature focuses on gender gaps during the school years, with the notable exception of Brenøe and Lundberg (2018), who consider educational outcomes at the secondary and tertiary levels, our approach provides new evidence on the heterogeneous effects of specific socioeconomic dimensions on gender gaps in education and on how these effects evolve over this extended and important developmental epoch.

Third, this study is among the first—following Autor *et al.* (2023)—to apply quantile regression methods to examine how socioeconomic factors contribute to gender gaps across the distribution of educational outcomes, offering new and policy-relevant insights into distributional heterogeneity.

This study presents four main findings. First, we document a gender gap favouring females that emerges as early as ages 5–6 across multiple developmental outcomes and persists into tertiary education, where females achieve higher Australian Tertiary Admission Rank (ATAR) scores and have a higher probability of attaining a higher education qualification. These gender gaps are broadly similar in population-wide analyses and in sibling- and twin-based analyses, suggesting that child gender is unlikely to be endogenous in the educational outcomes equation. Second, we find that the female advantage is most pronounced at the lower end of the educational attainment distribution.

Third, we show that socioeconomic advantage—measured by higher maternal or paternal education, home ownership, and attendance at higher-quality schools, but not household income or neighbourhood quality—disproportionately benefits males, particularly those at the lower tail of the distribution. However, this pattern is observed only for developmental outcomes measured in early primary school. For educational outcomes measured at the tertiary level, most indicators of socioeconomic advantage—particularly higher maternal education, greater household income, and home ownership—confer stronger benefits to females, especially those at the lower end of the educational attainment distribution. Fourth, the study identifies gender differences across siblings in school sector choice and early childhood health conditions—both favouring females—as potential mechanisms underlying the documented patterns.

The remainder of the paper is organised as follows. Section 2 briefly reviews the literature, and Section 3 describes the data and sample. Section 4 outlines the empirical models used to examine gender differences in educational outcomes and the differential effects of

socioeconomic conditions on gender gaps. Section 5 presents the main empirical results. Section 6 reports robustness checks and further investigates potential mechanisms, and Section 7 concludes.

2. A brief literature review

This study contributes to a large literature examining gender gaps in educational outcomes and the factors underlying these gaps (see, for example, Delaney and Devereux (2021), for a review). This literature consistently documents that, in many countries, females tend to exhibit higher educational attainment and achievement than males across most educational and developmental outcomes, with mathematics being a notable exception (Chapman & Ryan 2005; Fryer & Levitt 2010; Riphahn & Schwientek 2015; Cobb-Clark & Moschion 2017; Le & Nguyen 2018a; Cuff *et al.* 2025; OECD 2025).[1]

Within this broader literature, the present study is more closely related to research that seeks to identify potential mechanisms underlying gender gaps in educational achievement. Previously identified mechanisms include gender differences in non-cognitive skills (Jacob 2002; Gneezy *et al.* 2003; Duckworth & Seligman 2006; Christopher *et al.* 2013; Golsteyn & Schils 2014), willingness to guess during tests (Baldiga 2014; Iriberri & Rey-Biel 2021), performance under test pressure (Niederle & Vesterlund 2010; Cai *et al.* 2019; Montolio & Taberner 2021), teacher gender bias in grading (Lavy & Sand 2018; Carlana 2019), peer and school environment effects (Booth *et al.* 2018; Dustmann *et al.* 2018; Barrios-Fernández & Riudavets-Barcons 2024; García-Echalar *et al.* 2024), risk preferences (Croson & Gneezy 2009; Marianne 2011; Saygin & Atwater 2021), and time allocation, with females spending more time on activities that promote skill development (Nguyen *et al.* 2022).[2]

Among studies investigating the sources of gender differences in educational outcomes, this paper is most closely related to a growing body of research examining the effects of socioeconomic conditions on gender gaps in developmental outcomes. Buchmann and DiPrete (2006) show that, in the United States (U.S.), the female advantage in college completion is

[1] This pattern—often referred to as the "boy problem"—contrasts with a related and well-documented gender gap in fields of study, whereby females are more likely to choose lower-paying fields of specialization, sometimes labelled the "girl problem" (Chapman & Mulvey 1986; Francesconi & Parey 2018; Kahn & Ginther 2018; Landaud *et al.* 2020; Kugler *et al.* 2021; Delaney & Devereux 2025).

[2] While many of these mechanisms have been shown to be correlated with educational outcomes, data limitations—most notably the lack of datasets that jointly observe both the proposed mechanisms and educational outcomes—and identification challenges make it difficult to causally quantify the contribution of these factors to gender gaps in educational outcomes (Delaney & Devereux 2021). This study is also related to a large body of research examining gender gaps in other outcomes, particularly labour market outcomes (Blau & Kahn 2017), as well as several of the mechanisms discussed above (Goldin 2014).

largest among individuals from families with a low-educated or absent father. Bertrand and Pan (2013), also using U.S. data, show that family disadvantage—measured by single-mother households, low socioeconomic status, and young maternal age at childbirth—but not adverse school environments, disproportionately reduces the non-cognitive skills of boys relative to girls. Similarly, using administrative data from all public schools in Florida (U.S.), David *et al.* (2019) find that family disadvantage—measured by maternal education, marital status, Medicaid birth status, maternal age at birth, and lower school and neighbourhood quality—disproportionately impedes the educational and behavioural development of school-aged boys. Using the same dataset and applying quantile regression methods, Autor *et al.* (2023) further show that family socioeconomic status particularly affects boys' outcomes relative to girls' at the lower tails of the outcome distribution.

Evidence from outside the U.S. is more mixed. Using Danish administrative data, Brenøe and Lundberg (2018) show that boys benefit more than girls from an advantaged family environment—measured by parental education and marital status—at the grade-school level. In contrast, they document a markedly different pattern at the tertiary level, with maternal education benefiting daughters more strongly and paternal education benefiting sons.

Similarly, using survey data from the Longitudinal Study of Australian Children (LSAC) and focusing on numeracy—the only cognitive domain in which males outperform females in Australia at school age (Cobb-Clark & Moschion 2017; Le & Nguyen 2018a; Nguyen *et al.* 2022), Paterson *et al.* (2024) show that boys from lower socioeconomic backgrounds are more disadvantaged when socioeconomic status is measured using a composite index constructed from household income, parental education, and maternal labour force participation. However, when examining these components separately, they find mixed evidence regarding the role of individual socioeconomic factors in shaping gender gaps in numeracy at the primary and secondary school levels.[3]

[3] Three other Australian studies (Cobb-Clark & Moschion 2017; Le & Nguyen 2018a; Nguyen *et al.* 2022), which also draw on the LSAC to examine gender gaps in cognitive and non-cognitive outcomes among school-aged children, typically measure cognitive outcomes using National Assessment Program – Literacy and Numeracy (NAPLAN) test results administered in Years 3, 5, 7, and 9. These studies incorporate rich measures of family background—including maternal and paternal education and household income, similar to Paterson *et al.* (2024)—but do not focus explicitly on the differential effects of family background on gender gaps. Nevertheless, their reported Oaxaca–Blinder decompositions suggest that evidence on differential returns to family background is inconclusive and varies across developmental outcomes. Moreover, these studies examine outcome measures that differ from those considered in the present study and rely on substantially smaller samples. The LSAC began with approximately 5,000 observations per cohort, with effective sample sizes further reduced by attrition and missing data, thereby limiting statistical power relative to the population-wide administrative data used in this analysis. Importantly, the sampling design of the LSAC, which surveys only one child per family, does not allow

One major empirical challenge in identifying the robust relationship between socioeconomic status and the gender gap in educational outcomes is the potential endogeneity of socioeconomic variables. These variables are likely correlated with unobservable factors in the educational outcome equation, and failing to account for this endogeneity may result in biased estimates of the effect of socioeconomic status on the gender gap (Wooldridge 2010). To address this concern, studies in this literature have mainly employed two approaches. The first, and more commonly used, approach is to control for a rich set of explanatory variables (Bertrand & Pan 2013; Nguyen *et al.* 2022; Paterson *et al.* 2024). The second, arguably more robust, approach is to exploit sibling data and employ family fixed effects, which control for time-invariant unobservable factors within the same family. This approach is more data-intensive, as it requires substantial samples of siblings with variation in socioeconomic background over time, and consequently, only a few studies have been able to implement it, typically as a robustness check (Brenøe & Lundberg 2018; David *et al.* 2019; Autor *et al.* 2023).

A related concern in identifying the gender gap is the potential endogeneity of the gender variable itself, as gender may be correlated with unobserved factors in the educational outcome equation. For example, if parents have a son preference and continue having children until a son is born, girls may grow up in larger families (Almond & Edlund 2007; Dahl & Moretti 2008; Dossi *et al.* 2021). Such gender preferences can bias estimates of the gender gap in educational outcomes. Similar to approaches addressing the endogeneity of socioeconomic variables, this concern is typically addressed by using sibling data and family fixed effects, which account for time-invariant factors such as parental gender preferences (Brenøe & Lundberg 2018; David *et al.* 2019; Autor *et al.* 2023).

Overall, the brief review of the literature suggests mixed evidence regarding the contribution of socioeconomic status to gender gaps in educational outcomes. This heterogeneity likely reflects a range of factors, including differences in data sources, institutional and policy contexts, measures of educational outcomes, the age at which outcomes are observed, definitions of socioeconomic status, and empirical methods. These inconsistencies, together with the growing policy and academic emphasis on gender equality across a range of outcomes—including those closely linked to education, such as labour market outcomes (Blau

for the use of family or sibling fixed effects regressions, as employed in this study. Other Australian studies examining educational or developmental outcomes among young Australians generally include gender and socioeconomic variables in their empirical models but do not focus explicitly on gender gaps in educational outcomes, as the present study does (Cobb-Clark & Nguyen 2012; Nguyen *et al.* 2020; Nguyen *et al.* 2025).

& Kahn 2017; Blundell *et al.* 2025)—underscore the need for further analyses using newly available data and more rigorous empirical approaches (Delaney & Devereux 2021). In line with this motivation, the present study contributes by using population-wide linked data from the PLIDA to examine gender gaps in educational outcomes from early primary school through early adulthood in Australia, and the role of multiple socioeconomic factors in shaping these gaps.

3. Data and sample

3.1. Data

This study employs data from the Person Level Integrated Data Asset (PLIDA), developed and maintained by the Australian Bureau of Statistics (ABS 2026). PLIDA is a secure and comprehensive integrated data asset that combines longitudinal information across multiple domains, including education, government payments, income, health, and population demographics.

The analysis draws on three datasets. The first is the 2011 Census, which provides detailed socio-demographic information, including gender, age, educational attainment, family relationships, and household characteristics. Family relationship information reported in the 2011 Census is used to identify links between children and their co-residing parents, as well as relationships with other co-residing siblings, including twins. Based on these derived parent–child pairs, we construct parental characteristics such as age, migration background, and educational attainment. In addition, the Census provides household-level attributes, including income and tenure status, which serve as key explanatory variables in the empirical models. Residential location information from the 2011 Census is also used to link individuals to area-level Socio-Economic Indexes for Areas (SEIFA), capturing neighbourhood socioeconomic conditions.

The second dataset is the Australian Early Development Census (AEDC), previously referred to as the Australian Early Development Index (AEDI). The AEDC is a nationwide collection that assesses early childhood development at the point when children commence their first year of full-time schooling (AEDC 2025). First implemented in 2009, the AEDC has been administered on a triennial basis. Data are collected using the AEDC Instrument, which is built upon the Australian version of the Early Development Instrument (AvEDI). The AvEDI adapts the Early Development Instrument developed by the Offord Centre at McMaster University (Canada) to the Australian context. Teachers complete the AvEDI based on structured

observations of children in their class, generating information across five domains of early childhood development: physical health and wellbeing, social competence, emotional maturity, language and cognitive skills (school-based), and communication skills and general knowledge (Howells *et al.* 2024). Domain-specific responses are aggregated into scores ranging from 0 to 10, with higher values indicating more advanced developmental outcomes.

The third dataset is the Higher Education Information Management System (HEIMS), an administrative dataset provided by the Department of Education, Skills, and Employment. HEIMS contains nationwide information on higher education enrolments and completions from 2005 to 2021. From this dataset, we construct two outcome variables: Australian Tertiary Admission Rank (ATAR) scores and indicators of higher education qualification completion. In Australia, domestic students seeking admission to university typically require an ATAR, a percentile ranking ranging from 0.00 to 99.95, which is awarded upon completion of Year 12, the final year of secondary schooling. Most students are aged 17 or 18 at this stage (Nguyen *et al.* 2025). In HEIMS, ATAR scores are observed only for individuals who enrol in higher education and are therefore recorded in the system; these scores are bottom-coded, with values below 30 recorded as 30. To facilitate interpretation, all continuous developmental and academic outcomes used in the analysis are standardised to have a mean of zero and a standard deviation of one within the population-wide analytical sample, defined as individuals aged 18 years or younger at the time of the 2011 Census who were co-residing with both parents and had a valid observation for the relevant outcome (see below for further details on sample construction).

These datasets are linked through the Person Linkage Spine, a system developed to integrate records for individuals residing in Australia during the relevant reference periods. The linkage is performed using deterministic methods based on personal identifiers, including name, date of birth, gender, and address (ABS 2026).

3.2. Sample

This study investigates how family socioeconomic circumstances shape educational outcomes, which necessitates the use of detailed and reliable measures of family background. Accordingly, the baseline analytical population of interest is restricted to individuals aged 18 years or younger at the time of the 2011 Census. This restriction is imposed for two main reasons.

First, identification of family linkages relies on family relationship information recorded in the 2011 Census. In particular, parental characteristics can be observed only for individuals who were living with both biological or adoptive parents at the time of enumeration. As a result, parent–child linkages are feasible only for children who were co-residing with both parents in 2011. This design enables the direct observation of parental age, migration status, and educational attainment as self-reported by parents, thereby allowing a more accurate assessment of how these characteristics are associated with children's subsequent educational outcomes.

Second, limiting the sample to individuals aged 18 years or younger—an age group in which the majority remain in the parental household—reduces bias arising from selective residential independence among older children. Transitions out of the parental home for education-related reasons may occur at different rates by gender and could otherwise confound observed gender differences in educational trajectories.

In addition, we employ outcome-specific analytical samples. For analyses of AEDC outcomes, the sample is restricted to individuals aged 10 years or younger at the time of the 2011 Census who can be linked to at least one AEDC wave conducted in 2009, 2012, or 2015. This age restriction is chosen to align with the design of the AEDC, which is administered during the first year of primary school, when children are typically aged 5 or 6 years. Not all individuals in the cohort have AEDC observations, as the AEDC is conducted triennially and includes only children who are enrolled in their first year of full-time schooling in an AEDC collection year (AEDC 2025).

