BANKWEST CURTIN ECONOMICS CENTRE

WORKING PAPER SERIES

15/7: THE EVOLUTION OF THE GENDER TEST SCORE GAP THROUGH SEVENTH GRADE: NEW INSIGHTS FROM AUSTRALIA USING QUANTILE REGRESSION AND DECOMPOSITON

Ha Trong Nguyen

business.curtin.edu.au/bcec

Bankwest Curtin Economics Centre Working Paper Series
ISSN: 2202-2791
ISBN: 978-1-925083-34-7

Ha Trong Nguyen

Curtin University

Suggested Citation

Ha Trong Nguyen, November, 2015 "The evolution of the gender test score gap through seventh grade: New insights from Australia using quantile regression and decomposition." Bankwest Curtin Economics Centre Working Paper 15/7, Perth: Curtin University.

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

Ha Trong Nguyen[*]

Curtin University

Abstract

This paper documents the patterns of and examines factors contributing to a gender test score gap in five test subjects in early seven grades of schooling using a recent and nationally representative panel of Australian children. Regression results indicate that females excel at writing and grammar 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 writing and numeracy as the 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.

Keywords: Gender, Education, Quantile regression, Decomposition, Australia.
JEL classification: I20, J16.

---

[*] Corresponding author: Bankwest Curtin Economics Centre | Curtin Business School | Curtin University | Tel:+61 8 9266 5711 | Fax:+61 8 9266 2373 | Postal: GPO Box U1987, Perth WA 6845, Australia | Email: ha.nguyen@curtin.edu.au.

Acknowledgements: Research assistance from Christian Duplock and Huong Le and funding from Curtin Business School's Journal Publication Support Award are gratefully acknowledged. This paper uses unit record data from Growing Up in Australia, the Longitudinal Study of Australian Children. The study is conducted in partnership between the Department of Social Services (DSS), the Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics (ABS). The findings and views reported in this paper are those of the authors and should not be attributed to the DSS, the AIFS or the ABS.

1. Introduction

Gender differentials in educational outcomes 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). It is also widely acknowledged that achieving gender equality in education would reduce gender inequality in wages (Morton and Rufolo, 1990; Black *et al.*, 2008). While there is a rich international literature on a gender gap in educational outcomes, evidence from Australia is surprisingly rare. So far most Australian studies have documented a gender gap in educational outcomes by including a gender dummy variable in the educational outcome equations and overlooked factors contributing to the gender test score gap (Marks, 2008; Booth and Kee, 2011; Nghiem *et al.*, 2015). This paper aims to fill this gap in the Australian literature and contributes to the international literature as the first to explicitly examine the gender test score gap in Australia. It does so by documenting the evolution of and examining factors contributing to a gender gap in academic achievements in five test subjects in early seven grades of schooling using a recent and nationally representative Longitudinal Study of Australian Children (LSAC) survey.

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 methodological additions to the current literature. The first methodological contribution is that with remarkably rich panel data relative to the previous international literature - containing five assessments over the first seven years of schooling of the same children and an exhaustive list of home and school environments, we are able to test several socialisation theories. For example, one of the particular advantages of our data is that we observe pre-school cognitive skills of students, allowing us to investigate the way that initial academic endowments contribute to the gender test score over their first seven years of schooling. As another example, in our data, we observe 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 observe and 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). Our data thus allow us to examine the evolution of the gender test score through higher grades than that in the previous US studies. The second methodological contribution is that this paper is one of a few papers in this 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). To do so this paper applies a newly developed method called the unconditional quantile regression (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 of the distribution. The estimates from the unconditional quantile regression then can be directly applied to an Oaxaca-Blinder decomposition method to examine factors contributing to the gender test score gap across the entire distribution. We do just that and therefore make our third methodological contribution to the literature as one of a few papers applying a quantile decomposition method (Sohn, 2010; Gevrek and Seiberlich, 2014).

By using the first five waves of the LSAC survey, we document that females outperform in grade 5 and 7 writing and grade 7 grammar while males excel at numeracy at all grades, whether at means or along the distribution. We also uncover the heterogeneous patterns in the gender test score gap across the test score distribution, by test subjects and test grades. Our regression results also reveal a widening gender test score gap in writing and numeracy as the 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 our empirical regression 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 excelling 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 schooling (Coleman et al., 1966; Husain and Millimet, 2009; Fryer and Levitt, 2010).

Research has also been devoted to attempting to explain the recognised patterns in gender educational gap that came up with a wide range of different 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). There are also studies that have documented how gender differences in competition may contribute to the gender differences in educational outcomes (Gneezy et al., 2003; Niederle and Vesterlund, 2010). Differences in parental time investment in children may also explain a part of the observed gender differences in educational outcomes (Baker and Milligan, 2013). The gender gap may also be affected by school organisation and classroom factors (Tansel, 2002; Falch and Naper, 2013). In addition, social and cultural conditioning and gender-biased environments can have an impact on test performance (Guiso *et al.*, 2008; Bedard and Cho, 2010; Dickerson *et al.*, 2015). 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. This present paper contributes to the literature by assessing the role of pre-school cognitive skill endowments in contributing to the gender academic achievement gap and how that role evolves as the students advance their schooling.

Australian studies have also documented gender differences in academic outcomes at all educational levels. For example, Nghiem *et al.* (2015) use 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. As another example, Marks (2008) uses 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. At the tertiary educational level, Booth and Kee (2011) use aggregate data to report that Australian females were more likely than males to be enrolled at university since 1987. These studies attempt to capture the gender test score gap by including a gender dummy variable in a multivariate regression framework and only examine the mean gap. In addition, as we are aware of so far no Australian study has examined factors contributing to the gender gap in educational outcomes.

3. Data and descriptive statistics

3.1. Data and sample

We use data from the first five waves of the biannually nationally representative LSAC survey. The LSAC initiated in 2004 contains comprehensive information about children's test scores and other socio-economic and demographic background of children and their parents. The LSAC sampling frame consists of all children born between March 2003 and February 2004 (B-Cohort, infants aged 0–1 year in 2004), and between March 1999 and February 2000 (K-Cohort, children aged 4–5 years in 2004). In this study we focus on children of K-cohort because measures on student test scores are more widely available for this cohort in the first five waves of the survey.

We choose results from the National Assessment Program – Literacy and Numeracy (NAPLAN) test to indicate the academic achievements of students.[1] The NAPLAN test is required for 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 schools and over time (ACARA, 2014). The NAPLAN test results of children were collected via data linkage with LSAC data (Daraganova et al., 2013). At the time of this study, the linkage data for LSAC are mainly available for students in grades 3, 5, and 7. We thus focus on test results at these grades and use results of all test subjects in order to measure the academic achievements of students. 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.[2] This matching exercise shows that NAPLAN test scores in grades 3, 5, and 7 are merged with survey data in wave 2, 3, and 4, respectively.

To measure the initial stocks of students' academic achievements, 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 listening comprehension ability for spoken words in standard English (Dunn and Dunn, 1997). The PPVT test required a child to show the picture that best represents the meaning of a stimuli word spoken by the examiner. The Who Am I (WAI) test is also administered by an interviewer to measure the ability of pre-school age

[1] 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 we don't use them in our analysis.

[2] We address the differences in test dates and survey dates in our empirical models by including dummies for survey months and test and survey years (see Section 4).

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 test scores have been used widely to proxy child cognitive development in economic literature (Fiorini and Keane, 2014; Nghiem *et al.*, 2015). We use PPVT and WAI scores in Wave 1 when the student is 4 or 5 years old (i.e., before enrolling primary school).

3.2. Sample

In our analysis, we concentrate on K cohort students because test scores are more widely available for them. Furthermore, among students who took any test in any test grade, we focus on about 98 % of those who completed all five test subjects. We further restrict our sample to students without missing information on a list of important explanatory variables. To keep the results comparable over time, we use specifications that use variables which are available in all waves and contain the least missing information (see Table 1 and Section 4 for a list of variables included in our baseline models). These variables are commonly used in studies which use 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.[3]

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 2415, 2727, and 2771 students in Wave 2, 3, and 4, respectively. Note that we have a slightly smaller number of students in Wave 2 in our sample because the grade 3 NAPLAN tests were first introduced in 2008 when some K cohort students might have attended higher grades so did not take the test.

3.3. Summary statistics by gender

Summary statistics by gender for the student's background characteristics and home environment variables that we use in the analysis are presented in Table 1. Insignificant gender differences in parental characteristics (such as mother's nativity, education, work status, family size, income, and home ownership status) suggest that the gender of children in our sample is randomly assigned across families. There is also no significant difference in most of our measures of parental investment in children development such as mother working hours, mother parenting style, access to computer or school sectors. The only distinguishable

[3] To examine the impact of other important variables and check the robustness of the results, in extended specifications, we include a richer list of variables where possible. Our data include father information including age, education, work status, and nativity. However, due to a large number of missing occasions (13 % of our final sample has missing value), we do not use father information in our baseline specifications as some US studies did (Fryer Jr and Levitt, 2004; Fryer and Levitt, 2010; Bertrand and Pan, 2013).

gender difference is female students have more children books at home. We also notice significant gender differences in terms of initial cognitive and health endowments. In particular, female students have an academic advantage even before they started their school years because their PPVT or WAI scores measured at ages 4 or 5 are higher than male students'. In contrast, female students have an initially low stock of health as presented by a higher percentage of them having birth weight of 2500 g or lower. We also observe female students in the sample are slightly older (one month) than male students. This gender difference is also consistent with the pattern that girls' mothers are about four 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 about here]

4. Empirical regression models

We follow the prior literature to estimate the gender test score gap by regressing test scores (Y_i) of student i in each test grade and each subject on gender dummy variable ($Male_i$ which takes the value of one 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 gender test score gap from grade 3 of primary schools to either the final grade of primary schools or the first grade of secondary schools.[4] In particular, for each test subject and each test grade, we estimate the raw gender test score gap using the following basic model:

$$Y_i = \alpha + \beta Male_i + \varepsilon_i \qquad (1)$$

where ε_i represents idiosyncratic error terms.

