The evolution of the gender test score gap through seventh grade: new insights from Australia using unconditional quantile regression and decomposition

Huong Thu Le[1,2] and Ha Trong Nguyen[3*]

* Correspondence: ha.nguyen@
curtin.edu.au
[3]Bankwest Curtin Economics Centre,
Curtin University, GPO Box U1987,
Perth, WA 6845, Australia
Full list of author information is
available at the end of the article

Abstract

This paper documents the patterns and examines the factors contributing to a gender gap in educational achievements in early seventh grade of schooling using a recent and nationally representative panel of Australian children. Regression results indicate that females excel at non-numeracy subjects at later grades whereas males outperform females in numeracy in all grades, whether at the mean or along the distribution of the test score. Our results also reveal a widening gender test score gap in numeracy as students advance their schooling. Regression and decomposition results also highlight the importance of controlling for pre-school cognitive skills in examining the gender test score gap.

JEL Classification: I20, J16

Keywords: Gender, Education, Quantile regression, Decomposition, Australia

1 Introduction

Gender differentials in educational achievements have long been the focus of research. This is not surprising given that education has been shown to improve many life outcomes such as health and labour market outcomes (Card 1999; Schoeni et al. 2008). The underrepresentation of women in science, technology, engineering and mathematics (STEM) careers has resulted in research and policies focusing on gender gaps in test scores, particularly in maths-related subjects in the early years of schooling (Fryer and Levitt 2010; Justman and Méndez 2016). While there has been a rich literature on gender gaps in educational achievements, little consensus exists about the evolution as well as the factors contributing to the gaps in early childhood. One major issue plaguing researchers in documenting the evolution of the gaps is the lack of rich panel data. This study sets out to contribute to the literature by using a recent and nationally representative Longitudinal Study of Australian Children (LSAC) survey to document the evolution and examine factors contributing to gender gaps in academic achievements in early seventh grade of schooling.

This paper contributes to the international literature on the gender test score gap by not only introducing the Australian case study but also bringing three other additions to the current literature. The first addition is that with the remarkably rich panel data relative to

previous international literature—containing five assessments over the first 7 years of schooling of the same children, and an exhaustive list of home and school environments—enables the testing of several socialisation theories. For example, one of the particular advantages of the data is that pre-school cognitive skills[1] of students are observed, allowing investigation of the way that initial academic endowments contribute to the gender test score over their first 7 years of schooling. As another example, the data contain test scores of students up to the seventh grade while current US studies, which use a comparable US data set from the Early Childhood Longitudinal Study Kindergarten cohort, only examine the gender test score gap up to the fifth grade (Fryer Jr and Levitt 2004; Fryer and Levitt 2010; Sohn 2012; Bertrand and Pan 2013). These Australian data thus allow examination of the evolution of the gender test score gap through higher grades than that of the US studies.

The second addition is that this paper is one of a few papers in the literature applying a quantile regression to investigate the relative performance of male and female students along the whole distribution of test scores rather than at means (Husain and Millimet 2009; Sohn 2012; Gevrek and Seiberlich 2014). Analysis based solely on means may miss important information in other parts of the distribution (Firpo et al. 2009). This is especially relevant when policy concern is focused on the tail of the test score distribution, and when evaluating and decomposing the gender test score gap at different points of the test score distribution is of interest (Husain and Millimet 2009; Sohn 2012; Gevrek and Seiberlich 2014). To do so, this paper applies an unconditional quantile regression developed by Firpo et al. (2009). The advantage of the unconditional quantile regression over the traditional conditional quantile regression of Koenker and Bassett (1978) is that its estimates can be interpreted as the impact of changes in explanatory variables on the dependent variable for those at a specific point in the distribution.[2] The estimates from the unconditional quantile regression can then be directly applied to an Oaxaca-Blinder (OB) decomposition method to examine factors contributing to the gender test score gap across the entire distribution. Therefore, this study makes its third addition to the literature as one of a few papers (Sohn 2012; Gevrek and Seiberlich 2014) applying a quantile decomposition method to study the gender test score gap.

By using the first five waves of the LSAC survey, we find that males excel at numeracy at all grades, whether at means or along the distribution. Also, we uncover heterogeneous patterns in the gender test score gap across the test score distribution, by test subjects and test grades. The regression results also reveal a widening gender test score gap in numeracy as students advance their schooling. The decomposition results indicate that gender disparities in pre-school cognitive skills can explain a large part of the differences in academic performance.

The remainder of the paper is structured as follows. Section 2 summarises the most relevant literature while Section 3 describes the data. Section 4 presents this study's empirical regression and decomposition models and Section 5 discusses the regression results. Section 6 reports decomposition results of factors contributing to the gender test score gap, and, finally, Section 7 concludes.

2 Literature review

International literature has consistently shown significant gender test score gaps, with male students generally outperforming female students in maths and science while female students excel at literacy subjects (Wilder and Powell 1989; Marks 2008; Bedard and Cho

2010; Fryer and Levitt 2010; Christopher et al. 2013; Falch and Naper 2013; Stoet and Geary 2013; Dickerson et al. 2015). In addition, studies have often documented that the gender gap in a particular subject only appears at certain educational levels and tends to increase as students advance their schooling (Coleman et al. 1966; Husain and Millimet 2009; Fryer and Levitt 2010).

Research that has been devoted to attempting to explain the recognised patterns in the gender educational gap has proposed a wide range of different contributing factors. For example, some studies have demonstrated that differences in the brain between genders may explain these patterns as males tend to be better at analysing systems, while females tend to be better at reading the emotions of other people (Kimura 2000; Baron-Cohen 2007). Furthermore, gender differences in competition (Gneezy et al. 2003; Niederle and Vesterlund 2010), parental time investment in children (Baker and Milligan 2016), or social and cultural conditioning and gender-biased environments (Guiso et al. 2008; Bedard and Cho 2010; Dickerson et al. 2015) are possible explanations for the observed gender gaps in academic achievements. An emerging number of studies also highlight the roles of non-cognitive skills (Jacob 2002; Duckworth and Seligman 2006; Christopher et al. 2013; Golsteyn and Schils 2014) in contributing to the gender test score gap.[3] This present paper contributes to the literature by assessing the role of pre-school cognitive skills in contributing to the gender academic achievement gap and how that role evolves as students advance in their schooling.

Australian studies have documented gender differences in academic outcomes at all educational levels. For example, Nghiem et al. (2015) used the first four waves of the LSAC data to report that male students outperform their female counterparts in grade 3 and 5 numeracy. In contrast, female students outperform in grade 3 writing and grade 5 reading and grammar. More recently, Justman and Méndez (2016) used administrative data from Victoria to show that male students score higher than female students in mathematics and lower in reading in grades 7 and 9. As another example, Marks (2008) used the OECD's 2000 Programme for International Student Assessment (PISA) project to document that 15-year-old Australian females perform better than males in reading but worse in mathematics. Using various datasets, Homel et al. (2012) reported that 18-year-old Australian females are more likely to complete Year 12 than males. At the tertiary educational level, Booth and Kee (2011) used aggregate data to report that since 1987 Australian females were more likely than males to be enrolled at university. These studies often attempt to capture the gender educational achievement gap by including a gender dummy variable in a multivariate regression framework and only examine the mean gap.

3 Data and descriptive statistics

3.1 Data and sample

We use data from the first five waves of the biannual national representative LSAC survey. The LSAC, initiated in 2004, contains comprehensive information about children's test scores and other socio-economic and demographic background of the children and their parents. The LSAC sampling frame consists of all children born between March 2003 and February 2004 (the birth or "B cohort", infants aged 0–1 year in 2004), and between March 1999 and February 2000 (the kindergarten or "K cohort", children aged

4–5 years in 2004). In this study, children of K cohort are used because measures on student test scores are more widely available for this cohort in the first five waves of the survey.

To indicate the academic achievements of students, we employ results from the National Assessment Program – Literacy and Numeracy (NAPLAN) tests.[4] The NAPLAN test is required of all Australian students in grades 3, 5, 7 and 9 in the five domains of reading, writing, spelling, grammar and numeracy. The test scores range from 0 to 1000 and are comparable across students and over time (ACARA 2014). The NAPLAN test results of the children were collected via data linkage with the LSAC data (Daraganova et al. 2013). At the time of this study, the linkage data for LSAC were mainly available for students in grades 3, 5 and 7. Thus, we employ these test results at these grades to measure the academic achievements of students. Following the previous Australian literature (Justman and Méndez 2016; Cobb-Clark and Moschion 2017) and for brevity purposes, we focus on two main test subjects: reading and numeracy.[5] Since the NAPLAN test dates and LSAC survey dates are not the same, test results and survey data are merged in the way that test results are not pre-dated by survey data.[6] This matching exercise shows that NAPLAN test scores in grades 3, 5 and 7 are merged with survey data in waves 2, 3 and 4, respectively. As is generally done in the literature (Husain and Millimet 2009; Fryer and Levitt 2010; Sohn 2012; Golsteyn and Schils 2014), NAPLAN test scores are standardised (with mean 0 and standard deviation 1) by grade and domain in this paper.

To measure the initial stocks of students' cognitive skills, we use the Peabody Picture Vocabulary Test (PPVT) and Who Am I (WAI). The PPVT is an interviewer-administered test to assess a child's knowledge of the meaning of spoken words and his or her receptive vocabulary for standard English (Dunn and Dunn 1997). The PPVT test requires a child to show the picture that best represents the meaning of a stimuli word spoken by the examiner. The WAI test is also administered by an interviewer to measure the general cognitive ability of pre-school age children to perform literacy and numeracy tasks, such as reading, copying and writing letters, words, shapes and numbers (Lemos and Doig 1999). PPVT and WAI scores are used in wave 1 when the student is 4 or 5 years old (i.e., before enrolling in primary school). Similar to NAPLAN test scores, PPVT and WAI test scores are standardised for ease of interpretation.

3.2 Sample
As discussed in Section 3.1, this study focuses on K cohort children because test scores are more widely available for them. Furthermore, among students who took any test in any test grade, the focus is on about 96% of those who completed all five test subjects. Moreover, the sample is restricted to students without missing information on a list of important explanatory variables. To keep the results comparable over time, specifications that use variables which are available in all waves of the LSAC and contain the least missing information (see Table 1 and Section 4 for a list of variables included in our baseline models) are used. These variables are commonly used in studies which employ a popular and comparable US data set from the Early Childhood Longitudinal Study Kindergarten cohort (Fryer Jr and Levitt 2004; Fryer and Levitt 2010; Sohn 2012; Bertrand and Pan 2013) to study a gender test score gap of school students.[7]

Table 1 Summary statistics by gender

Variables	Males	Females	Males-females
Child age	106.17	107.03	− 0.86**
Native	0.97	0.96	0.00
Aboriginal	0.02	0.03	− 0.01*
Low birth weight	0.06	0.07	−0.01**
Breastfeed	0.73	0.76	−0.03***
Mother age	38.83	39.31	−0.48***
Mother native	0.65	0.65	0.00
Mother NESB	0.20	0.19	0.01
Mother ESB	0.15	0.15	0.00
Mother has no qualification	0.27	0.27	0.00
Mother has a certificate	0.30	0.30	0.00
Mother has an advanced diploma	0.11	0.09	0.02***
Mother has bachelor degree	0.17	0.18	−0.01
Mother has graduate diploma	0.07	0.08	−0.01
Mother has postgraduate degree	0.07	0.08	−0.01
Mother's weekly working hours	19.39	20.13	−0.73**
Home environment index	1.37	1.35	0.02
Out-of-home activity index	2.61	2.65	−0.04
Having a computer at home	0.93	0.94	−0.01
Public school	0.65	0.65	0.00
Catholic school	0.23	0.22	0.01
Other independent school	0.12	0.13	−0.01
Household size	4.61	4.57	0.04*
Number of siblings	1.63	1.60	0.03
Number of younger siblings	0.81	0.72	0.09***
Number of same age siblings	0.02	0.03	−0.01**
Living with both parents	0.82	0.82	0.00
Living in an owned home	0.77	0.78	−0.01
Household income	91.96	92.12	−0.16
Initial PPVT (s.d.)	−0.08	0.08	−0.16***
Initial WAI (s.d.)	−0.31	0.32	−0.64***

Notes: Statistics are calculated from all waves (pooled sample size: 8497 observations). Analysing each wave separately reveals similar patterns. Statistics are adjusted for sampling weights. Tests are performed on the significance of the difference between the sample mean for male and female students
Significance at the *10% level, **5% level, and ***1% level

The original sample sizes for the K cohort in waves 2, 3 and 4 are 4464, 4331 and 4169, respectively. The above restrictions result in final samples of 2471, 3225 and 2801 students in waves 2, 3 and 4, respectively. Appendix 1: Table 6 suggests that sample attritions are mainly attributed to the fact that students' NAPLAN test scores are not linked to the LSAC data. Reasons for original sample attrition are discussed in Norton and Monahan (2015), and seasons for not having NAPLAN test scores linked to the LSAC data are discussed in detail in a technical report by Daraganova et al. (2013). Note that there is a slightly smaller number of students in wave 2 in this sample because the grade 3 NAPLAN tests were first introduced in 2008 when some K cohort students might have attended higher grades, and as such did not take the tests.

Additionally, Appendix 1: Table 6 reveals that, conditional on having NAPLAN test scores linked to the LSAC data, sample attritions are mostly due to missing information on pre-school cognitive skills (i.e. PPVT and WAI) and household income. We dropped individuals with missing information on control variables rather than using the "dummy variable adjustment" method because deletion has been found to produce less-biased estimates (Allison 2001).

We investigate whether our sample selection criteria led to sample selection issues. One particular concern relating to our research design is that the child's gender may affect the probability that an individual child is included in the final sample. Therefore, we ran a probit model where the dependent variable is equal to one if the child is in our sample and zero otherwise. The explanatory variables are basic demographic characteristics, including the child's gender. Regression results (reported in Appendix 1: Table 7) suggest some evidence of statistically significant selection on some observables. For instance, children in our sample are more likely to come from more advantageous households with non-Aboriginal or native backgrounds or come from two-parent households or live in owned homes. However, the pseudo-R^2 values are relatively small, indicating that selection on observable characteristics is quantitatively weak. More importantly, in two out of three regressions by test grades, p values from a t test for statistical significance of the gender dummy included in the regression are greater than 0.05, alleviating concern that our results may be driven by sample selection.

3.3 Summary statistics by gender

Summary statistics by gender for students' background characteristics and home environment variables that are used in the analysis are presented in Table 1. Insignificant gender differences in parental characteristics (such as mother's ethnicity, education, work status, family size, income and home ownership status) suggest that the gender of children in this sample is randomly assigned across families.[8] There is also no significant difference in most of our measures of parental investment in child development, such as parental time with the child, children's access to computers or school sectors. The only distinguishable gender difference is that female students were more likely to be breast fed at 3 or 6 months old.

