TIMSS SOUTH AFRICA WORKING PAPER SERIES

Leaving boys behind: The impact of gendered repetition patterns on South Africa's TIMSS mathematics results

Heleen Hofmeyr

Enquiries related to this working paper can be sent to Heleen Hofmeyr at heleenh24@gmail.com.

Recommended citation
Hofmeyr, H. (2022). Subjective Leaving boys behind: The impact of gendered repetition patterns on South Africa's TIMSS mathematics results. TIMSS Working Paper Series. Human Sciences Research Council.

basic education
Department:
Basic Education
REPUBLIC OF SOUTH AFRICA

HSRC
Human Sciences
Research Council

Leaving boys behind: The impact of gendered repetition patterns on South Africa's TIMSS mathematics results

Heleen Hofmeyr

Abstract:

This paper analyses gender differences in South Africa's TIMSS Grade 9 mathematics achievement, with particular emphasis on the role of gendered grade repetition and dropout patterns in contributing to an apparent pro-girl gap in mathematics achievement. I make use of Oaxaca-Blinder decomposition analysis to decompose the observed pro-girl gap into its explained and unexplained components, separately by school socio-economic quintile. I present new evidence that there are important selection effects at play that influence the observed pro-girl gap in TIMSS mathematics achievement. In particular, comparing the size and age distributions of males and females in the Grade 9 TIMSS sample shows that boys are more likely to have dropped out of school and twice as likely as girls to have repeated a grade by Grade 9. Multivariate regression analysis suggests that males are more likely to be "weeded" out of their cohort as they progress through school, even when they are similar to females along a number of dimensions, including academic achievement.

1. Introduction

For the first time since South Africa's 20-year participation in the Trends in Mathematics and Science Study (TIMSS) at the Grade 8/9 level, girls outperformed boys in mathematics by a statistically significant margin in the 2019 round of the assessment (Mullis *et al.*, 2020). Given evidence that I present elsewhere (Hofmeyr, forthcoming) that part of South Africa's pro-girl gaps in the Progress in International Reading Literacy Study (PIRLS) and TIMSS Grade 5 is due to girls progressing through school faster than boys, the main focus of the analysis in this paper is investigating the extent to which the country's pro-girl gap in Grade 9 mathematics achievement is due to the same phenomenon. The main econometric strategy employed is Oaxaca-Blinder decomposition analysis, where I split the observed pro-girl gap into its explained and unexplained components, with particular emphasis on the contribution of prior grade repetition to the gender gap. Given evidence of differences in both the magnitude of the pro-girl gap and the extent of grade repetition across socio-economic status (SES), I also investigate SES differences in the contribution of prior grade repetition to the pro-girl achievement gap.

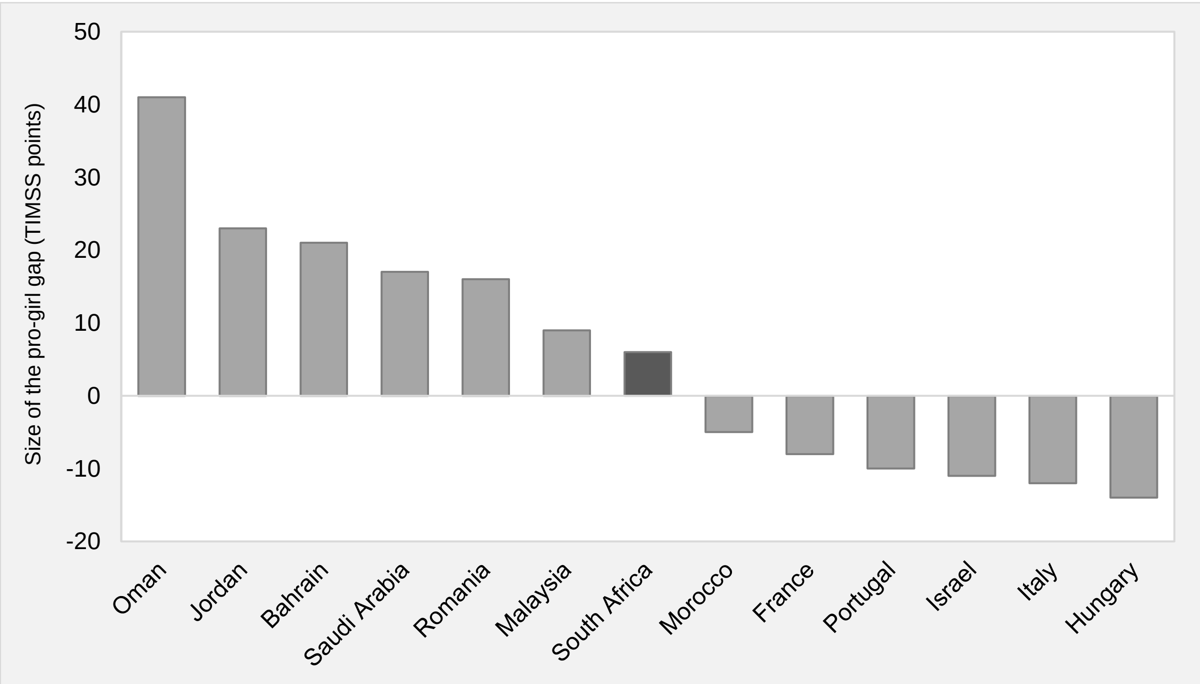
Contributing more evidence of South Africa's pro-girl advantage in schooling outcomes, I find that girls have a distinct advantage in terms of being retained in the system. Not only are there more girls in the TIMSS Grade 9 sample (making up 52% of the sample), but girls are also much less likely to be overage for their grade, indicating that they are less likely to have repeated a grade by the time they reach Grade 9. Regression analysis of the factors associated with grade repetition shows that boys are much more likely to have repeated, even when controlling for TIMSS mathematics achievement. This suggests that mastery of the curriculum is not the only factor that determines whether a student is promoted to the next grade, and moreover that boys may be unfairly discriminated against in grade promotion decisions.

2. Literature review

2.1. Gender gaps in international education assessments

While it used to be the case that boys tended to achieve better educational outcomes in low- and middle-income countries (LMICs) on average, a number of these countries exhibit pro-girl gaps in recent rounds of large-scale educational assessments such as the PIRLS (Mullis *et al.*, 2017), TIMSS (Mullis *et al.*, 2016), and PISA for Development (PISA-D) (OECD, 2018). In the 2019 round of TIMSS, 13 out of 39 participating countries exhibited statistically significant gender gaps in mathematics achievement at the Grade 8/9 level (shown in Figure 1). Of the seven countries with gender gaps favouring girls, four were LMICs (Jordan, Malaysia, Romania, and South Africa). This evidence from large-scale educational assessments points to changing gender differences in learning outcomes internationally. In particular, the evidence strongly indicates that it is no longer the case that schooling outcomes are skewed in favour of boys in LMICs, a trend that has been documented by a number of authors (see for example Grant and Behrman (2010); Zuze and Reddy (2014); Spaull and Makaluza (2019); and Buhl-Wiggers, Jones and Thornton (2021)).