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

$$Y_i = \alpha + \beta Male_i + X_i\gamma + \varepsilon_i \qquad (2)$$

where X_i include the student's characteristics (i.e., age, ethnicity, health status), household characteristics (i.e., mother's migration status[5], household size, parental education, and

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

[5] About 3.5 % of students in our sample were born overseas. We experimented with including students' migration status in their test score equations and found their impact in all equations statistically insignificant. This finding in line with an often found evidence that migration children arriving in the host country at young

household income), indicators of the parental investment in the student's education (e.g., the number of children books at home, access to computers, and an index of "quality time" that parents and children spend together), and indicators of neighbourhood characteristics (i.e., physical infrastructure or neighbourhood social-economic status). We also address the issues of students sitting the NAPLAN test in different years for the same grade by using information both on the age of students at the year they sat the test and dummy variables for test year. We also control for the differences in the survey time and test time by including the dummies for quarters of survey time in regressions. In model 2, we also include state dummy variables to control for differences in educational jurisdictions by states/territories.

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

$$Y_i = \alpha + \beta Male_i + X_i\gamma + E_{0Ki}\theta + \varepsilon_i \qquad (3)$$

We expect the student's initial stock of academic ability to be an important factor explaining the child development as genetic inheritance has been shown to transmit to future cognitive development of children (Todd and Wolpin, 2003; Cunha *et al.*, 2006; Cunha and Heckman, 2007; Cunha *et al.*, 2010).

We first apply an Ordinary Least Squares (OLS) method to estimate the mean gender test score gap using the three specifications described above. 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 along the whole distribution. The unconditional quantile regression (UQR) technique is employed to investigate the gender test score gap along the entire distribution.

The unconditional quantile regression technique is chosen over the (conditional) quantile regression method proposed by Koenker and Bassett (1978) because the latter does not allow us to interpret 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

---

ages have similar academic development as native children (Cortes, 2006; van Ours and Veenman, 2006). We therefore do not include the migration status of students in the final regressions. We do however include the migration status of their mothers in the regressions. English Speaking Background (ESB) countries include the United Kingdom (UK), New Zealand, Canada, US, Ireland and South Africa.

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.[6] The *RIF* for the quantile of interest q_τ is:

$$RIF(Y, q_\tau) = q_\tau + \frac{\tau - D(Y \leq q_\tau)}{f_Y(q_\tau)}, \quad (4)$$

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

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

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 us with 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 Oaxaca-Blinder decomposition method to examine factors contributing to the gender test score gap across the whole distribution without having to implement many simulations necessary in the alternative quantile regression-based decomposition method. This is what we will do in Section 6.

5. Empirical regression results

We report estimates on gender test score gaps at means in five test subjects over the three grade levels (3rd, 5th, and 7th) in Table 2. For each test subject and each grade level, we report results from three specifications and at four different points of test score distribution, one at means (using the OLS) and three selected quantiles (using the UQR).

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

Raw gender test score gaps at means (estimated from model 1, see the first row of each subject panel and the last column of each grade in Table 2) show the well-known gender gaps in both math and reading skills as observed in the international literature: male students outperform female students in maths but lag behind with respect to reading (Husain and Millimet, 2009; Fryer and Levitt, 2010). Raw figures in Table 2 additionally show that, in Australia, female students also outperform male students in writing, spelling and grammar. In

[6] 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).

addition, while excelling at all non-math tests, Australian female students appear to perform best in writing and grammar since the gender test score gaps in their favour are of the highest (20 points or greater). Furthermore, while the gender test score gap in reading, writing, spelling and grammar is already observed in all grades, the (reserve) gender gap in numeracy only presents in grade 5 and 7. Our finding that the gender test score gap in numeracy in favour of male students only presents at a certain educational levels is also in line with previous US findings that a gender math score only observed for US students at their first (Husain and Millimet, 2009) or third grade (Fryer and Levitt, 2010).[7] It is however interesting to note that while these raw figures suggest that a gender math score gap only appears at a certain grade, it takes from two to four more years to observe this pattern in Australia.

[Table 2 about here]

The raw gender test score gaps also show that, by test grades, except for reading where female students' advantage increases to 18 points in grade 5 and drops to 15 points in grade 7, the female students' advantage in writing, spelling and grammar widens as students advance to higher grades. In contrast, the male students' advantage in numeracy seems quite stable at around 11 points in both grade 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 neighbourhood does not change our earlier findings in terms of the magnitude as well as the statistical significant level. However, additionally including students' WAI and PPVT tests measured at ages 4 or 5 in the regression model 3 does. In particular, we observe a reversed and statistically significant (at the 5 % level) gender test score gap in favour of male students in third grade reading: male students now outperform female students by about seven points. Furthermore, the observed gender test score gap in grade 5 and 7 reading turns from statistically significant in model 2 to insignificant in model 3. This is also what we observe for spelling test scores since there is no statistically significant gender test score gap in spelling in all studied grades once students' prior academic endowment is controlled for. Moreover, including students' WAI and PPVT tests in the regressions turns the gender test score gap in third grade writing (grammar) from highly statistically significant to insignificant (marginally significant at the 10 % level) and noticeably reduces (by more than

[7] 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.

10 points) the magnitude of the gap in writing and grammar in grade 5 and 7. 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 twice) the magnitude of the gap in all studied grades.

In summary, above results suggest that including pre-school cognitive skills in the students' development equations while shrinking the gender gap in all non-numeracy subjects widens the gender gap in numeracy domain in terms of the statistical significance level and magnitude. Estimates of the above gender test score gaps also highlight the importance of controlling for students' initial academic endowment in the student development as shown in the literature (Todd and Wolpin, 2007; Bernal, 2008; Cunha *et al.*, 2010; Lai, 2010; Fortin *et al.*, 2013; Elder and Jepsen, 2014; 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 authors' data sets – this is 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).

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

To further explore the heterogeneity in gender test score gaps over the distribution of student performance, we estimate the value added gender gaps from the model 3 for different quantiles of the respective test score distribution. Figure 1 succinctly represents estimates of gender test score gaps (the thick solid orange line) and their respective 95 % confidence intervals[8] (the thin solid orange line) along the test score distribution for five test subjects. While the value added estimates are the focus of our analysis, for comparison purposes, Figure 1 also reports gender test score gap estimates (the thick dotted brown line) and their corresponding 95 % confidence intervals (the thin dotted brown line) obtained using regression model 2 which does not include initial endowment in cognitive skills.

[Figure 1 about here]

[8] 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.

Value added estimates for gender reading test score gaps (Panel A – Figure 1) show the 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 quantile) or top (above the 90th) of the distribution because estimates are statistically significant at these quantiles only. In contrast, females statistically significantly outperform males in grade 7 reading (by about seven points) roughly around the median of the distribution. Thus despite the mean test score gap being statistically indistinguishable from zero, the distributional investigation suggests a female students' statistical significant advantage in grade 7 reading. However, we do not observe any statistically significant difference in reading scores by gender at other remaining quantiles or test grades. We also note that controlling for pre-school cognitive ability at the ages of 4 or 5 reduces the gender reading test score gap favouring female students in terms of the magnitude and statistical significance at nearly all quantiles.

Turning to distributional value added estimates of the gender test score gap in writing (Figure 1 – Panel B) we observe that female students statistically significantly outperform male students at around the 20th to 40th quantiles in grade 3 and at almost all quantiles in grade 5 and 7. In terms of the magnitude, consistent with an earlier finding of a widening gap in writing at means, we also find that gender test score gap estimates are greatest in grade 7 and smallest in grade 3 over virtually the whole distribution. The magnitude of distributional estimates also suggests an opposite pattern of the gap along the distribution of writing test scores in grade 5 and 7. Specifically, in grade 5, female students' advantage decreases along the lower end of the distribution and increases along the higher end. In contrast, the female students' advantage in grade 7 writing first increases along the distribution until the median before starts to decrease. Again, as in the case of reading, we also observe that including students' pre-school cognitive skills in their test score equations noticeably reduces the gender test score gap favouring female students in writing at all quantiles.

Value added estimates on the gender test score gap in spelling (Figure 1 – Panel C) suggests the gap is negative (positive) and statistically significant below the 15th (above the 90th) quantile of the grade 7 spelling test score distribution. Thus although the mean gap is marginally significant (at the 10 % level) and in favour of females, the quantile regression estimates suggest that females statistically significantly (at the 5 % level) outperform males in the lower tail of the distribution whereas males excel in the upper tail. Figure 1 – Panel C additionally suggests that failing to account for pre-school cognitive ability would give upward biased estimates of a gender spelling test score gap in favour of female students.

Analysing the distributions for grammar (Figure 1 – Panel D), value added estimates suggest that female students outperform male students in some quantiles and in some grades only. Specifically, females now excel at grade 5 grammar at the middle (between the 30th and 70th quantile) of the distribution. In addition, in grade 7, female students outperform male students over virtually the whole distribution, and the gender gap appears to increase along the distribution. Comparing distributional estimates of the gender test score difference in grammar obtained from model 2 and model 3 also indicates that controlling for students' pre-school cognitive skills noticeably reduces the gap favouring female students in terms of the size as well as statistical significance level.

Turning to value added estimates on a gender test score gap in numeracy (Figure 1 – Panel E), we observe that males outperform females over virtually the whole distribution and in all grades. We additionally find that the gender numeracy test score gap is more pronounced at the upper end of the distribution. We also observe a widening gender test score gap in numeracy 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 grade 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 the magnitude and statistical significance level.