However, significant gender differences in terms of initial cognitive and health endowments are noticed. In particular, female students have an academic advantage even before they start their school years because their PPVT or WAI scores, measured at ages 4 or 5, are higher than male students of the same age. Our finding of a female advantage in pre-school reading test scores (as represented by PPVT) is consistent with that presented in the work by Fryer and Levitt (2010) for children in the USA. We additionally show that at ages 4 or 5 girls also display higher general cognitive ability (as measured by WAI) than boys.[9] In line with the Australian national birthweight pattern by gender reported in the medical literature (Dobbins et al. 2012), our data also show that female students are generally smaller than male students at birth, with females more likely to have birth weight of 2500 g or lower. Similarly, we observe female students in the

sample are slightly older (1 month) than male students. This gender difference is consistent with a pattern, observed in Table 1, that girls' mothers are about 4 months older than boys' mothers. Lastly, while male students appear to have a greater number of younger siblings than female students, the former have a lower number of same age siblings.

Table 1 displays that significant differences in verbal and general cognitive performance exist between boys and girls by the time they enter primary schools. Similar to the reasons behind the gender disparity in educational achievements discussed in Section 2, the origin of gender differences in pre-school cognitive skills remains largely unknown. Some suggest differences are due to the role of biological gender differences (Vandenberg 1967) while others suggest different treatments and expectations from parents or teachers may lead to pre-school gender cognitive differences (Lewis and Freedle 1972; Block 1976; Lewis and Brooks-Gunn 1979; Lavy and Sand 2015; Baker and Milligan 2016).

To have some ideas about how pre-school cognitive skills are formed, in a purely descriptive way, we follow the child development literature to run a regression of each of them (i.e. PPVT and WAI) on a list of factors contributing to the child's development (Currie 2009; Cunha et al. 2010). The list includes child characteristics (i.e. gender, age, ethnicity), early child outcomes (as measured by child birth weight), early parental investment (as measured by breastfeeding the child at 3 or 6 months), concurrent parental investment (as represented by a home environment index, an out-of-home activity index and access to computers)[10] and family environment (maternal age, migration background, health, number of siblings, maternal working hours, family income and living with both parents). The results (reported in Appendix 1: Table 8, column 1) show higher pre-school PPVT test scores are observed for girls, older children, children with normal birth weight, children of native or highly educated mothers, or children with more early or concurrent investment from parents. Appendix 1: Table 8 (column 2) additionally conveys that the characteristics associated with higher PPVT test scores are also factors explaining higher WAI test scores among 4- or 5-year-old children. An exception is that children of mothers migrating to Australia from a Non-English Speaking Background (NESB) country have higher WAI scores than children of native mothers. Overall, the results from this exercise highlight that significant differences in cognitive skills between boys and girls already exist before entering school and that pre-school cognitive skills may measure intergenerational genetic transmission or accrued parental investment in child development prior to school.

4 Empirical models

4.1 Regression models

Prior literature methods are followed to estimate the gender test score gap by regressing test scores (Y_i) of student i in each test grade and each subject on the gender dummy variable (Male$_i$ which takes the value of 1 if the student is male and 0 if female); therefore, the sign and magnitude of the gender coefficient estimate indicates the direction and magnitude of the gender test score gap. The

changes in the gender test score gaps estimated over the three school grades describe the evolution of the gender test score gap from grade 3 of primary school to either the final grade of primary school or the first grade of secondary school.[11] In particular, for each test subject and each test grade, the raw gender test score gap is estimated using the following basic model:

$$Y_i = \alpha + \beta \text{Male}_i + \varepsilon_i \tag{1}$$

where ε_i represents idiosyncratic error terms.

In addition to the raw test score gap, the gender test score gap conditional on a rich list of factors contributing to the student's development is examined using the following equation:

$$Y_i = \alpha + \beta \text{Male}_i + X_i \gamma + \varepsilon_i \tag{2}$$

where X_i include the student's characteristics (i.e. age, ethnicity, health status), household characteristics (i.e. mother's migration status,[12] household size, parents' education, and household income), indicators of the parental investment in the student's education (e.g. breastfeeding the child at 3 or 6 months, access to computers, and two indices of "quality time" that parents and children spend together), and indicators of neighbourhood characteristics (i.e. physical infrastructure or neighbourhood social-economic status). The issues of students sitting the NAPLAN test in different years for the same grade are addressed by using information both on the age of students at the year they sat the test and dummy variables for the test year. The differences in the survey time and test time are controlled for by including the dummies for quarters of survey time in regressions. In model 2, state dummy variables are included to control for differences in educational jurisdictions by states/territories.

The marginal gender test score gap after students entered primary schools is then examined by including the student's initial stock of academic ability as indicated by scores on WAI and PPVT tests (E_{0Ki}), which are administered prior to primary school entry, using the following "value-added" model:

$$Y_i = \alpha + \beta \text{Male}_i + X_i \gamma + E_{0Ki} \theta + \varepsilon_i \tag{3}$$

The value-added model is our preferred specification because it is in line with the dynamic theory of skill formation (Todd and Wolpin 2007; Cunha et al. 2010). As discussed in Section 3.3, pre-school cognitive skills may measure accrued parental investment in child development prior to primary school, so use of the value-added model also helps isolate effects of such investment on the gender test score gap observed during primary and early secondary school years.[13]

The ordinary least squares (OLS) method is first applied to estimate the mean gender test score gap using the three specifications described above. Unreported statistics from our data show that for both males and females the mean test score is usually not the same as the median, suggesting that the test score distribution is skewed and contains extreme values. This distributional characteristic suggests the need for examining the determinants of academic achievement not only at the mean but also along the whole distribution (Koenker and Bassett 1978; Firpo et al. 2009). The unconditional

quantile regression (UQR) technique is employed to investigate the gender test score gap along the entire distribution.

This technique is chosen over the (conditional) quantile regression method proposed by Koenker and Bassett (1978) because the latter does not allow interpretation of its estimates as the marginal impact of an explanatory variable on the outcome of interest unless the rank-preserving condition holds (Firpo 2007; Firpo et al. 2009). In contrast, the unconditional quantile regression technique introduced by Firpo et al. (2009) does. Technically, the unconditional quantile regression method runs a regression of the estimated re-centered influence function (RIF) on a set of explanatory variables (Firpo et al. 2009).[14] The *RIF* for the quantile of interest q_τ is:

$$\mathrm{RIF}(Y, q_\tau) = q_\tau + \frac{\tau - D(Y \le q_\tau)}{f_Y(q_\tau)}, \qquad (4)$$

where $f_Y(q_\tau)$ is the marginal density function of an outcome Y, and D is an indicator function. In practice, $\mathrm{RIF}(Y, q_\tau)$ is not observed so its sample counterpart is used instead:

$$\mathrm{RIF}(Y, \hat{q}_\tau) = \hat{q}_\tau + \frac{\tau - D(Y \le \hat{q}_\tau)}{\hat{f}_Y(q_\tau)}, \qquad (5)$$

where $\hat{q}_\tau$ is the sample quantile and $\hat{f}_Y(q_\tau)$ is the kernel density estimator. As mentioned above, one crucial distinguishing feature of the UQR method is that it provides a way to recover the marginal impact of the explanatory variables on the unconditional quantile of Y. Another appealing feature of the UQR method is that its regression results can be applied directly to an OB decomposition method to examine factors contributing to the gender test score gap across the whole distribution without having to implement many simulations that are necessary in the alternative quantile regression-based decomposition method.

4.2 Decomposition models

The factors contributing to the male-female test score gap at the mean and at selected percentiles are examined by following the literature on gender wage gaps (Blinder 1973; Oaxaca 1973; Fortin et al. 2011) in applying an OB type of decomposition of the form:

$$\hat{Y}_m - \hat{Y}_f = \underbrace{(\hat{Z}_m - \hat{Z}_f)\hat{\mu}^*}_{\text{``charateristic effect''}} + \left\{ \underbrace{\hat{Z}_m(\hat{\mu}_m - \hat{\mu}^*) + \hat{Z}_f\left(\hat{\mu}^* - \hat{\mu}_f\right)}_{\text{``return effect''}} \right\} \qquad (6)$$

where $\hat{Y}$ is the mean test score of males (m) or females (f), $\hat{Z}$ is a vector of the mean observed characteristics, $\hat{\mu}_m$ ($\hat{\mu}_f$) is a vector of the estimated coefficients in the regression of test score on the set of covariates, including the constant, for male (female) sample and $\hat{\mu}^*$ is a vector of the estimated coefficients from the pooled male and female sample with other covariates and the gender dummy. The gender dummy variable is included in estimating the reference structure $(\hat{\mu}^*)$ to obtain unbiased estimates of other variables (Neumark 1988; Fortin 2008; Jann 2008).[15]

In Eq. (6), the first term on the right-hand side is the component of the gender test score gap due to differences in observed characteristics—the "characteristic effect". The second term on the right hand-side is the difference in factors other than the observed characteristics—the "return effect", sometimes interpreted as "unexplained" or "discrimination". We focus on detailed decomposition of the characteristic effect because it is well-known that detailed decomposition results of the return effect are influenced by the arbitrary scaling of continuous variables (Jones 1983; Jones and Kelley 1984). To facilitate an interpretation of the results, variables contributing to the academic achievement of students are separated into four groups: (1) their characteristics, (2) their families' characteristics, (3) their initial cognitive skill endowments, and (4) other factors.

5 Empirical regression results

5.1 Estimates of gender test score gap at means of test score distribution

Estimates on gender test score gaps at means in reading and numeracy over the three grade levels (3rd, 5th, and 7th) from three specifications are reported in Table 2. Raw gender test score gaps at means (estimated from model 1, see the first row of each subject panel in Table 2) show the well-known gender gaps in both maths and reading skills as observed in the literature: male students outperform female students in maths but lag behind with respect to reading (Husain and Millimet 2009; Fryer and Levitt 2010; Nghiem et al. 2015; Justman and Méndez 2016). Furthermore, while the gender test score gap in reading is already observed in all grades, the (reverse) gender gap in numeracy only presents in grades 5 and 7. The finding of the gender test score gap in numeracy in favour of male students only being present at certain educational levels is also in line with previous US findings in that a gender maths score gap was only observed for US students at their first (Husain and Millimet 2009) or third grade tests

Table 2 Estimated gender score gap over the grades at mean

Subject	Model	Grade 3	Grade 5	Grade 7
Reading	(1)	−0.13***	− 0.23***	−0.22***
		(0.04)	(0.03)	(0.04)
	(2)	−0.13***	−0.21***	− 0.20***
		(0.04)	(0.03)	(0.03)
	(3)	0.07**	−0.03	−0.06*
		(0.04)	(0.03)	(0.03)
Numeracy	(1)	0.00	0.15***	0.15***
		(0.04)	(0.04)	(0.04)
	(2)	0.01	0.16***	0.17***
		(0.04)	(0.03)	(0.03)
	(3)	0.22***	0.38***	0.39***
		(0.04)	(0.03)	(0.03)

Notes: Females are the base group. Each estimate is obtained from a separate regression. Model 1 includes gender dummy only. Model 2 includes student characteristics (gender, age, Aboriginal status, and birth weight), household characteristics (mother's characteristics (age, migration background, completed qualification, and working hours), having computer at home, home environment index, out-of-home activity index, household size, number of siblings, living with both biological parents, living in an owned home, household income, and school sector), test states, test years, urban, local socio-economic background variables, and survey quarters. Model 3 includes all variables as in model 2 plus pre-school PPVT and WAI
Significance at the *10% level, **5% level, and ***1% level

(Fryer and Levitt 2010).[16] It is, however, interesting to note that while these raw figures suggest that a gender maths score gap only appears at a certain grade, it takes from two to four more years to observe this pattern in Australia. Table 2 additionally indicates that the raw gender test score gaps in reading and numeracy increase from grade 3 to grade 5 and are quite stable in both grades 5 and 7.
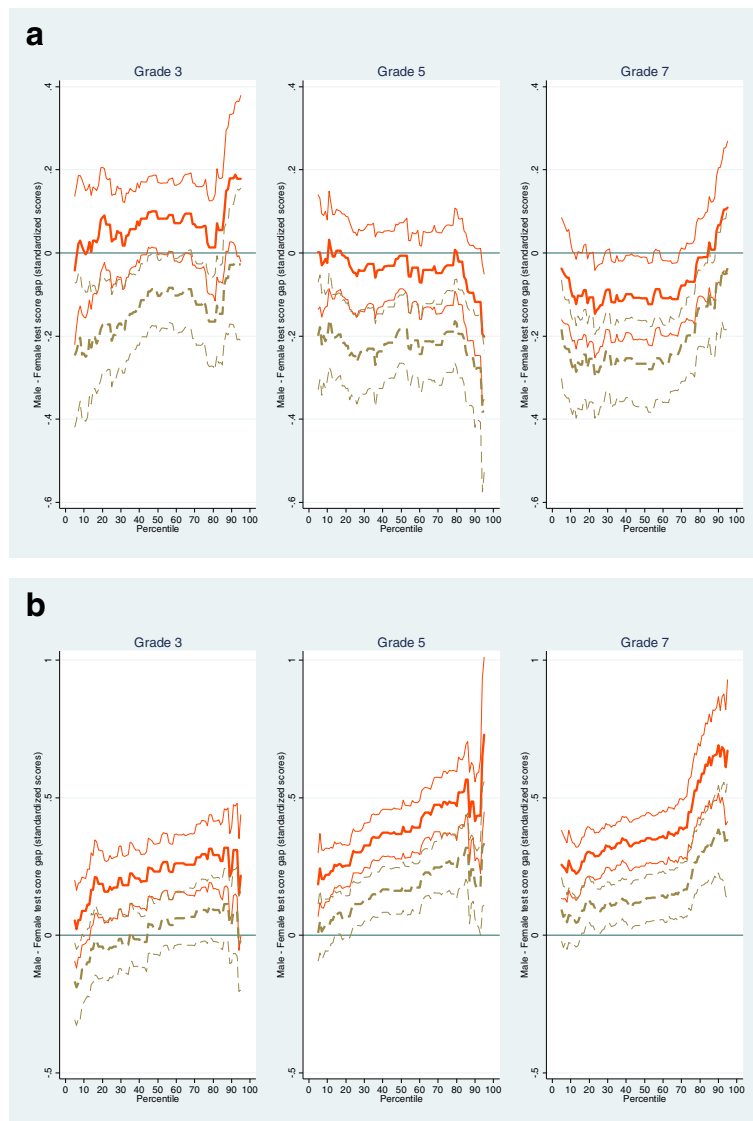
The gender test score gaps estimated from model 2 suggest that adjusting for a comprehensive list representing characteristics of students, their families and their neighbourhood does not change the earlier findings in terms of the magnitude as well as the statistical significance level. However, additionally including students' WAI and PPVT tests measured at ages 4 or 5 in the regression model 3 does. In particular, a reversed and statistically significant (at the 5% level) gender test score gap is observed in favour of male students in third grade reading, where male students outperform female students by about 0.07 standard deviations. Furthermore, the observed gender test score gap in grades 5 and 7 reading turns from statistically significant in model 2 to insignificant in model 3. In contrast, controlling for students' prior academic endowment turns the gender test score gap in numeracy in favour of male students from statistically insignificant to highly significant (at the 1% level) in grade 3 and substantially increases (by more than double) the magnitude of the gap in all studied grades.