Figure 1: Gender gaps in TIMSS 2019 mathematics achievement



Notes: Only statistically significant gender gaps are plotted. Negative values indicate a pro-boy gap. Source: TIMSS 2019.

While these country-level gender gaps are in themselves noteworthy, they mask differences in the magnitude and direction of gender gaps across socio-economic groupings in South Africa. Evidence of such differences are presented by Zuze and Beku (2019), who show that gender gaps in TIMSS 2015 Grade 9 mathematics achievement differ by school type, with the widest gaps being observed in no-fee (Quintile 1-3[1]) schools. These schools exhibited pro-girl gaps in mathematics achievement, while private schools (serving the wealthiest households) exhibited pro-boy gaps in mathematics achievement.

[1] Schools are assigned quintiles by the Department of Basic Education (DBE), based on the socio-economic status of the surrounding area. Quintile 1 schools are in the most disadvantaged areas, while Quintile 5 schools are in the wealthiest areas. Quintile 1-3 schools are typically no-fee schools, while Quintile 4 and 5 schools typically charge fees.

The reasons for these pro-girl gaps in large-scale education assessments are not well understood. One feature of education systems in LMICs which could influence gender gaps in achievement in international assessments, and which has received little attention in the literature, is differences in the rates at which boys and girls progress through school. Grade repetition tends to be common in LMICs (Crouch *et al.*, 2020), and gender differences in repetition rates are likely to influence achievement differentials in assessments such as TIMSS, which tests students in a particular grade. If boys and girls progress through school at different rates, a given Grade 8 or 9 class participating in TIMSS, for example, will consist of different proportions of boys and girls who have repeated a grade. If boys progress through school slower than girls, boys in a participating Grade 9 class will already be 'selected' to be weaker performing (since repeaters are generally weaker students). I present evidence of such effects in explaining South Africa's pro-girl gap in Grade 4 literacy and Grade 5 mathematics achievement (Hofmeyr, forthcoming). The aim of this paper is to extend that work by investigating the role of gender differences in prior grade repetition in contributing to South Africa's pro-girl gap in TIMSS mathematics achievement at the Grade 9 level[2].

2.2. Grade repetition in South Africa

Like many LMICs, South Africa's education system is characterised by persistent problems in internal efficiency, with grade repetition being extremely common in all grades, including the early grades (Van der Berg, Van Wyk, *et al.*, 2019). Holding back learners who do not meet the requirements of the curriculum is meant to function as a remedial strategy, whereby learners are given another opportunity to develop the skills that form the necessary foundation for further learning. Unfortunately, the existing evidence suggests that retaining learners in a grade rarely achieves these aims in resource-constrained contexts (Batista Gomez-Neto and Hanushek, 1994). Many educationalists contend that repetition is only effective if there is specialised support for the child who repeats (Van der Berg, Wills, *et al.*, 2019), which is seldom available in resource-constrained contexts like South Africa.

South Africa's grade repetition policy stipulates that students may not be held back more than once in a three-year school phase (Department of Basic Education, 2013). For example, if a student is held back in Grade 7 (the beginning of the Senior Phase), they must be automatically promoted to Grade 10 (the beginning of the Further Education and Training (FET) phase) over the next three years, even if they do not meet the curriculum requirements. Together, widespread grade repetition without additional remedial support combined with automatic grade progression of students who have not met the basic curriculum requirements contributes to large within-class heterogeneity in learner abilities. I present evidence of such within-class heterogeneity at the primary school level in earlier work (Hofmeyr, 2019): In an attempt to identify positive outliers in PIRLS 2016 results, I find that there are students who achieved Grade 4 reading comprehension scores 1.5 standard deviations higher than the median reading score in their class in the majority (71%) of participating schools. Using Evans and Yuan's (2019) metric to calculate equivalent years of schooling, this means the majority of primary schools in South Africa contain students who are between three and four years ahead of the median student in their class.

Both Fleisch and Shindler (2009) and Van der Berg *et al.* (2019) present evidence of a strong gender dimension to these progression patterns and resultant within-class heterogeneity, with boys being more likely to repeat almost all grades. Based on data from the 2016 Community Survey, Van der Berg *et al.* (2019) find that by age 11, 24% of boys are already one grade behind the appropriate age-for-grade, compared to 17% of girls. They also present evidence of a widening gender gap in repetition rates across

[2] Although TIMSS typically tests Grade 8 students, countries can choose to participate at the Grade 9 level if they suspect that the assessment will be too difficult for their Grade 8 learners. South Africa went with this option in 2003, 2011, 2015, and 2019.

grades, with 45% of males being older than the appropriate age-for-grade by age 15, compared with 30% of females.

3. Methodology

3.1. Hypotheses

The hypothesis that I aim to test in this study is whether these gender differences in grade progression patterns explain South Africa's pro-girl achievement gap in TIMSS. Broadly, I aim to test the hypothesis that the observed pro-girl gap in TIMSS mathematics achievement reflects a pro-girl advantage that is already present at the start of formal schooling, and in the absence of effective remedial support for boys, accumulates as learners progress through school. In this sense, I aim to test whether boys in a given Grade 9 class are selected to be weaker-performing than girls. My earlier work (Hofmeyr, forthcoming) showed that there is already evidence of this effect at the Grade 4 and 5 level, with the pro-girl advantage in grade progression explaining around 10% and 20% of the country's pro-girl achievement gaps in PIRLS Literacy and TIMSS Numeracy, respectively. I therefore expect the pro-girl advantage in grade repetition to play an even larger role in explaining South Africa's pro-girl gap in Grade 9 mathematics achievement.

3.2. The decomposition approach

Decomposition analysis has long been used by labour economists to identify the relative importance of various factors that contribute to gender, race or other gaps in labour market outcomes (Cobb-Clark and Moschion, 2017). More recently, a number of economists of education have used this approach to examine disparities in learning outcomes by, for example, urban-rural status (Burger, 2011), as well as gender (Kingdon, 2002; Cobb-Clark and Moschion (2017). In essence, this approach allows one to separate gender gaps into two components: (i) The component that can be explained due to differences in observable characteristics of males and females; and (ii) The unexplained component. My aim is to decompose the pro-girl gap in TIMSS mathematics achievement into these two components. Formally, assuming a linear model of achievement, boys' and girls' test scores can be expressed as

$$\bar{T}_G^j = \bar{X}_G \hat{\beta}_G^j + \bar{\varepsilon}_G, \qquad \bar{\varepsilon}_G = 0; G \in (M, F) \tag{1}$$

where $\bar{T}_G^j$ denotes the mean test score of students of gender G (male (M) or female (F)) in subject j, $\bar{X}_G$ denotes the mean endowments (observable characteristics) of students of that gender, $\hat{\beta}_G^j$ denotes the mean returns to those endowments in terms of mathematics scores (that is, how those endowments are translated into TIMSS test scores) for each gender, and $\bar{\varepsilon}_G$ denotes the error term, which we assume to be zero. The gender gap in test scores can therefore be expressed as

$$\bar{T}_M^j - \bar{T}_F^j = \bar{X}_M \hat{\beta}_M^j - \bar{X}_F \hat{\beta}_F^j \tag{2}$$