For educational outcomes measured at the post-secondary level and beyond, we restrict the sample to individuals who were aged 10 to 18 years at the time of the 2011 Census. This age range allows longitudinal observation through to the potential attainment of ATAR scores and the completion of higher education within the coverage period of the higher education dataset, which extends to 2021. As ATAR scores are awarded at the completion of Year 12, when students are typically aged 17 or 18, and a standard undergraduate degree generally requires approximately four years to complete, individuals in this cohort would be between 20 and 28 years old by 2021. This timeframe is therefore sufficient to capture both higher education participation and completion outcomes.

We further refine the analytical sample by excluding individuals with missing information on key variables, including age and educational attainment of both parents (see the following

sections for detailed variable definitions). We restrict the sample to individuals co-residing with both parents in order to observe the characteristics of both mothers and fathers. Although this more restrictive sampling strategy does not allow us to examine the effects of single-parent status on gender gaps, as in prior U.S. studies (Bertrand & Pan 2013; David *et al.* 2019; Autor *et al.* 2023), it is feasible given the large underlying population datasets and enables a more granular analysis of heterogeneity in parental background—particularly parental educational attainment—and its association with individuals' educational outcomes. This approach is central to examining the potential moderating role of parental characteristics in the relationship between socioeconomic background and educational attainment, and is consistent with the literature on the intergenerational transmission of advantage, particularly with respect to their potentially differential effects on gender gaps in children's educational outcomes (Bertrand & Pan 2013; Figlio *et al.* 2016; Brenøe & Lundberg 2018; David *et al.* 2019; Mogstad & Torsvik 2023).

Applying these sample restrictions to the full population yields over 350,000 observations for the AEDC outcomes, approximately 330,000 observations for ATAR outcomes, and around 810,000 observations for higher education qualification attainment. In selected analyses, we further restrict the sample to sibling or twin subsamples drawn from these population-wide cohorts. Siblings are defined as individuals who share the same mother and father, while twins are identified as siblings born in the same month and year within the same family.[4] Sample sizes for the sibling and twin analyses vary by outcome and specification and are reported where relevant.

4. Empirical models

To examine gender differences in a given educational outcome Y, we estimate the following regression model:

$$Y_{i,t=2011+k} = \alpha + Male_i\beta + X_{i(t=2011)}\gamma + \varepsilon_i \quad (1)$$

where $Male_i$ is an indicator equal to one if individual i is male, and $X_{i(t=2011)}$ denotes a vector of explanatory variables measured at the time of the 2011 Census. These covariates are measured at baseline for two main reasons. First, many are constructed from Census data and

[4] The data do not permit separate identification of biological versus adoptive children, nor do they allow identification of siblings who were not co-residing at the time of the 2011 Census. In addition, it is not possible to distinguish between dizygotic and monozygotic twins, which limits our ability to fully disentangle genetic influences from shared family environmental factors, as would be possible in analyses restricted to monozygotic twins (Royer 2009; Bhalotra & Clarke 2023). Moreover, month and year of birth are the most detailed date-of-birth information available in PLIDA (ABS 2026).

are not consistently available in other administrative sources. Second, for time-varying characteristics such as household income, baseline measurement helps mitigate —although does not entirely eliminate—concerns about reverse causality; for example, that children's educational outcomes may influence parental labour supply or parenting behaviours, which in turn affect household income.

As documented in Session 3, our main analysis focuses on three sets of outcomes. The first comprises five indicators of early childhood development from the AEDC. The remaining two outcomes relate to later educational attainment: ATAR scores and completion of a higher education qualification.

Outcomes are observed at different points in time, denoted by $t = 2011 + k$, where $k = -2$ corresponds to AEDC outcomes measured in 2009, $k = 1$ to AEDC outcomes measured in 2012, $k = 4$ to AEDC outcomes measured in 2015, and $k = 10$ to higher education outcomes measured at the end of the observation window in 2021. As noted above, most outcomes are measured after the baseline covariates are observed, reflecting data availability and further reducing concerns related to reverse causality. In Equation (1), ε_i denotes the idiosyncratic error term, and α, β and γ are parameters to be estimated.

Following the approach commonly adopted in the literature on gender gaps in developmental and educational outcomes (Bertrand & Pan 2013; Figlio *et al.* 2016; Brenøe & Lundberg 2018; David *et al.* 2019), the vector X_i includes a rich set of individual- and household-level characteristics that may be correlated with educational outcomes. These covariates include individual attributes (such as migration background and age, along with its square), parental characteristics (including age and age squared, migration background, and highest educational attainment for both mothers and fathers), and household characteristics (such as household income[5] and housing tenure). State and territory fixed effects are also included to account for geographic and institutional differences across jurisdictions.

In addition to these individual- and household-level controls, we incorporate two measures that proxy for socioeconomic background at the school and neighbourhood levels. To capture school-level socioeconomic conditions, and following an approach used in related U.S. studies (Figlio *et al.* 2016; David *et al.* 2019; Autor *et al.* 2023), we exploit the population-wide AEDC data to construct a school-level measure defined as the average AEDC score across all five

[5] Household income in the Census is reported in income brackets. For ease of presentation and interpretation, we follow standard practice by treating income as a continuous variable and assigning the midpoint of each bracket.

domains for each school in a given year.[6] This school-level average is then assigned to each student attending that school in the corresponding year. By construction, schools with higher average AEDC scores are interpreted as serving populations with more advantaged socioeconomic backgrounds (Nghiem *et al.* 2015; Nghiem *et al.* 2016; Le *et al.* 2025). This measure is included only in regressions of AEDC outcomes, as it is derived directly from AEDC data. In addition, to account for potential gender differences in the age at school entry (McCrary & Royer 2011; Crawford *et al.* 2014; Hanly *et al.* 2019), we additionally control for age and age squared at the time of AEDC assessment in all regressions of AEDC outcomes.

To capture neighbourhood-level socioeconomic conditions, we use the Socio-Economic Indexes for Areas (SEIFA), linked to individuals' residential location at the 2011 Census at the Statistical Area Level 2 (SA2).[7] Measuring neighbourhood characteristics during childhood allows for a more credible assessment of neighbourhood effects on developmental outcomes, while reducing concerns related to endogenous residential sorting that may be more pronounced in adulthood. This approach is consistent with the neighbourhood-effects literature, which emphasises early-life exposure as a key determinant of later outcomes (Chetty & Hendren 2018; Chetty *et al.* 2020).

In Equation (1), the coefficient β is of primary interest, as it captures gender differences in the outcome of interest. A positive estimate of β indicates a male advantage in the relevant educational outcome, whereas a negative estimate implies a female advantage.

Consistent with the existing literature, we exploit the richness of the linked Census–administrative data to control for a comprehensive set of covariates commonly associated with educational outcomes, thereby reducing the influence of omitted variable bias on the estimated gender gap. Several of these covariates have not been jointly considered in prior studies in the international literature. For example, U.S.-based studies (Bertrand & Pan 2013; Figlio *et al.* 2016; David *et al.* 2019; Autor *et al.* 2023) typically lack information on paternal education. While Brenøe and Lundberg (2018), using Danish administrative data, examine the separate roles of maternal and paternal education as well as marital status at birth, they are unable to

[6] Unfortunately, the available data do not allow us to use gains in test scores as a proxy for school-level socioeconomic background, as has been done in prior U.S. studies (Figlio *et al.* 2016; David *et al.* 2019; Autor *et al.* 2023).

[7] Specifically, we employ the Index of Relative Socioeconomic Disadvantage (IRSD), an area-level composite measure produced by the Australian Bureau of Statistics that captures relative disadvantage based on a range of socioeconomic characteristics, including income, educational attainment, employment status, and housing conditions, with higher values indicating more advantaged areas. For presentation purposes, the IRSD score is scaled by a factor of 1/100.

incorporate other key socioeconomic dimensions, such as household income, home ownership, and measures of school and neighbourhood quality, which are included in the present study.

In addition, we exploit the panel structure of the linked datasets to measure selected time-varying explanatory variables at the beginning of the study period, thereby mitigating concerns about reverse causality. By combining baseline measurement of key covariates with a rich set of controls, the analysis provides robust evidence on gender gaps in educational outcomes conditional on observable characteristics. Nevertheless, as in most studies examining differential relationships between socioeconomic status and gender gaps in educational and developmental outcomes (Bertrand & Pan 2013; Figlio *et al.* 2016; Brenøe & Lundberg 2018; David *et al.* 2019; Paterson *et al.* 2024), the coefficients on time-varying socioeconomic variables in Equation (1) should not be interpreted as causal, as these variables may be correlated with unobserved factors captured by the error term ε_i.

To further strengthen the analysis, and consistent with several international studies in this literature (Brenøe & Lundberg 2018; David *et al.* 2019; Autor *et al.* 2023), we conduct additional robustness checks using sibling samples and family fixed-effects models, which control for unobserved time-invariant characteristics shared within families that may be correlated with both time-varying covariates and educational outcomes.[8] Moreover, to assess potential endogeneity of the gender variable in the educational outcome equations, we exploit twin samples—for the first time in this literature—and apply family fixed-effects regressions. This approach provides a more robust analysis of gender gaps in educational outcomes, as a child's gender within twin pairs is plausibly exogenous relative to gender variation among non-twin siblings (Bhalotra & Clarke 2023).

To examine whether socioeconomic background variables differentially influence gender gaps in educational outcomes, we extend Equation (1) by incorporating interaction terms between the $Male_i$ indicator and selected socioeconomic characteristics, following standard practice in the literature (Brenøe & Lundberg 2018; David *et al.* 2019; Autor *et al.* 2023; Paterson *et al.*

[8] In the family or twin fixed-effects regressions, all time-invariant variables—including most socioeconomic characteristics measured in the 2011 Census—are absorbed by the fixed effects and therefore cannot be included as regressors. In addition, these specifications substantially reduce the sample size, as they require at least two siblings with valid information per family. Moreover, as discussed below, it is not possible to introduce interaction terms between gender, which is time-invariant, and other time-invariant variables, because the effects of both time-invariant variables (i.e., gender and the socioeconomic characteristics) cannot be identified simultaneously. Consequently, as in most prior studies in this literature (David *et al.* 2019; Autor *et al.* 2023), we primarily use family fixed-effects regressions as robustness checks. Moreover, while family FE regressions appear to be the most robust specification in this literature, they are not without limitations. One potential concern—also applicable to the twin FE approach—is that siblings may compete for resources, which may complicate the interpretation of the estimates for some included socioeconomic variables.

2024). Several of these characteristics are also included as main effects in the baseline specification. The coefficients on these interaction terms are of primary interest, as they indicate whether the association between a given socioeconomic factor and educational outcomes differs by gender.

We focus on six key dimensions of socioeconomic background: maternal education, paternal education, household income, home ownership status, school quality, and neighbourhood quality (measured by SEIFA). Rather than summarising these dimensions into a single composite index—such as one derived from principal components analysis, as in some previous studies (David *et al.* 2019; Autor *et al.* 2023)—we examine each factor separately by interacting the $Male_i$ indicator with each socioeconomic variable. This approach, made feasible by the large sample size and the richness of the linked data, allows for a more transparent and nuanced assessment of the contribution of individual socioeconomic factors to the gender gap. In particular, it facilitates an assessment of their relative importance, for example by distinguishing the potentially differential roles of maternal and paternal education, as well as of household income and home ownership—with the latter serving as a proxy for household wealth—in shaping gender disparities in educational outcomes.

We begin by estimating Ordinary Least Squares (OLS) regressions to examine the determinants of educational outcomes at the mean. For continuous outcomes, such as AEDC domain scores and ATAR results, we additionally explore heterogeneity in the gender gap and its contributing factors across the outcome distribution. Following prior studies on gender gaps in educational and developmental outcomes (Le & Nguyen 2018a; Nguyen *et al.* 2022; Autor *et al.* 2023), we employ an unconditional quantile regression (UQR) framework to estimate Equation (1) and its augmented specification at selected points of the outcome distribution.[9]

Specifically, we apply the UQR approach proposed by Firpo *et al.* (2009), which allows for the estimation of the marginal effects of explanatory variables on unconditional quantiles of the outcome variable without requiring the rank-preservation assumption inherent in conditional

[9] We focus on selected quantiles of the outcome distribution—the 20th, 40th, 60th, and 80th percentiles—rather than estimating effects over a finer grid. This choice reflects the nature of several educational outcomes considered in the analysis, particularly the AEDC measures, which are designed to identify children at risk of developmental vulnerability (AEDC 2025). Consequently, although the outcome distributions are sufficiently rich to support quantile regression, AEDC measures exhibit discreteness and mass points in the upper tail of the distribution, causing multiple upper quantiles to coincide. In such settings, quantile regression yields identical estimates across adjacent quantiles. Restricting attention to selected percentiles therefore avoids redundancy and facilitates clearer interpretation of distributional patterns.

quantile regression.[10] This feature makes the method particularly well suited for analysing distributional effects in educational outcomes.

5. Empirical results

5.1. Descriptive results

Table 1 reports the mean statistics of key explanatory and outcome variables used in this study by child gender, revealing two salient patterns. First, explanatory variables capturing children's characteristics, along with parental and household attributes and neighbourhood socioeconomic characteristics, are remarkably similar across male and female subgroups. This similarity is corroborated by the test results in Column 3, which show no statistically significant differences between genders. There are, however, two notable exceptions. (1) Male children in the sample are older than female children both at the time of the 2011 Census and at the time of the AEDC assessment, consistent with existing evidence that boys are more likely than girls to enrol in school at a later age (Hanly *et al.* 2019), as well as with the fact that our descriptive sample includes only individuals who enrolled in primary school and have a valid AEDC measure. (2) Fathers of sons in the sample have, on average, slightly higher educational qualifications than fathers of daughters. This is consistent with our sample restriction to children living with both parents and with U.S. evidence indicating that marriages are less likely to persist following the birth of girls than boys (Bedard & Deschênes 2005). Aside from these differences, socioeconomic characteristics are broadly comparable across genders, suggesting that child gender is unlikely to be systematically correlated with observed characteristics in this context.

Second, males attend schools with lower average AEDC scores and exhibit poorer outcomes across all five AEDC domains, as well as lower ATAR scores and a lower probability of attaining a higher education qualification.

[10] For a technical exposition of the UQR methodology, see Firpo *et al.* (2009). Autor *et al.* (2023) provide a detailed and highly relevant discussion of practical considerations when applying this approach to examine the role of socioeconomic status in shaping gender gaps in educational outcomes across the distribution. As noted by Autor *et al.* (2023), a key identifying assumption underlying this approach is that child gender is as good as randomly assigned with respect to family socioeconomic status. As shown below, this assumption is likely to hold in our context for two reasons. First, descriptive statistics reveal no systematic gender differences across most socioeconomic variables used in the analysis. Second, estimated gender gaps are broadly similar in population-wide analyses and in sibling- and twin-based specifications, suggesting that child gender is unlikely to be endogenous in the educational outcomes equation.