In summary, the above results suggest that including pre-school cognitive skills in students' development equations shrinks the gender gap in reading while widening the gender gap in numeracy in terms of the statistical significance level and magnitude. This finding is consistent with our previously observed pattern of girls having higher pre-school cognitive skills. Estimates of the above gender test score gaps also highlight the importance of controlling for students' pre-school cognitive skills, which is the summary of genetic and early childhood investment in the formation of human capital, in the student development as shown in the literature (Todd and Wolpin 2007; Bernal 2008; Cunha et al. 2010; Lai 2010; Elder and Jepsen 2014; Fortin et al. 2015; Nghiem et al. 2015). As previous studies in this literature were unable to control for pre-school cognitive skills—due to the unavailability of such measures in the researchers' data sets—this is a novel empirical result.

The estimated gender test score gaps, where statistically significant, are largely in line with international literature; however, the gender gap in a particular subject only appears at certain educational levels and tends to increase as students progress through school (Coleman et al. 1966; Husain and Millimet 2009; Fryer and Levitt 2010). Our results additionally show that the pattern of a widening gender test score gap as students advance through school persists even conditioning on pre-school cognitive skills. Two observations from the full results of test score regressions (reported in Appendix 1: Tables 9 to 11) help explain why including pre-school cognitive scores does not change the above observed pattern. First, the impact of pre-cognitive skills on subsequent academic achievements is relatively stable across school grades, so including pre-cognitive skills which are in favour of females in the regressions tends to change the estimate of the male dummy by the same magnitude. Second, including pre-school cognitive skills in the test score regressions while improving the explanatory power of all included explanatory variables leaves a substantial part of students' academic achievements unexplained (the maximum R^2 is 0.35, as shown in Appendix 1: Tables 9 to 11).

5.2 Estimates of gender test score gap along the test score distribution

We next explore the heterogeneity in gender test score gaps over the distribution of student performance. Figure 1 succinctly represents estimates of gender test score gaps (the thick solid orange line) and their respective 95% confidence intervals[17] (the thin solid orange line) along the test score distribution for reading and numeracy. While the value-added estimates are the focus of this analysis, Fig. 1 also reports gender test score gap estimates (the thick dotted brown line) for comparison purposes and their corresponding 95% confidence intervals (the

Notes: Gender test score gap: Male - Female, standardized scores. Estimates are based on unconditional quantile regressions. Thick (thin) solid orange line indicates test score gap estimates (95 % confidence intervals) using model 3. Thick (thin) dotted brown line shows test score gap estimates (95 % confidence intervals) using model 2. Confidence intervals are obtained using 500 bootstrap replications.

Fig. 1 Gender test score gaps along the distribution by test subject and grade. Panel **a**: Reading. Panel **b**: Numeracy

thin dotted brown line) obtained using regression model 2, which does not include initial endowment in cognitive skills.

Value-added estimates for gender reading test score gaps (panel A, Fig. 1) show male students' statistically significant advantage in grade 3 reading observed earlier at means may have been driven by those in the middle (around the 50th percentile) or top (above the 90th percentile) of the distribution because estimates are statistically significant at these percentiles only. In contrast, females statistically significantly outperform males in grade 7 reading roughly around the median of the distribution. Thus, despite the mean test score gap being statistically indistinguishable from zero, the distributional investigation suggests female students' statistically significant advantage in grade 7 reading. However, statistically significant differences in reading scores by gender are not observed at any other remaining percentiles or test grades. Also it is noted that controlling for pre-school cognitive skills reduces the gender reading test score gap favouring female students in terms of the magnitude and statistical significance in nearly all percentiles.

Turning to value-added estimates on a gender test score gap in numeracy (panel B, Fig. 1), males outperform females over virtually the whole distribution and in all grades. Additionally, the gender numeracy test score gap is more pronounced at the upper end of the distribution. A widening gender test score gap in numeracy is also observed as students advance through school. Furthermore, the steeper slope of the gender test score gap line at the higher end of the distribution (more visible for grades 5 and 7) suggests that the observed widening gender numeracy test score gap favouring male students may have been driven by top performing students. Finally, including students' pre-school cognitive ability is found to increase the gender numeracy test score gap favouring male students in terms of magnitude and statistical significance.

In summary, the above analysis of the gender test score gap across the distribution indicates that focusing on mean gap could overlook important policy relevant heterogeneity across the distribution. Furthermore, this analysis highlights the importance of controlling for pre-school cognitive skills in analysing the gender test score gap. In particular, the results from quantile regressions indicate that controlling for pre-school cognitive skills closes down the gender gap favouring females in reading, while increasing the gender gap favouring males in numeracy, and this pattern holds at all points of the test score distribution.

6 Empirical decomposition results

We next discuss about the decomposition results using the methods outlined in Section 4.2. Tables 3 and 4 report the estimated total male-female test score gap, together with its contributing factors at the mean and selected percentiles, separated by grades for reading and numeracy, respectively. Figure 2 displays concise estimates of total gender test score gap (with their 95% confidence intervals) and the characteristic and return effect along the whole distribution for reading and numeracy.[18] Estimates of the total gender gap (results are reported on the first row of Tables 3 and 4) are largely similar to those obtained from regression model 1 (results are reported in Table 2 and Fig. 1). Tables 3 and 4 show that the estimated total gender gaps are statistically insignificant at some points of the test score

Table 3 Contributions to the male-female test score gap at mean and selected percentiles by grade—reading

	Grade 3				Grade 5				Grade 7			
	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total gap	−0.26***	−0.11**	0.00	−0.13***	−0.22***	−0.21***	−0.28***	−0.23***	−0.29***	−0.22***	−0.06	−0.22***
Explained part												
Child	−0.02	−0.01	−0.00	−0.01	−0.01	−0.01	0.00	−0.01	−0.02*	0.00	0.00	−0.01
	[8]	[9]	[n/a]	[8]	[5]	[5]	[0]	[4]	[7]	[0]	[0]	[5]
Household	0.00	0.00	0.00	0.01	−0.01	−0.01	−0.02	−0.01	−0.01	−0.01	−0.00	−0.01
	[0]	[0]	[n/a]	[−8]	[5]	[5]	[7]	[4]	[3]	[5]	[0]	[5]
Others	0.01	0.01	0.02	0.01	0.02	0.02**	0.04**	0.03***	0.01	0.02	0.01	0.01
	[−4]	[−9]	[n/a]	[−8]	[−9]	[−10]	[−14]	[−13]	[−3]	[−9]	[−17]	[−5]
Initial	−0.26***	−0.20***	−0.22***	−0.22***	−0.19***	−0.20***	−0.19***	−0.20***	−0.18***	−0.17***	−0.15***	−0.15***
	[100]	[182]	[n/a]	[169]	[86]	[95]	[68]	[87]	[62]	[77]	[250]	[68]
Total	−0.27***	−0.19***	−0.20***	−0.21***	−0.20***	−0.20***	−0.16***	−0.20***	−0.20***	−0.16***	−0.14***	−0.16***
	[104]	[173]	[n/a]	[162]	[91]	[95]	[57]	[87]	[69]	[73]	[233]	[73]
Unexplained part												
Child	−0.68	−2.07	2.67	−0.85	−1.37	−2.33	4.38	−1.19	−3.72	−4.67*	−6.41*	−4.04**
	[262]	[1882]	[n/a]	[654]	[623]	[1110]	[−1564]	[517]	[1283]	[2123]	[10683]	[1836]
Household	−0.48	−0.09	1.24	0.46	0.47	−0.04	−1.13	−0.09	0.13	0.84	1.11	0.42
	[185]	[82]	[n/a]	[−354]	[−214]	[19]	[404]	[39]	[−45]	[−382]	[−1850]	[−191]
Others	6.32	−0.20	−5.99	−2.03	6.74	−1.58	−0.58	−0.40	1.35	−0.81	−4.51	0.28
	[−2431]	[182]	[n/a]	[1562]	[−3064]	[752]	[207]	[174]	[−466]	[368]	[7517]	[−127]
Initial	0.02	−0.00	0.00	0.00	−0.00	−0.01	0.00	−0.00	0.01	−0.02	−0.00	−0.00
	[−8]	[0]	[n/a]	[0]	[0]	[5]	[0]	[0]	[−3]	[9]	[0]	[0]
Constant	−5.17	2.44	2.28	2.49	−5.86	3.95	−2.79	1.66	2.15	4.59	9.89*	3.28

Table 3 Contributions to the male-female test score gap at mean and selected percentiles by grade—reading (Continued)

	Grade 3				Grade 5				Grade 7			
	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	[1988]	[−2218]	[n/a]	[−1915]	[2664]	[−1881]	[996]	[−722]	[−741]	[−2086]	[−16,483]	[−1491]
	0.01	0.08*	0.20***	0.07**	−0.02	−0.01	−0.11*	−0.03	−0.08	−0.06	0.08	−0.06*
Total	[−4]	[−73]	[n/a]	[−54]	[9]	[5]	[39]	[13]	[28]	[27]	[−133]	[27]

Notes: Females are the base group. Standard errors (not reported for brevity) are obtained using 500 bootstrap replications. Estimates from model 3 are used. Values in squared brackets are percentage of the estimated total gap. Percentages may not add up to 100% due to rounding. "n/a" denotes "not applicable" because of dividing by zero. Grouped variables: Student: age, Aboriginal status, and birth weight; Household: mother's characteristics (age, migration background, completed qualification and working hours), having computer at home, home environment index, out-of-home activity index, household size, number of siblings, living with both biological parents, living in an owned home, household income, and school sector; Others: test states, test years, urban, local socio-economic background variables, and survey quarters; Initial: pre-school PPVT and WAI

Significance at the *10% level, **5% level, and ***1% level

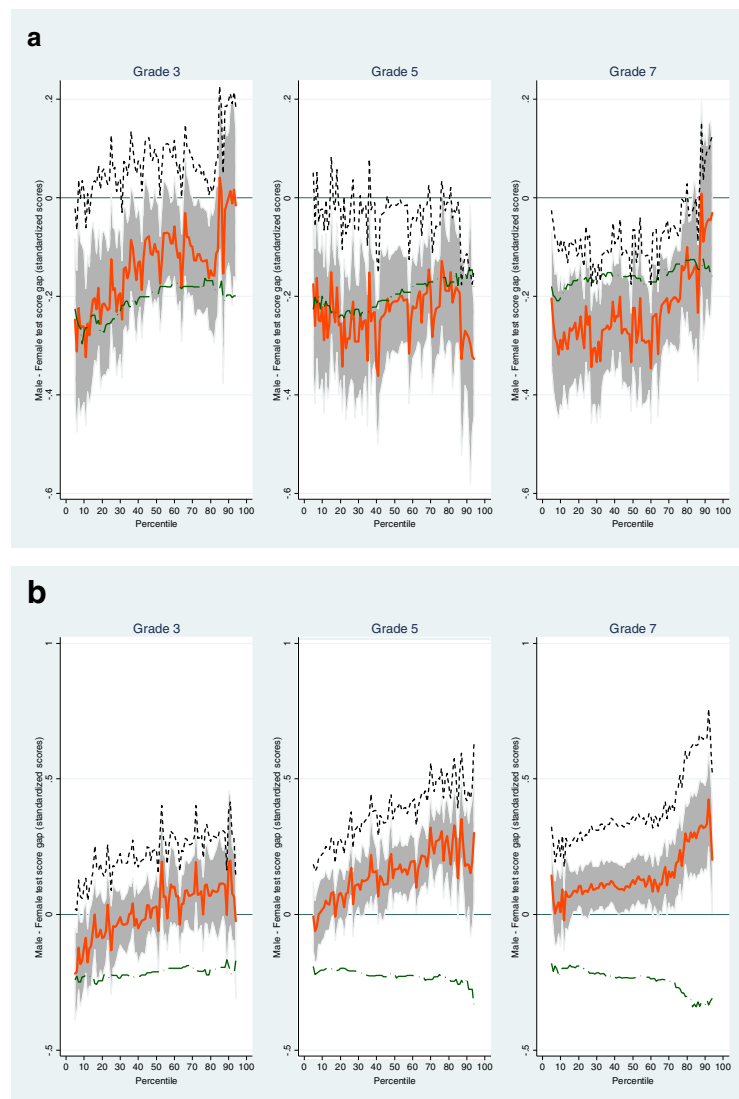**Table 4** Contributions to the male-female test score gap at mean and selected percentiles by grade—numeracy

	Grade 3				Grade 5				Grade 7			
	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total gap	-0.12*	0.03	0.19	0.00	0.02	0.17***	0.19*	0.15***	0.01	0.13***	0.32***	0.15***
Explained part												
Child	-0.02	0.00	0.00	-0.00	-0.01	-0.00	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01
	[17]	[0]	[0]	[n/a]	[-50]	[0]	[-11]	[-7]	[-100]	[-8]	[-3]	[-7]
Household	0.00	0.01	0.01	0.01	-0.01	-0.02*	-0.01	-0.01	-0.02	-0.01	-0.02	-0.02
	[0]	[33]	[5]	[n/a]	[-50]	[-12]	[-5]	[-7]	[-200]	[-8]	[-6]	[-13]
Others	0.01	-0.00	0.01	0.00	0.02	0.03**	0.04**	0.03***	0.01	0.02*	0.03	0.02*
	[-8]	[0]	[5]	[n/a]	[100]	[18]	[21]	[20]	[100]	[15]	[9]	[13]
Initial	-0.23***	-0.22***	-0.20***	-0.23***	-0.20***	-0.23***	-0.25***	-0.24***	-0.19***	-0.23***	-0.32***	-0.24***
	[192]	[-733]	[-105]	[n/a]	[-1000]	[-135]	[-132]	[-160]	[-1900]	[-177]	[-100]	[-160]
Total	-0.23***	-0.21***	-0.19***	-0.22***	-0.20***	-0.22***	-0.23***	-0.23***	-0.21***	-0.23***	-0.33***	-0.25***
	[192]	[-700]	[-100]	[n/a]	[-1000]	[-129]	[-121]	[-153]	[-2100]	[-177]	[-103]	[-167]
Unexplained part												
Child	-3.40	1.71	-4.09	-1.27	-2.64	-3.06	-3.66	-2.03	-2.67	-5.70**	0.44	-3.26*
	[2833]	[5700]	[-2153]	[n/a]	[-13,200]	[-1800]	[-1926]	[-1353]	[-26,700]	[-4385]	[138]	[-2173]
Household	0.56	-0.09	1.12	0.23	0.09	0.19	0.32	0.16	-0.44	0.14	-0.19	0.25
	[-467]	[-300]	[589]	[n/a]	[450]	[112]	[168]	[107]	[-4400]	[108]	[-59]	[167]
Others	4.26	-0.90	4.79	1.42	4.13	-0.14	5.09	1.41	5.08	0.67	-0.23	2.19
	[-3550]	[-3000]	[25521]	[n/a]	[20650]	[-82]	[2679]	[940]	[508000]	[515]	[-72]	[1460]
Initial	-0.01	-0.01	-0.01	-0.00	0.01	-0.00	-0.06**	-0.00	-0.01	-0.00	0.01	0.00
	[8]	[-33]	[-5]	[n/a]	[50]	[0]	[-32]	[0]	[-100]	[0]	[3]	[0]
Constant	-1.31	-0.47	-1.44	-0.16	-1.37	3.40	-1.27	0.85	-1.75	5.25	0.63	1.22