In summary, the above analysis of the gender test score gap across the distribution points out that focusing on mean gap could overlook important policy relevant heterogeneity across the distribution. Our analysis also highlights that failing to control for pre-school cognitive skills would over-estimate a gender test score gap in favour of female students in non-numeracy subjects and under-estimate a gap in favour of male students in numeracy.

6. Explaining the gender academic achievement gap

In this section, we examine the factors contributing to the male-female test score gap at the mean and at selected quantiles. We do so by following the literature on gender wage gaps (Blinder, 1973; Oaxaca, 1973; Fortin *et al.*, 2011) in applying an Oaxaca-Blinder (OB) type of decomposition of the form:

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

where $\hat{Y}$ is the mean test score of males (m) or females (f), $\hat{Z}$ is a vector of the mean observed characteristics, $\hat{\mu}$ is a vector of the estimated coefficients in the regression of test score on the set of covariates, including the constant, and $\hat{\mu}^{*}$ is a vector of the estimated coefficients from the pooled male and female sample with other covariates and the gender dummy.[9]

In equation (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". Since 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), we do not perform a detailed decomposition of this component. We therefore focus on detailed decomposition of the characteristic effect. To facilitate an interpretation of the results, we group variables contributing to the academic achievement of students into four groups (1) their characteristics, (2) their families' characteristics, (3) their initial cognitive skill endowments, and (4) other factors.

Table 3 reports the estimated total male-female test score gap, together with its contributing factors at the mean and selected quantiles, separated by test subjects and grades. Figure 2 reports concise estimates of total gender test score gap (with their 95 % confidence intervals) and the characteristic and return effect along the whole distribution.[10] Estimates of the total gender gap (results are reported on the first row of each panel in Table 3) are largely similar to those obtained from the regression model 1 (results are reported in Table 2). Table 3 shows that the estimated total gender gaps are statistically insignificant at some points of the test score distribution for some test subjects or grades (for instance, at the 90th quantile of grade 3 and 7 reading, at the median of grade 3 spelling and the 90th quantile of grade 5 and 7 spelling, at the 90th quantile of grade 3 grammar or at means and all quantiles of grade 3 numeracy and at the 10th quantile of grade 5 and 7 numeracy). As it is not meaningful to explain the total gender gaps which are statistically insignificant, in what follows we focus on decomposition results where the gaps are statistically significant.

[9] We include the gender dummy variable in estimating the reference structure to get unbiased estimates of other variables (Neumark, 1988; Fortin, 2008; Jann, 2008). In this paper, we focus on decomposition results of grouped variables so our results are not sensitive to the choice of reference group for categorical variables (Fortin *et al.*, 2011).

[10] 95 % confidence interval estimates for the total characteristic and return effect are not reported to keep the figures discernible.

[Table 3 and Figure 2 about here]

Decomposition results for reading (Panel A in Table 3 and Figure 2) 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 we examine the total gap 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 quantiles) 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 - Figure 2). Regarding the contributions of the characteristic effect, estimates from Table 3 indicate that gender differences in pre-school cognitive skills play the most significant role since their estimates are statistically significant and of the same sign and largely similar magnitude as those of the total characteristic effect. Estimates for factors rather than pre-school cognitive skills, in contrast, 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 the students advance to higher grades.[11] This is consistent with the declining contribution of initial cognitive skill endowments to the total characteristic effect when the students progress school.[12]

Decomposition results for writing scores (Panel B in Table 3 and Figure 2) also suggests that females possess more of the characteristics associated with high writing scores because estimates of the characteristic effect are statistically significant and have the same sign as the total gap estimates. Again, pre-school cognitive skills contribute the most to the characteristic effect since their estimates are also negative and statistically significant. Furthermore, while

[11] In Panel A of Figure 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 the 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.

[12] This trend can be explained as follows. When the students advance 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 decrease. Estimation results (not reported for brevity but will be available upon request) confirm diminishing (but still positive) returns to pre-school cognitive skills along grades.

the characteristic effect contributes to a larger share of the total gap at grade 3 the return effect dominates at grade 5 and 7. Panel B - Figure 2 additionally shows that while the magnitude of the characteristic effect is quite stable along the distribution at grade 3 and 5, it decreases at grade 7. This pattern suggests that gender differences in observable characteristics (dominated by pre-school cognitive skills) have a more important role in explaining the overall gender test score gap at the higher end of the grade 7 writing test score distribution.

Turning to spelling test scores (Panel C in Table 3 and Figure 2) we also observe similar patterns to the gap in reading scores. Specifically, the characteristic effect is negative and statistically significant at the 1 % level, implying that gender differences in observable characteristics suggest an advantage of females over males in spelling test scores, whether at the mean or along the distribution. In addition, pre-school cognitive skills are by far the most important explanatory factors contributing to the characteristic effect. The return effect is statistically insignificant at almost all points of the test score distribution, suggesting that there is no discernible gender difference in transforming observable characteristics into spelling test scores. Panel C in Figure 2 further indicates that as the students advance school, the role of the characteristic effect in explaining the total gap tends to decline while that of the return effect to increase.

Decomposition results of grammar test scores (Panel D in Table 3 and Figure 2) also show that gender differences in observable characteristics (especially pre-school cognitive skills) favouring female students predict their advantage over male students in this subject since estimates of the characteristic effect is negative and highly statistically significant. Furthermore, the greater (in absolute terms) estimates for the characteristic effect over the return effect indicate that the former contributes a larger share to the total gap that the latter, whether at means or along the distribution. However, the relative contribution of the characteristic effect (over the return effect) to the total gender gap appears to decline as the students advance school.[13]

Finally, Panels E in Table 3 and Figure 2 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. Again, similar to the gap in all non-numeracy subjects, pre-school cognitive skills account for the most of the characteristic effect

[13] Visually, from Panel D in Figure 2, this can be seen as the characteristic effect line moves further away from the total gap line while the return effect line moves closer to the total gap line as the students advance school.

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 grade 5 and 7 only. Panel E in Figure 2 additionally shows that, at grade 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. The return effect line also diverts from 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 pre-school cognitive skills in explaining the gap. Our decomposition results further suggest that failing to account for initial education inputs would considerably limit our ability to explain factors contributing to the gender test score gap. This prediction is supported by a finding reported in previous US studies (Husain and Millimet, 2009; Fryer and Levitt, 2010; Sohn, 2012) that characteristics rather than pre-school cognitive skills play an insignificant role in explaining the gender test score gap.[14] The decomposition analysis also suggests that focusing on only the mean gap would overlook important policy relevant heterogeneity across the distribution.

7. Conclusion

Drawing on the recent and nationally representative panel of Australian children, we have examined the patterns of and factors contributing to gender test score gap in five test subjects over the first 7 years of schooling. Regression results reveal that females outperform in grade

[14] In unreported robustness analyses, we include 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). 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 will be available upon request.

5 and 7 writing and grade 7 grammar while males excel at numeracy at all grades, whether at means or along the distribution. While mean regression results show a male advantage in grade 3 reading or a female advantage in grade 5 grammar, quantile regression results suggest that those gender test score gaps at means may have been driven by the academic achievements of students at some particular points of the distribution only. In addition, while mean regressions do not show noticeable gender differences in grade 7 reading and spelling, quantile regression results suggest females do outperform males at some points of the test score distribution. Our regression results also reveal a widening gender test score gap in writing and numeracy as the students advance 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 Oaxaca-Blinder decomposition method, we have examined the impacts of gender differences in resources and their returns on academic achievements. 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. However, at higher grades, especially in numeracy, male students are more efficient in converting education inputs into higher test scores. To this end, further studies on factors contributing to the male students' greater efficiency in transferring education inputs into higher test scores would be worthwhile.

From a policy perspective, it is important to understand the patterns of 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 our results is that pre-school cognitive skills play a significant role in explaining the gender test score gap observed up to grade seven. This result suggests that policies aiming at reducing the gender test score gap should be implemented even prior to students enrolling at school. While this policy implication is not new to the skill development literature which usually shows early intervention is more beneficial than late intervention (Heckman, 2000), it appears to be novel to the gender test score gap literature. 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.

References

ACARA, 2014. National Assessment Program – Literacy and Numeracy 2013: Technical Report. Sydney: Australian Curriculum, Assessment and Reporting Authority (ACARA).

Baker, M., Milligan, K., 2013. Boy-girl differences in parental time investments: Evidence from three countries. National Bureau of Economic Research Working Paper number 18893.

Baron-Cohen, S., 2007. The Essential Difference: Men, Women, and the Extreme Male Brain. UK: Penguin Books Limited.

Bedard, K., Cho, I., 2010. Early gender test score gaps across OECD countries. Economics of Education Review 29, 348-363.

Bernal, R., 2008. The Effect of Maternal Employment and Child Care on Children's Cognitive Development*. International Economic Review 49, 1173-1209.

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

Black, D.A., Haviland, A.M., Sanders, S.G., Taylor, L.J., 2008. Gender Wage Disparities among the Highly Educated. Journal of Human Resources 43, 630-659.

Blinder, A.S., 1973. Wage Discrimination: Reduced Form and Structural Estimates. The Journal of Human Resources 8, 436-455.

Booth, A.L., Kee, H.J., 2011. A long-run view of the university gender gap in Australia*. Australian Economic History Review 51, 254-276.

Card, D., 1999. Chapter 30 - The Causal Effect of Education on Earnings, In: Orley, C.A., David, C. (Eds.), Handbook of labor economics: Elsevier, pp. 1801-1863.

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

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., York, R., 1966. Equality of educational opportunity. Washington D.C.

Cortes, K.E., 2006. The effects of age at arrival and enclave schools on the academic performance of immigrant children. Economics of Education Review 25, 121-132.

Cunha, F., Heckman, J., 2007. The Technology of Skill Formation. The American Economic Review 97, 31-47.