5.2. Regression results on the gender gaps in educational outcomes

5.2.1. Gender gaps at the mean using different samples and specifications

Table 2 reports estimates of gender gaps in multiple educational outcomes across various samples (i.e., the full population, siblings, and twins) and alternative model specifications (OLS and family fixed effects (FE)). Overall, the results indicate that, on average, females achieve higher educational outcomes than males, as evidenced by negative and statistically significant coefficients on the male indicator at the 1% level. This pattern holds across all educational outcomes, samples, and specifications considered.

For example, the OLS regression results using the full population samples (Column 1) show that, on average, females outperform males in all AEDC domains, with the largest gender gap observed in emotional maturity (0.51 standard deviations [SD]), followed by social competence (0.39 SD), communication skills and general knowledge (0.27 SD), physical health and wellbeing (0.27 SD), and language and cognitive skills (school-based) (0.23 SD). At the tertiary level, the female advantage persists but is smaller in magnitude: females exceed males by 0.09 SD in ATAR scores. Additionally, females have a 13 percentage point higher probability of attaining a higher education qualification.

The results in Table 2 show that estimates from OLS regressions using the full population sample (Column 1), as well as those based on sibling samples (Columns 2–4) and twin samples (Columns 5–7), are largely similar in both magnitude and statistical significance. Moreover, comparisons between OLS and FE regressions within the same sibling (Columns 3 and 4) or twin (Columns 6 and 7) samples yield highly consistent estimates. This consistency is particularly encouraging given the substantially smaller sample sizes for twins.

Taken together with the evidence presented in Figure 1—which shows that estimates of the gender gap, particularly for AEDC outcomes, remain largely unchanged after controlling for an extensive set of covariates, including variables that may be endogenous in the educational outcome equation—these findings suggest that child gender in this Australian context is likely exogenous. The results are consistent with evidence from sibling-based analyses in Denmark (Brenøe & Lundberg 2018) and the U.S. (David *et al.* 2019), supporting the interpretation that the estimated gender gaps in this study reflect genuine differences rather than estimation bias.

Mean regression results for other key variables are reported in Appendix Table A1 and are largely consistent with expectations. For example, children of parents with higher educational qualifications exhibit better outcomes, for both maternal and paternal education, across all

outcomes considered. While family income shows a mixed association with educational outcomes, children residing in owner-occupied homes achieve higher outcomes. Similarly, students attending schools with higher average AEDC scores tend to have higher AEDC scores. Moreover, residing in areas with higher SEIFA scores is associated with lower AEDC scores but higher ATAR scores and an increased probability of attaining a higher education qualification. Finally, as with the consistent estimates of gender gaps across different samples, the regression results for other variables are largely similar between the full population sample (reported in odd columns of Appendix Table A1) and the sibling sample (even columns), highlighting the robustness of our findings to sample choice.

5.2.2. Gender gaps along the distribution

Figure 1 reports estimates of gender gaps at selected percentiles, together with mean gender gaps, based on OLS regressions using the full population sample. In this analysis, and in subsequent exercises unless otherwise stated, we follow a common approach in the literature (David *et al.* 2019; Autor *et al.* 2023) and focus on OLS estimates from the full population sample for four main reasons. First, as shown above, estimated gender gaps are largely similar across samples and model specifications (OLS and FE). Second, using the full population sample maximises statistical power. Third, as discussed in more detail below, OLS specifications allow us to examine differential effects of time-invariant socioeconomic variables by gender, which is not feasible in FE models. Fourth, where possible, we complement these baseline results with more restrictive samples (such as sibling or twin samples) and FE regressions to demonstrate the robustness of our findings.

Figure 1 shows that gender gaps favouring females—while statistically significant at the 1% level across all examined percentiles—are most pronounced at the lower end of the distribution for all continuous educational outcomes. That is, the estimated coefficient on the male indicator is most negative at lower percentiles. For example, for emotional maturity, the outcome for which the female advantage at the mean is largest, the adjusted gender gap in favour of females is 0.20 SD at the 80th percentile but increases to 0.83 SD at the 20th percentile. Similarly, for ATAR scores—where the female advantage at the mean is relatively small—the adjusted gender gap is only 0.03 SD at the 80th percentile but rises to 0.13 SD at the 20th percentile.

Figure 1 also shows that controlling for additional covariates has different implications across educational stages. For AEDC outcomes, the inclusion of controls slightly reduces the female advantage at both the mean and across the distribution. In contrast, for ATAR outcomes,

controlling for additional variables substantially increases the estimated gender gap in favour of females. At the mean, the adjusted ATAR gap is approximately three times larger than the raw gap (0.09 SD compared with 0.03 SD). Moreover, at the 80th percentile, the gender gap reverses sign after adjustment: while the raw estimate of the male indicator is positive (0.01 SD), the adjusted estimate is negative (–0.03 SD). Importantly, however, across all outcomes and specifications, the results consistently indicate that gender gaps favouring females are largest at the lower end of the outcome distribution.

In summary, our finding that females outperform males across all developmental outcomes considered is consistent with prior Australian studies using LSAC data, which document female advantages in all NAPLAN domains except numeracy from Years 3 to 9 (Cobb-Clark & Moschion 2017; Le & Nguyen 2018a; Nguyen *et al.* 2022; Paterson *et al.* 2024). Moreover, our results indicate that the gender gap in favour of females emerges as early as ages 5–6 (i.e., Year 1) and persists into tertiary education, where females achieve higher ATAR scores and exhibit a higher probability of attaining a higher education qualification.

Furthermore, our finding that the female advantage is more pronounced at the lower end of the distribution of AEDC and ATAR scores is also consistent with Australian evidence showing that gender gaps favouring females in non-numeracy NAPLAN test scores are largest at the lower tail of the distribution (Le & Nguyen 2018a; Nguyen *et al.* 2022).[11] In this respect, the Australian evidence aligns closely with U.S. findings reported by Autor *et al.* (2023), who show that female advantages in reading achievement and behavioural outcomes at school age are driven primarily by lower-tail gaps.

5.3. Differential effects of socioeconomic factors on gender gaps in educational outcomes

5.3.1. Regression results at the mean

We next examine the potential differential effects of socioeconomic factors on gender gaps in educational outcomes using the augmented OLS results for the full population sample. Table 3 reports estimates of the interaction terms between the male indicator and six selected socioeconomic variables, providing evidence of gender-differentiated effects for some of these factors.[12] For example, the effects of parental education differ by child gender and vary across

[11] Australian evidence indicates that males' advantage in numeracy is concentrated at the upper tail of the mathematics score distribution (Le & Nguyen 2018a; Nguyen *et al.* 2022), consistent with patterns observed among school-aged students in the United States (Autor *et al.* 2023).

[12] In this analysis, we transform two categorical variables—parental education and home ownership status—that are interacted with the gender indicator into binary measures to facilitate interpretation and presentation. Specifically, although the estimated coefficients for the original categorical variables vary in magnitude across

educational stages. At the early childhood level, as measured by AEDC outcomes, boys benefit more than girls from higher parental education, for both maternal and paternal education, as indicated by positive and statistically significant interaction terms between the male indicator and parental education variables for most outcomes. This effect is somewhat more pronounced for paternal education, with interaction estimates that are generally larger in magnitude or more precisely estimated than those for maternal education.

In contrast, the role of parental education changes markedly at the tertiary level. Maternal education benefits daughters more strongly, as reflected in negative and statistically significant interaction terms between the male indicator and maternal education for both ATAR scores and higher education attainment. Conversely, paternal education benefits sons more strongly, as indicated by positive and statistically significant interaction terms for higher education attainment. Overall, these patterns in Australia closely mirror evidence from Denmark reported by Brenøe and Lundberg (2018), who similarly find that boys benefit more from both parents' education at the school level, while maternal education disproportionately benefits daughters and paternal education disproportionately benefits sons at the tertiary level.

Our results further show that two additional socioeconomic factors—home ownership and school quality—which are not examined in Brenøe and Lundberg (2018), also exhibit gender-differentiated effects on educational outcomes, and that these effects vary by educational stage. At the early primary school level, boys benefit more than girls from both home ownership and school quality across all five AEDC domains. In contrast, at the tertiary level, girls benefit more from home ownership for both educational outcomes considered.

For household income and local area socioeconomic conditions, we find limited evidence of gender differences in their effects on educational outcomes. Two exceptions emerge for language and cognitive skills (school-based) and higher education attainment, where the interaction between the male indicator and local socioeconomic conditions or household income, respectively, is negative and statistically significant at the 5% level. This indicates that daughters benefit more than sons from higher-quality neighbourhoods or greater family income for these outcomes.

Taken together, the mean regression results suggest that, on average, males benefit more from socioeconomic advantage during the early school years. This finding is consistent with

categories, their signs and levels of statistical significance are generally consistent, as shown in the preceding regression results. Accordingly, we construct binary indicators capturing whether either parent holds a tertiary qualification and whether the child resides in an owner-occupied dwelling (as opposed to rental housing).

Australian evidence reported by Paterson *et al.* (2024), who show that boys benefit more from favourable socioeconomic conditions—measured using a composite index constructed from household income, parental education, and maternal labour force participation—in terms of improvements in numeracy scores measured in Years 3 and 9.[13] This pattern is also in line with U.S. evidence from David *et al.* (2019), who show that family advantage—measured by higher maternal education, coupled households, higher socioeconomic status, and older maternal age at childbirth—as well as higher school and neighbourhood quality, disproportionately benefits the educational and behavioural development of boys relative to girls at school age. However, our results reveal a reversal of this pattern at the tertiary level, where females generally benefit more from most socioeconomic advantages considered, with the notable exception of paternal education.

5.3.2. Regression results along the distribution

Turning to the quantile regression results reported in Figure 2, we find that socioeconomic advantages disproportionately benefit males at the lower end of the educational outcome distributions for early primary school outcomes (see Panels a to e). This pattern is evidenced by the interaction terms between the male indicator and various socioeconomic variables, which are generally larger in magnitude and more statistically significant at the lower tail of the distribution.

For example, for emotional maturity, the estimated interaction between the male indicator and paternal education is 0.09 SD and statistically significant at the 1% level at the 20th percentile, but only 0.01 SD and significant at the 10% level at the 80th percentile. Similarly, the interaction between the male indicator and school quality is positive and statistically significant at the 1% level at the 20th percentile (0.27 SD), but becomes negative (–0.07 SD) at the 80th percentile, despite remaining statistically significant. A largely similar pattern is observed across the other AEDC domains. Taken together, these results suggest that among early primary school outcomes, boys at the lower end of the outcome distribution benefit more from favourable socioeconomic conditions than boys at the upper end.

By contrast, for the ATAR outcome—where socioeconomic advantages tend to benefit females more than males at the mean—we find that females at the lower end of the ATAR distribution benefit more from favourable socioeconomic conditions than those at the upper end. This

[13] As noted earlier, our results regarding the role of individual socioeconomic factors are not directly comparable to those reported by Paterson *et al.* (2024) due to substantial differences in data sources, measures of educational outcomes, the ages at which outcomes are observed, definitions of socioeconomic status, and empirical methods.

pattern is reflected in more negative interaction terms between the male indicator and maternal education, as well as home ownership, at lower percentiles. Conversely, females at the higher end of the distribution also benefit more from paternal education, as the interaction between the male indicator and paternal education is negative and statistically significant at the 5% level or better for the 60th or 80th percentiles (see Panel f in Figure 2).

Overall, our finding that socioeconomic advantages disproportionately benefit males at the lower tail of the distribution for early primary school outcomes is consistent with U.S. evidence from Autor *et al.* (2023), who show that family socioeconomic advantage has particularly strong effects on boys' outcomes relative to girls' at the lower end of the distribution for school-age outcomes. Extending this literature, we provide the first evidence that for an educational outcome measured at the end of secondary school in Australia—namely ATAR—females appear to benefit more from favourable socioeconomic conditions at the lower end of the outcome distribution.

6. Robustness checks and potential mechanisms

6.1. Robustness checks

This subsection assesses the robustness of our findings on the differential impacts of socioeconomic factors on gender gaps in educational outcomes using alternative samples and model specifications.[14] First, we re-estimate the baseline OLS specifications using a sibling sample in place of the full population sample. The results from this exercise, reported in Appendix Table A4, largely replicate the baseline findings, albeit with slightly lower levels of statistical significance, which is likely attributable to the reduced sample size.

Second, we estimate family FE models in place of the baseline OLS specifications to examine whether unobserved time-invariant family characteristics confound the estimated gender gaps and their interactions with selected socioeconomic factors. As in prior studies in this literature (Brenøe & Lundberg 2018; David *et al.* 2019), this approach addresses concerns that unobserved, time-invariant family-level factors may be correlated with both educational outcomes and socioeconomic characteristics, thereby biasing the estimated coefficients on these variables and their interactions with the male indicator.

[14] For brevity and clarity of exposition, these robustness checks focus on the mean regression results. The corresponding quantile regression estimates, which lead to the same qualitative conclusions, are therefore not reported.

Because family fixed effects models do not simultaneously identify the effects of individual gender and its interactions with within-family time-invariant characteristics, this robustness exercise can be implemented only for socioeconomic variables that vary within families over time across siblings. In our data, this restriction confines the analysis to school and neighbourhood quality indicators, which are measured at the time of the AEDC assessment and may differ across siblings or twins.

We first apply the family FE model to a sample of siblings, following the existing literature (Brenøe & Lundberg 2018; David *et al.* 2019). The results, reported in Panel A of Appendix Table A5, are broadly consistent with the baseline findings. In particular, the interaction term between the male indicator and school quality is positive and statistically significant at the 1% level across all five AEDC outcomes, indicating that boys benefit more than girls from attending higher-quality schools. In addition, while the baseline OLS results show little evidence of gender differences in the effects of neighbourhood quality, the family FE estimates suggest that boys benefit more from higher neighbourhood quality for emotional maturity and communication skills and general knowledge, with the corresponding interaction terms positive and statistically significant at the 10% and 5% levels, respectively.

We then estimate the same family FE specification using a sample of twins, with results reported in Panel B of Appendix Table A5. Despite the substantially smaller sample size—approximately 4,000 twins compared with over 84,000 siblings—the estimates similarly indicate that males benefit more from higher school quality. The consistency of these results across sibling and twin samples is particularly reassuring, as twin FE models provide a more stringent control for family-level and genetic confounders and better address concerns regarding the exogeneity of child gender (Royer 2009; Bhalotra & Clarke 2023).

6.2. Potential mechanisms

The results above indicate that returns to selected socioeconomic conditions—many of which are commonly shared among family members, including siblings and twins—differ by child gender. While it remains challenging to causally identify the sources of these differences, particularly given that they appear to reverse over an extended period, this study seeks to shed light on potential mechanisms by documenting gender differences in selected variables that are typically associated with educational outcomes and are available in our data (Nghiem *et al.* 2015; Le & Nguyen 2017, 2018b; Delaney & Devereux 2021).