Table 4 Contributions to the male-female test score gap at mean and selected percentiles by grade—numeracy (Continued)

	Grade 3				Grade 5				Grade 7			
	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	[1092]	[−1567]	[−758]	[n/a]	[−6850]	[2000]	[−668]	[567]	[−17,500]	[4038]	[197]	[813]
Total	0.11	0.24***	0.38***	0.22***	0.22***	0.39***	0.42***	0.38***	0.21***	0.36***	0.65***	0.39***
	[−92]	[800]	[200]	[n/a]	[1100]	[229]	[221]	[253]	[2100]	[277]	[203]	[260]

Notes: Females are the base group. Standard errors (not reported for brevity) are obtained using 500 bootstrap replications. Estimates from model 3 are used. Values in squared brackets are percentage of the estimated total gap. Percentages may not add up to 100% due to rounding. "n/a" denotes "not applicable" because of dividing by zero. Grouped variables: Student: age, Aboriginal status, and birth weight; Household: mother's characteristics (age, migration background, completed qualification and working hours), having computer at home, home environment index, out-of-home activity index, household size, number of siblings, living with both biological parents, living in an owned home, household income, and school sector; Others: test states, test years, urban, local socio-economic background variables, and survey quarters; Initial: pre-school PPVT and WAI

Significance at the *10% level, **5% level, and ***1% level

Notes: Gender test score gap: Male - Female, standardized scores. Results are obtained using unconditional quantile regression model 3. Thick solid orange line (grey shaded area) indicates total test score gap estimates (95 % confidence intervals). Green long dash dot (black short dash) line shows the characteristic (return) effect. Confidence intervals are obtained using 500 bootstrap replications. 95 % confidence interval estimates for the aggregate characteristic and return effect are not reported to keep the figures discernible.

Fig. 2 Decomposition of test score gap along the distribution by test subject and grade. Panel **a**: Reading. Panel **b**: Numeracy

distribution for some test subjects or grades (for instance, at the 90th percentile of grades 3 and 7 reading, at means and all percentiles of grade 3 numeracy and at the 10th percentile of grades 5 and 7 numeracy). As it is not meaningful to explain the total gender gaps which are statistically insignificant, the focus is on the decomposition results where the gaps are statistically significant.

Decomposition results for reading (Table 3 and Fig. 2, panel A) show that estimates for the characteristic effect are negative and statistically significant, implying that gender differences in observable characteristics predict an advantage favouring female students in reading scores. In addition, estimates of the

characteristic effect are of the same sign and largely similar magnitude as those for the total gap, indicating that female students' advantages in reading are greatly attributable to their more favourable endowments of characteristics promoting reading scores. This is the case when the total gap is examined either at means or along the distribution. In contrast, the return effect plays a smaller role in contributing to the total gap since its estimates are statistically insignificant (at almost all selected percentiles) or of an opposite sign to the total gap estimates (at virtually the entire distribution of grade 3 reading test scores as can be seen in the first graph in panel A of Fig. 2). Regarding the contributions of the characteristic effect, estimates from Table 3 indicate that gender differences in preschool cognitive skills play the most significant role since their estimates are statistically significant, of the same sign and largely similar magnitude as those of the total characteristic effect. In contrast, estimates for factors other than preschool cognitive skills suggest that they contribute little to the total characteristic effect since their estimates are usually statistically insignificant or small in size. The aggregate decomposition results (either at means or along the distribution) additionally suggest a decreasing role of the characteristic effect in contributing to the total gap as students advance to higher grades.[19] This is consistent with the declining contribution of initial cognitive skill endowments to the total characteristic effect as students progress through school.[20]

Table 4 and Fig. 2 (panel B) show the characteristic effect is negative and statistically significant, indicating that gender differences in observable characteristics predict an advantage in favour of female students in numeracy. Similar to the gap in reading, preschool cognitive skills account for most of the characteristic effect in the case of the numeracy gap. In contrast, the return effect is positive and statistically significant, suggesting that male students are better able to convert educational inputs into higher numeracy test scores. Since the return effect dominates the characteristic effect, whether at the mean or along the distribution, the total gender numeracy score gap is positive, suggesting that male students outperform female students in numeracy. However, consistent with the regression results from regression model 1, estimates of the total gap are statistically significant in grades 5 and 7 only. Panel B in Fig. 2 additionally shows that at grades 5 and 7, the characteristic effect line diverts from the zero horizontal line along the test score distribution (i.e. the effect is more negative), suggesting that female students at the higher end of the distribution possess more of the characteristics associated with higher numeracy scores. In addition, the return effect line diverts from the zero horizontal line along the test score distribution, indicating that male students at the higher end of the distribution are more efficient in transforming education inputs into higher numeracy test scores. The combination of these two opposite trends explains the widening gender numeracy test score gap in favour of male students along the distribution.

In sum, consistent with the regression results presented in Section 5, the above decomposition analysis of the gender test score gap highlights the role of preschool cognitive skills in explaining the gap. These decomposition results further suggest that failing to account for initial academic skills would considerably limit the ability to explain factors contributing to the gender test score gap.[21,22] However, a large part of the gender test score gap remains unexplained in this study,

as has also been reported in the previous international studies (Sohn 2012; Gevrek and Seiberlich 2014; Golsteyn and Schils 2014). Similarly, our finding of an insignificant role of the return part in explaining the total gender test score gap in reading (grades 5 and 7) is in line with findings from previous studies of primary school students from the Netherlands (Golsteyn and Schils 2014) or the USA (Sohn 2012). Unfortunately, why the large part of the gender test score gap remains unexplained and why the return part plays an insignificant role in explaining the total gender test score gap remain open questions, suggesting a need for more research on factors driving the gender test score gaps. The decomposition analysis additionally suggests that focusing on only the mean gap overlooks important policy relevant heterogeneity across the distribution. It is interesting to observe that while the test score gap favouring females (i.e. in reading) is mostly due to differences in pre-school cognitive skills, the test score gap favouring males (i.e. in numeracy) is mainly due to differences in returns (i.e. the unexplained part). The significant female advantage in pre-school cognitive skills suggests the test score gap favouring females is usually due to differences in pre-school cognitive skills; however, the test score gap favouring males is largely due to differences in returns, which remains unanswered in this study, consistent with previous studies (Sohn 2012; Golsteyn and Schils 2014). To this end, further research into factors contributing to male students' greater efficiency in transferring education inputs into higher test scores would be worthwhile.

7 Conclusions

Drawing on the recent and nationally representative panel of Australian children, the patterns and factors contributing to the gender test score gap in academic achievements over the first 7 years of schooling have been examined. Regression results reveal that males excel at numeracy across all grades, whether at means or along the distribution. While mean regression results indicate a male advantage in grade 3 reading, quantile regression results show this gender test score gap is generally driven by those in the middle or top of the distribution. In addition, while mean regressions do not show noticeable gender differences in grade 7 reading, quantile regression results suggest females do outperform males at the lower end of the test score distribution. The regression results herein also reveal a widening gender test score gap in numeracy as students advance in their schooling. Quantile regression results additionally suggest that the widening gender numeracy test score gap favouring male students may have been driven by top performing students.

Applying an OB decomposition method, the impacts of gender differences in resources and their returns on academic achievements have been examined. The main results are that gender disparities in pre-school cognitive skills can explain a considerable part of the differences in academic performance. Female students are better endowed with pre-school cognitive skills and they use them to achieve better scores or reduce their score disadvantages relative to male students.

This paper has documented that differences in pre-school cognitive skills considerably help explain the gender test score gaps observed during primary and

early secondary school years. While these findings cannot be interpreted as causal, given the descriptive nature of the paper, they contribute to understanding gender test score gaps, with results useful in informing the direction of future interventions aimed at reducing the gender test score gap. Many questions remain unanswered, with a large part of the gender test score gap remaining unexplained, and no increased understanding in why the test score gap favouring males is largely due to differences in returns, indicating more research on the relationship between gender and educational achievement is warranted.

From a policy perspective, it is important to understand the patterns as well as the factors contributing to the gender test score gap, not only at the mean but along the distribution of the test score. One of the results from this study is the finding that pre-school cognitive skills play a significant role in explaining the gender test score gap observed up to seventh grade. This result suggests that policies aiming at reducing the gender test score gap should be implemented even prior to students enrolling at school. This policy implication is in line with that from the skill development literature, which usually shows early intervention is more beneficial than late intervention (Heckman 2000). Another finding of the heterogeneity of the gender test score gap across the distribution indicates that such policies should be targeted at some particular student groups.

Endnotes

[1]We use scores from the Peabody Picture Vocabulary Test (PPVT) and the Who Am I (WAI) test. These were administered prior to primary school enrolment to measure the pre-school cognitive skills (see Section 3 for details). Following the child development literature (Heckman and Kautz 2013), we term scores from these tests as "pre-school cognitive skills" and use "pre-school cognitive skills", "initial academic skills", and "initial cognitive endowment" alternatively in this study.

[2]Specifically, estimates obtained from the traditional conditional quantile regression can only be interpreted with the respect to the distribution of test score, conditional on test score determinants—i.e. only among individuals with the same observed characteristics such as gender, age, ethnicity, or parental education. As such, in most cases, the traditional conditional quantile regression may produce results that are often not generalisable or interpretable in a policy or population context (Firpo et al. 2009; Borah and Basu 2013).

[3]In the current study, exam papers are blind evaluated so results from these tests are thought to be independent of teacher assessments of non-cognitive traits of students (Lavy 2008; Hinnerich et al. 2011; Christopher et al. 2013; Simon and Greaves 2013; Heckman and Kautz 2014; Botelho et al. 2015).

[4]LSAC data also have other indirect measures of students' academic performance assessed by a class teacher and a parent. These assessments are based on a relative comparison with the student's classmates and therefore might differ across parents, teachers and schools (Daraganova et al. 2013). Because of this, they were not used in this analysis.

[5]Unreported results for writing, spelling and grammar are largely similar to the results of reading reported in this paper. The results for other non-numeracy test subjects are available upon request.

[6]The differences in test dates and survey dates in the empirical models are addressed by including dummies for survey months and test and survey years (see Section 4) (see Appendix 1: Table 5 for variable description and summary statistics).

[7]To examine the impact of other important variables and check the robustness of the results, a richer list of variables is included in extended specifications, where possible. The data contain father information including age, education, work status and ethnicity. However, due to a large number of missing data (13% of the final sample has missing data), father information is not used in our baseline specifications like US studies (Fryer Jr and Levitt 2004; Fryer and Levitt 2010; Bertrand and Pan 2013).

[8]The child's gender is implicitly assumed to be exogenous in this study, as has been assumed in the extant literature (Husain and Millimet 2009; Fryer and Levitt 2010; Sohn 2012). This assumption tends to hold in our case because sex selection is banned in Australia and there is no statistical evidence against such an assumption (Australian Health Ethics Committee 2004).

[9]Fryer and Levitt (2010) also documented no statistical difference in math scores between boys and girls upon entry to school. Unfortunately, our data do not contain a good measure of math ability of pre-school age children. As such, we are unable to compare the pre-school numeracy ability between boys and girls.

[10]The home environment index (on a scale of 0 to 3) is created from information about the frequency of activities the family do together at home such as reading, games or drawing pictures. The out-of-home activity index is measured by the number of "yes" answers to questions about activities that the family do together, such as going to a movie, sporting event, library or religious service.

[11]In Australia, secondary schools in Queensland, South Australia and Western Australia usually serve students from grade 8 while those in remaining states/territories from grade 7.

[12]About 3.5% of students in the sample were born overseas. Thus students' migration status was experimented with in their test score equations, however, their impact in all equations is statistically insignificant. This finding is in line with often found evidence that migrant children arriving in the host country at young ages have similar academic development as native children (Cortes 2006; van Ours and Veenman 2006; Cobb-Clark and Nguyen 2012). Therefore, the migration status of students is not included in the final regressions. However, the migration status of their mothers is included in the regressions. English Speaking Background (ESB) countries include the United Kingdom (UK), New Zealand, Canada, US, Ireland and South Africa.

[13]An alternative "value-added" model would condition the current outcome on the last outcome. Following this approach, one would condition grade 3 scores of all test subjects on pre-school scores of PPVT and WAI and condition grade 5 (7) scores of each test subject on respective grade 3 (5) scores. Regression results from this approach are presented in Appendix 2. As this approach reduces the sample size significantly and makes the results across grades less comparable, the results from model (3) are the focus of this paper.

[14]See Firpo et al. (2009) for a technical treatment of this method. This method has been applied in other economic literature strands (Fortin 2008; Le and Booth 2013; Fisher and Marchand 2014; Hirsch and Winters 2014; Kassenboehmer and Sinning 2014; Morin 2015). We use the rifreg command in Stata programmed by Firpo et al. (2009).

[15]In this paper, the focus is on decomposition results of grouped variables so the results are not sensitive to the choice of reference group for categorical variables (Fortin et al. 2011).

[16]Both US studies (Husain and Millimet 2009; Fryer and Levitt 2010) use a comprehensive set of characteristics without students' pre-school cognitive skills (like those in model 2 in this paper). They also note that controlling for covariates does not qualitatively change the results.

[17]95% confidence intervals are obtained using 500 bootstrap repetitions. Visually, 95% confidence intervals which do not include zero indicate a statistically significant (at the 5% level) estimate. Full regression results at three selected percentiles are presented in Appendix 1: Tables 9, 10 and 11.

[18]95% confidence interval estimates for the total characteristic and return effect are not reported to keep the figures discernible. For demonstration purposes, Appendix 1: Table 12 reports a full list of coefficient estimates for reading and numeracy test scores at grade 5, separately for males and females.

[19]In panel A of Fig. 2, the decreasing role of the characteristic effect can be seen as the line representing this effect approaches the zero horizontal line from below when students advance to higher grades. In contrast, the increasing contribution of the return effect can be viewed as the return effect line first approaches the zero horizontal line from above then gets closer to the total gap line which is always below the zero horizontal line.

[20]This trend can be explained as follows. As students advance through school, the first term of the characteristic effect, representing the male-female difference in pre-school cognitive skills $(\hat{Z}_m - \hat{Z}_f)$, is largely unchanged while the second term $(\hat{\mu}^*)$ describing returns to pre-school cognitive skills decreases. Estimation results (reported in Appendix 1: Tables 9, 10 and 11) confirm diminishing (but still positive) returns to pre-school cognitive skills along grades.