Cunha, F., Heckman, J.J., Lochner, L., Masterov, D.V., 2006. Chapter 12 Interpreting the Evidence on Life Cycle Skill Formation, In: Hanushek, E., Welch, F. (Eds.), Handbook of the Economics of Education: Elsevier, pp. 697-812.

Cunha, F., Heckman, J.J., Schennach, S.M., 2010. Estimating the Technology of Cognitive and Noncognitive Skill Formation. Econometrica 78, 883-931.

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.

Dickerson, A., McIntosh, S., Valente, C., 2015. Do the maths: An analysis of the gender gap in mathematics in Africa. Economics of Education Review 46, 1-22.

Duckworth, A.L., Seligman, M.E.P., 2006. Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. Journal of Educational Psychology 98, 198-208.

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? Journal of Urban Economics 80, 28-38.

Falch, T., Naper, L.R., 2013. Educational evaluation schemes and gender gaps in student achievement. Economics of Education Review 36, 12-25.

Fiorini, M., Keane, M.P., 2014. How the Allocation of Children’s Time Affects Cognitive and Non-Cognitive Development. Journal of Labor Economics 3, 787-836.

Firpo, S., 2007. Efficient Semiparametric Estimation of Quantile Treatment Effects. Econometrica 75, 259-276.

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

Fisher, J., Marchand, J., 2014. Does the retirement consumption puzzle differ across the distribution? The Journal of Economic Inequality 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: Elsevier, pp. 1-102.

Fortin, N.M., 2008. The Gender Wage Gap among Young Adults in the United States. Journal of Human Resources 43, 884-918.

Fortin, N.M., Oreopoulos, P., Phipps, S., 2013. Leaving boys behind: Gender disparities in high academic achievement. National Bureau of Economic Research working paper 19331.

Fryer Jr, R.G., Levitt, S.D., 2004. Understanding the black-white test score gap in the first two years of school. Review of Economics and Statistics 86, 447-464.

Fryer, R.G., Levitt, S., 2010. An Empirical Analysis of the Gender Gap in Mathematics. American Economic Journal: Applied Economics 2, 210-240.

Gevrek, Z.E., Seiberlich, R.R., 2014. Semiparametric decomposition of the gender achievement gap: An application for Turkey. Labour Economics 31, 27-44.

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

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

Guiso, L., Monte, F., Sapienza, P., Zingales, L., 2008. Culture, Gender, and Math. Science 320, 1164-1165.

Heckman, J.J., 2000. Policies to foster human capital. Research in Economics 54, 3-56.

Hirsch, B.T., Winters, J.V., 2014. An Anatomy Of Racial and Ethnic Trends in Male Earnings in the U.S. Review of Income and Wealth 60, 930-947.

Husain, M., Millimet, D.L., 2009. The mythical ‘boy crisis’? Economics of Education Review 28, 38-48.

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

Jann, B., 2008. The Blinder-Oaxaca decomposition for linear regression models. Stata Journal 8, 453-479.

Jones, F.L., 1983. On Decomposing the Wage Gap: A Critical Comment on Blinder's Method. The Journal of Human Resources 18, 126-130.

Jones, F.L., Kelley, J., 1984. Decomposing Differences between Groups: A Cautionary Note on Measuring Discrimination. Sociological Methods & Research 12, 323-343.

Kassenboehmer, S.C., Sinning, M.G., 2014. Distributional changes in the gender wage gap. Industrial & Labor Relations Review 67, 335-361.

Kimura, D., 2000. Sex and cognition: MIT press.

Koenker, R., Bassett, G., Jr., 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. Economics of Education Review 29, 383-399.

Le, H.T., Booth, A.L., 2013. Inequality in Vietnamese Urban–Rural Living Standards, 1993–2006. Review of Income and Wealth 60, 862–886.

Lemos, M.d., Doig, B., 1999. Who Am I? Developmental Assessment Manual. Melbourne: ACER Press.

Marks, G.N., 2008. Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries. Oxford Review of Education 34, 89-109.

Morin, L.-P., 2015. Do Men and Women Respond Differently to Competition? Evidence from a Major Education Reform. Journal of Labor Economics 33, 443-491.

Morton, P., Rufolo, A.M., 1990. Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences. Journal of Labor Economics 8, 123-144.

Neumark, D., 1988. Employers' Discriminatory Behavior and the Estimation of Wage Discrimination. The Journal of Human Resources 23, 279-295.

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

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

Oaxaca, R., 1973. Male-Female Wage Differentials in Urban Labor Markets. International Economic Review 14, 693-709.

Schoeni, R.F., House, J.S., Kaplan, G.A., Pollack, H., 2008. Making Americans healthier: Social and economic policy as health policy: Russell Sage Foundation.

Sohn, K., 2010. The dynamics of the evolution of the Black–White test score gap. Education Economics 20, 175-188.

Sohn, K., 2012. A new insight into the gender gap in math. Bulletin of Economic Research 64, 135-155.

Stoet, G., Geary, D.C., 2013. Sex differences in mathematics and reading achievement are inversely related: Within-and across-nation assessment of 10 years of PISA data. PloS one 8, e57988.

Tansel, A., 2002. Determinants of school attainment of boys and girls in Turkey: individual, household and community factors. Economics of Education Review 21, 455-470.

Todd, P.E., Wolpin, K.I., 2003. On the specification and estimation of the production function for cognitive achievement*. The Economic Journal 113, F3-F33.

Todd, P.E., Wolpin, K.I., 2007. The production of cognitive achievement in children: Home, school, and racial test score gaps. Journal of Human Capital 1, 91-136.

van Ours, J.C., Veenman, J., 2006. Age at immigration and educational attainment of young immigrants. Economics Letters 90, 310-316.

Wilder, G.Z., Powell, K., 1989. Sex differences in test performance: A survey of the literature. ETS Research Report Series 1989, i-50.

Table 1: Summary statistics by gender

Variables		Males	Females	Males-Females
Reading scores				
	Grade 3	419.43	429.83	-10.4***
	Grade 5	499.64	518.08	-18.44***
	Grade 7	546.42	561.27	-14.85***
Writing scores				
	Grade 3	421.37	440.60	-19.24***
	Grade 5	481.31	508.45	-27.14***
	Grade 7	511.32	550.16	-38.84***
Spelling scores				
	Grade 3	412.41	426.99	-14.58***
	Grade 5	487.56	503.01	-15.45***
	Grade 7	537.26	556.99	-19.73***
Grammar				
	Grade 3	421.82	442.06	-20.25***
	Grade 5	505.37	529.61	-24.24***
	Grade 7	541.71	567.34	-25.64***
Numeracy				
	Grade 3	419.64	420.27	-0.63
	Grade 5	506.66	498.03	8.63***
	Grade 7	555.49	546.45	9.04***
Explanatory variables				
Initial PPVT		63.93	64.86	-0.93***
Initial WAI		61.88	66.72	-4.84***
Child age (months)		106.60	107.52	-0.92*
Native		0.97	0.96	0.00
Aboriginal		0.02	0.03	-0.01
Low birth weight		0.07	0.08	-0.02***
Mother age (years)		38.74	39.11	-0.38***
Mother native		0.64	0.63	0.00
Mother from Non-English Speaking Background country		0.22	0.22	0.00
Mother from English Speaking Background country		0.14	0.15	-0.01
Mother has no qualification		0.30	0.30	0.00
Mother has a certificate		0.33	0.32	0.01
Mother has an advanced diploma		0.11	0.09	0.01**
Mother has bachelor degree		0.14	0.15	-0.01
Mother has graduate diploma		0.07	0.07	0.00
Mother has postgraduate degree		0.06	0.07	-0.01
Mother's weekly working hours		18.79	19.34	-0.55
Mother's warm parenting		0.79	0.80	0.00
Having more than 30 children books at home		0.77	0.81	-0.04***

Variables	Males	Females	Males-Females
Having a computer at home	0.93	0.93	0.00
Public school	0.66	0.67	0.00
Catholic school	0.22	0.21	0.01
Other independent school	0.12	0.12	-0.01
Household size	4.60	4.57	0.03
Number of siblings	1.63	1.61	0.02
Number of younger siblings	0.80	0.72	0.08***
Number of same age siblings	0.02	0.03	-0.01**
Living with both parents	0.80	0.79	0.01
Living in an owned home	0.76	0.76	0.00
Household income	88.18	87.24	0.94

Notes: Test scores are measured by test grades while all other variables are calculated from all waves. Analysing each wave separately also 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. The symbol *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

Table 2: Estimated gender score gap over the grades at selected quantiles and at mean