To this end, we employ family FE regressions using sibling samples (and, where possible, twin samples) to examine within-family gender differences. Specifically, our data allow us to explore gender differences in two such dimensions: (i) school sector choice, measured by whether a child attends a Catholic or independent school either at the time of the AEDC assessment or at 2011 Census (with public schools as the reference category); and (ii) special educational needs, captured by an indicator of whether teachers report that a child has special needs due to a chronic medical, physical, or intellectual disabling condition. With the exception of the school sector choice variables obtained from the 2011 Census—which are measured at the Census date and cover both primary and secondary school attendance[15]—all other variables are measured at the time of the AEDC assessment, when children are typically aged 5–6, and are drawn from the AEDC dataset.

School sector choice is of particular interest in the Australian context because non-government schools—including Catholic and independent schools—generally charge substantial fees and tend to enrol students from more socioeconomically advantaged families, whereas public schools are ostensibly free to attend (Dearden *et al.* 2011; Le *et al.* 2025). Moreover, students attending non-government schools, especially independent schools, exhibit higher average developmental outcomes than their counterparts in public schools (Nghiem *et al.* 2015; Nghiem *et al.* 2016). As such, school sector choice may reflect gender differences in parental investment decisions in early childhood (for AEDC-based outcomes) and later childhood (for Census-based outcomes).

The special educational needs indicator is also informative in this context, as it is explicitly based on diagnoses provided by a parent or guardian, a health professional, or another formal source and is recorded separately from the other AEDC measures used above (AEDC 2025). Consequently, this measure may reflect gender differences in early childhood health conditions.

Results from the family FE regressions using sibling samples—our preferred specifications, reported in Column 4 of Table 4—indicate that males are less likely to be enrolled in

[15] This variable is derived from a Census question on the type of educational institution attended by individuals who were studying at the time of the 2011 Census. Attendance at "Other non-government" schools is classified as independent schooling, while "Government" and "Catholic" schools are classified as public and Catholic schooling, respectively. For outcomes measured using the 2011 Census, the analysis is restricted to individuals aged 5 to 18 at the time of the Census, thereby covering both primary and secondary school students. For AEDC-based outcomes, the analysis is restricted to individuals aged 18 or younger at the time of the Census. We do not include the school-average AEDC score as a control variable in regressions of AEDC-based outcomes in this exercise, as this measure of school quality is closely related to school sector, which is treated as the outcome variable in these regressions.

independent schools at the time of the AEDC assessment (Panel B), as well as in Catholic or independent schools at the time of the 2011 Census (Panels C and D, respectively), and are approximately three percentage points more likely to be identified as having special educational needs due to chronic medical, physical, or intellectual disabling conditions at the time of the AEDC assessment. Estimates obtained using the full population sample (Column 1) or OLS regressions applied to the same sibling samples (Columns 2 and 3) yield qualitatively similar conclusions, albeit with some variation in magnitude and statistical significance.[16] To the best of our knowledge, there is little existing evidence documenting within-family gender differences in school sector choice and early childhood health conditions—both favouring daughters—which our analysis is able to uncover due to the scale and richness of the data.

Although the observed gender differences in school sector choice and early childhood health conditions—both favouring females—are consistent with the female advantages in educational outcomes documented above, it remains unclear whether these differences account for the observed gender heterogeneity in returns to selected socioeconomic factors. To assess this possibility, and following the approach applied to educational outcomes, we estimate an augmented empirical model that introduces interaction terms between the male indicator and selected socioeconomic variables, using OLS specifications and the population sample.

The results provide mixed evidence. For Catholic school attendance (Columns 1 and 3 of Table 5), socioeconomic advantage appears to benefit boys more than girls. Specifically, the interaction terms between the male indicator and selected socioeconomic variables—such as paternal education (for Catholic school attendance measured at the AEDC stage) and family income or neighbourhood quality (for Catholic school attendance measured at the Census)—are positive and statistically significant at the 5% level or better. Similarly, for the special educational needs outcome (Column 5), boys appear to benefit more from favourable socioeconomic conditions than girls: the interaction terms between the male indicator and four of the five included socioeconomic variables (maternal education, family income, home ownership, and neighbourhood quality) are negative and statistically significant at the 5% level or better, indicating that boys from more advantaged backgrounds are less likely to have early childhood health conditions than their female counterparts.

[16] Unreported results from family FE regressions using twin samples, despite the substantially smaller sample size, provide additional insight. While these estimates suggest no statistically significant gender differences in school sector choice among twins—likely reflecting limited within-twin variation in school sector attendance—they indicate that male twins are more likely to have special educational needs.

This pattern suggests that gender differences in parental investment—particularly through Catholic school choice—or in early childhood health conditions may help explain the observed differential returns to selected socioeconomic factors, favouring sons over daughters in early primary school educational outcomes. However, these mechanisms do not appear to account for the finding that most indicators of socioeconomic advantage confer stronger benefits to females at the tertiary level.

In contrast, the results for independent school attendance are less conclusive. For example, boys from more socioeconomically advantaged families, as measured by paternal education and family income, are slightly less likely than girls to attend independent schools at the AEDC stage (Column 2), with interaction effects statistically significant at the 10% level. At the Census stage, boys from higher-income families are also less likely than girls to attend independent schools (Column 4), as indicated by a negative and statistically significant interaction term at the 1% level. However, boys from owner-occupied households are more likely than girls to attend independent schools at the same stage, with a positive and statistically significant interaction term at the 1% level.

Taken together, these findings confirm the presence of within-family gender differences in school sector choice and early childhood health conditions that favour females and are consistent with the female advantages in educational outcomes documented earlier. While there is some evidence that gender differences in parental investment—particularly with respect to Catholic school choice—and early health conditions may help explain differential returns to selected socioeconomic factors, the evidence is less consistent for independent school attendance. This mixed pattern, together with the continuing difficulty of quantifying the mechanisms underlying the reversal in the role of socioeconomic factors in shaping gender gaps from early primary school to tertiary education, suggests that further research is needed to clarify the mechanisms underlying gender differences in educational attainment (Delaney & Devereux 2021).

7. Conclusion

This study pioneers the use of high-quality population-wide linked data and employs some of the most robust empirical models in the current literature to document gender gaps in educational outcomes from early primary school through early adulthood in Australia, and to examine how socioeconomic factors contribute to these gaps. We document a persistent gender gap favouring females that emerges as early as ages 5–6 and continues into early adulthood,

with the female advantage most pronounced at the lower end of the educational attainment distribution.

We further show that family socioeconomic advantage disproportionately benefits males, particularly those at the lower tail of the distribution, but only for developmental outcomes observed in early primary school—a finding consistent with international evidence at school levels (Brenøe & Lundberg 2018; Autor *et al.* 2023; Paterson *et al.* 2024). However, we provide new evidence that this pattern reverses post-school, with socioeconomic advantages tending to benefit females, particularly those at the lower end of the educational attainment distribution, more strongly. Supporting evidence also suggests that within-family gender differences in school sector choice and early childhood health conditions—both favouring females—may serve as mechanisms underlying these patterns.

While this study identifies the reversal of most socioeconomic factors in contributing to gender gaps between the first year of primary school and the end of secondary school (i.e., ATAR), the available data do not allow us to pinpoint exactly when this reversal occurs. Additionally, although our empirical methods are among the most rigorous currently feasible, the findings—particularly regarding the role of socioeconomic factors—cannot be interpreted as causal. Further research using alternative datasets or more robust empirical approaches to examine educational outcomes over time, or related outcomes not covered here, would be valuable.

Despite these limitations, this study demonstrates that gender differences in various socioeconomic factors help explain observed gaps in educational outcomes, including the reversal of their effects over an extended and critical developmental period. These findings improve our understanding of gender-based developmental disparities and provide evidence to guide interventions aimed at promoting better outcomes for both genders.

References

ABS, 2026. Person Level Integrated Data Asset (PLIDA). Australian Bureau of Statistics (ABS) (Ed.), https://www.abs.gov.au/about/data-services/data-integration/integrated-data/person-level-integrated-data-asset-plida

AEDC, 2025. AEDC Data Guidelines 2025. Australian Early Development Census (AEDC) (Ed.), https://www.aedc.gov.au/resources/detail/aedc-data-guidelines

Almond, D., Edlund, L., 2007. Trivers–Willard at birth and one year: evidence from US natality data 1983–2001. Proceedings of the Royal Society B: Biological Sciences 274, 2491-2496

Autor, D., Figlio, D., Karbownik, K., Roth, J., Wasserman, M., 2023. Males at the Tails: How Socioeconomic Status Shapes the Gender Gap. The Economic Journal 133, 3136-3152

Baldiga, K., 2014. Gender Differences in Willingness to Guess. Management Science 60, 434-448

Barrios-Fernández, A., Riudavets-Barcons, M., 2024. Teacher value-added and gender gaps in educational outcomes. Economics of Education Review 100, 102541

Bedard, K., Deschênes, O., 2005. Sex Preferences, Marital Dissolution, and the Economic Status of Women. Journal of Human Resources 40, 411

Bertrand, M., Pan, J., 2013. The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior. American Economic Journal: Applied Economics 5, 32-64

Bhalotra, S., Clarke, D., 2023. Analysis of twins. In: Zimmermann KF (ed.) Handbook of Labor, Human Resources and Population Economics. pp. 1-37.

Blau, F.D., Kahn, L.M., 2017. The gender wage gap: Extent, trends, and explanations. Journal of Economic Literature 55, 789-865

Blundell, J., Duchini, E., Simion, Ş., Turrell, A., 2025. Pay Transparency and Gender Equality. American Economic Journal: Economic Policy 17, 418–45

Blundell, R., Dearden, L., Meghir, C., Sianesi, B., 1999. Human capital investment: the returns from education and training to the individual, the firm and the economy. Fiscal studies 20, 1-23

Booth, A.L., Cardona-Sosa, L., Nolen, P., 2018. Do single-sex classes affect academic achievement? An experiment in a coeducational university. Journal of Public Economics 168, 109-126

Brenøe, A.A., Lundberg, S., 2018. Gender gaps in the effects of childhood family environment: Do they persist into adulthood? European Economic Review 109, 42-62

Buchmann, C., DiPrete, T.A., 2006. The growing female advantage in college completion: The role of family background and academic achievement. American sociological review 71, 515-541

Cai, X., Lu, Y., Pan, J., Zhong, S., 2019. Gender Gap Under Pressure: Evidence from China's National College Entrance Examination. Review of Economics and Statistics 101, 249–263

Carlana, M., 2019. Implicit Stereotypes: Evidence from Teachers' Gender Bias. The Quarterly Journal of Economics 134, 163–1224

Chapman, B., Ryan, C., 2005. The access implications of income-contingent charges for higher education: lessons from Australia. Economics of Education Review 24, 491-512

Chapman, B.J., Mulvey, C., 1986. An analysis of the origins of sex differences in Australian wages. Journal of Industrial Relations 28, 504-520

Chetty, R., Hendren, N., 2018. The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. The Quarterly Journal of Economics 133, 1107–1162

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, 711–783

Christopher, C., David, B.M., Jessica Van, P., 2013. Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School. Journal of Human Resources 48, 236-264

Clark, D., Royer, H., 2013. The effect of education on adult mortality and health: Evidence from Britain. American Economic Review 103, 2087-2120

Cobb-Clark, D.A., Moschion, J., 2017. Gender gaps in early educational achievement. Journal of Population Economics 3, 1093–1134

Cobb-Clark, D.A., Nguyen, T.-H., 2012. Educational Attainment Across Generations: The Role of Immigration Background. Economic Record 88, 554-575

Crawford, C., Dearden, L., Greaves, E., 2014. The Drivers of Month-of-Birth Differences in Children's Cognitive and Non-Cognitive Skills. Journal of the Royal Statistical Society Series A: Statistics in Society 177, 829-860

Croson, R., Gneezy, U., 2009. Gender Differences in Preferences. Journal of Economic Literature 47, 448-74

Cuff, K., Gamarra Rondinel, A., Payne, A.A., 2025. Gender Differences in University Enrollment and STEM Major: The Role of Tuition Policy in Australia. Melbourne Institute Working Paper No. 10/25

Dahl, G.B., Moretti, E., 2008. The Demand for Sons. The Review of Economic Studies 75, 1085-1120

David, H., Figlio, D.N., Karbownik, K., Roth, J., Wasserman, M., 2019. Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes. American Economic Journal: Applied Economics 11, 338-81

Dearden, L., Ryan, C., Sibieta, L., 2011. What Determines Private School Choice? A Comparison between the United Kingdom and Australia. Australian Economic Review 44, 308-320

Delaney, J.M., Devereux, P.J., 2021. The Economics of Gender and Educational Achievement: Stylized Facts and Causal Evidence. In: Oxford Research Encyclopedia of Economics and Finance. Oxford University Press

Delaney, J.M., Devereux, P.J., 2025. Gender differences in graduate degree choices. Journal of Economic Behavior & Organization 230, 106882

Dossi, G., Figlio, D., Giuliano, P., Sapienza, P., 2021. Born in the family: Preferences for boys and the gender gap in math. Journal of Economic Behavior & Organization 183, 175-188

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, 198-208

Dustmann, C., Ku, H., Kwak, D.W., 2018. Why Are Single-Sex Schools Successful? Labour Economics 54, 79-99

Figlio, D., Karbownik, K., Roth, J., Wasserman, M., 2016. School quality and the gender gap in educational achievement. American Economic Review: Papers & Proceedings 106, 289-95

Firpo, S., Fortin, N.M., Lemieux, T., 2009. Unconditional Quantile Regressions. Econometrica 77, 953-973

Francesconi, M., Parey, M., 2018. Early gender gaps among university graduates. European Economic Review 109, 63-82

Fryer, R.G., Levitt, S., 2010. An Empirical Analysis of the Gender Gap in Mathematics. American Economic Journal: Applied Economics 2, 210-40

Galama, T., Lleras-Muney, A., Kippersluis, H.v., 2018. The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence. In: Oxford Research Encyclopedia of Economics and Finance. Oxford University Press.

García-Echalar, A., Poblete, S., Rau, T., 2024. Teacher value-added and the test score gender gap. Labour Economics 89, 102588

Gneezy, U., Niederle, M., Rustichini, A., 2003. Performance in competitive environments: Gender differences. Quarterly Journal of Economics 118, 1049-1074

Goldin, C., 2014. A Grand Gender Convergence: Its Last Chapter. American Economic Review 104, 1091-1119

Golsteyn, B.H.H., Schils, T., 2014. Gender gaps in primary school achievement: A decomposition into endowments and returns to IQ and non-cognitive factors. Economics of Education Review 41, 176-187

Gunderson, M., Oreopolous, P., 2020. Chapter 3 - Returns to education in developed countries. In: Bradley S & Green C (eds.) The Economics of Education (Second Edition). Academic Press, pp. 39-51.