[21]The decomposition results using the model 2, which does not account for preschool cognitive skills, indicate that characteristics other than the student's preschool cognitive skills play an insignificant role in explaining the gender test score gap (i.e. visually, Appendix 1: Figure 3 shows the characteristic effect line virtually overlaps the zero horizontal line and this is the case for all test subjects). This finding is consistent with the previous finding of insignificant differences in parental characteristics and parental investment in child development by gender of the child (Section 3.3). Our finding that household and student characteristics, other than the student's pre-school cognitive skills, are not important in explaining the gender test score gap is in line with that reported in previous US studies (Husain and Millimet 2009; Fryer and Levitt 2010; Sohn 2012).

[22]In unreported robustness analyses, a wider range of school characteristics such as school quality (as measured by student/teacher ratios and school resources) and peer impact (gender, ESB ratio, NAPLAN test score by grade, subject and year) are included. These additional school characteristics are most widely available in grade 5. Regression and decomposition results from this robustness check suggest that these school characteristics play an insignificant role in explaining the gender test score gap in all grade 5 test subjects. Similarly, students' fathers' characteristics including age, migration status, education and work status contribute little to explain the gender test score gap. Results from these robustness checks are available upon request.

Appendix 1

Table 5 Variable description and summary statistics

Variable	Description	Mean	Min	Max	SD
Male	Dummy = 1 if male, = 0 if female	0.51	0.00	1.00	0.50
Child age	Child age (months)	106.59	75.00	140.00	19.51
Native	Dummy: = if child was born in Australia, = 0 otherwise	0.96	0.00	1.00	0.18
Aboriginal	Dummy: = 1 if child has Aboriginal/Torres Strait Islander origin, = 0 otherwise	0.02	0.00	1.00	0.15
Low birth weight	Dummy: = 1 if child's birth weight is 2500 g or less, = 0 otherwise	0.07	0.00	1.00	0.25
Breastfeed	Dummy: = 1 if child was breastfed at 3 or 6 month, = 0 otherwise	0.74	0.00	1.00	0.44
Mother age	Mother age (years)	39.07	18.00	69.00	5.24
Mother NESB[a]	Dummy: = 1 if mother was born in a Non-English Speaking Background (NESB) country, = 0 otherwise	0.20	0.00	1.00	0.40
Mother ESB[a]	Dummy: = 1 if mother was born in an English Speaking Background (ESB) country, = 0 otherwise	0.15	0.00	1.00	0.36
Mother has a certificate[b]	Dummy: = 1 if other has a certificate, = 0 otherwise	0.30	0.00	1.00	0.46
Mother has an advanced diploma[b]	Dummy: = 1 if mother has advanced diploma/diploma, = 0 otherwise	0.10	0.00	1.00	0.30
Mother has bachelor degree[b]	Dummy: = 1 if mother has a bachelor degree, = 0 otherwise	0.17	0.00	1.00	0.38
Mother has graduate diploma[b]	Dummy: = 1 if mother has graduate diploma/certificate, = 0 otherwise	0.08	0.00	1.00	0.27
Mother has postgraduate degree[b]	Dummy: = 1 if mother has a postgraduate degree, = 0 otherwise	0.08	0.00	1.00	0.26
Mother's weekly working hours	Mother's average weekly working hours (hours)	19.76	0.00	120.00	16.44
Home environment index	A scale between 0 and 3 describing the frequency of activities the family do together at home such as reading, games, or drawing pictures.	1.36	0.00	3.00	0.60
Out-of-home activity index	The number of "yes" answers to questions about activities that the family do together, such as going to a movie, sporting event, library, or religious service	2.63	0.00	5.00	1.21
Having a computer at home	Dummy: = 1 if having a computer at home, = 0 otherwise	0.94	0.00	1.00	0.24
Catholic school[c]	Dummy: = 1 if child attends a Catholic school, = 0 otherwise	0.22	0.00	1.00	0.42
Other independent school[c]	Dummy: = 1 if child attends an Independent school, = 0 otherwise	0.13	0.00	1.00	0.33
Household size	Household size	4.59	2.00	13.00	1.15
Number of siblings	Number of siblings	1.61	0.00	9.00	1.02
Number of younger siblings	Number of younger siblings	0.77	0.00	6.00	0.85
Number of same age siblings	Number of same age siblings	0.03	0.00	2.00	0.18
Living with both parents	Dummy: = 1 if child lives with both parents, = 0 otherwise	0.82	0.00	1.00	0.38
Living in an owned home	Dummy: = 1 if child attends a Catholic school, = 0 otherwise	0.78	0.00	1.00	0.42
Household income	Yearly household income (AUD 1000)	92.04	1.69	2792.92	75.72

Notes: Figures are calculated from a pooled sample of wave 2, 3 and 4 (8497 observations). Statistics are adjusted for sampling weights
[a]Native as the base group
[b]No qualification as the base group
[c]Public school as the base group

Table 6 Sample attritions

Sample selection criteria	Grade 3 (wave 2)	Grade 5 (wave 3)	Grade 7 (wave 4)	All
All children in the original sample	4464	4331	4169	12,964
Children having any NAPLAN test	2979	3940	3531	10,450
Children having all five NAPLAN tests	2865	3785	3358	10,008
Final sample with no missing information on basic controls	2471	3225	2801	8497
Sample attrition due to missing information on basic controls	394	560	557	1511
Sample attrition due to missing pre-school PPVT	286	370	303	959
Sample attrition due to missing pre-school WAI	30	41	30	101
Sample attrition due to missing pre-school PPVT and WAI	296	382	313	991
Sample attrition due to missing household income	57	121	181	359
Sample attrition due to missing pre-school PPVT and WAI and house hold income	347	485	470	1302

Notes: the number of children in respective samples is reported

Table 7 Differences between original and selected samples

	Grade 3 (wave 2)	Grade 5 (wave 3)	Grade 7 (wave 4)	All
	(1)	(2)	(3)	(4)
Male	0.02	− 0.02*	− 0.06***	− 0.02**
Child age	− 0.04***	0.00	0.02***	− 0.01***
Native	0.05	0.09***	0.02	0.06***
Aboriginal	− 0.14***	− 0.11***	− 0.07	− 0.11***
Low birth weight	−0.01	− 0.06**	− 0.09***	− 0.05***
Breastfeed	0.04**	0.02	0.04**	0.03***
Mother age	0.00*	−0.00	− 0.00*	− 0.00
Mother NESB[a]	− 0.08***	− 0.07***	− 0.07***	− 0.08***
Mother ESB[a]	−0.04**	− 0.01	− 0.00	−0.02
Mother has a certificate[b]	0.00	−0.01	−0.03*	− 0.01
Mother has an advanced diploma[b]	0.07**	0.03	−0.02	0.03*
Mother has bachelor degree[b]	0.05**	−0.01	− 0.03	0.00
Mother has graduate diploma[b]	−0.01	− 0.00	− 0.01	−0.00
Mother has postgraduate degree[b]	0.03	−0.01	−0.01	0.01
Mother's weekly working hours	−0.00	0.00	0.00	0.00
Household size	−0.00	− 0.01	− 0.03*	− 0.02**
Number of siblings	− 0.01	0.00	0.01	0.00
Number of younger siblings	0.02**	0.01	0.02*	0.02***
Number of same age siblings	−0.01	0.02	0.06	0.03
Living with both parents	0.08***	0.09***	0.06***	0.08***
Living in an owned home	0.05***	0.06***	0.05**	0.05***
Wave 3[c]				0.34***
Wave 4[c]				0.40***
Observations	4372	4228	4038	12,638
Number in selected sample	2471	3225	2801	8497
Pseudo R^2	0.08	0.03	0.03	0.04
P t test	0.16	0.07	0.00	0.02

Notes: Results (marginal effects) are from a probit model. Marginal effects are calculated at the means of continuous variables. The dependent variable is equal to one if the child is in our sample and zero otherwise
P t test P value of a t test for whether male estimate is equal to zero
Significance at the *10% level, **5% level, and ***1% level
[a]Native as the base group
[b]No qualification as the base group
[c]Wave 2 as the base group

Table 8 Associations between initial PPVT and WAI with characteristics

Variable	PPVT	WAI
	(1)	(2)
Male	−0.12***	− 0.58***
Child age	0.05***	0.12***
Native	0.02	−0.16**
Aboriginal	−0.13	−0.17**
Low birth weight	−0.22***	−0.20***
Breastfeeding at 3 or 6 months	0.12***	0.06**
Mother age	0.02***	0.00
Mother NESB[a]	−0.51***	0.22***
Mother ESB[a]	−0.05	0.03
Mother has a certificate[b]	0.04	0.10***
Mother has an advanced diploma[b]	0.13**	0.14***
Mother has bachelor degree[b]	0.23***	0.19***
Mother has graduate diploma[b]	0.21***	0.15**
Mother has postgraduate degree[b]	0.19***	0.24***
Mother's weekly working hours	−0.00	0.00
Home environment index	0.10***	0.00
Out-of-home activity index	0.06***	0.06***
Having a computer at home	0.18***	0.19***
Household size	− 0.10***	− 0.05*
Number of siblings	−0.06*	− 0.05*
Number of younger siblings	0.08***	0.07***
Number of same age siblings	−0.12	− 0.09
Living with both parents	0.18***	0.12***
Living in an owned home	0.02	0.04
Household income (log)	0.07***	0.04*
Observations	3923	4312
R^2	0.18	0.25

Notes: Results from ordinary least squared (OLS) regressions. Sample: K cohort children surveyed in wave 1. All variables are measured in wave 1. PPVT and WAI scores are standardised

Significance at the *10% level, **5% level, and ***1% level

[a]Native as the base group

[b]No qualification as the base group

Table 9 Full regression results at selected percentiles and at mean—grade 3

	Reading				Numeracy			
	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean
	(1)	(2)	(3)	(4)	(17)	(18)	(19)	(20)
Male	0.00	0.08*	0.18**	0.07**	0.09	0.23***	0.23***	0.22***
Child test age	− 0.05***	− 0.01	0.00	− 0.01	− 0.04**	0.00	0.01	− 0.00
Aboriginal	−0.80**	− 0.19	− 0.04	−0.23*	− 0.61*	− 0.16	0.01	− 0.31**
Low birth weight	−0.01	−0.11	− 0.13	−0.13*	− 0.03	− 0.19**	− 0.33***	−0.15**
Breastfeed	0.11	0.08	0.12*	0.11***	0.10	0.06	0.09	0.07
Mother age	0.00	0.01**	0.01	0.00	0.00	−0.00	0.01	0.00
Mother NESB[a]	0.04	0.07	0.27**	0.13**	0.21***	0.14**	0.23**	0.17***
Mother ESB[a]	0.10	0.03	0.06	0.07	0.08	0.00	0.08	0.04
Mother has a certificate[b]	−0.00	− 0.06	−0.07	− 0.09*	− 0.12	− 0.06	− 0.05	−0.06
Mother has an advanced diploma[b]	0.12	−0.08	− 0.09	− 0.04	−0.02	− 0.01	−0.11	− 0.02
Mother has bachelor degree[b]	0.14	0.12*	0.19	0.13**	0.03	0.18**	0.11	0.14**
Mother has graduate diploma[b]	0.13	0.18*	0.14	0.10	−0.00	0.11	0.09	0.12*
Mother has postgraduate degree[b]	0.15	0.23**	0.24	0.25***	−0.09	0.18*	0.34*	0.18**
Mother's weekly working hours	−0.00	−0.00**	− 0.00*	−0.00*	0.00*	0.00	−0.00	0.00
Home environment index	−0.10	− 0.08*	− 0.04	−0.07*	− 0.11*	−0.10**	− 0.10	−0.10***
Out-of-home activity index	0.05	0.02	−0.01	0.02	0.04	0.00	0.02	0.03
Having a computer at home	0.16	−0.03	−0.05	−0.01	0.16	0.01	−0.14	0.03
Catholic school[c]	0.06	−0.06	− 0.10	− 0.02	−0.15*	− 0.02	−0.12	− 0.09**
Other independent school[c]	0.09	−0.02	−0.19	− 0.01	0.06	0.03	−0.13	− 0.02
Household size	0.06	0.05	−0.04	0.02	0.03	0.00	−0.02	−0.00
Number of siblings	−0.12	−0.13***	0.01	−0.08*	0.01	0.01	0.03	0.01
Number of younger siblings	0.07	0.12***	0.08	0.10***	0.03	0.03	0.06	0.04
Number of same age siblings	0.08	0.01	0.01	0.04	0.40***	0.01	0.18	0.08
Living with both parents	−0.04	0.04	−0.01	0.01	0.10	0.02	0.05	0.05
Living in an owned home	0.07	0.05	0.03	0.05	0.09	−0.02	−0.03	−0.00
Household income (log)	−0.00	0.05	0.11*	0.05	0.02	0.12**	0.08	0.08**
Initial PPVT	0.28***	0.24***	0.30***	0.26***	0.24***	0.20***	0.17***	0.21***
Initial WAI	0.35***	0.26***	0.28***	0.28***	0.30***	0.31***	0.28***	0.31***
Observations	2471	2471	2471	2471	2471	2471	2471	2471
R-squared	0.14	0.20	0.11	0.29	0.12	0.18	0.09	0.27

Notes: Estimates at selected percentiles are based on unconditional quantile regressions. Standard errors (not reported for brevity) are obtained using 500 bootstrap replications. Each estimate is obtained from a separate regression using model 3. Estimates for other variables (test states, test years, urban, local socio-economic background variables, and survey quarters) are not reported for brevity
Significance at the *10% level, **5% level, and ***1% level
[a]Native as the base group
[b]No qualification as the base group
[c]Public school as the base group