I adopt Cobb-Clark and Moschion's (2017) approach of introducing a gender neutral coefficient vector ($\hat{\beta}_P^j$) to determine the contribution of the gender differences in endowments such that

$$\bar{X}_M \hat{\beta}_M^j - \bar{X}_F \hat{\beta}_F^j = (\bar{X}_M - \bar{X}_F)\, \hat{\beta}_P^j + \{\bar{X}_M(\hat{\beta}_M^j - \hat{\beta}_P^j) + \bar{X}_F(\hat{\beta}_P^j - \hat{\beta}_F^j)\} \tag{3}$$

where $\hat{\beta}_P^j$ is the coefficient from a pooled ordinary least squares regression of test scores on the full set of covariates over both males and females, and $\hat{\beta}_M^j$ and $\hat{\beta}_F^j$ are coefficients from separate regressions for males and females, respectively (Jann, 2008). Thus the first term on the right-hand side of equation (3) is the explained component of the gender gap in test scores, that is, the difference in boys' and girls' test scores that arises because boys and girls have different endowments of the characteristics that matter for achievement (the 'endowment effect'). These characteristics are evaluated (i.e. weighted) using the vector of gender-neutral responses ($\hat{\beta}_P^j$) (Jann, 2008). The second term on the right-hand side of equation (3) is the unexplained component of the gender gap (the 'response effect'). This term can be interpreted as the part of the gender gap that arises because boys' and girls' endowments are not translated into test scores in a gender-neutral way (Cobb-Clark and Moschion, 2017). However, given that this interpretation requires the strong assumption that all factors that matter for achievement are included in the model (Jann, 2008), I focus instead on the results relating to the explained component of the pro-girl gap.

4. Data

4.1. Estimation sample

TIMSS is an international large-scale mathematics and science assessment conducted by the International Association of the Evaluation of Educational Achievement (IEA). TIMSS 2019 was administered in South Africa by the Human Science Resources Council (HSRC). Students were sampled according to a two-stage stratified cluster sampling design, where schools were sampled at the first stage according to province and the quintile ranking (a measure of socio-economic status (SES)) of the school (Reddy *et al.*, 2020). Within the sampled schools, classes were randomly selected for participation. Sampled classes thus constitute the second-stage sampling units. All Grade 9 students in sampled classes who were present on the day of the assessment wrote the test. Two of the higher-achieving provinces, Gauteng and the Western Cape, were oversampled since they sought more precise provincial achievement estimates. As a result, 150 schools were sampled in each of these provinces, while the sample size in the remaining seven provinces remained at 30 schools each (Reddy *et al.*, 2020). Results presented in this paper are weighted according to this sampling design so that the oversampling of schools in Gauteng and the Western Cape is accounted for. In 2019 the realised TIMSS sample consisted of 20,829 students from 519 schools. Girls made up 51.9% of the realised sample.

3.2. Description of measures

Educational achievement measure

TIMSS is designed to align broadly with the mathematics and science curricula of participating countries (Reddy *et al.*, 2020). One of the major goals of TIMSS is to provide valid comparisons of achievement across student populations, based on broad coverage of the achievement domain (von Davier, 2020). This translates into several hundred achievement domains, only a fraction of which can be assessed in the available testing time. As a result, TIMSS makes use of an assessment design whereby achievement items are arranged in blocks that are then assembled into student booklets. Each booklet therefore contains different but systematically overlapping sets of item blocks – in other words, no student sees and answers all the achievement tasks. Item response theory is used to link the different student booklets together so that student proficiency can be reported as imputed estimates commonly known as plausible values. Although five plausible values are estimated for each student, I make use of only the first of these in this paper[3].

Grade repetition measure

Unfortunately, the TIMSS data does not include direct measures of whether a student has repeated a

[3] Unfortunately, it is not possible to conduct decomposition analysis using Stata's pv module, or using the IDB Analyser. The results using the first plausible value were checked against results obtained from using the other four plausible values, and are robust to the use of different plausible values.

grade. Because of this, I use students' age as a proxy for grade repetition. I code this as a dummy variable which takes a value of 1 if a student is one or more years older than they would be if they were "on-track" in terms of age-for-grade. I also create an ordinal variable of the number of years that a student is overage (1, 2, and 3 or more years) to investigate how the likely number of times that a student has repeated impacts on the main results.

School SES measure

Given evidence presented elsewhere of SES differences in the nature of South Africa's gender gap in educational outcomes (Zuze and Reddy, 2014), I investigate whether the pro-girl gap in TIMSS Grade 9 mathematics achievement differs by school SES. I make use of the DBE's quintile ranking system to classify schools according to SES, with Quintile 1 schools being the most socio-economically disadvantaged, and Quintile 5 schools being the wealthiest. Quintile 1 to 3 schools serve roughly 75% of South African learners, Quintile 4 and 5 schools serve about 20% of learners, and private schools serve the remaining 5% of learners in South Africa (Branson, Hofmeyr and Lam, 2015). Given the small size of the private school sector in South Africa, a decision was made to limit the analysis to public schools only.

Covariates

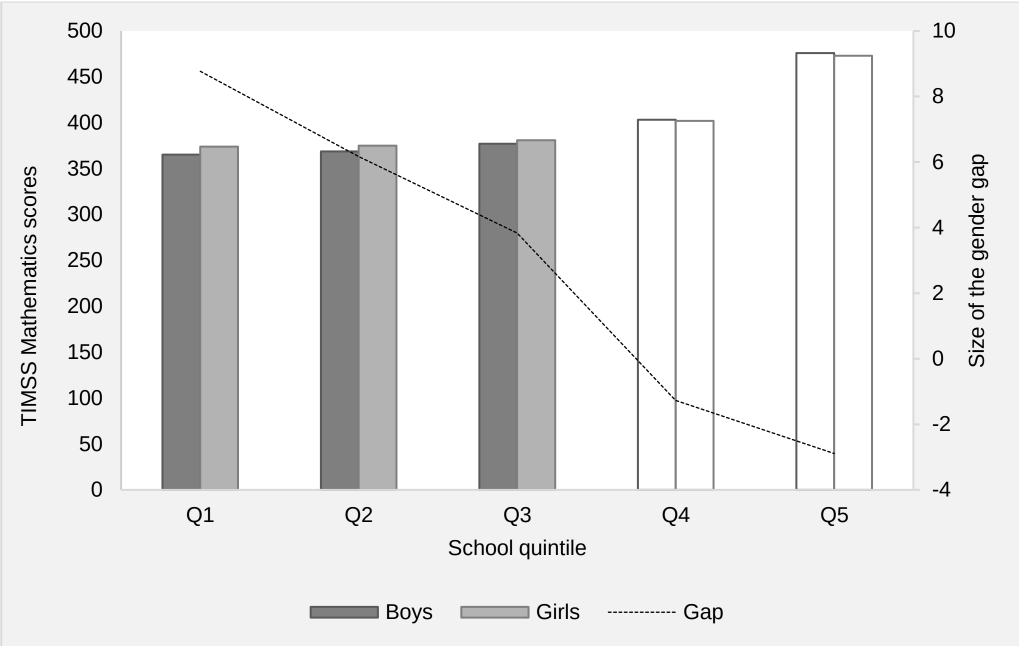
In addition to student age, I include a number of measures to capture individual characteristics and behaviours of students that have been shown to contribute to gender gaps in achievement elsewhere in the literature. These include student attitudes toward mathematics and school in general (based on items in the student background questionnaire), the frequency with which they experience bullying, whether they wrote the assessment in their home language, the frequency with which they did homework (self-reported), and student SES. Student SES is included to investigate whether there are gender differences in how home wealth is translated into academic achievement, as has been shown by for example Entwisle, Alexander and Olson (2007) and Bertrand and Pan (2011). Student SES is measured as scores on an asset index created using principal components analysis (PCA) on nine home possessions that students indicated having in their homes. The asset index has an alpha reliability coefficient of 0.67. The multivariate analysis also includes controls for the number of books students indicated having in their homes, as an additional measure of home resources. A number of school-level covariates are also controlled for, given existing evidence of gender differences in how school characteristics are translated into achievement (see for example Autor *et al.* (2016)), namely teacher sex, whether their school has a library, and whether their school has computers.