Hanly, M., Edwards, B., Goldfeld, S., Craven, R.G., Mooney, J., Jorm, L., Falster, K., 2019. School starting age and child development in a state-wide, population-level cohort of children

in their first year of school in New South Wales, Australia. Early Childhood Research Quarterly 48, 325-340

Howells, S., Nguyen, H.T., Brinkman, S., Mitrou, F., 2024. Investigating the Validity of the Australian Early Development Census. Child Psychiatry & Human Development 55, 1564–1581

Iriberri, N., Rey-Biel, P., 2021. Brave boys and play-it-safe girls: Gender differences in willingness to guess in a large scale natural field experiment. European Economic Review 131, 103603

Jacob, B.A., 2002. Where the boys aren't: non-cognitive skills, returns to school and the gender gap in higher education. Economics of Education Review 21, 589-598

Kahn, S., Ginther, D., 2018. Women and Science, Technology, Engineering, and Mathematics (STEM). In: Averett SL, Argys LM & Hoffman SD (eds.) The oxford handbook of women and the economy. Oxford University Press.

Kugler, A.D., Tinsley, C.H., Ukhaneva, O., 2021. Choice of majors: are women really different from men? Economics of Education Review 81, 102079

Landaud, F., Ly, S.T., Maurin, E., 2020. Competitive schools and the gender gap in the choice of field of study. Journal of Human Resources 55, 278-308

Lavy, V., Sand, E., 2018. On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases. Journal of Public Economics 167, 263-279

Le, H., Blyth, C.C., Schlegel, C., Morgan, J.-A., Mitrou, F., Nguyen, H., Foong, R., Carlson, S., Hughes, C., Liu, B., Moore, H.C., 2025. Socio-economic inequality underpins inequity in influenza vaccination uptake between public and private secondary schools: an Australian population-based study. The Lancet Regional Health - Western Pacific 65, 101761

Le, H.T., Nguyen, H.T., 2017. Parental health and children's cognitive and noncognitive development: New evidence from the longitudinal survey of Australian children. Health Economics 26, 1767–1788

Le, H.T., Nguyen, H.T., 2018a. The evolution of the gender test score gap through seventh grade: New insights from Australia using unconditional quantile regression and decomposition. IZA Journal of Labor Economics 7, 2

Le, H.T., Nguyen, H.T., 2018b. The Impact of Maternal Mental Health Shocks on Child Health: Estimates from Fixed Effects Instrumental Variables Models for two Cohorts of Australian Children. American Journal of Health Economics 4, 185-225

Marianne, B., 2011. Chapter 17 - New Perspectives on Gender. In: Orley A & David C (eds.) Handbook of labor economics. Elsevier, Burlington, pp. 1543-1590.

McCrary, J., Royer, H., 2011. The Effect of female education on fertility and infant health: Evidence from school entry policies using exact date of birth. American Economic Review 101, 158-95

Mogstad, M., Torsvik, G., 2023. Family background, neighborhoods, and intergenerational mobility. Handbook of the Economics of the Family 1, 327-387

Montolio, D., Taberner, P.A., 2021. Gender differences under test pressure and their impact on academic performance: A quasi-experimental design. Journal of Economic Behavior & Organization 191, 1065-1090

Nghiem, H.S., Nguyen, H.T., Khanam, R., Connelly, L.B., 2015. Does School Type Affect Cognitive and Non-Cognitive Development in Children? Evidence from Australian Primary Schools. Labour Economics 33, 55-65

Nghiem, S., Nguyen, H.T., Connelly, L.B., 2016. The Efficiency of Australian Schools: A Nationwide Analysis Using Gains in Test Scores of Students as Outputs. Economic Papers: A journal of applied economics and policy 35, 256-268

Nguyen, H.T., Brinkman, S., Le, H.T., Zubrick, S.R., Mitrou, F., 2022. Gender differences in time allocation contribute to differences in developmental outcomes in children and adolescents. Economics of Education Review 89, 102270

Nguyen, H.T., Connelly, L., Le, H.T., Mitrou, F., Taylor, C., Zubrick, S., 2020. Ethnicity differentials in academic achievements: The role of time investments. Journal of Population Economics 33, 1381-1418

Nguyen, H.T., Zając, T., Tomaszewski, W., Mitrou, F., 2025. Educational pathways and earnings trajectories of second-generation immigrants in Australia: New insights from linked census-administrative data. Economics of Education Review 109, 102716

Niederle, M., Vesterlund, L., 2010. Explaining the Gender Gap in Math Test Scores: The Role of Competition. The Journal of Economic Perspectives 24, 129-144

OECD, 2025. Education at a Glance 2025 OECD Indicators. Organisation for Economic Cooperation and Development (OECD) Publishing, Paris

Paterson, M., Parasnis, J., Rendall, M., 2024. Gender, socioeconomic status, and numeracy test scores. Journal of Economic Behavior & Organization 227, 106751

Riphahn, R.T., Schwientek, C., 2015. What drives the reversal of the gender education gap? Evidence from Germany. Applied Economics 47, 5748-5775

Royer, H., 2009. Separated at Girth: US Twin Estimates of the Effects of Birth Weight. American Economic Journal: Applied Economics 1, 49–85

Saygin, P.O., Atwater, A., 2021. Gender differences in leaving questions blank on high-stakes standardized tests. Economics of Education Review 84, 102162

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, Mass.

Table 1: Summary statistics by gender

Variable	Male	Female	Difference (Male – Female)
	(1)	(2)	(3)
Child age at 2011 Census (years)	4.613	4.553	0.060***
Child ESB migrant [(a)]	0.025	0.024	0.000
Child NESB migrant [(a)]	0.030	0.029	0.000
Child age at AEDC time (years)	5.637	5.583	0.054***
Mother age at 2011 Census (years)	35.726	35.701	0.025
Mother ESB migrant [(a)]	0.086	0.085	0.001
Mother NESB migrant [(a)]	0.192	0.193	-0.001
Mother education: Year 12 [(a)]	0.218	0.219	-0.001
Mother education: Certificate [(a)]	0.128	0.127	0.001
Mother education: Diploma [(a)]	0.123	0.124	-0.001
Mother education: Bachelor degree or higher [(a)]	0.362	0.363	-0.001
Father age at 2011 Census (years)	38.243	38.219	0.024
Father ESB migrant [(a)]	0.102	0.101	0.001
Father NESB migrant [(a)]	0.189	0.189	0.000
Father education: Year 12 [(a)]	0.150	0.150	0.000
Father education: Certificate [(a)]	0.296	0.296	0.000
Father education: Diploma [(a)]	0.100	0.098	0.002**
Father education: Bachelor degree or higher [(a)]	0.294	0.297	-0.003**
Family income at 2011 Census ($1,000)	82.761	82.823	-0.061
Mortgage homeowner [(a)]	0.651	0.650	0.001
Outright homeowner [(a)]	0.097	0.097	-0.001
School average AEDC score	8.526	8.569	-0.043***
SEIFA score at 2011 Census (/100)	10.104	10.105	-0.001
Physical health and wellbeing (standardised)	-0.123	0.160	-0.283***
Social competence (standardised)	-0.179	0.225	-0.404***
Emotional maturity (standardised)	-0.241	0.278	-0.519***
Language and cognitive skills (school-based) (standardised)	-0.094	0.144	-0.238***
Communication skills and general knowledge (standardised)	-0.117	0.173	-0.290***
ATAR (standardised)	-0.006	0.022	-0.028***
Higher education qualification [(a)]	0.261	0.391	-0.130***
Number of individuals	184,690	174,555	

Notes: Figures are sample means. [(a)] indicates a binary variable. Statistics are reported for the population sample using physical health and wellbeing as an educational outcome. Tests are performed on the significance of the difference between the sample mean for males and females. The symbol * indicates statistical significance at 10% level, ** at 5% level, and *** at 1% level.

Table 2: Gender gaps in educational outcomes from different samples and specifications

Sample:	Population	Siblings			Twins		
Specification:	OLS 1	OLS 1	OLS 2	FE	OLS 1	OLS 2	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Physical health and wellbeing							
Male	-26.92***	-25.76***	-26.25***	-25.85***	-30.74***	-29.68***	-29.89***
	(0.30)	(0.33)	(0.65)	(0.82)	(2.53)	(2.49)	(2.47)
Observations	359,245	287,422	76,538	76,538	3,887	3,554	3,554
R-squared	0.168	0.169	0.170	0.642	0.199	0.193	0.737
Panel B: Social competence							
Male	-39.02***	-37.47***	-37.39***	-36.70***	-36.39***	-36.13***	-35.92***
	(0.30)	(0.32)	(0.62)	(0.77)	(2.41)	(2.41)	(2.40)
Observations	359,245	287,422	76,538	76,538	3,887	3,554	3,554
R-squared	0.188	0.186	0.186	0.668	0.224	0.213	0.743
Panel C: Emotional maturity							
Male	-50.58***	-49.06***	-49.04***	-48.66***	-46.78***	-46.81***	-46.60***
	(0.30)	(0.33)	(0.63)	(0.80)	(2.46)	(2.48)	(2.47)
Observations	359,245	287,422	76,538	76,538	3,887	3,554	3,554
R-squared	0.172	0.171	0.172	0.650	0.203	0.199	0.721
Panel D: Language and cognitive skills (school-based)							
Male	-23.14***	-22.80***	-22.03***	-20.61***	-20.37***	-20.33***	-19.39***
	(0.29)	(0.32)	(0.62)	(0.76)	(2.31)	(2.30)	(2.25)
Observations	359,245	287,422	76,538	76,538	3,887	3,554	3,554
R-squared	0.222	0.226	0.221	0.694	0.246	0.229	0.784

Sample:	Population	Siblings			Twins		
Specification:	OLS 1	OLS 1	OLS 2	FE	OLS 1	OLS 2	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel E: Communication skills and general knowledge							
Male	-27.36***	-27.06***	-26.35***	-24.56***	-26.26***	-26.12***	-25.88***
	(0.29)	(0.32)	(0.62)	(0.76)	(2.26)	(2.27)	(2.24)
Observations	359,245	287,422	76,538	76,538	3,887	3,554	3,554
R-squared	0.222	0.224	0.220	0.686	0.252	0.244	0.780
Panel F: ATAR							
Male	-8.52***	-8.55***	-10.36***	-13.45***	-9.62***	-16.30***	-15.87***
	(0.32)	(0.34)	(0.51)	(0.57)	(2.81)	(3.26)	(3.24)
Observations	334,592	299,811	122,136	122,136	3,206	1,770	1,770
R-squared	0.174	0.176	0.157	0.719	0.187	0.177	0.708
Panel G: Higher education qualification							
Male	-12.94***	-12.97***	-13.41***	-13.26***	-13.12***	-13.00***	-13.08***
	(0.09)	(0.10)	(0.12)	(0.15)	(0.85)	(0.88)	(0.88)
Observations	818,848	731,159	466,836	466,836	7,886	7,194	7,194
R-squared	0.218	0.218	0.219	0.645	0.208	0.208	0.680

Notes: Results are estimated using either Ordinary Least Squares (OLS) or family fixed-effects (FE) models, based on the respective sibling or twin sample. "OLS 1" indicates results using the full sample from the respective sample while "OLS 2" indicates results using the sample from FE regressions. Results (estimates and standard errors) are multiplied by 100 for aesthetic purposes. Additional covariates include child age at the 2011 Census (and its square), child migrant status, child age at the time of the AEDC assessment (and its square, for AEDC outcomes), maternal age at the 2011 Census (and its square), maternal education, maternal migrant status, paternal age at the 2011 Census (and its square), paternal education, paternal migrant status, total family income, residential tenure, SEIFA, school-level average AEDC score (for AEDC outcomes only), and state/territory fixed effects. Robust standard errors clustered at the family level are reported in parentheses. The symbol * indicates statistical significance at 10% level, ** at 5% level, and *** at 1% level.

Table 3: Differential effects of socioeconomic conditions on gender gaps in educational outcomes – Results at the mean

Outcome:	Physical health and wellbeing	Social competence	Emotional maturity	Language and cognitive skills (school-based)	Communication skills and general knowledge	ATAR	Higher education qualification
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male	-95.04***	-149.64***	-144.73***	-131.63***	-112.99***	-0.78	-7.37***
	(6.16)	(5.88)	(5.62)	(6.55)	(5.59)	(5.34)	(1.34)
Male x Mother has a Diploma, Bachelor degree or higher	1.24*	1.83***	0.60	4.57***	2.73***	-3.86***	-1.36***
	(0.66)	(0.66)	(0.67)	(0.62)	(0.64)	(0.73)	(0.22)
Male x Father has a Diploma, Bachelor degree or higher	1.33**	3.85***	4.34***	6.06***	3.29***	-0.88	2.08***
	(0.67)	(0.67)	(0.68)	(0.62)	(0.65)	(0.73)	(0.23)
Male x Family income	1.64	7.45	-3.92	-0.96	14.24*	-8.59	-12.77***
	(7.75)	(7.59)	(7.61)	(7.71)	(7.65)	(6.73)	(1.86)
Male x Homeowner	1.71**	3.11***	3.07***	2.45***	1.76**	-3.50***	-4.33***
	(0.75)	(0.74)	(0.73)	(0.75)	(0.73)	(1.07)	(0.23)
Male x School average AEDC score	7.63***	12.75***	10.61***	13.16***	10.20***		
	(0.59)	(0.54)	(0.49)	(0.67)	(0.46)		
Male x SEIFA	0.03	-0.37	-0.05	-1.03**	-0.66	-0.06	-0.10
	(0.49)	(0.48)	(0.48)	(0.49)	(0.48)	(0.53)	(0.14)
Observations	359,245	359,245	359,245	359,245	359,245	334,592	818,848
R-squared	0.168	0.188	0.173	0.219	0.219	0.156	0.206

Notes: Results are obtained from an OLS regression using a population sample. Results (estimates and standard errors) are multiplied by 100 for aesthetic purposes. Additional covariates include child age at the 2011 Census (and its square), child migrant status, child age at the time of the AEDC assessment (and its square, for AEDC outcomes), maternal age at the 2011 Census (and its square), maternal education, maternal migrant status, paternal age at the 2011 Census (and its square), paternal education, paternal migrant status, total family income, residential tenure, SEIFA, school-level average AEDC score (for AEDC outcomes only), and state/territory fixed effects. Robust standard errors clustered at the family level are reported in parentheses. The symbol * indicates statistical significance at 10% level, ** at 5% level, and *** at 1% level.