		Grade 3				Grade 5				Grade 7			
Subject	Model	Q10th	Q50th	Q90th	Mean	Q10th	Q50th	Q90th	Mean	Q10th	Q50th	Q90th	Mean
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. Reading	(1)	-19.48***	-8.69**	-1.35	-10.03***	-11.70**	-16.47***	-23.75***	-18.35***	-19.13***	-18.17***	-4.97	-14.88***
		(6.37)	(3.87)	(6.22)	(3.49)	(5.43)	(4.07)	(5.37)	(2.99)	(4.35)	(3.83)	(4.54)	(2.59)
	(2)	-17.06***	-8.12**	-2.71	-9.91***	-8.89*	-15.52***	-25.34***	-17.64***	-17.13***	-17.80***	-4.52	-13.77***
		(6.46)	(3.82)	(5.78)	(3.35)	(4.78)	(3.84)	(5.46)	(2.71)	(4.25)	(3.50)	(4.09)	(2.46)
	(3)	2.88	7.34**	14.53**	6.92**	6.06	-1.95	-11.45**	-3.59	-5.32	-7.00**	5.68	-3.52
		(6.12)	(3.53)	(6.26)	(3.01)	(5.35)	(3.98)	(5.62)	(2.91)	(4.08)	(3.17)	(4.28)	(2.42)
B. Writing	(1)	-35.76***	-19.85***	-8.46	-19.31***	-28.72***	-20.40***	-25.95***	-25.34***	-39.80***	-45.07***	-33.84***	-39.76***
		(7.32)	(3.82)	(5.17)	(2.98)	(4.43)	(2.94)	(4.98)	(2.59)	(4.83)	(3.91)	(5.83)	(2.56)
	(2)	-34.09***	-19.80***	-8.46*	-19.00***	-27.65***	-19.28***	-26.57***	-24.99***	-38.22***	-43.83***	-32.41***	-38.45***
		(7.50)	(3.47)	(4.98)	(2.54)	(4.49)	(2.86)	(4.76)	(2.52)	(4.07)	(3.96)	(5.74)	(2.69)
	(3)	-12.80*	-5.11	6.82	-4.59*	-15.21***	-9.35***	-15.44***	-13.00***	-30.89***	-32.29***	-17.85***	-27.25***
		(7.00)	(3.70)	(5.34)	(2.66)	(4.01)	(3.03)	(5.47)	(2.31)	(4.41)	(3.61)	(5.62)	(2.49)
C. Spelling	(1)	-21.87***	-8.96**	-14.31**	-13.73***	-14.80***	-12.79***	-8.20**	-13.65***	-31.09***	-16.31***	-6.66	-18.05***
		(6.55)	(3.56)	(6.30)	(3.14)	(4.17)	(3.40)	(4.00)	(2.60)	(5.67)	(3.10)	(4.98)	(2.50)
	(2)	-21.24***	-9.45***	-13.78**	-13.27***	-14.26***	-12.69***	-7.25	-13.23***	-30.90***	-16.91***	-6.30	-18.16***
		(6.55)	(3.40)	(5.86)	(2.74)	(4.17)	(3.46)	(4.41)	(2.55)	(5.38)	(2.97)	(4.71)	(2.60)
	(3)	-2.73	6.29*	6.19	3.59	-1.30	1.02	5.49	0.81	-15.43***	-3.49	9.95*	-4.01*
		(6.00)	(3.48)	(6.54)	(2.88)	(3.93)	(3.14)	(4.07)	(2.46)	(5.29)	(2.98)	(5.77)	(2.38)
D. Grammar	(1)	-14.42***	-21.44***	-15.07**	-19.66***	-30.21***	-27.51***	-16.14***	-23.41***	-24.13***	-21.45***	-44.16***	-26.22***
		(5.54)	(4.54)	(5.97)	(3.48)	(6.09)	(3.67)	(5.41)	(3.11)	(4.80)	(3.69)	(9.55)	(2.94)
	(2)	-14.17**	-21.96***	-15.20***	-19.74***	-27.67***	-27.15***	-15.78**	-23.12***	-22.86***	-21.05***	-44.42***	-25.73***
		(5.58)	(4.55)	(5.82)	(3.36)	(5.95)	(3.44)	(6.23)	(3.06)	(4.56)	(3.76)	(9.43)	(2.81)
	(3)	2.10	-4.12	3.54	-1.83	-9.13	-12.42***	-2.97	-7.36**	-9.56**	-7.83**	-20.25***	-11.08***
		(5.25)	(4.50)	(6.08)	(3.08)	(5.75)	(3.54)	(5.78)	(2.95)	(4.32)	(3.42)	(7.47)	(2.78)
E. Numeracy	(1)	-8.23*	1.65	3.44	0.58	2.45	11.29***	13.04**	10.89***	3.18	8.71**	25.06***	10.58***
		(4.67)	(3.60)	(5.08)	(2.88)	(4.06)	(3.42)	(5.66)	(2.76)	(3.72)	(3.63)	(6.03)	(2.71)
	(2)	-6.47	1.83	2.87	0.93	4.64	11.83***	13.40***	11.52***	4.64	9.51***	27.41***	12.15***
		(4.55)	(3.48)	(4.92)	(2.81)	(3.65)	(3.22)	(5.10)	(2.50)	(3.68)	(3.21)	(5.50)	(2.50)
	(3)	7.99*	16.39***	16.00***	15.70***	16.69***	27.09***	29.93***	27.22***	18.11***	25.60***	50.12***	28.94***
		(4.20)	(3.55)	(5.59)	(2.77)	(4.14)	(3.28)	(6.09)	(2.49)	(4.16)	(3.65)	(6.81)	(2.23)

Notes: Females are the base group. Standard errors obtained using 500 bootstrap replications are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 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, working hours, and parenting style), number of books at home, having computer at home, 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.

Table 3: Contributions to the male-female test score gap at mean and selected quantiles by subject and grade

	Grade 3				Grade 5				Grade 7			
	Q10th	Q50th	Q90th	Mean	Q10th	Q50th	Q90th	Mean	Q10th	Q50th	Q90th	Mean
Panel A: Reading												
Predicted total gap	-15.32**	-8.68**	0.60	-10.03***	-13.63**	-17.03***	-24.65***	-18.35***	-16.94***	-15.10***	-4.03	-14.88***
Characteristic effect												
Student	-1.30	-0.08	0.28	-0.14	-1.68	-0.64	-0.33	-0.46	-1.50**	0.31	-0.00	-0.38
	[8.49]	[0.92]	[46.67]	[1.4]	[12.33]	[3.76]	[1.34]	[2.51]	[8.85]	[-2.05]	[0]	[2.55]
Household	-0.60	-0.17	0.15	-0.03	-0.75	-0.29	-0.41	-0.42	-0.66	-1.17	-0.79	-0.95
	[3.92]	[1.96]	[25]	[0.3]	[5.5]	[1.7]	[1.66]	[2.29]	[3.9]	[7.75]	[19.6]	[6.38]
Others	0.66	0.64	2.05	1.09	1.65	1.87*	4.32**	2.11**	0.92	1.20	1.00	0.89
	[-4.31]	[-7.37]	[341.67]	[-10.87]	[-12.11]	[-10.98]	[-17.53]	[-11.5]	[-5.43]	[-7.95]	[-24.81]	[-5.98]
Initial	-21.12***	-16.42***	-18.36***	-17.87***	-16.98***	-15.46***	-15.87***	-15.98***	-12.56***	-11.51***	-10.86***	-10.92***
	[137.86]	[189.17]	[-3060]	[178.17]	[124.58]	[90.78]	[64.38]	[87.08]	[74.14]	[76.23]	[269.48]	[73.39]
Characteristic effect total	-22.36***	-16.03***	-15.88***	-16.95***	-17.76***	-14.52***	-12.30***	-14.76***	-13.81***	-11.17***	-10.65***	-11.36***
	[145.95]	[184.68]	[-2646.67]	[168.99]	[130.3]	[85.26]	[49.9]	[80.44]	[81.52]	[73.97]	[264.27]	[76.34]
Return effect total	7.03	7.35*	16.48***	6.92**	4.14	-2.51	-12.35	-3.59	-3.13	-3.92	6.62	-3.52
	[-45.89]	[-84.68]	[2746.67]	[-68.99]	[-30.37]	[14.74]	[50.1]	[19.56]	[18.48]	[25.96]	[-164.27]	[23.66]
Panel B: Writing												
Predicted total gap	-37.05***	-23.51***	-14.69*	-19.31***	-26.15***	-16.03***	-23.38***	-25.34***	-44.85***	-41.62***	-34.02***	-39.76***
Characteristic effect												
Student	-1.03	-0.41	-0.02	-0.19	-0.19	-0.31	0.97	0.06	-0.53	-0.36	0.03	-0.24
	[2.78]	[1.74]	[0.14]	[0.98]	[0.73]	[1.93]	[-4.15]	[-0.24]	[1.18]	[0.86]	[-0.09]	[0.6]
Household	-0.42	0.39	-0.58	0.06	-0.23	-0.37	-0.31	-0.23	-0.88	-0.79	-0.47	-0.77
	[1.13]	[-1.66]	[3.95]	[-0.31]	[0.88]	[2.31]	[1.33]	[0.91]	[1.96]	[1.9]	[1.38]	[1.94]
Others	0.95	0.80	1.51	0.64	0.87	0.81	1.46	1.33*	0.29	0.61	-0.12	0.38
	[-2.56]	[-3.4]	[-10.28]	[-3.31]	[-3.33]	[-5.05]	[-6.24]	[-5.25]	[-0.65]	[-1.47]	[0.35]	[-0.96]
Initial	-22.46***	-15.50***	-16.19***	-15.24***	-13.96***	-11.17***	-12.64***	-13.50***	-7.79***	-12.24***	-15.42***	-11.89***
	[60.62]	[65.93]	[110.21]	[78.92]	[53.38]	[69.68]	[54.06]	[53.28]	[17.37]	[29.41]	[45.33]	[29.9]
Characteristic effect total	-22.96***	-14.73***	-15.28***	-14.72***	-13.51***	-11.05***	-10.51***	-12.34***	-8.91***	-12.78***	-15.99***	-12.52***
	[61.97]	[62.65]	[104.02]	[76.23]	[51.66]	[68.93]	[44.95]	[48.7]	[19.87]	[30.71]	[47]	[31.49]
Return effect total	-14.08*	-8.78*	0.59	-4.59*	-12.64***	-4.98	-12.87**	-13.00***	-35.94***	-28.85***	-18.04***	-27.25***
	[38]	[37.35]	[-4.02]	[23.77]	[48.34]	[31.07]	[55.05]	[51.3]	[80.13]	[69.32]	[53.03]	[68.54]

Table 3: Contributions to the male-female test score gap at mean and selected quantiles by subject and grade (cont.)