5. Descriptive statistics

4.1. Gender differences in mathematics achievement by school and province

Figure 2 shows boys' and girls' scores in the TIMSS mathematics assessment, separately by school quintile. The dotted line represents the size of the gender gap in mathematics scores. The unshaded bars for Quintile 4 and 5 schools indicates that the gender gaps in mathematics achievement in these schools are not statistically significant. The fact that the size of the gender gap decreases by school quintile is immediately apparent from the figure – so much so that the pro-girl gap is negative in Quintile 4 and 5 schools (that is, there is a pro-boy gap), however these gaps are not statistically significant. The figure therefore shows that the national pro-girl gap in TIMSS mathematics achievement is driven by statistically significant pro-girl gaps in Quintiles 1-3.

Figure 2: Gender differences in mathematics scores by school quintile



4.2. Descriptive differences between boys and girls

Descriptive differences between boys and girls are reported in Table 1, both nationally and by school quintile. The results in the table indicate that boys are roughly twice as likely as girls to be overage (that is, to have repeated at least once): While 20% of girls are overage nationally, this proportion is 39% for boys. 21% of boys are one year overage, and the remaining 18% are two or more years overage. This implies that almost one in five Grade 9 boys have repeated two or more grades. The proportion of girls who have repeated twice is less than half of this (7%). If being overage is indeed associated with weaker performance, these results already provide some support for the hypothesis that Grade 9 boys are selected to be weaker-performing than girls, which could contribute to the pro-girl gap in South Africa's TIMSS mathematics results. The association between being overage and achievement on the TIMSS mathematics assessment is explored further below.

The results in Table 1 further point to statistically significant differences between boys and girls both in terms of their attitudes towards learning mathematics, and their sense of safety and belonging at school. Interestingly, boys report having more positive attitudes toward learning mathematics in almost all school quintiles (barring Quintile 5), despite their inferior performance in the TIMSS mathematics assessment. Boys also report experiencing bullying less often than girls. Conversely, boys report having less of a sense of belonging at school on average. In terms of differences across school quintiles, it is interesting to note that students in Quintile 5 schools report much worse attitudes toward learning mathematics, less of a sense of belonging at school, and much more frequent bullying than their counterparts in Quintile 1-4 schools, on average.

Table 1: Descriptive differences between boys and girls

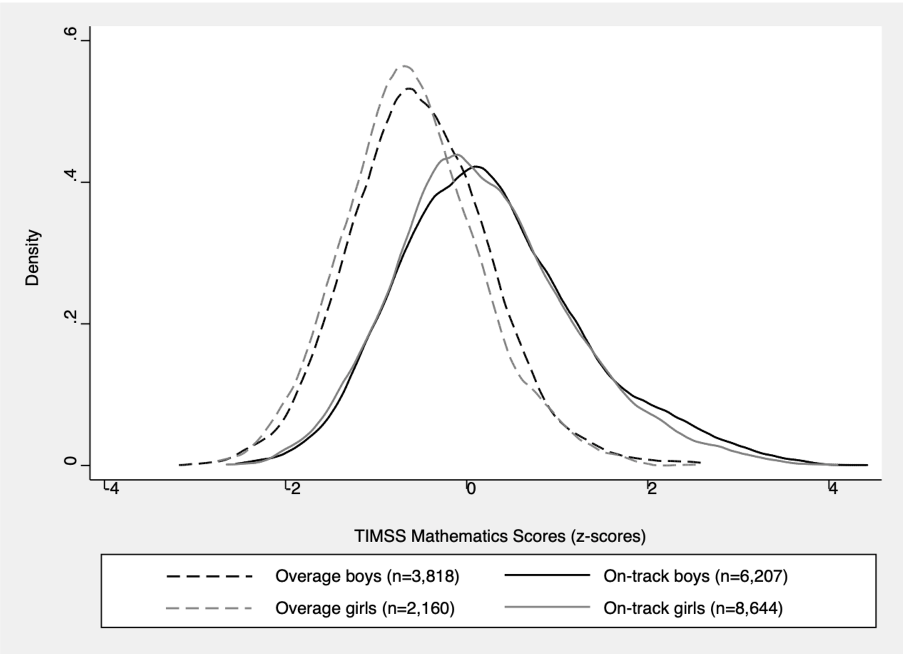
	National		Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Student age												
On-track	61%	80%***	54%	76%***	57%	77%***	56%	76%***	64%	81%***	77%	91%***
1 year overage	21%	13%***	24%	14%***	22%	15%***	23%	16%***	22%	12%***	16%	7%***
2 years overage	12%	5%***	13%	7%***	13%	5%***	15%	6%***	10%	5%***	6%	1%***
3 or more years overage	6%	2%***	9%	3%***	8%	3%***	5%	2%***	4%	1%***	1%	0%
Student attitudes toward learning mathematics												
Confidence in mathematics	0.06	-0.06***	0.09	-0.02***	0.09	-0.02***	0.04	-0.01*	0.02	-0.19***	0.07	-0.07***
Like learning mathematics	0.04	-0.03***	0.14	0.09	0.19	0.15	0.14	0.12	0.02	-0.14***	-0.35	0.40
Instructional clarity in mathematics lessons	0.03	-0.03***	0.14	0.09	0.18	0.15	0.14	0.12	0.02	-0.14***	-0.35	0.40
School belonging and safety												
Sense of school belonging	-0.05	0.05***	-0.01	0.15***	0.08	0.18***	0.08	0.18***	-0.11	-0.02***	-0.34	-0.18***
Frequency of experiencing bullying	-0.02	0.02***	-0.27	0.21**	-0.18	-0.14	-0.11	-0.03***	0.08	0.06	0.37	0.39
Observations	10,025	10,804	1,582	1,821	1,784	1,937	2,513	2,753	1,821	2,112	1,664	1,902

Notes: Student attitude and school safety measures are standardised versions of the scales provided in the TIMSS data. Statistically significant gender differences are indicated by asterisks such that *** $p < 0.01$; ** $p < 0.05$; and * $p < 0.10$.