Table 4: Gender differences in school sector choice and early health conditions

Sample:	Population	Siblings		
Specification:	OLS 1	OLS 1	OLS 2	FE
	(1)	(2)	(3)	(4)
Panel A: Catholic school at AEDC time				
Male	-0.74***	-0.64***	-0.39	0.10
	(0.14)	(0.15)	(0.29)	(0.15)
Observations	362,674	290,215	77,982	77,982
R-squared	0.029	0.031	0.033	0.932
Mean dependent variable	0.22	0.23	0.23	0.23
Panel B: Independent school at AEDC time				
Male	-0.87***	-0.71***	-0.59***	-0.34**
	(0.10)	(0.11)	(0.20)	(0.14)
Observations	362,674	290,215	77,982	77,982
R-squared	0.029	0.029	0.031	0.882
Mean dependent variable	0.10	0.10	0.09	0.09
Panel C: Catholic school at 2011 Census				
Male	-0.81***	-0.73***	-0.63***	-0.28***
	(0.08)	(0.08)	(0.09)	(0.05)
Observations	1,268,567	1,145,556	867,926	867,926
R-squared	0.027	0.028	0.029	0.917
Mean dependent variable	0.25	0.25	0.25	0.25
Panel D: Independent school at 2011 Census				
Male	-0.61***	-0.50***	-0.43***	-0.30***
	(0.06)	(0.07)	(0.08)	(0.05)
Observations	1,268,567	1,145,556	867,926	867,926
R-squared	0.065	0.064	0.064	0.899
Mean dependent variable	0.16	0.16	0.16	0.16
Panel E: Special need status at AEDC time				
Male	3.04***	2.90***	2.89***	3.01***
	(0.06)	(0.07)	(0.14)	(0.18)
Observations	362,674	290,215	77,982	77,982
R-squared	0.023	0.023	0.025	0.549
Mean dependent variable	0.04	0.04	0.04	0.04

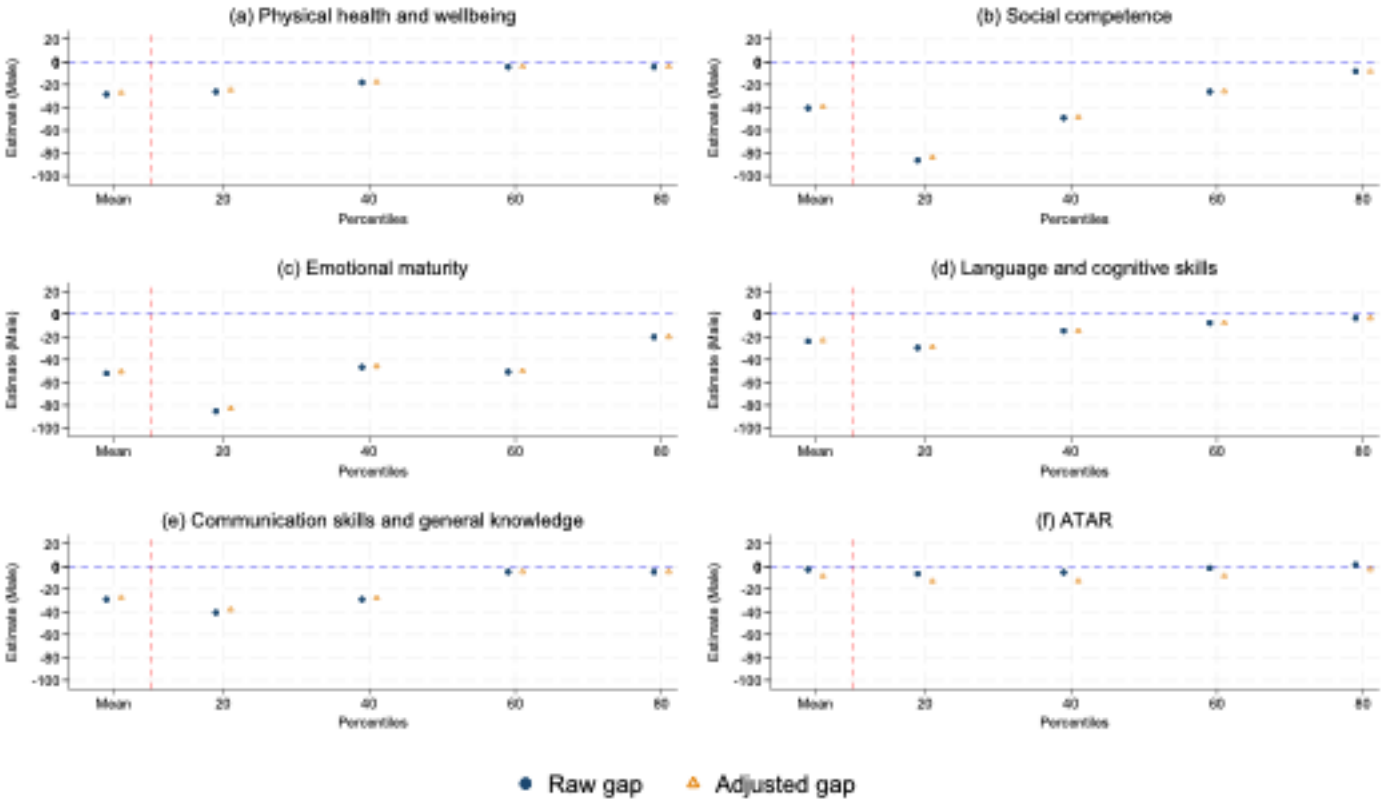
Notes: Results are obtained from Ordinary Least Squared (OLS) regression or family Fixed Effects (FE) regression. "OLS 1" indicates results using the full sample from the respective sample while "OLS 2" indicates results using the sample from FE regressions. Results (estimates and standard errors) are multiplied by 100 for aesthetic purposes. Additional explanatory variables include child age at 2011 Census (and its square), child migrant status, child age at the time of the AEDC assessment (and its square, for AEDC based outcomes), maternal age at 2011 Census (and its square), maternal education, maternal migration status, paternal age at 2011 Census (and its square), paternal education, paternal migration status, total family income, family residential tenure status, SEIFA, and state/territory dummy variables. Robust standard errors clustered at the family level are reported in parentheses. The symbol * indicates statistical significance at 10% level, ** at 5% level, and *** at 1% level.

Table 5: Differential effects of socioeconomic conditions on gender gaps in school sector choice and early health conditions

Outcome:	Catholic school at AEDC time	Independent school at AEDC time	Catholic school at 2011 Census	Independent school at 2011 Census	Special need status at AEDC time
	(1)	(2)	(3)	(4)	(5)
Male	-4.02**	1.22	-3.71***	0.33	6.98***
	(2.02)	(1.47)	(1.11)	(0.91)	(1.01)
Male x Mother has a Diploma, Bachelor degree or higher	-0.20	0.08	-0.10	-0.21	-0.51***
	(0.32)	(0.23)	(0.18)	(0.15)	(0.15)
Male x Father has a Diploma, Bachelor degree or higher	0.95***	-0.43*	0.21	0.10	-0.05
	(0.33)	(0.24)	(0.18)	(0.16)	(0.15)
Male x Family income	1.46	-4.60*	3.71**	-4.79***	-6.71***
	(3.42)	(2.67)	(1.63)	(1.44)	(1.63)
Male x Homeowner	0.19	-0.04	-0.03	0.42***	-0.60***
	(0.30)	(0.22)	(0.18)	(0.15)	(0.16)
Male x SEIFA	0.27	-0.15	0.25**	-0.08	-0.26**
	(0.21)	(0.15)	(0.11)	(0.09)	(0.10)
Observations	362,674	362,674	1,268,567	1,268,567	362,674
R-squared	0.025	0.027	0.023	0.060	0.023
Mean dependent variable	0.22	0.10	0.25	0.16	0.04

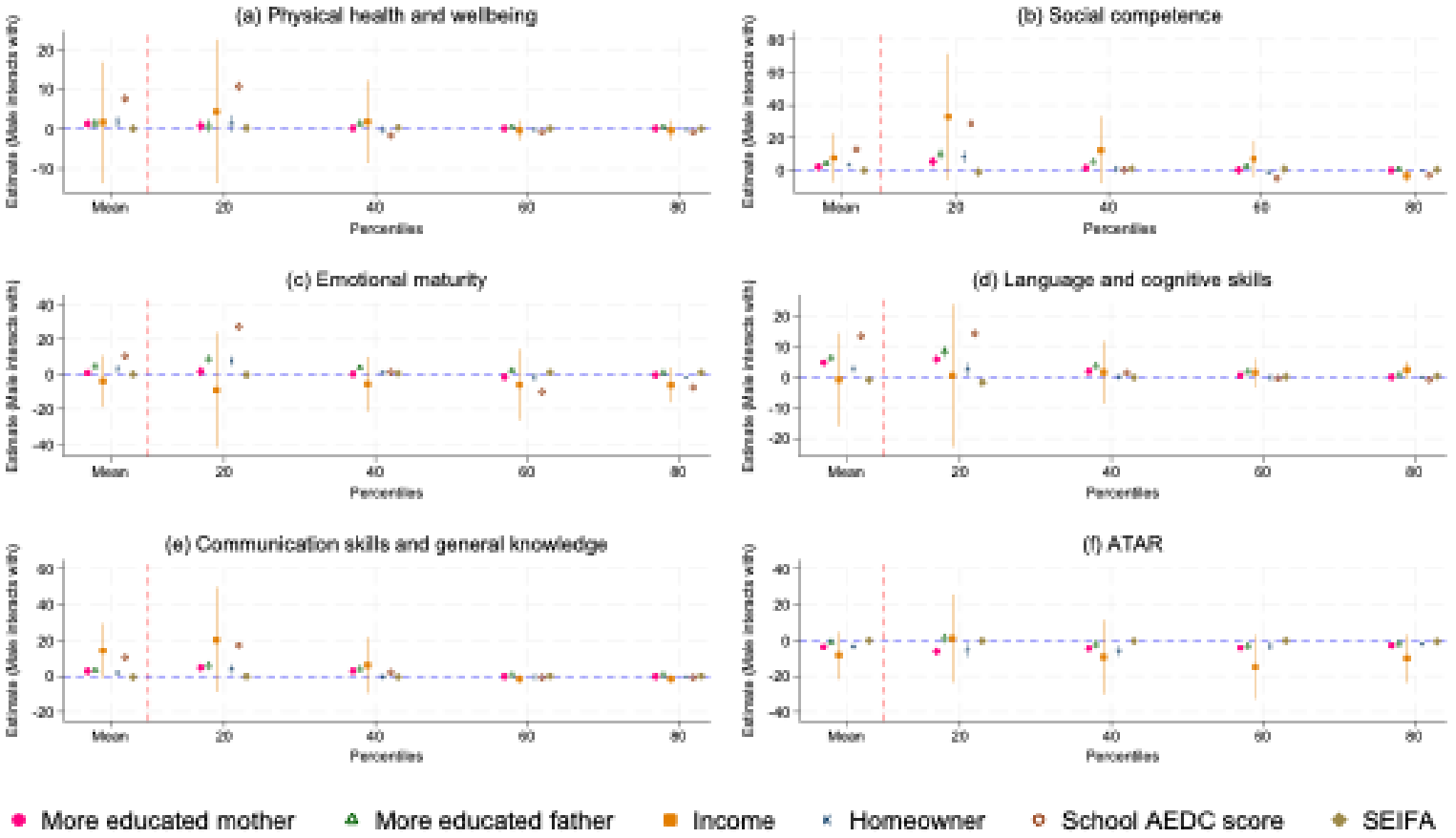
Notes: Results are obtained from an Ordinary Least Squared (OLS) regression using a population sample. Results (estimates and standard errors) are multiplied by 100 for aesthetic purposes. Additional explanatory variables include child age at 2011 Census (and its square), child migrant status, child age at the time of the AEDC assessment (and its square, for AEDC based outcomes), maternal age at 2011 Census (and its square), maternal education, maternal migration status, paternal age at 2011 Census (and its square), paternal education, paternal migration status, total family income, family residential tenure status, SEIFA, and state/territory dummy variables. Robust standard errors clustered at the family level are reported in parentheses. The symbol * indicates statistical significance at 10% level, ** at 5% level, and *** at 1% level.

Figure 1: Gender gaps in educational outcomes at the mean and selected percentiles



Notes: This figure reports estimates of the male indicator at the mean, obtained from an OLS regression, and at selected percentiles, obtained from UQR. Estimates and corresponding 95% confidence intervals (displayed as whiskers) are multiplied by 100 for presentation purposes. "Raw gap" refers to estimates from regressions without additional controls, whereas "Adjusted gap" refers to estimates from regressions that include additional covariates. Additional covariates include child age at 2011 Census (and its square), child migrant status, child age at the time of the AEDC assessment (and its square, for AEDC outcomes), maternal age at 2011 Census (and its square), maternal education, maternal migration status, paternal age at 2011 Census (and its square), paternal education, paternal migration status, total family income, family residential tenure status, SEIFA, school-level average AEDC score (for AEDC outcomes only), and state/territory dummy variables. Detailed regression results are reported in Appendix Table A2.

Figure 2: Differential effects of socioeconomic conditions on gender gaps in educational outcomes at the mean and selected percentiles



Notes: This figure reports estimates of the interaction terms between the male indicator and the selected socioeconomic variables (as indicated in the figure legend). Estimates at the mean are obtained from an OLS regression, while those at selected percentiles are obtained from UQR. Coefficients and corresponding 95% confidence intervals (displayed as whiskers) are multiplied by 100 for presentation purposes. Detailed regression results are presented in Appendix Table A3.