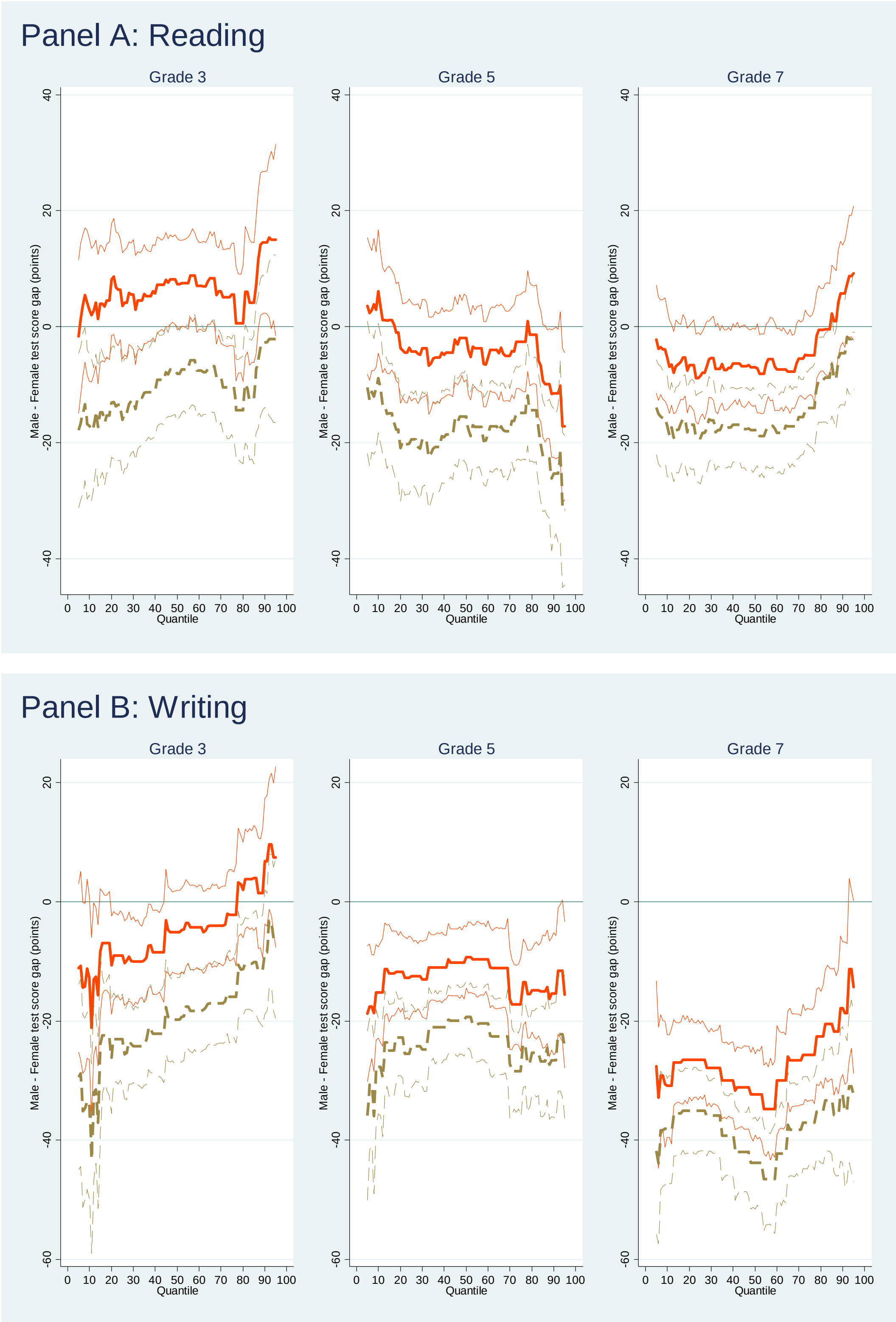
	Grade 3				Grade 5				Grade 7			
	Q10th	Q50th	Q90th	Mean	Q10th	Q50th	Q90th	Mean	Q10th	Q50th	Q90th	Mean
Panel C: Spelling												
Predicted total gap	-21.89***	-4.77	-29.83***	-13.73***	-17.60***	-11.62***	-0.29	-13.65***	-28.97***	-16.32***	-6.57	-18.05***
Characteristic effect												
Student	-0.87	-0.33	-0.11	-0.50	0.48	-0.76	-0.61	-0.23	-1.30*	0.07	0.11	-0.36
	[3.97]	[6.92]	[0.37]	[3.64]	[-2.73]	[6.54]	[210.34]	[1.68]	[4.49]	[-0.43]	[-1.67]	[1.99]
Household	0.39	0.44	0.33	0.31	0.31	-0.10	-0.41	-0.00	-0.05	-0.45	-0.70	-0.24
	[-1.78]	[-9.22]	[-1.11]	[-2.26]	[-1.76]	[0.86]	[141.38]	[0]	[0.17]	[2.76]	[10.65]	[1.33]
Others	0.82	1.16	0.21	0.59	0.25	2.41**	1.56	1.49*	2.08*	1.76**	1.17	1.55**
	[-3.75]	[-24.32]	[-0.7]	[-4.3]	[-1.42]	[-20.74]	[-537.93]	[-10.92]	[-7.18]	[-10.78]	[-17.81]	[-8.59]
Initial	-19.48***	-16.52***	-20.93***	-17.72***	-14.54***	-15.36***	-14.23***	-15.72***	-16.39***	-14.21***	-17.20***	-14.99***
	[88.99]	[346.33]	[70.16]	[129.06]	[82.61]	[132.19]	[4906.9]	[115.16]	[56.58]	[87.07]	[261.8]	[83.05]
Characteristic effect total	-19.14***	-15.25***	-20.50***	-17.32***	-13.50***	-13.81***	-13.69***	-14.46***	-15.66***	-12.83***	-16.61***	-14.04***
	[87.44]	[319.71]	[68.72]	[126.15]	[76.7]	[118.85]	[4720.69]	[105.93]	[54.06]	[78.62]	[252.82]	[77.78]
Return effect total	-2.75	10.48**	-9.33	3.59	-4.10	2.18	13.40**	0.81	-13.32**	-3.50	10.04	-4.01*
	[12.56]	[-219.71]	[31.28]	[-26.15]	[23.3]	[-18.76]	[-4620.69]	[-5.93]	[45.98]	[21.45]	[-152.82]	[22.22]
Panel D: Grammar												
Predicted total gap	-15.48*	-21.12***	-11.30	-19.66***	-36.82***	-22.39***	-36.48***	-23.41***	-26.99***	-23.15***	-36.67***	-26.22***
Characteristic effect												
Student	-1.56	0.02	0.19	-0.12	-2.11	-0.26	-0.88	-0.81	-0.72	-0.24	-0.25	-0.46
	[10.08]	[-0.09]	[-1.68]	[0.61]	[5.73]	[1.16]	[2.41]	[3.46]	[2.67]	[1.04]	[0.68]	[1.75]
Household	-0.08	0.43	-0.40	-0.01	-0.08	-0.17	-0.69	-0.62	-1.31	-1.22	-1.81	-1.05
	[0.52]	[-2.04]	[3.54]	[0.05]	[0.22]	[0.76]	[1.89]	[2.65]	[4.85]	[5.27]	[4.94]	[4]
Others	2.28	1.10	1.38	1.23	2.06	1.98*	2.83	3.16***	1.57	1.87*	3.77**	1.92**
	[-14.73]	[-5.21]	[-12.21]	[-6.26]	[-5.59]	[-8.84]	[-7.76]	[-13.5]	[-5.82]	[-8.08]	[-10.28]	[-7.32]
Initial	-17.17***	-18.87***	-19.77***	-18.92***	-20.95***	-16.65***	-14.44***	-17.78***	-14.11***	-14.04***	-25.62***	-15.53***
	[110.92]	[89.35]	[174.96]	[96.24]	[56.9]	[74.36]	[39.58]	[75.95]	[52.28]	[60.65]	[69.87]	[59.23]
Characteristic effect total	-16.52***	-17.32***	-18.61***	-17.82***	-21.08***	-15.09***	-13.17***	-16.05***	-14.57***	-13.63***	-23.91***	-15.13***
	[106.72]	[82.01]	[164.69]	[90.64]	[57.25]	[67.4]	[36.1]	[68.56]	[53.98]	[58.88]	[65.2]	[57.7]
Return effect total	1.04	-3.80	7.31	-1.83	-15.73*	-7.29*	-23.32*	-7.36**	-12.42**	-9.52**	-12.76*	-11.08***
	[-6.72]	[17.99]	[-64.69]	[9.31]	[42.72]	[32.56]	[63.93]	[31.44]	[46.02]	[41.12]	[34.8]	[42.26]

Table 3: Contributions to the male-female test score gap at mean and selected quantiles by subject and grade (cont.)