4.3. Descriptive differences between overage and on-track students

Given the gender differences in the proportions of students that are overage presented above, it is instructive to consider whether there are differences in the TIMSS mathematics scores between students who are on-track in terms of age-for-grade versus those who are overage. Figure 3 shows the distributions of TIMSS mathematics scores by gender and overage. Mathematics scores are expressed as z-scores, where a score of zero is equal to the mean mathematics score of the total South African TIMSS sample of public schools. Overage students are represented by the dotted lines and on-track students are represented by the solid lines. Boys' mathematics scores are plotted in black and girls' in grey. The fact that the distributions of mathematics scores for overage students (both boys and girls) clearly lie to the left of the distributions of mathematics scores for on-track students indicates that being overage is indeed associated with weaker academic performance. In fact, there appears to be very little difference between the distributions of mathematics scores of boys and girls when considering overage and on-track students separately – the distributions of mathematics scores for boys and girls (the solid lines in the figure) are almost identical. Given that a much larger number of boys are overage (3,818 compared with 2,160 girls), the lower mathematics scores of overage boys could be driving the overall pro-girl gap in TIMSS mathematics achievement.

Figure 3: Mathematics scores by gender and overage



Notes: Mathematics scores are expressed as z-scores. Kernel density estimation used to plot the distributions of mathematics scores (bandwidth: 0.7). Source: TIMSS 2019.

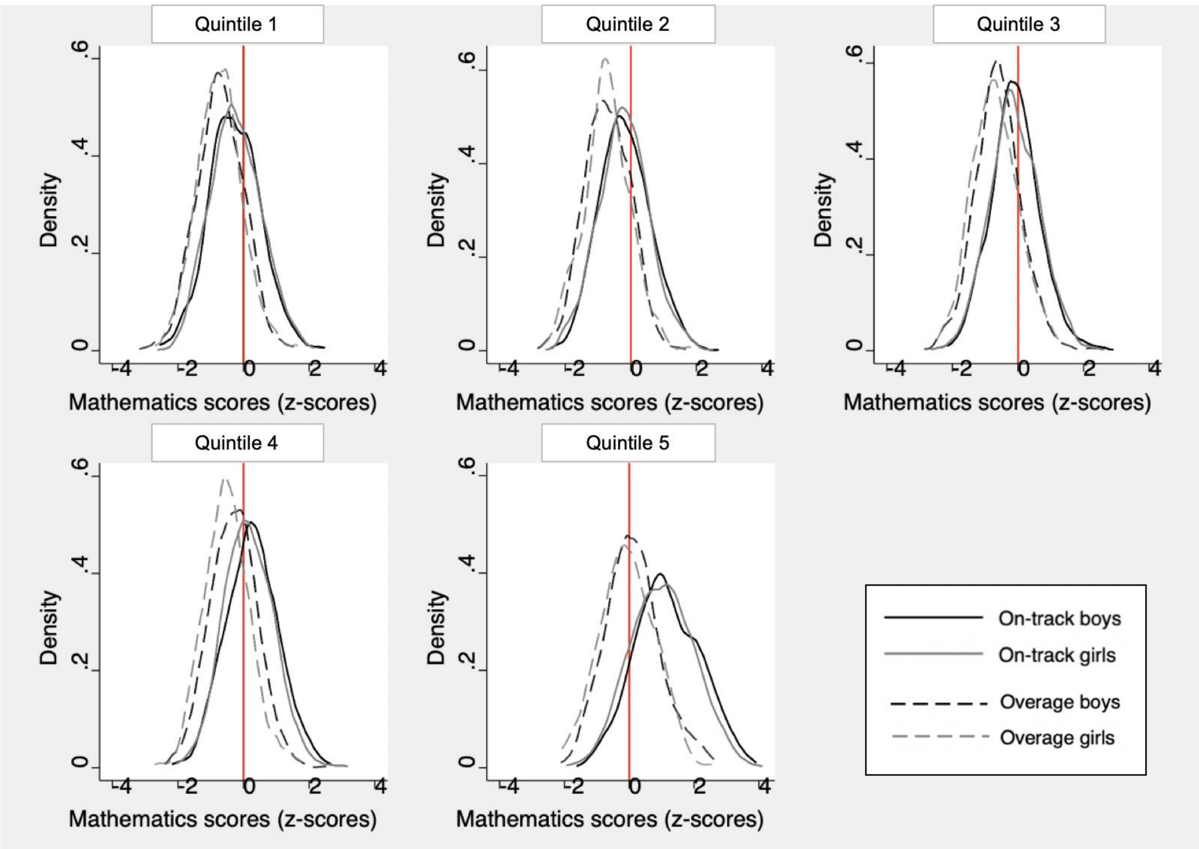
Figure 4 shows the same information as above, this time by school quintile. Mathematics scores are again expressed as z-scores, and the sample mean is indicated by the vertical red lines in each graph. The figure shows that overage students, regardless of gender, had lower mathematics scores than on-track students in all school quintiles.

The figure also shows that in Quintiles 1-3, the distributions of mathematics scores of on-track boys and girls are nearly identical, as in Figure 3 above. That is, when splitting the sample into overage and on-track students, it appears that the pro-girl advantage in mathematics scores disappears. Standard t-tests reveal that the mathematics scores of on-track boys and girls in Quintile 1-3 schools are not statistically significantly different from each other, indicating that indeed the pro-girl gap in mathematics scores disappears when considering only students who are on-track in terms of age-for-grade.

In Quintiles 4 and 5, the distributions of boys' mathematics scores lie slightly to the right of girls' (both for overage and on-track students). This is suggestive of a pro-*boy* gap in TIMSS mathematics achievement in Quintile 4 and 5 schools, once the effect of being overage has been accounted for. T-tests confirm that on-track boys in Quintile 4 and 5 schools did in fact outperform on-track girls by statistically significant margins.

The results from both Figure 3 and Figure 4 also indicate that South Africa's overall pro-girl gap in TIMSS mathematics performance is driven entirely by the disproportionate number of boys in Grade 9 that are overage for their grade, that is, those who have repeated one or more grades. Considering only on-track students brings up the average mathematics scores of boys across all school quintiles so that in Quintiles 1-3 the pro-girl gap that exists when considering all Grade 9 students disappears, and in Quintiles 4 and 5, where there is no statistically significant pro-girl gap when considering all Grade 9 students, a pro-boy advantage appears when considering only on-track students. This evidence provides strong support in favour of the hypothesis that the country's pro-girl gap in TIMSS mathematics is due to a pro-girl advantage in prior grade progression. This result is explored further in the decomposition analysis that follows.

Figure 4: Mathematics scores by gender and overage, separately by school quintile



Notes: The red reference lines indicate the mean mathematics score for the South African TIMSS sample (expressed as z-scores). Kernel density estimation used to plot the distributions of z-scores (bandwidth: 0.7). Source: TIMSS 2019.