Online Appendixes

For refereeing purposes and to be published online

Appendix Table A1: Determinants of educational outcomes – Remaining results

Outcome:	Physical health and wellbeing		Social competence		Emotional maturity		Language and cognitive skills (school-based)	
Sample:	Population	Siblings	Population	Siblings	Population	Siblings	Population	Siblings
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Male	-26.92***	-25.76***	-39.02***	-37.47***	-50.58***	-49.06***	-23.14***	-22.80***
	(0.30)	(0.33)	(0.30)	(0.32)	(0.30)	(0.33)	(0.29)	(0.32)
Child age at 2011 Census	1.84***	1.56***	3.39***	2.55***	6.14***	4.99***	-3.91***	-2.28***
	(0.32)	(0.36)	(0.32)	(0.35)	(0.32)	(0.35)	(0.31)	(0.35)
Child age at 2011 Census squared	0.01	0.03	-0.17***	-0.12***	-0.52***	-0.45***	0.15***	0.01
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Child ESB migrant [(a)]	3.97***	4.20***	1.78*	1.88*	1.82*	1.89	0.05	0.40
	(1.03)	(1.12)	(1.05)	(1.14)	(1.08)	(1.17)	(1.02)	(1.11)
Child NESB migrant [(a)]	4.87***	5.80***	1.45	3.09***	0.52	2.28**	-0.56	-1.26
	(0.96)	(1.13)	(1.00)	(1.16)	(0.96)	(1.12)	(1.00)	(1.21)
Child age at AEDC time	227.84***	211.11***	228.02***	212.08***	187.49***	175.90***	268.38***	252.08***
	(10.92)	(11.54)	(10.60)	(11.19)	(9.32)	(9.84)	(12.30)	(13.30)
Child age at AEDC time squared	-18.71***	-17.21***	-18.81***	-17.40***	-15.66***	-14.58***	-20.94***	-19.40***
	(0.97)	(1.02)	(0.94)	(0.99)	(0.83)	(0.87)	(1.09)	(1.17)
Mother age at 2011 Census	1.92***	2.21***	2.82***	2.79***	2.47***	2.41***	2.05***	2.59***
	(0.39)	(0.48)	(0.42)	(0.47)	(0.35)	(0.40)	(0.34)	(0.43)
Mother age at 2011 Census squared	-0.03***	-0.03***	-0.04***	-0.04***	-0.03***	-0.03***	-0.03***	-0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
Mother ESB migrant [(a)]	0.43	0.61	-0.14	-0.04	-0.52	-0.76	0.61	0.91
	(0.60)	(0.66)	(0.59)	(0.65)	(0.61)	(0.68)	(0.56)	(0.63)
Mother NESB migrant [(a)]	4.29***	4.51***	-2.01***	-1.79***	-2.70***	-2.44***	-2.43***	-2.87***
	(0.53)	(0.61)	(0.55)	(0.62)	(0.54)	(0.61)	(0.52)	(0.61)
Mother education: Year 12 [(b)]	8.28***	8.10***	7.28***	7.42***	5.15***	5.00***	16.64***	16.85***
	(0.55)	(0.61)	(0.54)	(0.59)	(0.54)	(0.58)	(0.58)	(0.64)
Mother education: Certificate [(b)]	5.80***	5.64***	3.94***	4.35***	1.62***	1.71**	15.03***	14.95***
	(0.62)	(0.69)	(0.61)	(0.68)	(0.61)	(0.67)	(0.64)	(0.72)
Mother education: Diploma [(b)]	8.97***	8.99***	6.10***	6.48***	2.37***	2.47***	19.60***	19.72***
	(0.62)	(0.68)	(0.62)	(0.67)	(0.62)	(0.67)	(0.62)	(0.69)
Mother education: Bachelor degree or higher [(b)]	11.61***	11.62***	11.76***	11.80***	6.18***	5.74***	28.75***	28.78***
	(0.54)	(0.60)	(0.54)	(0.59)	(0.54)	(0.59)	(0.54)	(0.61)
Father age at 2011 Census	0.01	0.15	0.02	-0.30	0.28	-0.06	-0.68***	-0.60*

Outcome:	Physical health and wellbeing		Social competence		Emotional maturity		Language and cognitive skills (school-based)	
Sample:	Population	Siblings	Population	Siblings	Population	Siblings	Population	Siblings
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	(0.25)	(0.32)	(0.26)	(0.30)	(0.25)	(0.29)	(0.24)	(0.31)
Father age at 2011 Census squared	-0.00	-0.01*	-0.00	-0.00	-0.01**	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Father ESB migrant [(a)]	1.07*	0.94	0.41	0.23	-0.66	-0.80	1.38***	1.48**
	(0.55)	(0.62)	(0.55)	(0.61)	(0.57)	(0.63)	(0.51)	(0.58)
Father NESB migrant [(a)]	4.58***	4.22***	-1.30**	-1.73***	0.48	0.08	-4.48***	-4.79***
	(0.54)	(0.60)	(0.54)	(0.61)	(0.54)	(0.61)	(0.53)	(0.60)
Father education: Year 12 [(b)]	6.80***	7.63***	8.14***	8.18***	5.82***	5.94***	14.27***	14.15***
	(0.61)	(0.68)	(0.60)	(0.67)	(0.59)	(0.66)	(0.62)	(0.70)
Father education: Certificate [(b)]	8.25***	8.48***	8.49***	8.32***	5.72***	5.58***	12.49***	12.93***
	(0.54)	(0.60)	(0.52)	(0.58)	(0.52)	(0.57)	(0.55)	(0.62)
Father education: Diploma [(b)]	8.01***	8.91***	9.42***	9.73***	6.15***	6.14***	17.14***	18.22***
	(0.66)	(0.74)	(0.66)	(0.73)	(0.67)	(0.73)	(0.65)	(0.73)
Father education: Bachelor degree or higher [(b)]	7.97***	8.70***	13.48***	13.62***	8.76***	8.70***	21.75***	22.59***
	(0.58)	(0.64)	(0.57)	(0.63)	(0.57)	(0.63)	(0.57)	(0.64)
Family income at 2011 Census	6.26	-0.01	-10.48***	-11.00**	-6.26	-4.53	1.42	-4.36
	(4.02)	(4.51)	(3.97)	(4.44)	(3.98)	(4.43)	(4.05)	(4.62)
Mortgage homeowner [(c)]	9.31***	9.78***	8.92***	8.62***	6.63***	6.40***	10.62***	10.74***
	(0.41)	(0.47)	(0.40)	(0.46)	(0.40)	(0.45)	(0.41)	(0.47)
Outright homeowner [(c)]	10.40***	10.87***	10.50***	10.49***	8.47***	8.28***	11.39***	11.69***
	(0.60)	(0.67)	(0.60)	(0.66)	(0.60)	(0.66)	(0.59)	(0.67)
School average AEDC score	52.53***	52.09***	51.42***	50.69***	44.01***	43.62***	45.94***	45.82***
	(0.31)	(0.35)	(0.28)	(0.32)	(0.26)	(0.29)	(0.36)	(0.40)
SEIFA	-5.43***	-5.42***	-6.32***	-6.29***	-4.53***	-4.47***	-2.63***	-2.91***
	(0.27)	(0.30)	(0.26)	(0.29)	(0.26)	(0.29)	(0.27)	(0.30)
Constant	-1,125.04***	-1,083.29***	-1,118.03***	-1,057.86***	-931.70***	-887.19***	-1,229.61***	-1,198.28***
	(31.26)	(33.32)	(30.43)	(32.34)	(26.75)	(28.43)	(35.18)	(38.35)
Observations	359,245	287,422	359,245	287,422	359,245	287,422	359,245	287,422
R-squared	0.168	0.169	0.188	0.186	0.172	0.171	0.222	0.226

Notes: Results are obtained from an Ordinary Least Squared (OLS) regressions. Results (estimates and standard errors) are multiplied by 100 for aesthetic purposes. [(a)], [(b)], and [(c)] indicates "Australia born", "Year 11 or lower", and "Renter" as the respective reference groups. Additional explanatory variables include state/territory dummy variables. Robust standard errors clustered at the family level are reported in parentheses. The symbol * indicates statistical significance at 10% level, ** at 5% level, and *** at 1% level.

Appendix Table A1: Determinants of educational outcomes – Remaining results (continued)

Outcome:	Communication skills and general knowledge		ATAR		Higher education qualification	
Sample:	Population	Siblings	Population	Siblings	Population	Siblings
	(9)	(10)	(11)	(12)	(13)	(14)
Male	-27.36***	-27.06***	-8.52***	-8.55***	-12.94***	-12.97***
	(0.29)	(0.32)	(0.32)	(0.34)	(0.09)	(0.10)
Child age at 2011 Census	-2.52***	-1.87***	-14.09***	-14.35***	34.57***	34.47***
	(0.31)	(0.35)	(0.90)	(0.94)	(0.25)	(0.26)
Child age at 2011 Census squared	0.32***	0.28***	0.45***	0.46***	-1.05***	-1.04***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)	(0.01)
Child ESB migrant [(a)]	3.42***	3.38***	1.48*	0.96	-1.44***	-1.48***
	(1.02)	(1.11)	(0.84)	(0.89)	(0.27)	(0.28)
Child NESB migrant [(a)]	-15.26***	-15.38***	-5.17***	-6.71***	-0.68***	-1.01***
	(1.14)	(1.34)	(0.71)	(0.77)	(0.24)	(0.25)
Child age at AEDC time	252.05***	237.76***				
	(10.86)	(11.61)				
Child age at AEDC time squared	-20.98***	-19.71***				
	(0.96)	(1.03)				
Mother age at 2011 Census	1.58***	2.07***	7.24***	7.53***	-0.58***	-0.82***
	(0.30)	(0.41)	(0.57)	(0.65)	(0.12)	(0.14)
Mother age at 2011 Census squared	-0.02***	-0.03***	-0.07***	-0.07***	0.01***	0.01***
	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Mother ESB migrant [(a)]	-1.48***	-1.18*	-1.97***	-1.72**	-1.34***	-1.30***
	(0.56)	(0.62)	(0.63)	(0.67)	(0.19)	(0.20)
Mother NESB migrant [(a)]	-19.84***	-20.13***	13.55***	13.20***	8.19***	8.00***
	(0.55)	(0.63)	(0.59)	(0.63)	(0.19)	(0.20)
Mother education: Year 12 [(b)]	12.55***	12.59***	11.66***	11.20***	6.84***	6.57***
	(0.54)	(0.60)	(0.59)	(0.63)	(0.14)	(0.15)
Mother education: Certificate [(b)]	12.75***	12.84***	4.52***	4.13***	3.56***	3.57***
	(0.60)	(0.67)	(0.71)	(0.75)	(0.16)	(0.17)
Mother education: Diploma [(b)]	15.79***	16.10***	17.40***	16.70***	9.14***	9.10***
	(0.60)	(0.67)	(0.62)	(0.67)	(0.17)	(0.18)
	23.77***	23.94***	34.59***	34.23***	14.56***	14.38***

Mother education: Bachelor degree or higher [(b)]	(0.53)	(0.59)	(0.54)	(0.58)	(0.16)	(0.17)
Father age at 2011 Census	-0.35	-0.19	4.66***	5.27***	1.16***	1.30***
	(0.23)	(0.30)	(0.36)	(0.42)	(0.09)	(0.10)
Father age at 2011 Census squared	-0.00	-0.00	-0.05***	-0.05***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Father ESB migrant [(a)]	1.12**	0.92	-2.51***	-2.43***	-0.96***	-0.94***
	(0.51)	(0.58)	(0.60)	(0.64)	(0.18)	(0.19)
Father NESB migrant [(a)]	-14.93***	-14.71***	6.65***	5.55***	6.86***	6.62***
	(0.55)	(0.62)	(0.59)	(0.63)	(0.18)	(0.20)
Father education: Year 12 [(b)]	9.97***	10.57***	14.75***	14.32***	8.82***	8.62***
	(0.59)	(0.67)	(0.72)	(0.76)	(0.18)	(0.19)
Father education: Certificate [(b)]	11.03***	11.52***	4.24***	4.26***	3.98***	3.93***
	(0.52)	(0.58)	(0.63)	(0.67)	(0.13)	(0.14)
Father education: Diploma [(b)]	12.99***	14.21***	18.16***	17.83***	11.80***	11.60***
	(0.65)	(0.73)	(0.73)	(0.78)	(0.20)	(0.21)
Father education: Bachelor degree or higher [(b)]	16.45***	17.42***	41.70***	41.55***	18.99***	18.88***
	(0.56)	(0.63)	(0.63)	(0.68)	(0.17)	(0.18)
Family income at 2011 Census	37.51***	28.85***	20.78***	24.44***	-2.92***	-3.67***
	(4.00)	(4.52)	(3.62)	(3.93)	(0.99)	(1.07)
Mortgage homeowner [(c)]	8.27***	8.80***	11.99***	12.22***	6.64***	6.67***
	(0.40)	(0.46)	(0.59)	(0.63)	(0.13)	(0.14)
Outright homeowner [(c)]	6.20***	7.14***	29.43***	29.73***	12.84***	12.94***
	(0.60)	(0.67)	(0.65)	(0.70)	(0.17)	(0.18)
School average AEDC score	52.88***	52.84***				
	(0.24)	(0.27)				
SEIFA	-5.07***	-5.19***	16.29***	16.22***	4.08***	4.02***
	(0.26)	(0.29)	(0.29)	(0.31)	(0.08)	(0.08)
Constant	-1,179.48***	-1,152.88***	-427.02***	-444.75***	-322.93***	-319.97***
	(30.96)	(33.37)	(11.96)	(13.29)	(2.74)	(3.04)
Observations	359,245	287,422	334,592	299,811	818,848	731,159
R-squared	0.222	0.224	0.174	0.176	0.218	0.218

Notes: Results are obtained from an Ordinary Least Squared (OLS) regressions. Results (estimates and standard errors) are multiplied by 100 for aesthetic purposes. [(a)], [(b)], and [(c)] indicates "Australia born", "Year 11 or lower", and "Renter" as the respective reference groups. Additional explanatory variables include state/territory dummy variables. Robust standard errors clustered at the family level are reported in parentheses. The symbol * indicates statistical significance at 10% level, ** at 5% level, and *** at 1% level.

Appendix Table A2: Gender gap in educational outcomes: Specification without and with additional control variables

Regression at:	Mean		Percentile 20th		Percentile 40th		Percentile 60th		Percentile 80th	
Additional control:	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Physical health and wellbeing										
Male	-28.34***	-26.92***	-26.04***	-24.53***	-17.87***	-17.49***	-4.26***	-4.23***	-4.26***	-4.23***
	(0.32)	(0.30)	(0.42)	(0.41)	(0.22)	(0.22)	(0.05)	(0.05)	(0.05)	(0.05)
Observations	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245
R-squared	0.021	0.168	0.011	0.103	0.019	0.121	0.022	0.109	0.022	0.109
Panel B: Social competence										
Male	-40.43***	-39.02***	-86.12***	-83.48***	-48.95***	-48.23***	-25.98***	-25.85***	-8.14***	-8.15***
	(0.32)	(0.30)	(1.14)	(1.09)	(0.51)	(0.49)	(0.24)	(0.23)	(0.08)	(0.08)
Observations	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245
R-squared	0.042	0.188	0.029	0.115	0.034	0.119	0.031	0.107	0.023	0.081
Panel C: Emotional maturity										
Male	-51.95***	-50.58***	-85.05***	-82.78***	-46.57***	-45.72***	-50.61***	-49.91***	-20.00***	-19.81***
	(0.32)	(0.30)	(0.98)	(0.94)	(0.50)	(0.49)	(0.47)	(0.46)	(0.29)	(0.28)
Observations	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245
R-squared	0.068	0.172	0.046	0.108	0.049	0.116	0.042	0.102	0.026	0.071
Panel D: Language and cognitive skills (school-based)										
Male	-23.84***	-23.14***	-29.42***	-28.96***	-14.54***	-14.64***	-7.62***	-7.75***	-3.33***	-3.48***
	(0.32)	(0.29)	(0.55)	(0.52)	(0.23)	(0.21)	(0.11)	(0.10)	(0.08)	(0.07)
Observations	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245
R-squared	0.015	0.222	0.011	0.143	0.012	0.179	0.014	0.174	0.007	0.122

Regression at:	Mean		Percentile 20th		Percentile 40th		Percentile 60th		Percentile 80th	
Additional control:	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel E: Communication skills and general knowledge										
Male	-28.99***	-27.36***	-40.52***	-38.09***	-29.04***	-27.97***	-4.89***	-4.79***	-4.89***	-4.79***
	(0.32)	(0.29)	(0.77)	(0.71)	(0.41)	(0.38)	(0.06)	(0.05)	(0.06)	(0.05)
Observations	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245	359,245
R-squared	0.022	0.222	0.013	0.145	0.018	0.157	0.020	0.133	0.020	0.133
Panel F: ATAR										
Male	-2.82***	-8.52***	-6.45***	-13.04***	-5.30***	-12.86***	-1.63***	-8.44***	1.37***	-2.88***
	(0.35)	(0.32)	(0.68)	(0.69)	(0.51)	(0.50)	(0.44)	(0.42)	(0.36)	(0.47)
Observations	334,592	334,592	334,592	334,592	334,592	334,592	334,592	334,592	334,592	334,592
R-squared	0.000	0.174	0.000	0.099	0.000	0.117	0.000	0.119	0.000	0.091

Notes: Results for the mean and selected quantiles are estimated using OLS and UQR, respectively. Coefficients and standard errors are multiplied by 100 for presentation purposes. Additional covariates include child age at the 2011 Census (and its square), child migrant status, child age at the time of the AEDC assessment (and its square, for AEDC outcomes), maternal age at the 2011 Census (and its square), maternal education, maternal migrant status, paternal age at the 2011 Census (and its square), paternal education, paternal migrant status, total family income, residential tenure, SEIFA, school-level average AEDC score (for AEDC outcomes only), and state/territory fixed effects. Robust standard errors clustered at the family level are reported in parentheses. The symbol * indicates statistical significance at 10% level, ** at 5% level, and *** at 1% level.