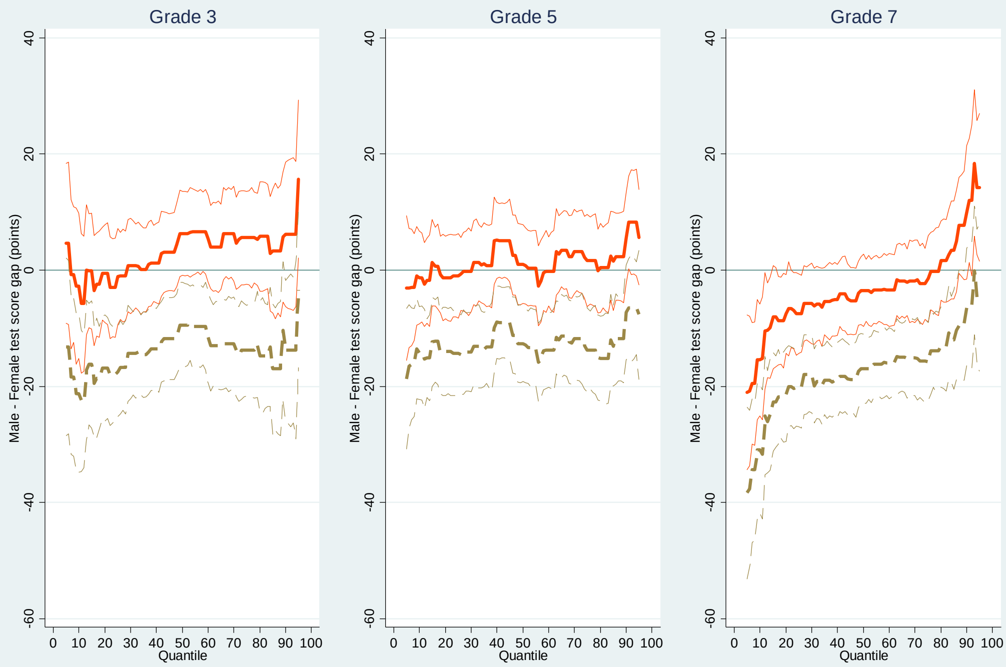
	Grade 3				Grade 5				Grade 7			
	Q10th	Q50th	Q90th	Mean	Q10th	Q50th	Q90th	Mean	Q10th	Q50th	Q90th	Mean
Panel E: Numeracy												
Predicted total gap	-6.97	2.41	13.77	0.58	3.44	11.98***	13.78**	10.89***	0.62	8.89**	28.75***	10.58***
Characteristic effect												
Student	-0.78	0.43	0.52	0.20	-0.60	-0.55	-0.93	-0.51	-0.28	-0.67	-1.03	-0.52
	[11.19]	[17.84]	[3.78]	[34.48]	[-17.44]	[-4.59]	[-6.75]	[-4.68]	[-45.16]	[-7.54]	[-3.58]	[-4.91]
Household	-0.56	0.29	0.01	0.07	-0.62	-0.17	0.11	-0.35	-0.98	-0.69	-1.86	-1.20
	[8.03]	[12.03]	[0.07]	[12.07]	[-18.02]	[-1.42]	[0.8]	[-3.21]	[-158.06]	[-7.76]	[-6.47]	[-11.34]
Others	0.44	-0.08	0.78	0.23	0.54	2.09*	2.44	2.16**	0.61	1.51*	1.86	1.15
	[-6.31]	[-3.32]	[5.66]	[39.66]	[15.7]	[17.45]	[17.71]	[19.83]	[98.39]	[16.99]	[6.47]	[10.87]
Initial	-15.33***	-15.38***	-13.87***	-15.62***	-13.54***	-17.18***	-18.49***	-17.63***	-14.27***	-17.04***	-24.03***	-17.79***
	[219.94]	[-638.17]	[-100.73]	[-2693.1]	[-393.6]	[-143.41]	[-134.18]	[-161.89]	[-2301.61]	[-191.68]	[-83.58]	[-168.15]
Characteristic effect total	-16.23***	-14.74***	-12.56***	-15.12***	-14.23***	-15.81***	-16.88***	-16.33***	-14.93***	-16.89***	-25.06***	-18.36***
	[232.86]	[-611.62]	[-91.21]	[-2606.9]	[-413.66]	[-131.97]	[-122.5]	[-149.95]	[-2408.06]	[-189.99]	[-87.17]	[-173.53]
Return effect total	9.25*	17.15***	26.33***	15.70***	17.67***	27.79***	30.66***	27.22***	15.55***	25.78***	53.82***	28.94***
	[-132.71]	[711.62]	[191.21]	[2706.9]	[513.66]	[231.97]	[222.5]	[249.95]	[2508.06]	[289.99]	[187.2]	[273.53]

Notes: Females are the base group. Standard errors (not reported for brevity) are obtained using 500 bootstrap replications. *** p<0.01, ** p<0.05, * p<0.1.
Estimates from model 3 are used. Values in brackets are percentage of the predicted total gap.
Grouped variables: Student: age, Aboriginal status, and birth weight; Household: mother's characteristics (age, migration background, completed qualification, working hours, and parenting style), number of books at home, having computer at home, 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.

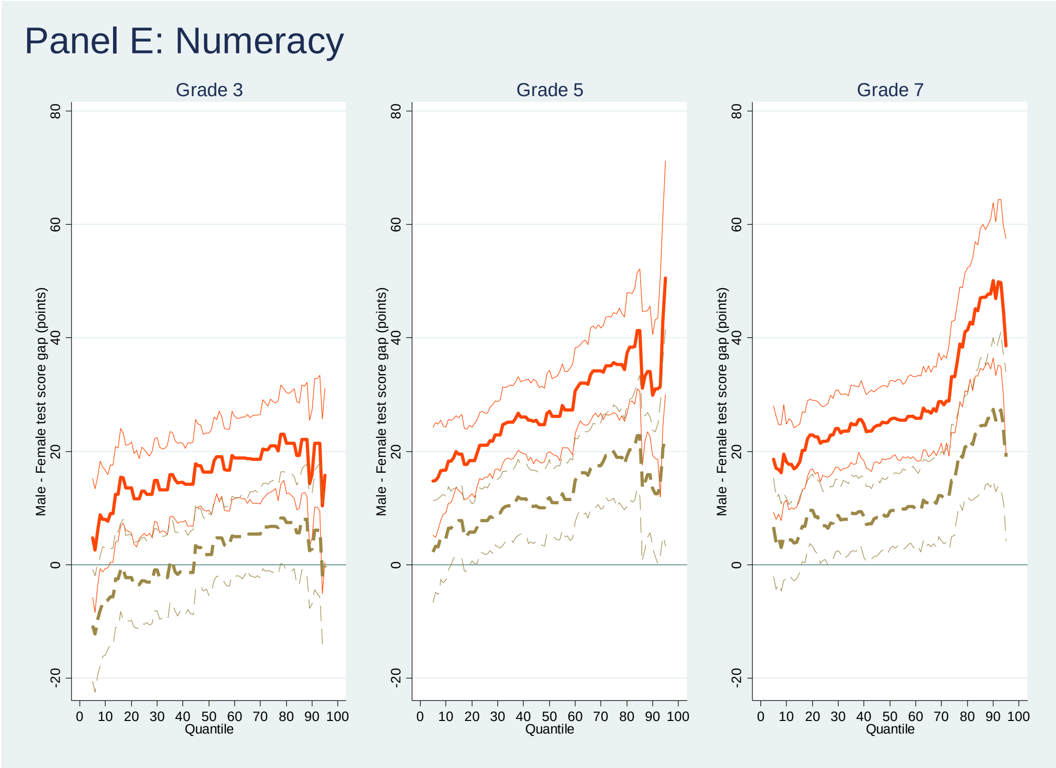
Figure 1: Gender test score gaps along the distribution by test subject and grade