6. Results from the decomposition analysis

Given the evidence of both differences in the nature of the gender gap in mathematics scores by school quintile presented in the previous section, I conduct decomposition analyses separately by school quintile. The results from the decompositions are presented in

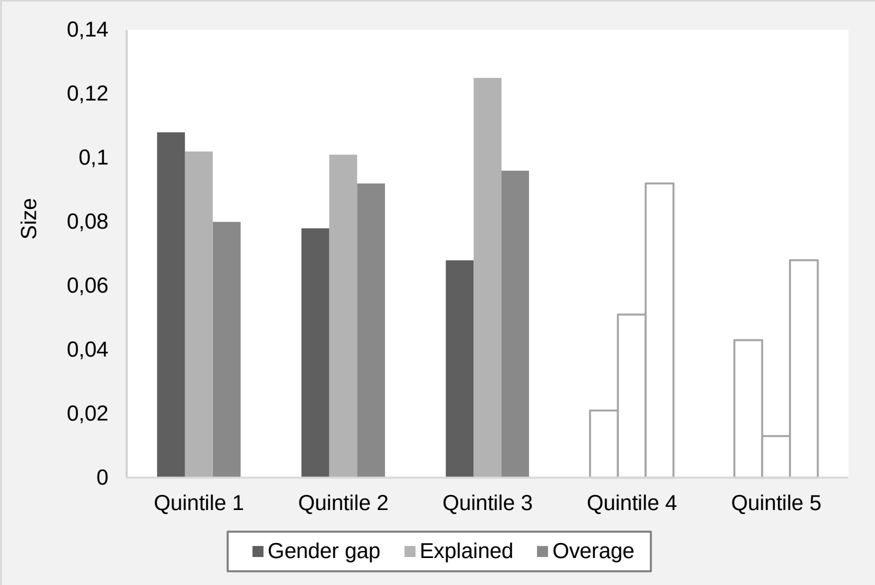
Table 2. The table shows boys' and girls' average standardised TIMSS scores and the magnitude of the gender gap in test scores, by school quintile. Negative values indicate a pro-girl advantage. The results in the table confirm those from the descriptive statistics in terms of differences in the pro-girl advantage across school quintiles, with only Quintile 1-3 schools exhibiting statistically significant pro-girl gaps in mathematics scores.

The share of the gender gap in test scores attributable to differences in boys' and girls' characteristics (the 'endowment effect') is reported in the "Explained" row. Differences between boys and girls in how endowments are translated into achievement (the 'response effect') are reported in the "Unexplained" row. Dividing the explained component by the total gender gap provides an estimate of the proportion of the observed gender gap that can be explained due to differences in boys' and girls' observable characteristics. Put differently, this proportion indicates by how much boys' mathematics scores would increase if they had the same characteristics as girls. For example, the explained component for students in Quintile 1 schools is 0.102. Dividing this by the size of the total gender gap (1.08 standard deviations) gives a value of 94%, indicating that 94% of the total gender gap in mathematics scores in these schools is due to observable differences between boys and girls. This means that if boys had the same characteristics as girls, their mathematics scores would increase by 94% (0.102 points).

The coefficients on the different covariates in the second panel of

Table 2 indicate the proportion of the total gender gap that can be explained due to gender differences in boys' and girls' endowments of that characteristic. For example, the coefficient on 0.08 on 'average' among students in Quintile 1 schools indicates that 74% of the pro-girl gap in mathematics achievement (0.08 divided by the overall gender gap of 0.108) is due to boys being more likely to be overage for their grade. Put differently, if the same proportions of boys were overage as girls in these schools, boys' mathematics scores would increase by 74%. The fact that this proportion is greater than one in Quintile 2 and 3 schools (with ratios of 1.12 and 1.41, respectively) indicates that if the same proportions of boys were overage as girls in these schools, boys would perform *better* than girls in mathematics. These results are presented graphically in Figure 5, which plots the overall pro-girl gap, the size of the explained component, and the coefficients on 'overage', for each school quintile. The figure makes clear that for both Quintile 2 and 3 schools, the coefficient on 'overage' is larger than the gender gap in TIMSS mathematics scores itself, indicating that if boys were equally likely as girls to be overage, their TIMSS scores would increase such that they achieved better results than girls.

Figure 5: Magnitudes of the gender gaps, explained components, and coefficients on 'overage', by school quintile



Notes: Unshaded bars plotted for Quintile 4 and 5 schools since there are no statistically significant gender gaps in mathematics performance in these two quintiles. Source: TIMSS 2019.

The only other covariate that consistently contributes significantly to the pro-girl achievement gap across school quintiles is the frequency with which students indicated doing mathematics homework. The fact that girls reported doing mathematics homework more often than boys accounts for between 16% and 24% of the overall gender gap in Quintile 1-3 schools. Students' self-reported confidence in mathematics contributes significantly to the observed gender gap in Quintile 1 and 2 schools. The fact that the coefficients on this variable are positive indicates that boys' higher confidence in mathematics adds to the pro-girl gap (by around 15%). In other words, if boys had the same levels of confidence in mathematics as girls, their TIMSS scores would *decrease* by about 15%.

Table 2: Results from the decomposition analysis by school quintile

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Gender gap	-0.108***	-0.078***	-0.068***	0.021	0.043
	(0.031)	(0.030)	(0.025)	(0.032)	(0.056)
Boys' average	-0.464***	-0.418***	-0.310***	0.027	0.965***
	(0.040)	(0.040)	(0.033)	(0.044)	(0.076)
Girls' average	-0.356***	-0.339***	-0.242***	0.006	0.921***
	(0.040)	(0.040)	(0.036)	(0.044)	(0.074)
Explained	-0.102***	-0.101***	-0.125***	-0.051***	-0.013
	(0.017)	(0.016)	(0.015)	(0.019)	(0.046)
Unexplained	-0.007	0.023	0.058***	0.072**	0.056
	(0.027)	(0.026)	(0.020)	(0.028)	(0.032)
Overage	-0.080***	-0.092***	-0.096***	-0.092***	-0.068***
	(0.009)	(0.009)	(0.010)	(0.011)	(0.011)
Homework	-0.017***	-0.015***	-0.016***	-0.011***	-0.008
	(0.004)	(0.004)	(0.003)	(0.003)	(0.005)
First language	-0.009**	-0.004	-0.001	-0.003	-0.002
	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)
Confidence in maths (z-scores)	0.016***	0.017***	0.009	0.043***	0.049***
	(0.006)	(0.005)	(0.005)	(0.009)	(0.016)

Like maths (z-scores)	-0.068	-0.048	-0.028	-0.291***	-0.130
	(0.075)	(0.048)	(0.047)	(0.086)	(0.127)
Instructional clarity (z-scores)	0.069	0.047	0.027	0.296***	0.126
	(0.078)	(0.051)	(0.048)	(0.087)	(0.124)
Sense of belonging (z-scores)	-0.000	0.002	0.001	0.007**	0.004
	(0.003)	(0.002)	(0.001)	(0.003)	(0.003)
Bullying (z-scores)	-0.008	-0.002	-0.006**	0.001	-0.001
	(0.004)	(0.002)	(0.002)	(0.003)	(0.001)
Asset index (z-scores)	0.003	0.001	-0.003	0.003	0.010
	(0.003)	(0.002)	(0.002)	(0.002)	(0.006)
No. of books at home	-0.000	-0.000	-0.001	-0.000	-0.003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.008)
Female maths teacher	0.000	-0.003	0.000	-0.001	-0.002
	(0.001)	(0.003)	(0.000)	(0.002)	(0.002)
School has a library	-0.006	-0.002	-0.006	0.000	-0.000
	(0.004)	(0.003)	(0.004)	(0.001)	(0.001)
School has computers	-0.002	0.001	0.001	-0.001	0.006
	(0.004)	(0.002)	(0.002)	(0.002)	(0.007)
Constant	-0.125	-0.110	0.052	0.028	-0.029
	(0.091)	(0.079)	(0.088)	(0.123)	(0.128)
N	3,403	3,721	5,266	3,933	3,566