Appendix Table A3: The differential effects of socioeconomic status on gender gap in educational outcomes at selected percentiles

Regression at:	Percentile 20th	Percentile 40th	Percentile 60th	Percentile 80th
	(1)	(2)	(3)	(4)
Panel A: Physical health and wellbeing				
Male	-120.57***	-8.13**	1.54*	1.54*
	(6.87)	(3.81)	(0.81)	(0.81)
Male x Mother has a Diploma,	0.74	0.19	0.14	0.14
Bachelor degree or higher	(0.86)	(0.46)	(0.10)	(0.10)
Male x Father has a Diploma,	0.70	1.33***	0.38***	0.38***
Bachelor degree or higher	(0.86)	(0.46)	(0.11)	(0.11)
Male x Family income	4.44	1.79	-0.33	-0.33
	(9.15)	(5.31)	(1.26)	(1.26)
Male x Homeowner	1.31	-0.19	-0.08	-0.08
	(1.09)	(0.52)	(0.11)	(0.11)
Male x School average AEDC score	10.67***	-1.59***	-0.77***	-0.77***
	(0.65)	(0.34)	(0.07)	(0.07)
Male x SEIFA	0.29	0.36	0.07	0.07
	(0.62)	(0.33)	(0.07)	(0.07)
Observations	359,245	359,245	359,245	359,245
R-squared	0.103	0.120	0.108	0.108
Panel B: Social competence				
Male	-327.15***	-58.74***	12.31***	17.88***
	(14.38)	(7.33)	(4.24)	(1.40)
Male x Mother has a Diploma,	4.97***	1.10	-0.10	-0.31
Bachelor degree or higher	(1.74)	(0.95)	(0.57)	(0.22)
Male x Father has a Diploma,	9.68***	4.78***	1.94***	0.43**
Bachelor degree or higher	(1.81)	(0.94)	(0.55)	(0.21)
Male x Family income	32.71*	12.32	6.83	-3.53*
	(19.82)	(10.44)	(5.73)	(2.03)
Male x Homeowner	8.13***	0.47	-1.48**	-0.36*
	(2.11)	(1.05)	(0.60)	(0.20)
Male x School average AEDC score	28.27***	-0.05	-5.05***	-3.12***
	(1.27)	(0.57)	(0.36)	(0.13)
Male x SEIFA	-1.29	0.71	0.48	0.12
	(1.22)	(0.66)	(0.39)	(0.14)
Observations	359,245	359,245	359,245	359,245
R-squared	0.115	0.117	0.106	0.082
Panel C: Emotional maturity				
Male	-317.93***	-60.90***	26.90***	36.88***
	(12.33)	(5.95)	(6.87)	(3.47)
Male x Mother has a Diploma,	1.45	0.06	-1.52	-0.51
Bachelor degree or higher	(1.48)	(0.77)	(0.94)	(0.48)
Male x Father has a Diploma,	8.70***	3.63***	2.10**	0.76*
Bachelor degree or higher	(1.43)	(0.70)	(0.93)	(0.43)
Male x Family income	-8.97	-5.61	-5.87	-6.07
	(16.72)	(8.05)	(10.39)	(4.98)
Male x Homeowner	7.63***	0.76	-1.62*	-1.78***
	(1.55)	(0.78)	(0.95)	(0.47)
Male x School average AEDC score	26.96***	1.08**	-9.87***	-7.41***
	(1.01)	(0.48)	(0.60)	(0.32)
Male x SEIFA	-0.44	0.44	0.91	0.84***
	(1.03)	(0.52)	(0.60)	(0.31)
Observations	359,245	359,245	359,245	359,245
R-squared	0.109	0.116	0.102	0.072

Regression at:	Percentile 20th	Percentile 40th	Percentile 60th	Percentile 80th
	(1)	(2)	(3)	(4)
Panel D: Language and cognitive skills (school-based)				
Male	-137.19***	-23.52***	-2.28	3.55***
	(8.09)	(3.35)	(1.65)	(1.02)
Male x Mother has a Diploma,	5.63***	1.69***	0.34	-0.17
Bachelor degree or higher	(0.97)	(0.44)	(0.23)	(0.16)
Male x Father has a Diploma,	8.27***	3.32***	1.52***	0.48***
Bachelor degree or higher	(1.01)	(0.45)	(0.23)	(0.17)
Male x Family income	0.28	1.38	1.29	2.14
	(11.91)	(5.24)	(2.46)	(1.45)
Male x Homeowner	2.28**	-0.12	-0.34	-0.21
	(1.09)	(0.50)	(0.23)	(0.14)
Male x School average AEDC score	13.92***	1.11***	-0.64***	-0.99***
	(0.66)	(0.28)	(0.14)	(0.09)
Male x SEIFA	-1.83***	-0.28	-0.06	0.13
	(0.67)	(0.32)	(0.16)	(0.10)
Observations	359,245	359,245	359,245	359,245
R-squared	0.140	0.174	0.170	0.118
Panel E: Communication skills and general knowledge				
Male	-193.31***	-42.37***	2.60***	2.60***
	(11.25)	(6.52)	(0.98)	(0.98)
Male x Mother has a Diploma,	4.68***	2.86***	-0.19	-0.19
Bachelor degree or higher	(1.15)	(0.71)	(0.13)	(0.13)
Male x Father has a Diploma,	5.60***	3.79***	0.60***	0.60***
Bachelor degree or higher	(1.27)	(0.78)	(0.12)	(0.12)
Male x Family income	20.26	5.95	-1.47	-1.47
	(14.96)	(8.05)	(1.22)	(1.22)
Male x Homeowner	3.89***	-0.38	-0.40***	-0.40***
	(1.44)	(0.84)	(0.13)	(0.13)
Male x School average AEDC score	17.25***	1.94***	-0.82***	-0.82***
	(0.89)	(0.53)	(0.09)	(0.09)
Male x SEIFA	-0.13	-0.52	-0.01	-0.01
	(0.94)	(0.58)	(0.08)	(0.08)
Observations	359,245	359,245	359,245	359,245
R-squared	0.144	0.153	0.130	0.130
Panel F: ATAR				
Male	-1.62	2.12	2.69	9.09**
	(9.28)	(8.38)	(7.05)	(4.56)
Male x Mother has a Diploma,	-6.32***	-4.55***	-4.30***	-2.77***
Bachelor degree or higher	(1.34)	(1.18)	(0.95)	(0.73)
Male x Father has a Diploma,	1.11	-2.25*	-3.38***	-1.88**
Bachelor degree or higher	(1.37)	(1.20)	(1.06)	(0.78)
Male x Family income	1.02	-9.62	-15.03	-10.32
	(12.54)	(10.61)	(9.24)	(6.82)
Male x Homeowner	-5.55***	-6.08***	-3.38***	-1.94**
	(2.10)	(1.65)	(1.24)	(0.87)
Male x SEIFA	-0.30	-0.40	-0.17	-0.61
	(0.93)	(0.83)	(0.70)	(0.46)
Observations	334,592	334,592	334,592	334,592
R-squared	0.092	0.105	0.103	0.078

Notes: Results for each percentile are estimated separately using UQR. Results (estimates and standard errors) are multiplied by 100 for aesthetic purposes. Additional covariates include child age at the 2011 Census (and its square), child migrant status, child age at the time of the AEDC assessment (and its square, for AEDC outcomes), maternal age at the 2011 Census (and its square), maternal education, maternal migrant status, paternal age at the 2011 Census (and its square), paternal education, paternal migrant status, total family income, residential tenure, SEIFA, school-level average AEDC score (for AEDC outcomes only), and state/territory fixed effects. Robust standard errors clustered at the family level are reported in parentheses. The symbol * indicates statistical significance at 10% level, ** at 5% level, and *** at 1% level.

Appendix Table A4: Robustness check - The differential effects of socioeconomic status on gender educational outcome gap using a sibling sample

Outcome:	Physical health and wellbeing	Social competence	Emotional maturity	Language and cognitive skills (school-based)	Communication skills and general knowledge	ATAR	Higher education qualification
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male	-93.69***	-145.33***	-141.31***	-124.79***	-108.84***	0.69	-7.03***
	(6.79)	(6.45)	(6.17)	(7.24)	(6.21)	(5.65)	(1.41)
Male x Mother has a Diploma, Bachelor degree or higher	1.26*	1.75**	0.42	4.21***	2.89***	-3.75***	-1.22***
	(0.73)	(0.72)	(0.74)	(0.70)	(0.71)	(0.77)	(0.24)
Male x Father has a Diploma, Bachelor degree or higher	0.73	3.45***	3.99***	6.26***	3.29***	-1.38*	1.91***
	(0.74)	(0.73)	(0.75)	(0.69)	(0.72)	(0.77)	(0.24)
Male x Family income	-0.93	3.83	-4.15	-6.61	11.36	-3.32	-11.55***
	(8.65)	(8.42)	(8.42)	(8.72)	(8.61)	(7.25)	(2.00)
Male x Homeowner	1.38	2.32***	2.78***	2.57***	1.34	-3.21***	-4.29***
	(0.85)	(0.83)	(0.82)	(0.86)	(0.84)	(1.13)	(0.25)
Male x School average AEDC score	7.26***	12.39***	10.28***	12.24***	9.70***		
	(0.65)	(0.60)	(0.54)	(0.74)	(0.51)		
Male x SEIFA	0.39	-0.23	0.09	-0.83	-0.58	-0.26	-0.15
	(0.55)	(0.53)	(0.53)	(0.55)	(0.54)	(0.56)	(0.14)
Observations	289,383	289,148	288,218	289,169	289,305	299,811	731,159
R-squared	0.168	0.187	0.172	0.222	0.222	0.158	0.206

Notes: Results are obtained from an Ordinary Least Squared (OLS) regression using a sibling sample. Results (estimates and standard errors) are multiplied by 100 for aesthetic purposes. Additional covariates include child age at the 2011 Census (and its square), child migrant status, child age at the time of the AEDC assessment (and its square, for AEDC outcomes), maternal age at the 2011 Census (and its square), maternal education, maternal migrant status, paternal age at the 2011 Census (and its square), paternal education, paternal migrant status, total family income, residential tenure, SEIFA, school-level average AEDC score (for AEDC outcomes only), and state/territory fixed effects. Robust standard errors clustered at the family level are reported in parentheses. The symbol * indicates statistical significance at 10% level, ** at 5% level, and *** at 1% level.

Appendix Table A5: Robustness check - The differential effects of socioeconomic status on gender gap in AEDC outcomes using a family fixed effects model

Outcome:	Physical health and wellbeing	Social competence	Emotional maturity	Language and cognitive skills (school-based)	Communication skills and general knowledge
	(1)	(2)	(3)	(4)	(5)
Panel A: Sibling sample					
Male	-72.94***	-136.26***	-144.15***	-123.61***	-107.72***
	(13.63)	(12.31)	(12.26)	(14.09)	(12.08)
School average AEDC score	49.22***	45.41***	37.48***	38.87***	48.12***
	(1.27)	(1.10)	(1.04)	(1.36)	(1.03)
Male x School average AEDC score	4.04***	10.59***	9.43***	10.44***	7.45***
	(1.51)	(1.30)	(1.27)	(1.53)	(1.23)
SEIFA at AEDC time	-3.74***	-5.07***	-4.54***	-5.48***	-5.15***
	(0.98)	(0.93)	(0.91)	(0.98)	(0.94)
Male x SEIFA at AEDC time	1.26	0.91	1.50*	1.35	1.93**
	(0.95)	(0.90)	(0.91)	(0.96)	(0.92)
Observations	84,151	84,151	84,151	84,151	84,151
R-squared	0.644	0.669	0.652	0.698	0.689
Panel B: Twin sample					
Male	-143.11***	-196.06***	-241.26***	-137.91***	-152.05***
	(41.52)	(39.05)	(37.85)	(42.07)	(37.53)
School average AEDC score	45.93***	54.71***	29.32***	68.58***	52.34***
	(11.60)	(7.07)	(9.53)	(9.40)	(7.84)
Male x School average AEDC score	16.13***	20.41***	22.39***	18.91***	15.38***
	(4.75)	(4.34)	(4.17)	(4.16)	(3.85)
SEIFA at AEDC time	-6.38	4.80	2.29	-4.39	2.71
	(7.56)	(7.03)	(6.99)	(5.33)	(6.58)
Male x SEIFA at AEDC time	-2.52	-1.48	0.21	-4.34	-0.65
	(3.24)	(2.89)	(3.01)	(2.95)	(2.94)
Observations	3,970	3,970	3,970	3,970	3,970
R-squared	0.732	0.746	0.730	0.787	0.781

Notes: Results are obtained from a family fixed effects regression using a sibling or twin sample. Results (estimates and standard errors) are multiplied by 100 for aesthetic purposes. Additional covariates include child age at the 2011 Census (and its square), child age at the time of the AEDC assessment (and its square), and state/territory at the time of AEDC assessment fixed effects. Robust standard errors clustered at the family level are reported in parentheses. The symbol * indicates statistical significance at 10% level, ** at 5% level, and *** at 1% level.