Table 10 Full regression results at selected percentiles and at mean—grade 5

	Reading				Numeracy			
	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean
	(1)	(2)	(3)	(4)	(17)	(18)	(19)	(20)
Male	−0.02	− 0.01	− 0.12*	− 0.03	0.22***	0.37***	0.41***	0.38***
Child test age	−0.02	− 0.01	0.00	−0.01	− 0.02*	−0.01	− 0.02	− 0.01**
Aboriginal	− 0.73**	− 0.20	− 0.10	−0.36***	− 0.96***	−0.39***	− 0.26***	−0.48***
Low birth weight	0.05	−0.07	− 0.32***	− 0.05	− 0.04	− 0.15*	− 0.34***	− 0.15***
Breastfeed	0.03	0.09*	0.08	0.08**	0.01	0.06	0.16**	0.08**
Mother age	0.01	0.01**	0.02**	0.01***	0.01	0.00	0.00	0.01*
Mother NESB[a]	−0.02	0.14**	0.12	0.09**	−0.01	0.10	0.24**	0.12***
Mother ESB[a]	0.08	0.08	− 0.06	0.04	0.08	−0.02	− 0.12	0.01
Mother has a certificate[b]	0.00	−0.00	−0.06	− 0.03	0.01	− 0.02	−0.04	− 0.05
Mother has an advanced diploma[b]	0.21**	0.10	−0.12	0.08	0.14*	0.01	0.05	0.04
Mother has bachelor degree[b]	0.19**	0.28***	0.40***	0.25***	0.19***	0.25***	0.24**	0.23***
Mother has graduate diploma[b]	0.19**	0.11	0.34**	0.20***	0.15*	0.11	−0.07	0.04
Mother has postgraduate degree[b]	0.03	0.34***	0.51***	0.23***	0.09	0.31***	0.59***	0.27***
Mother's weekly working hours	− 0.00	− 0.00*	− 0.01***	− 0.00**	0.00	− 0.00	− 0.00	− 0.00
Home environment index	− 0.15***	− 0.14***	− 0.16***	− 0.14***	− 0.10**	− 0.15***	− 0.17***	− 0.15***
Out-of-home activity index	0.04	0.07***	0.05*	0.06***	0.02	0.07***	0.07**	0.06***
Having a computer at home	0.15	0.11	0.16	0.11*	0.27**	0.10	0.05	0.11*
Catholic school[c]	0.05	−0.10**	− 0.24***	−0.09**	− 0.01	− 0.05	− 0.30***	− 0.15***
Other independent school[c]	0.18***	0.07	−0.02	0.06	0.03	0.06	0.04	0.00
Household size	0.03	0.04	0.14**	0.04	− 0.06	0.05	0.00	0.02
Number of siblings	−0.16**	− 0.09**	− 0.22***	− 0.12***	0.02	− 0.07	− 0.05	− 0.06
Number of younger siblings	0.07	0.10***	0.13***	0.10***	0.07*	0.05	0.11**	0.07***
Number of same age siblings	0.00	0.02	−0.14	0.02	−0.10	− 0.02	0.30*	0.09
Living with both parents	0.07	0.04	−0.16	0.03	0.17*	0.00	0.01	0.06
Living in an owned home	0.08	0.01	0.01	0.01	0.08	0.08	−0.01	0.03
Household income (log)	0.04	− 0.00	−0.00	0.01	0.03	0.02	0.04	0.05*
Initial PPVT	0.22***	0.25***	0.29***	0.24***	0.14***	0.15***	0.08**	0.13***
Initial WAI	0.24***	0.23***	0.21***	0.24***	0.27***	0.31***	0.36***	0.33***
Observations	3225	3225	3225	3225	3225	3225	3225	3225
R^2	0.13	0.20	0.11	0.30	0.13	0.20	0.12	0.31

Notes: Estimates at selected percentiles are based on unconditional quantile regressions. Standard errors (not reported for brevity) are obtained using 500 bootstrap replications. Each estimate is obtained from a separate regression using model 3. Estimates for other variables (test states, test years, urban, local socio-economic background variables, and survey quarters) are not reported for brevity

Significance at the *10% level, **5% level, and ***1% level

[a]Native as the base group

[b]No qualification as the base group

[c]Public school as the base group

Table 11 Full regression results at selected percentiles and at mean—grade 7

	Reading				Numeracy			
	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean
	(1)	(2)	(3)	(4)	(17)	(18)	(19)	(20)
Male	−0.09	−0.11**	0.07	−0.06*	0.25***	0.34***	0.69***	0.39***
Child test age	−0.03***	0.00	−0.01	−0.01*	−0.01	−0.02**	−0.03**	−0.02***
Aboriginal	−0.54*	−0.33**	−0.25***	−0.35***	−0.59**	−0.37***	−0.12	−0.38***
Low birth weight	0.13	−0.14	−0.20*	−0.02	−0.03	−0.11	−0.21	−0.14**
Breastfeed	0.03	0.12**	0.01	0.09**	0.11*	0.10**	0.04	0.08**
Mother age	0.02***	0.01**	0.02**	0.02***	0.00	0.01	0.01*	0.01**
Mother NESB[a]	−0.08	0.15**	0.15	0.08	−0.05	0.15**	0.29**	0.14***
Mother ESB[a]	0.09	0.07	0.13	0.06	0.04	0.06	−0.15	0.01
Mother has a certificate[b]	0.09	0.06	−0.03	0.02	0.02	−0.09	−0.17*	−0.05
Mother has an advanced diploma[b]	0.03	0.14	0.22*	0.11*	0.01	−0.06	−0.08	−0.01
Mother has bachelor degree[b]	0.18*	0.32***	0.39***	0.26***	0.19***	0.18**	0.20	0.23***
Mother has graduate diploma[b]	0.23**	0.40***	0.21	0.24***	0.15*	0.18*	0.17	0.17***
Mother has postgraduate degree[b]	0.05	0.30***	0.57***	0.26***	0.07	0.22**	0.49**	0.27***
Mother's weekly working hours	0.00*	−0.00	−0.00	0.00	0.00**	−0.00	0.00	0.00
Home environment index	−0.15***	−0.14***	−0.07	−0.12***	−0.07	−0.19***	−0.15**	−0.13***
Out-of-home activity index	0.04	0.08***	0.03	0.05***	0.06**	0.06***	0.04	0.05***
Having a computer at home	0.38	0.13	0.04	0.15	0.22	0.29**	0.12	0.22***
Catholic school[c]	0.05	−0.13**	−0.20**	−0.10**	−0.10*	−0.11**	−0.20**	−0.13***
Other independent school[c]	0.05	−0.04	−0.09	0.00	0.02	0.07	0.39***	0.11**
Household size	−0.09	0.08	0.08	0.03	0.02	0.01	0.06	0.02
Number of siblings	0.00	−0.17***	−0.23***	−0.12***	−0.05	−0.03	−0.13	−0.04
Number of younger siblings	0.13***	0.12***	0.21***	0.13***	0.03	0.07*	0.13**	0.07***
Number of same age siblings	0.29**	0.02	0.05	0.06	0.30**	0.08	0.46*	0.22**
Living with both parents	0.10	−0.07	0.04	0.01	0.25***	−0.01	0.02	0.05
Living in an owned home	0.12	0.06	0.08	0.07	0.08	0.12**	0.19**	0.13***
Household income (log)	0.05	0.11***	0.09*	0.09***	0.04	0.09**	−0.10	0.06**
Initial PPVT	0.26***	0.30***	0.24***	0.26***	0.13***	0.11***	0.12***	0.14***
Initial WAI	0.23***	0.20***	0.19***	0.19***	0.28***	0.35***	0.49***	0.35***
Observations	2801	2801	2801	2801	2801	2801	2801	2801
R^2	0.13	0.21	0.12	0.31	0.14	0.22	0.13	0.32

Notes: Estimates at selected percentiles are based on unconditional quantile regressions. Standard errors (not reported for brevity) are obtained using 500 bootstrap replications. Each estimate is obtained from a separate regression using model 3. Estimates for other variables (test states, test years, urban, local socio-economic background variables, and survey quarters) are not reported for brevity

Significance at the *10% level, **5% level, and ***1% level

[a]Native as the base group

[b]No qualification as the base group

[c]Public school as the base group

Table 12 Full regression results at selected percentiles and at mean—grade 5 reading and numeracy—by gender

	Reading								Numeracy							
	P10th		P50th		P90th		Mean		P10th		P50th		P90th		Mean	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Child test age	−0.02	−0.03	−0.01	−0.02*	−0.02	0.02	−0.01	−0.01	−0.01	−0.03**	0.00	−0.02*	−0.01	−0.04	−0.01	−0.03**
Aboriginal	−0.75*	−0.57	−0.49***	0.10	−0.11	−0.06	−0.56***	−0.07	−1.40***	−0.22	−0.48***	−0.21	−0.33**	−0.27*	−0.68***	−0.25
Low birth weight	0.13	−0.33	−0.02	0.01	−0.23*	−0.35***	−0.04	−0.07	0.03	−0.17	−0.19	−0.11	−0.36***	−0.35**	−0.16**	−0.18**
Breastfeed	0.08	−0.08	0.08	0.10	−0.00	0.15**	0.09*	0.07	−0.01	0.02	0.06	0.06	0.17*	0.17*	0.05	0.09*
Mother age	0.00	0.01	0.01*	0.01*	0.02	0.02**	0.01*	0.01***	0.00	0.01	0.00	0.01	−0.00	0.01	0.01	0.01
Mother NESB[a]	−0.01	−0.00	0.10	0.18**	0.16	0.03	0.07	0.11*	−0.01	−0.01	0.15*	0.09	0.23	0.38**	0.13**	0.11*
Mother ESB[a]	0.19	0.06	−0.04	0.13	−0.03	0.03	−0.01	0.10	0.20**	−0.01	0.09	−0.08	−0.16	−0.04	0.06	−0.03
Mother has a certificate[b]	0.05	0.08	−0.04	0.04	−0.09	−0.04	−0.07	0.02	0.09	−0.03	−0.03	−0.03	0.05	−0.07	−0.02	−0.04
Mother has an advanced diploma[b]	0.26*	0.29*	0.07	0.06	−0.18	−0.06	0.08	0.09	0.21*	0.13	0.04	−0.02	0.13	0.11	0.06	0.05
Mother has bachelor degree[b]	0.31***	0.31**	0.20**	0.31***	0.33**	0.30**	0.20***	0.31***	0.18*	0.22**	0.25***	0.25**	0.29*	0.31*	0.17**	0.32***
Mother has graduate diploma[b]	0.11	0.26*	0.12	0.22	0.34*	0.18	0.20**	0.19**	0.17	0.10	0.17	0.01	0.10	−0.23	0.10	−0.01
Mother has postgraduate degree[b]	0.05	0.06	0.29**	0.36***	0.53**	0.40*	0.20*	0.26***	0.20*	0.01	0.20	0.31**	0.90***	0.38	0.26***	0.29***
Mother's weekly working hours	0.00	0.00	−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
Home environment index	−0.12	−0.21***	−0.10*	−0.14***	−0.12	−0.15**	−0.12***	−0.16***	−0.05	−0.13**	−0.13**	−0.15***	−0.33***	−0.10	−0.14***	−0.15***
Out-of-home activity index	0.04	0.07*	0.08***	0.04*	0.07*	0.04	0.08***	0.04**	0.02	0.00	0.07***	0.06**	0.11**	0.07	0.06***	0.05**
Having a computer at home	0.21	0.20	0.16	−0.08	0.26**	0.01	0.17*	0.08	0.23	0.27	0.07	0.12	0.17	0.07	0.12	0.12
Catholic school[c]	−0.10	0.01	−0.19**	−0.02	−0.18	−0.23**	−0.13**	−0.03	0.06	−0.07	−0.12*	0.01	−0.36***	−0.22*	−0.19***	−0.10*
Other independent school[c]	0.14	0.19**	0.10	0.04	0.09	−0.07	0.06	0.05	0.01	0.06	0.01	0.14	0.03	−0.04	−0.02	0.04
Household size	−0.05	0.10	0.05	0.09	0.11	0.12*	0.05	0.05	−0.10	−0.00	0.07	0.03	0.09	−0.01	0.01	0.04
Number of siblings	−0.06	−0.21**	−0.10	−0.17**	−0.14	−0.26***	−0.12**	−0.14***	0.04	−0.01	−0.07	−0.06	−0.01	−0.09	−0.02	−0.04
Number of younger siblings	0.12*	0.04	0.15***	0.04	0.15*	0.13*	0.15***	0.06*	0.09*	0.02	0.08*	0.03	0.15**	0.13*	0.10***	0.04
Number of same age siblings	−0.12	0.25	−0.02	−0.21	−0.10	−0.01	0.02	0.02	−0.17	0.03	0.04	−0.03	0.38	0.08	0.09	0.10

Table 12 Full regression results at selected percentiles and at mean—grade 5 reading and numeracy—by gender (Continued)
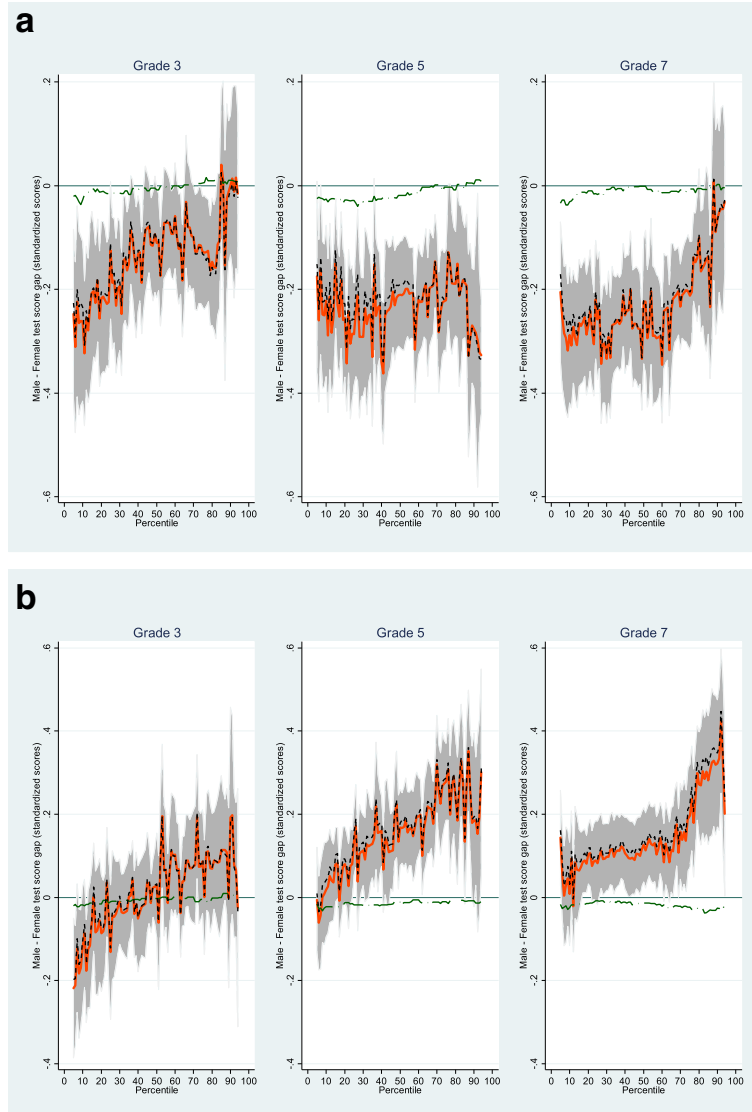
	Reading								Numeracy							
	P10th		P50th		P90th		Mean		P10th		P50th		P90th		Mean	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Living with both parents	0.26	−0.08	0.03	0.06	−0.11	−0.19	0.07	−0.02	0.19	0.11	−0.01	−0.02	0.06	−0.15	0.07	0.02
Living in an owned home	0.11	−0.01	0.02	−0.08	−0.03	0.08	0.03	0.01	0.11	0.05	0.17**	−0.03	0.00	0.08	0.08	−0.02
Household income	0.02	0.05	−0.01	0.05	0.08	−0.09	0.01	0.01	0.05	−0.01	0.01	0.06	−0.02	0.12	0.02	0.08*
Initial PPVT	0.24***	0.23***	0.23***	0.27***	0.21***	0.26***	0.22***	0.27***	0.14***	0.13***	0.11***	0.16***	−0.02	0.16***	0.09***	0.16***
Initial WAI	0.28***	0.19***	0.25***	0.25***	0.27***	0.15***	0.27***	0.22***	0.21***	0.33***	0.34***	0.29***	0.53***	0.35***	0.37***	0.31***
Observations	1596	1629	1596	1629	1596	1629	1596	1629	1596	1629	1596	1629	1596	1629	1596	1629
R^2	0.14	0.13	0.20	0.22	0.13	0.13	0.29	0.31	0.15	0.15	0.23	0.19	0.17	0.12	0.35	0.30

Notes: Estimates at selected percentiles are based on unconditional quantile regressions. Standard errors (not reported for brevity) are obtained using 500 bootstrap replications. Each estimate is obtained from a separate regression using model 3. Estimates for other variables (test states, test years, urban, local socio-economic background variables, and survey quarters) are not reported for brevity

Significance at the *10% level, **5% level, and ***1% level

[a]Native as the base group

[b]No qualification as the base group

[c]Public school as the base group

Notes: Gender test score gap: Male - Female, standardized scores. Results are obtained using unconditional quantile regression model 2. Thick solid orange line (grey shaded area) indicates total test score gap estimates (95 % confidence intervals). Green long dash dot (black short dash) line shows the characteristic (return) effect. Confidence intervals are obtained using 500 bootstrap replications. 95 % confidence interval estimates for the aggregate characteristic and return effect are not reported to keep the figures discernible.