Notes: Only coefficients of the explained component of the pro-girl gap are reported, given that not all factors that matter for mathematics achievement are available in the TIMSS data. All models include provincial controls. Standard errors are calculated at the school level and reported in parentheses. Statistical significance is indicated by asterisks such that *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

7. Robustness checks

The results presented thus far suggest that when considering gender gaps within school quintile groupings in South Africa, there are no statistically significant pro-girl gaps in TIMSS Grade 9 mathematics achievement among students who are in the correct age-for-grade. In fact, there is a pro-*boy* gap in TIMSS Grade 9 mathematics scores among on-track students in Quintile 4 and 5 schools. These findings are noteworthy when considered against existing evidence of a pro-girl gap in TIMSS mathematics achievement among on-track students at the Grade 5 level (Hofmeyr, forthcoming). The fact that this pro-girl gap among on-track students within school quintiles disappears by Grade 9 suggests there may be some catch-up happening among boys who manage to progress to Grade 9 without repeating or dropping out. It is worth investigating further the extent to which there is evidence of such catch-up, since existing evidence suggests that early learning deficits (especially in mathematics) only become more pronounced over time (see for example Spaull and Kotze (2015)). Whether there is evidence of catch-up among boys is also relevant for policy purposes: If the pro-girl gap in Grade 5 mathematics does indeed disappear by Grade 9, we might attribute less research and policy importance to the observed pro-girl gap in the primary school grades.

Importantly, there may be selection effects at play which could account for these results. These need to be investigated to bolster the results presented thus far. There are two possible sources of selection bias that need to be investigated, namely gender bias in (1) repetition, and (2) dropout. This section examines the extent to which each of these potential selection effects may be contributing to the relative performance of boys and girls in the TIMSS sample, and attempts to correct for this bias in order to investigate further whether there is evidence of true catch-up among boys in TIMSS mathematics achievement between Grade 5 and Grade 9.

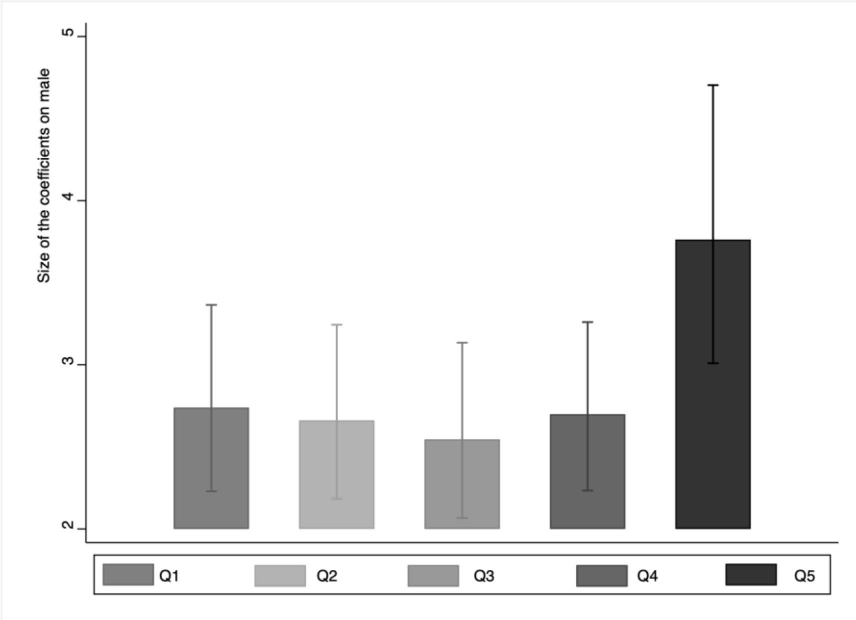
6.1. Evidence of selection bias

Gender bias in grade repetition

The presentation and discussion of results thus far has assumed that boys are more likely to have repeated a grade by the time they reach Grade 9 because they did not meet the curriculum requirements for being promoted to the next grade, and by implication that girls who are on-track in terms of age-for-grade did meet these curriculum requirements. There is existing evidence from the local literature on grade repetition that deserves to be considered here, however, namely that there is an element of randomness to grade repetition in South Africa, especially in low-quality schools (Lam, Ardington and Leibbrandt, 2011), and that teachers take into account student characteristics other than mastery of the curriculum when deciding whether to hold learners back (Walton, 2018). For example, Walton (2018) presents evidence, based on interviews with teachers in Johannesburg, that many teachers use learner 'maturity' as a metric that informs their grade promotion decisions. While neither of these studies consider potential gender differences in teachers' decisions to hold learners back, their results point to the important conclusion that boys' relative lack of mastery of the curriculum may not be the only factor that causes them to be more likely to repeat than girls. That is, it may be that boys are more likely to be held back, even if they exhibit the same mastery of the curriculum as girls.

Given this evidence, it is instructive to investigate whether boys are more likely than girls to have repeated a grade, all other things being equal. To do this, I run ordinary least squares logistic regressions where the outcome of interest is the dummy variable indicating whether students are one or more years overage for their grade in the TIMSS sample, separately by school quintile. I control for all the same factors that are included in the decomposition models above, including TIMSS mathematics scores. The coefficients on being male are all statistically significant at $p<0.01$ and are presented graphically in Figure 6. The figure shows that in all school quintiles, boys are at least twice as likely to have repeated a grade than girls, all other things being equal (including mathematics achievement in Grade 9). This is a noteworthy result, since it suggests that boys who have the same characteristics as girls in Grade 9 are more likely to have repeated by the time they reach this grade. Of course, there may be a temporal element to these results, since being overage is an indication of having repeated at some point in the past, while the characteristics controlled for here are measured at a later point in time. It is therefore at least conceivable that students' achievement levels may have changed since repeating a grade. It is nonetheless interesting to note the large coefficients on being male across all school quintiles in regressions of the probability of being overage that control for a host of factors at the individual, home, and school level, including TIMSS Grade 9 mathematics achievement. These results are at the very least suggestive of a pro-girl bias in grade progression, whereby girls are more likely to be promoted to the next grade, all else being equal.