Panel C: Spelling

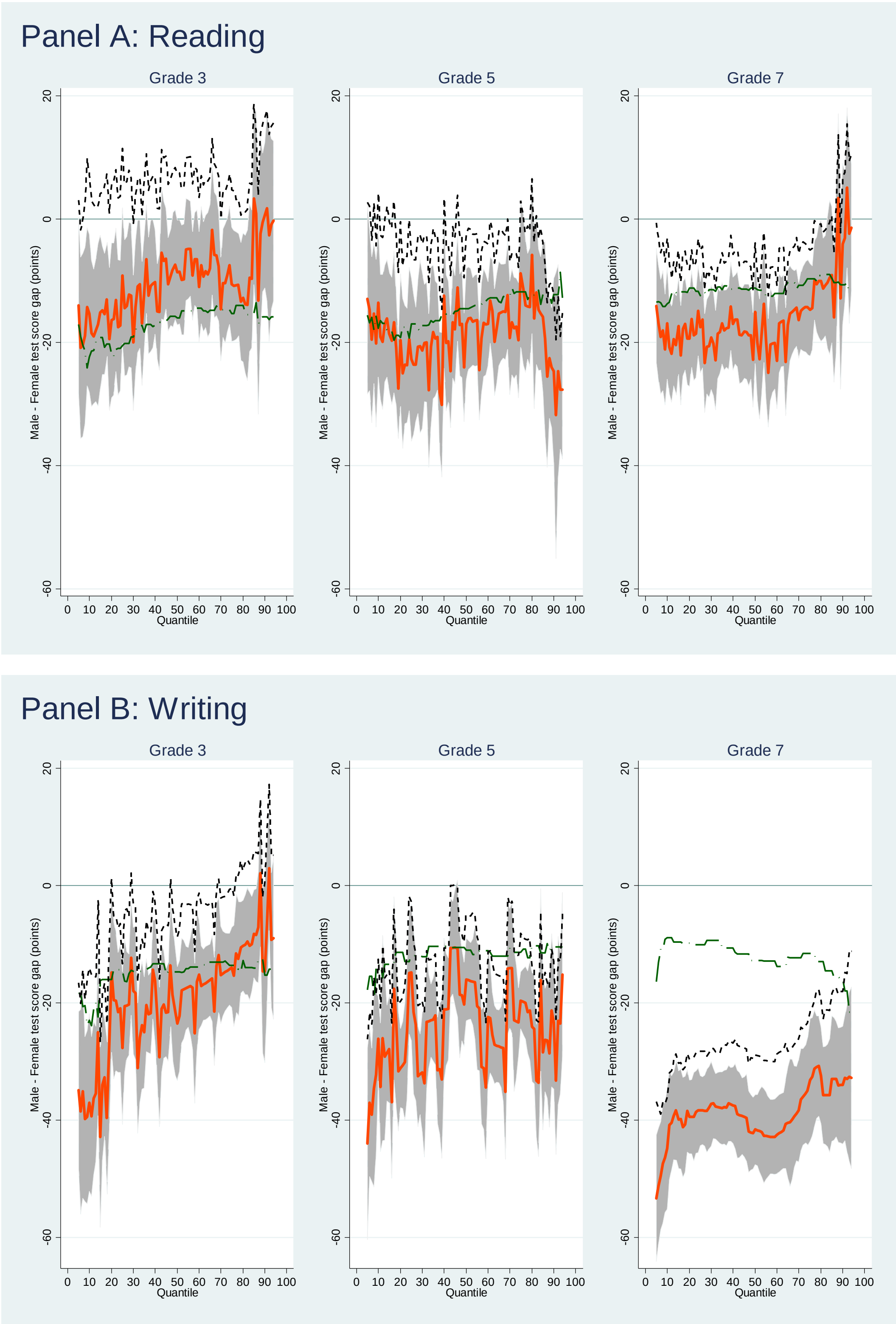


Panel D: Grammar

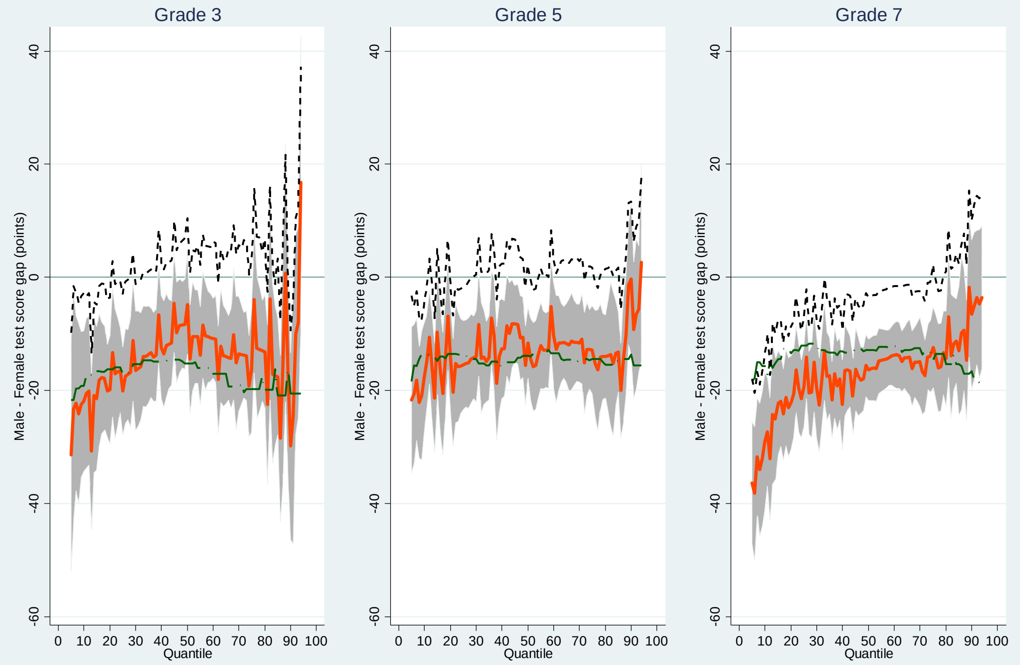


Notes: Gender test score gap: Male - Female, points. Thick (thin) solid orange line indicates test score gap estimates (95 % confidence intervals) using model 3. Thick (thin) dotted brown line shows test score gap estimates (95 % confidence intervals) using model 2. Confidence intervals are obtained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

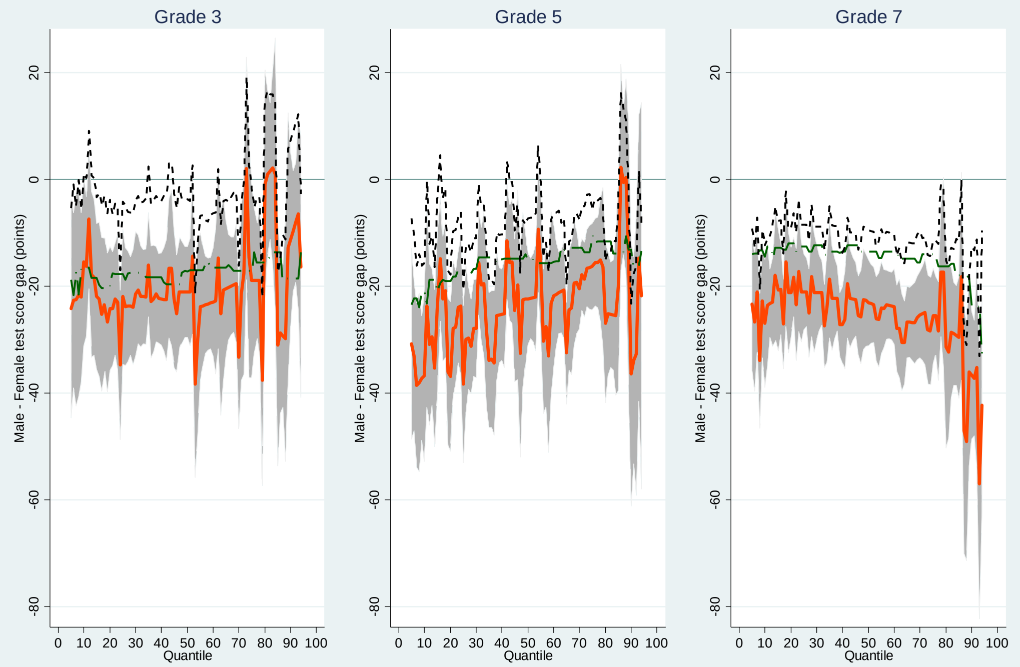
Figure 2: Decomposition of test score gap along the distribution by test subject and grade

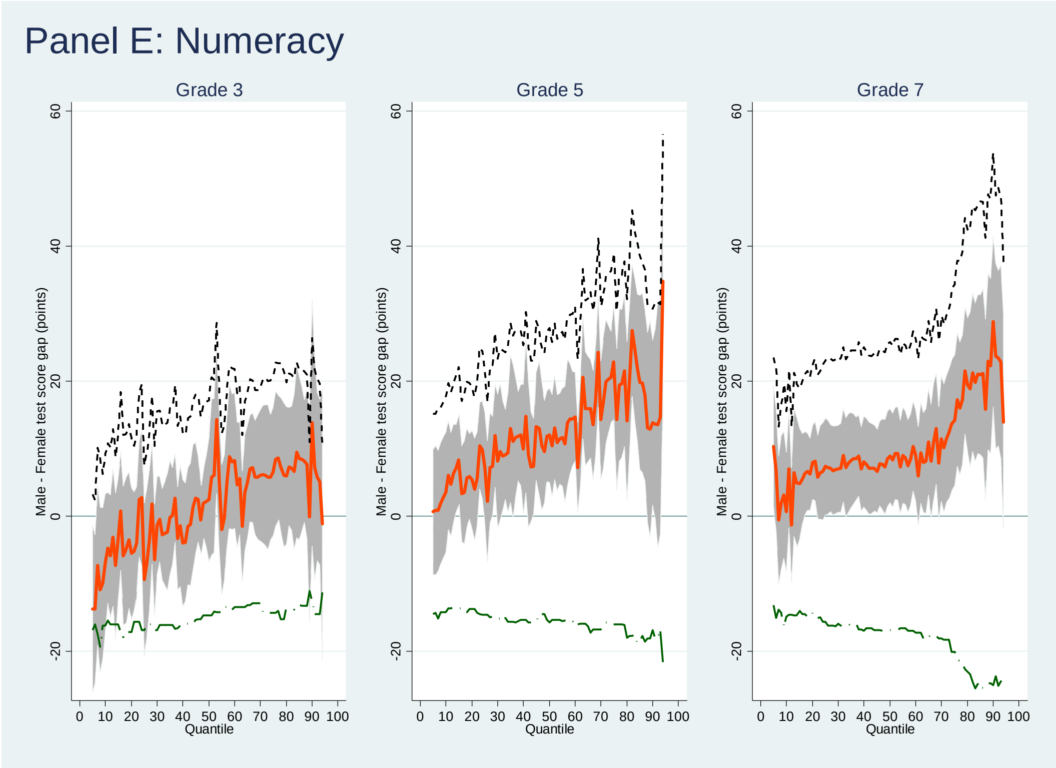


Panel C: Spelling



Panel D: Grammar





Notes: Gender test score gap: Male - Female, points. 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.

The Bankwest Curtin Economics Centre is an independent economic and social research organisation located within the Curtin Business School at Curtin University. The Centre was established in 2012 through the generous support from Bankwest (a division of the Commonwealth Bank of Australia), with a core mission to undertake high quality, objective research on the key economic and social issues of relevance to Western Australia.

The Centre's research and engagement activities are designed to influence economic and social policy debates in state and Federal Parliament, regional and national media, and the wider Australian community. Through high quality, evidence-based research and analysis, our research outcomes inform policy makers and commentators of the economic challenges to achieving sustainable and equitable growth and prosperity both in Western Australia and nationally.

The Centre capitalises on Curtin University's reputation for excellence in economic modelling, forecasting, public policy research, trade and industrial economics and spatial sciences. Centre researchers have specific expertise in economic forecasting, quantitative modelling, microdata analysis and economic and social policy evaluation.

A suite of tailored and national economic models and methods are maintained within the Centre to facilitate advanced economic policy analysis: these include macroeconomic and time series models, micro(simulation) models, computable general equilibrium (CGE) models, spatial modelling methods, economic index analysis, and behavioural modelling methods.

CONTACT

Bankwest Curtin Economics Centre
Curtin University
Kent Street Bentley WA 6102
GPO Box U1987 Perth WA 6845

Tel: +61 8 9266 2873

bcec@curtin.edu.au
business.curtin.edu.au/bcec