Fig. 3 Decomposition of test score gap along the distribution by test subject and grade—model 2. Panel **a**: Reading. Panel **b**: Numeracy

Appendix 2

7.1 Supplemental materials for refereeing purposes and on-line publication

This Appendix reports results from an alternative "value-added" model which conditions the current outcome on the last outcome (denoted model (4)). Following this approach, one would condition grade 3 scores of all test subjects on pre-school scores of PPVT and WAI and condition grade 5 (7) scores of each test subject on respective grade 3 (5) scores. Appendix 2 contains three tables and one figure.

Appendix 2: Table 13 reports estimated gender score gap over the grades at mean (similar to Table 2).

Appendix 2: Tables 14 and 15 reports decomposition results on contributions to the male-female test score gap at mean and selected percentiles by grade for Reading (Numeracy). These tables are similar to Tables 3 and 4.

Appendix 2: Figure 4 represents gender test score gaps along the distribution by test subject and grade (similar to Fig. 1). Figure 5 reports decomposition results of test score gaps along the distribution by test subject and grade (the same as Fig. 2).

Table 13 Estimated gender score gap over the grades at mean

Subject	Model	Grade 3	Grade 5	Grade 7
Reading	(1)	−0.13***	− 0.22***	− 0.22***
		(0.04)	(0.04)	(0.04)
	(2)	−0.13***	− 0.22***	− 0.20***
		(0.04)	(0.04)	(0.03)
	(4)	0.07**	−0.13***	− 0.04*
		(0.04)	(0.03)	(0.02)
Numeracy	(1)	0.00	0.15***	0.15***
		(0.04)	(0.04)	(0.04)
	(2)	0.01	0.15***	0.17***
		(0.04)	(0.04)	(0.03)
	(4)	0.22***	0.14***	0.04*
		(0.04)	(0.03)	(0.02)
Observations		2471	2509	2857

Notes: Females are the base group. Each estimate is obtained from a separate regression. Model 1 includes gender dummy only. Model 2 includes student characteristics (gender, age, Aboriginal status, and birth weight), household characteristics (mother's characteristics (age, migration background, completed qualification, and working hours), having computer at home, home environment index, out-of-home activity index, household size, number of siblings, living with both biological parents, living in an owned home, household income, and school sector), test states, test years, urban, local socio-economic background variables, and survey quarters. Model 4 includes all variables as in model 2 plus pre-school PPVT and WAI for grade 3 test score regressions and respective grade 3 (5) test scores for grade 5 (7) test score regressions
Significance at the *10% level, **5% level, and ***1% level

**Table 14** Contributions to the male-female test score gap at mean and selected percentiles by grade—reading

	Grade 3				Grade 5				Grade 7			
	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total gap	−0.26***	−0.11**	0.00	−0.13***	−0.04	−0.21***	−0.31***	−0.22***	−0.30***	−0.21***	−0.07	−0.22***
Explained part												
Child	−0.02	−0.01	−0.00	−0.01	−0.01	0.01	0.01	0.00	−0.01	0.00	0.00	−0.00
	[8]	[9]	[n/a]	[8]	[25]	[−5]	[−3]	[0]	[3]	[0]	[0]	[0]
Household	−0.26***	−0.20***	−0.22***	−0.22***	−0.08***	−0.09***	−0.09***	−0.09***	−0.15***	−0.19***	−0.17***	−0.16***
	[100]	[182]	[n/a]	[169]	[200]	[43]	[29]	[41]	[50]	[90]	[243]	[73]
Others	0.01	0.01	0.02	0.01	−0.01	−0.01	0.01	−0.00	−0.01	0.01	0.01	0.00
	[−4]	[−9]	[n/a]	[−8]	[25]	[5]	[−3]	[0]	[3]	[−5]	[−14]	[0]
Initial	0.00	0.00	0.00	0.01	0.00	0.00	−0.00	0.00	−0.01	−0.01	0.00	−0.01
	[0]	[0]	[n/a]	[−8]	[0]	[0]	[0]	[0]	[3]	[5]	[0]	[5]
Total	−0.27***	−0.19***	−0.20***	−0.21***	−0.10***	−0.09***	−0.07**	−0.08***	−0.18***	−0.18***	−0.16***	−0.17***
	[104]	[173]	[n/a]	[162]	[250]	[43]	[23]	[36]	[60]	[86]	[229]	[77]
Unexplained part												
Child	−0.68	−2.07	2.67	−0.85	−1.25	−4.41*	5.04	−2.04	−2.77	−2.58	−3.04	−1.55
	[262]	[1882]	[n/a]	[654]	[3125]	[2100]	[−1626]	[927]	[923]	[1229]	[4343]	[705]
Household	0.02	−0.00	0.00	0.00	−0.18*	−0.15**	−0.06	−0.06*	0.24	0.12	0.35	0.23***
	[−8]	[0]	[n/a]	[0]	[450]	[71]	[19]	[27]	[−80]	[−57]	[−500]	[−105]
Others	6.32	−0.20	−5.99	−2.03	9.05**	−0.14	−2.91	0.78	−0.46	−1.95	−6.09	−0.78
	[−2431]	[182]	[n/a]	[1562]	[−22,625]	[67]	[939]	[−355]	[153]	[929]	[8700]	[355]
Initial	−0.48	−0.09	1.24	0.46	0.09	0.27	−0.63	−0.30	−0.97	0.13	0.12	−0.33
	[185]	[82]	[n/a]	[−354]	[−225]	[−129]	[203]	[136]	[323]	[−62]	[−171]	[150]
Total	−5.17	2.44	2.28	2.49	−7.65	4.31	−1.67	1.49	3.85	4.26	8.75	2.39

Table 14 Contributions to the male–female test score gap at mean and selected percentiles by grade—reading *(Continued)*

	Grade 3				Grade 5				Grade 7			
	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	[1988]	[−2218]	[n/a]	[−1915]	[19125]	[−2052]	[539]	[−677]	[−1283]	[−2029]	[−12,500]	[−1086]
Unexplained Total	0.01	0.08	0.20***	0.07**	0.06	−0.12*	−0.24**	−0.13***	−0.11**	−0.02	0.09	−0.04**
	[−4]	[−73]	[n/a]	[−54]	[−150]	[57]	[77]	[59]	[37]	[10]	[−129]	[18]

Notes: Females are the base group. Standard errors (not reported for brevity) are obtained using 500 bootstrap replications. Estimates from model 4 are used. Values in squared brackets are percentage of the estimated total gap. Percentages may not add up to 100% due to rounding. "n/a" denotes "not applicable" because of dividing by zero. Grouped variables: Student: age, Aboriginal status, and birth weight; Household: mother's characteristics (age, migration background, completed qualification and working hours), having computer at home, home environment index, out-of-home activity index, household size, number of siblings, living with both biological parents, living in an owned home, household income, and school sector; Others: test states, test years, urban, local socio-economic background variables, and survey quarters; Initial: pre-school PPVT and WAI for grade 3 and previous reading test scores for grades 5 and 7

Significance at the *10% level, **5% level, and ***1% level

Table 15 Contributions to the male-female test score gap at mean and selected percentiles by grade—numeracy

	Grade 3				Grade 5				Grade 7			
	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total gap	-0.12	0.03	0.19	0.00	0.04	0.09	0.17**	0.15***	0.04	0.15***	0.38***	0.15***
Explained part												
Child	-0.02	0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	-0.00	-0.01*	-0.01	-0.01
	[17]	[0]	[0]	[n/a]	[0]	[0]	[0]	[0]	[0]	[-7]	[-3]	[-7]
Household	-0.23***	-0.22***	-0.20***	-0.23***	0.00	0.00	0.00	0.00	0.09***	0.12***	0.17***	0.13***
	[192]	[-733]	[-105]	[n/a]	[0]	[0]	[0]	[0]	[225]	[80]	[45]	[87]
Others	0.01	-0.00	0.01	0.00	-0.01	0.00	0.01	0.00	-0.01	0.00	0.00	-0.00
	[-8]	[0]	[5]	[n/a]	[-25]	[0]	[6]	[0]	[-25]	[0]	[0]	[0]
Initial	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	-0.01	-0.00	-0.00	-0.01
	[0]	[33]	[5]	[n/a]	[0]	[0]	[0]	[0]	[-25]	[0]	[0]	[-7]
Total	-0.23***	-0.21***	-0.19***	-0.22***	-0.00	0.00	0.02	0.01	0.07***	0.11***	0.16***	0.11***
	[192]	[-700]	[-100]	[n/a]	[0]	[0]	[12]	[7]	[175]	[73]	[42]	[73]
Unexplained part												
Child	-3.40	1.71	-4.09	-1.27	-4.69	-4.05	-2.80	-3.48**	-1.49	-3.32*	6.13*	-0.03
	[2833]	[5700]	[-2153]		[-11,725]	[-4500]	[-1647]	[-2320]	[-3725]	[-2213]	[1613]	[-20]
Household	-0.00	-0.01	-0.02	-0.00	0.15	0.06	-0.12	0.02	0.13	0.14	0.08	0.09
	[0]	[-33]	[-11]	[n/a]	[375]	[67]	[-71]	[13]	[325]	[93]	[21]	[60]
Others	4.26	-0.90	4.79	1.42	4.28	-0.77	-0.38	-0.01	2.91	-1.09	-4.68	-0.02
	[-3550]	[-3000]	[2521]	[n/a]	[10700]	[-856]	[-224]	[-7]	[7275]	[-727]	[-1232]	[-13]
Initial	0.56	-0.09	1.12	0.23	-0.99	-0.21	0.68	-0.17	-0.57	-0.64	-0.58	-0.26
	[-467]	[-300]	[589]	[n/a]	[-2475]	[-233]	[400]	[-113]	[-1425]	[-427]	[-153]	[-173]
Constant	-1.31	-0.47	-1.43	-0.16	1.29	5.05	2.77	3.78	-1.00	4.95	-0.74	0.26

Table 15 Contributions to the male–female test score gap at mean and selected percentiles by grade—numeracy (Continued)

	Grade 3				Grade 5				Grade 7			
	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean	P10th	P50th	P90th	Mean
	[1092]	[−1567]	[−753]	[n/a]	[3225]	[5611]	[1629]	[2520]	[−2500]	[3300]	[−195]	[173]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total	0.11	0.24***	0.38***	0.22***	0.05	0.09*	0.15*	0.14***	−0.03	0.03	0.22***	0.04*
	[−92]	[800]	[200]	[n/a]	[125]	[100]	[88]	[93]	[−75]	[20]	[58]	[27]

Notes: Females are the base group. Standard errors (not reported for brevity) are obtained using 500 bootstrap replications. Estimates from model 4 are used. Values in squared brackets are percentage of the estimated total gap. Percentages may not add up to 100% due to rounding. "n/a" denotes "not applicable" because of dividing by zero. Grouped variables: Student: age, Aboriginal status, and birth weight; Household: mother's characteristics (age, migration background, completed qualification and working hours), having computer at home, home environment index, out-of-home activity index, household size, number of siblings, living with both biological parents, living in an owned home, household income, and school sector; Others: test states, test years, urban, local socio-economic background variables, and survey quarters; Initial: pre–school PPVT and WAI for grade 3 and previous numeracy test scores for grades 5 and 7
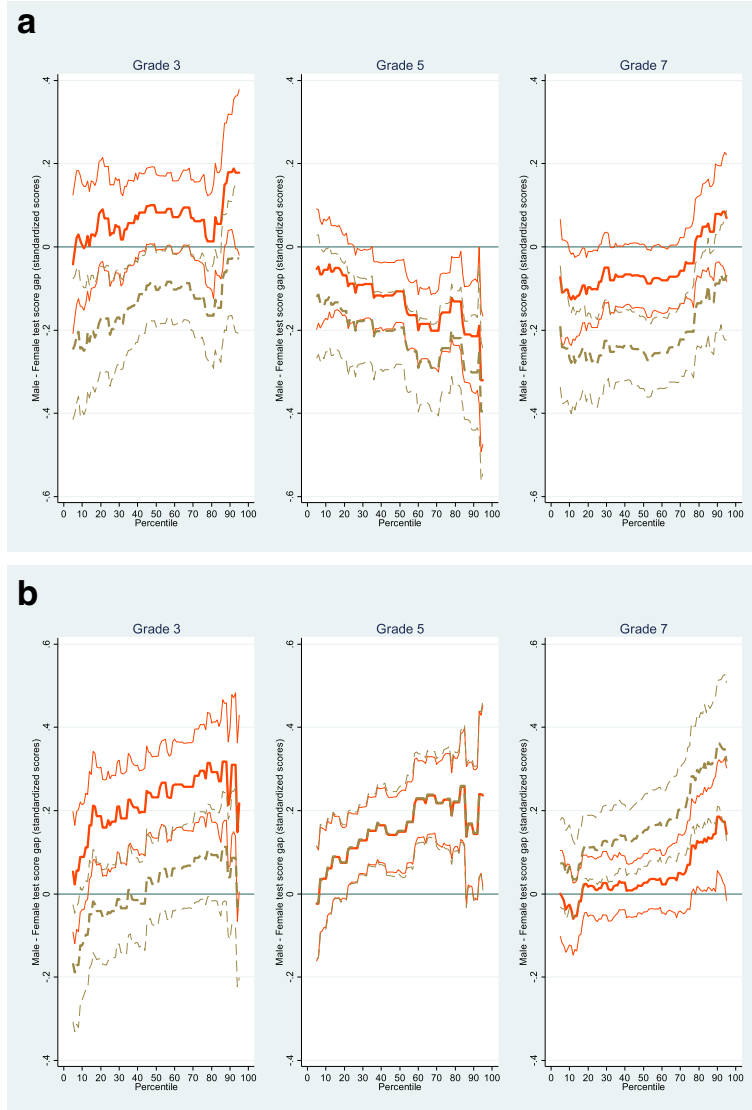
Significance at the *10% level, **5% level, and ***1% level

Notes: Gender test score gap: Male - Female, standardized scores. Estimates are based on unconditional quantile regressions. Thick (thin) solid orange line indicates test score gap estimates (95 % confidence intervals) using model 4. Thick (thin) dotted brown line shows test score gap estimates (95 % confidence intervals) using model 2. Confidence intervals are obtained using 500 bootstrap replications.

Fig. 4 Gender test score gaps along the distribution by test subject and grade. Panel **a**: Reading. Panel **b**: Numeracy
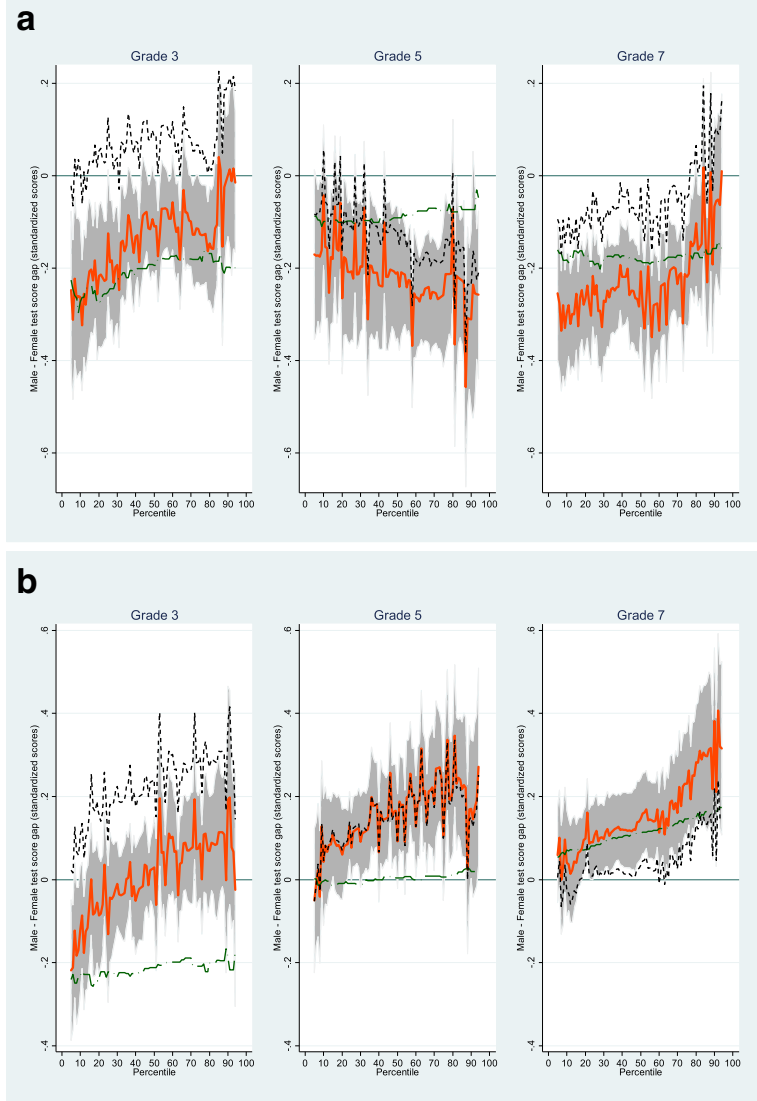
Notes: Gender test score gap: Male - Female, standardized scores. Results are obtained using unconditional quantile regression model 4. Thick solid orange line (grey shaded area) indicates total test score gap estimates (95 % confidence intervals). Green long dash dot (black short dash) line shows the characteristic (return) effect. Confidence intervals are obtained using 500 bootstrap replications. 95 % confidence interval estimates for the aggregate characteristic and return effect are not reported to keep the figures discernible.

Fig. 5 Decomposition of test score gap along the distribution by test subject and grade. Panel **a**: Reading. Panel **b**: Numeracy

Availability of data and materials
This is an empirical paper using a data set but the data are confidential and cannot be published. However, the computer programs (STATA) to replicate the results will be made available upon request.

Competing interests
The IZA Journal of Labor Economics is committed to the IZA Guiding Principles of Research Integrity. The authors declare that they have observed these principles.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia. [2]School of Population and Global Health, University of Western Australia, Perth, Australia. [3]Bankwest Curtin Economics Centre, Curtin University, GPO Box U1987, Perth, WA 6845, Australia.

References
ACARA (2014) National Assessment Program – literacy and numeracy 2013: technical report. Australian Curriculum, Assessment and Reporting Authority (ACARA), Sydney

Allison PD (2001) Missing Data. SAGE Publications, Inc., Thousand Oaks, CA

Australian Health Ethics Committee (2004) Ethical guidelines on the use of assisted reproductive technology in clinical practice and research, National Health and Medical Research Council, Canberra

Baker M, Milligan K (2016) Boy-girl differences in parental time investments: evidence from three countries. J Hum Cap 10:399–441

Baron-Cohen S (2007) The essential difference: men, women, and the extreme male brain. Penguin Books Limited, London

Bedard K, Cho I (2010) Early gender test score gaps across OECD countries. Econ Educ Rev 29:348–363

Bernal R (2008) The effect of maternal employment and child care on children's cognitive development*. Int Econ Rev 49:1173–1209

Bertrand M, Pan J (2013) The trouble with boys: social influences and the gender gap in disruptive behavior. Am Econ J Appl Econ 5:32–64

Blinder AS (1973) Wage discrimination: reduced form and structural estimates. J Hum Resour 8:436–455

Block JH (1976) Issues, problems, and pitfalls in assessing sex differences: a critical review of "the psychology of sex differences". Merrill-Palmer Qu Behav Dev 22:283–308

Booth AL, Kee HJ (2011) A long-run view of the university gender gap in Australia. Aust Econ Hist Rev 51:254–276

Borah BJ, Basu A (2013) Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence. Health Econ 22:1052–1070

Botelho F, Madeira R, Rangel MA (2015) Racial discrimination in grading: evidence from Brazil. Am Econ J Appl Econ 7: 37–52

Card D (1999) Chapter 30 - the causal effect of education on earnings. In: Orley CA, David C (eds) Handbook of labor economics. Elsevier, Amsterdam. pp 1801–1863

Christopher C, David BM, Jessica Van P (2013) Noncognitive skills and the gender disparities in test scores and teacher assessments: evidence from primary school. J Hum Resour 48:236–264

Cobb-Clark DA, Moschion J (2017) Gender gaps in early educational achievement. J Popul Econ 3:1093–1134

Cobb-Clark DA, Nguyen T-H (2012) Educational attainment across generations: the role of immigration background. Econ Rec 88:554–575

Coleman JS, Campbell EQ, Hobson CJ, McPartland J, Mood AM, Weinfeld FD, York R (1966) Equality of educational opportunity. U.S. Government Printing Office, Washington D.C.

Cortes KE (2006) The effects of age at arrival and enclave schools on the academic performance of immigrant children. Econ Educ Rev 25:121–132

Cunha F, Heckman JJ, Schennach SM (2010) Estimating the technology of cognitive and noncognitive skill formation. Econometrica 78:883–931

Currie J (2009) Healthy, wealthy, and wise: socioeconomic status, poor health in childhood, and human capital development. J Econ Lit 47:87–122

Daraganova G, Edwards B, Sipthorp M (2013) Using National Assessment Program—Literacy and numeracy (NAPLAN) data in the longitudinal study of Australian children (LSAC), LSAC technical paper no. 8. Australian Institute of Family Studies, Canberra

Dickerson A, McIntosh S, Valente C (2015) Do the maths: an analysis of the gender gap in mathematics in Africa. Econ Educ Rev 46:1–22

Dobbins TA, Sullivan EA, Roberts CL, Simpson JM (2012) Australian national birthweight percentiles by sex and gestational age, 1998-2007. Med J Aust 197:291

Duckworth AL, Seligman MEP (2006) Self-discipline gives girls the edge: gender in self-discipline, grades, and achievement test scores. J Educ Psychol 98:198–208

Dunn, L.M., Dunn, L.M., 1997. Examiner's manual for the PPVT-III Peabody picture vocabulary test: form IIIA and form IIIB: AGS

Elder T, Jepsen C (2014) Are Catholic primary schools more effective than public primary schools? J Urban Econ 80:28–38

Falch T, Naper LR (2013) Educational evaluation schemes and gender gaps in student achievement. Econ Educ Rev 36:12–25

Firpo S (2007) Efficient semiparametric estimation of quantile treatment effects. Econometrica 75:259–276

Firpo S, Fortin NM, Lemieux T (2009) Unconditional quantile regressions. Econometrica 77:953–973

Fisher J, Marchand J (2014) Does the retirement consumption puzzle differ across the distribution? J Econ Inequal 12: 279–296

Fortin N, Lemieux T, Firpo S (2011) Chapter 1 - decomposition methods in economics. In: Orley A, David C (eds) Handbook of labor economics. Amsterdam: Elsevier, pp 1–102

Fortin NM (2008) The gender wage gap among young adults in the United States. J Hum Resour 43:884–918

Fortin NM, Oreopoulos P, Phipps S (2015) Leaving boys behind: gender disparities in high academic achievement. J Hum Resour 50:549–579

Fryer Jr RG, Levitt SD (2004) Understanding the black-white test score gap in the first two years of school. Rev Econ Stat 86:447–464

Fryer RG, Levitt S (2010) An empirical analysis of the gender gap in mathematics. Am Econ J Appl Econ 2:210–240

Gevrek ZE, Seiberlich RR (2014) Semiparametric decomposition of the gender achievement gap: an application for Turkey. Labour Econ 31:27–44

Gneezy U, Niederle M, Rustichini A (2003) Performance in competitive environments: gender differences. Q J Econ 118: 1049–1074

Golsteyn BHH, Schils T (2014) Gender gaps in primary school achievement: a decomposition into endowments and returns to IQ and non-cognitive factors. Econ Educ Rev 41:176–187

Guiso L, Monte F, Sapienza P, Zingales L (2008) Culture, gender, and math. Science 320:1164–1165

Heckman JJ (2000) Policies to foster human capital. Res Econ 54:3–56

Heckman JJ, Kautz T (2013) Fostering and measuring skills: interventions that improve character and cognition. In: Heckman JJ, Humphries JE, Kautz T (eds) The myth of achievement tests: the GED and the role of character in American life. The University of Chicago Press, Chicago, IL, pp 341–430

Heckman JJ, Kautz T (2014) Fostering and measuring skills: interventions that improve character and cognition. In: Heckman JJ, Humphries JE, Kautz T (eds) The myth of achievement tests: the GED and the role of character in American life. University of Chicago Press, Chicago, pp 341–430

Hinnerich BT, Höglin E, Johannesson M (2011) Are boys discriminated in Swedish high schools? Econ Educ Rev 30:682–690

Hirsch BT, Winters JV (2014) An anatomy of racial and ethnic trends in male earnings in the U.S. Rev Income Wealth 60:930–947

Homel, J., Mavisakalyan, A., Nguyen, H.T., Ryan, C., 2012. School completion: what we learn from different measures of family background. Longitudinal surveys of Australian youth research report number 59

Husain M, Millimet DL (2009) The mythical 'boy crisis'? Econ Educ Rev 28:38–48

Jacob BA (2002) Where the boys aren't: non-cognitive skills, returns to school and the gender gap in higher education. Econ Educ Rev 21:589–598

Jann B (2008) The Blinder-Oaxaca decomposition for linear regression models. Stata J 8:453–479

Jones FL (1983) On decomposing the wage gap: a critical comment on Blinder's method. J Hum Resour 18:126–130

Jones FL, Kelley J (1984) Decomposing differences between groups: a cautionary note on measuring discrimination. Sociol Methods Res 12:323–343

Justman, M., Méndez, S.J., 2016. Gendered selection of STEM subjects for matriculation. Melbourne institute working paper no. 10/16

Kassenboehmer SC, Sinning MG (2014) Distributional changes in the gender wage gap. Ind Labor Relat Rev 67:335–361

Kimura D (2000) Sex and cognition. Cambridge, Massachusetts: MIT press

Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46:33–50

Lai F (2010) Are boys left behind? The evolution of the gender achievement gap in Beijing's middle schools. Econ Educ Rev 29:383–399

Lavy V (2008) Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. J Public Econ 92:2083–2105

Lavy, V., Sand, E., 2015. On the origins of gender human capital gaps: short and long term consequences of teachers' stereotypical biases. National Bureau of economic research working paper number 20909

Le HT, Booth AL (2013) Inequality in Vietnamese urban–rural living standards, 1993–2006. Rev Income Wealth 60:862–886

Lemos Md, Doig B (1999) Who am I? Developmental assessment manual. ACER Press, Melbourne

Lewis M, Brooks-Gunn J (1979) Toward a theory of social cognition: the development of self. New Dir Child Adolesc Dev 1979:1–20

Lewis M, Freedle R (1972) Mother-infant dyad: the cradle of meaning, ETS research bulletin series 1972, pp i–43

Marks GN (2008) Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries. Oxf Rev Educ 34:89–109

Morin L-P (2015) Do men and women respond differently to competition? Evidence from a major education reform. J Labor Econ 33:443–491

Neumark D (1988) Employers' discriminatory behavior and the estimation of wage discrimination. J Hum Resour 23:279–295

Nghiem HS, Nguyen HT, Khanam R, Connelly LB (2015) Does school type affect cognitive and non-cognitive development in children? Evidence from Australian primary schools. Labour Econ 33:55–65

Niederle M, Vesterlund L (2010) Explaining the gender gap in math test scores: the role of competition. J Econ Perspect 24:129–144

Norton, A., Monahan, K., 2015. Wave 6 weighting and non-response, LSAC technical paper no. 15: National Centre for longitudinal data, Canberra

Oaxaca R (1973) Male-female wage differentials in urban labor markets. Int Econ Rev 14:693–709

Schoeni RF, House JS, Kaplan GA, Pollack H (2008) Making Americans healthier: social and economic policy as health policy. Russell Sage Foundation, New York

Simon B, Greaves E (2013) Test scores, subjective assessment, and stereotyping of ethnic minorities. J Labor Econ 31: 535–576

Sohn K (2012) A new insight into the gender gap in math. Bull Econ Res 64:135–155

Stoet G, Geary DC (2013) Sex differences in mathematics and reading achievement are inversely related: within-and across-nation assessment of 10 years of PISA data. PLoS One 8:e57988

Todd PE, Wolpin KI (2007) The production of cognitive achievement in children: home, school, and racial test score gaps. J Hum Cap 1:91–136

van Ours JC, Veenman J (2006) Age at immigration and educational attainment of young immigrants. Econ Lett 90: 310–316

Vandenberg SG (1967) Primary mental abilities or general intelligence? Evidence from twin studies. Eugen Soc Symp 4: 146–160

Wilder GZ, Powell K (1989) Sex differences in test performance: a survey of the literature, ETS research report series 1989, pp i–50