Figure 6: Size of the coefficients on being male from logistic regressions of the probability of being overage



These results also suggest that there may be weaker-performing girls in the TIMSS Grade 9 sample who would have been held back if they had been boys. Given the results presented above that boys are at least twice as likely to have repeated as girls (even when controlling for TIMSS mathematics scores), it may be that weaker girls are more likely to have progressed to Grade 9 without interruption than boys who achieved similar results. It is important to investigate this possibility since it would mean that the disappearance of the pro-girl gap among students who are on-track in terms of age-for-grade in TIMSS mathematics achievement between Grade 5 and Grade 9 is due to weaker boys being held back while equally weak girls were not. This, in turn, would mean that the disappearance of the pro-girl gap is not due to boys catching up to girls, but rather that weaker boys were less likely to make it into the Grade 9 TIMSS sample than equally weak girls. The holding back of weaker boys when equally weak girls were promoted to the next grade could therefore contribute to the disappearance of the pro-girl gap in TIMSS mathematics achievement between Grade 5 and Grade 9.

Gender bias in dropout

Given evidence presented elsewhere that South African boys are more likely to drop out of school during secondary school (other things being equal) (Branson, Hofmeyr and Lam, 2014; De Wet and Mkwananzi, 2014), the weakest boys may have already dropped out of school before reaching Grade 9, while the weakest girls may still be in school and captured in the Grade 9 TIMSS sample. This would result in a selection effect that would pull down the average TIMSS maths scores of girls. A comparison of the proportions of boys and girls in the TIMSS Grade 5 versus Grade 9 data provides evidence in support of this hypothesis: Even though there are virtually equal numbers of males and females at birth, the TIMSS 2019 Grade 9 sample consisted of 9,404 males and 10,586 females. That is, there were 12.6% more females than males. If we assume (as a number of other authors do, such as for example Spaull and Makaluza (2019) and Van der Berg *et al.* (2019)) that the weakest boys were the ones who dropped out, this would mean that the realised TIMSS Grade 9 sample contains low-achieving girls who would have dropped out, had they been boys. As is the case with gender bias in grade repetition, this selection effect would pull down the average TIMSS achievement of girls.

6.2. Correcting for selection bias

To account for this selection bias, I adopt Spaull and Makaluza's (2019) strategy of creating a comparable sample of equal numbers of boys and girls writing the TIMSS assessment by dropping the weakest 10% of females from the TIMSS sample. This results in a balanced sample where girls make up exactly 50% of observations.

I next run the same set of decompositions (separately by school quintile) presented in Section 5 on the restricted TIMSS sample to investigate whether there is still evidence that the pro-girl gap in Mathematics achievement disappears by Grade 9. The results are presented in Table 3 below. The models include the same controls as those in Section 5, but are not reported here[4]. Notably, restricting the sample to account for the potential selection effects discussed above results in pro-girl gaps in mathematics achievement in Quintile 1-3 schools that are comparable in magnitude with those observable in the TIMSS Grade 5 data for South Africa, with girls performing between around 10% and 20% of a standard deviation better than boys, depending on the school quintile ranking (Hofmeyr, forthcoming). These results therefore suggest that the disappearance of the pro-girl gap in TIMSS mathematics achievement between Grade 5 and Grade 9 among students who are on-track in terms of age-for-grade is not due to boys catching up to girls. Rather, the results presented here suggest this is due to weaker boys being more likely to be held back and to drop out of school entirely, while equally low-performing girls progress through the schooling system without interruption.

Table 3: Decomposition results, restricted TIMSS sample, on-track students only

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Gender gap	-0.198***	-0.128***	-0.104***	0.044	0.139**
	(0.035)	(0.037)	(0.030)	(0.039)	(0.056)
Boys' average	-0.264***	-0.178***	-0.085**	0.259***	1.186***
	(0.046)	(0.047)	(0.039)	(0.050)	(0.073)
Girls' average	-0.067	-0.050	0.019	0.215***	1.047***
	(0.034)	(0.031)	(0.033)	(0.042)	(0.069)
Explained	-0.022	-0.002	-0.021	0.047***	0.100**
	(0.017)	(0.017)	(0.013)	(0.017)	(0.044)
Unexplained	-0.176***	-0.126***	-0.083***	-0.003	0.039
	(0.033)	(0.031)	(0.026)	(0.037)	(0.035)
Constant	-0.400***	-0.315***	-0.286**	0.018	-0.072
	(0.111)	(0.104)	(0.119)	(0.167)	(0.139)
N	2,091	2,289	3,359	2,643	3,025

Notes: Samples are restricted to exclude the weakest-performing 10% of girls. Models include all the same controls as those in Table 2, but are not reported here. Standard errors are calculated at the school level and reported in parentheses. Statistical significance is indicated by asterisks such that *** p < 0.01; ** p < 0.05; * p < 0.1.

.

8. Discussion

7.1. Summary of main results

The results presented in this paper contribute to our understanding of gender differences in mathematics achievement in South Africa. Importantly, the analysis shows that gender gaps in international educational assessments such as TIMSS should not be taken at face value. In particular, the results point to various selection effects that arise from gender differences in grade progression and dropout that should be taken

[4] The full set of results are presented in Table A2 of the Appendix.

into account when considering gender gaps in international assessments.

One major source of selection bias in South African TIMSS data is the higher proportion of boys that are overage for their grade, indicating that boys are much more likely to have repeated a grade than girls. The analysis in this paper shows that, as one might expect, this bias is much more pronounced at the Grade 9 level than the Grade 5 level. Interestingly, when correcting for this bias by considering only students who are on-track in terms of age-for-grade, the observed pro-girl gap in TIMSS mathematics achievement disappears completely, across all school quintiles. Since a pro-girl gap in mathematics achievement is observable at the Grade 5 level, even when considering only learners who are on-track in terms of age-for-grade (Hofmeyr, forthcoming), this result is suggestive of catch-up among boys between Grade 5 and Grade 9. This is a surprising result when considered against existing evidence that early learning deficits, especially in mathematics, are difficult to surmount without proper remedial efforts.

Analysis of the relative numbers of males and females in the TIMSS Grade 9 sample and the relative probabilities of having repeated a grade for boys and girls suggests, however, that boys are more likely to be held back and to drop out of school, all else being equal. This would result in weaker boys not making it into the TIMSS Grade 9 sample, while equally weak girls do. Put differently, the evidence presented in this paper suggests that the realised TIMSS Grade 9 sample consists of weaker-performing girls who would not have been in the sample, had they been boys. This selection effect pulls down the average mathematics marks of girls, resulting in the lack of a pro-girl gap in Grade 9 mathematics achievement, and suggesting that boys manage to catch up to girls in mathematics between Grade 5 and Grade 9. Accounting for this selection effect by removing the worst-performing 10% of girls from the sample shows, however, that significant pro-girl gaps are observable in the bottom three school quintiles. This, in turn, suggests that the initial evidence of catch-up among boys is the result of a pro-girl advantage in grade promotion, and that the pro-girl gap in TIMSS mathematics achievement observed in Grade 5 is still evident in Grade 9. In other words, the evidence presented here suggests that South Africa's pro-girl gap in mathematics achievement in primary school extends to high school, once selection effects are accounted for.

7.2. Limitations

While the results of this paper highlight the importance of gendered repetition and dropout patterns in contributing to South Africa's observed pro-girl achievement gap in Grade 9 mathematics achievement, this result only shifts the question of the sources of the pro-girl advantage to the earlier grades. That is, the results presented here do not bring us any closer to understanding *why* boys are underachieving relative to girls in the grades leading up to Grade 9. My earlier finding that the effect of gendered repetition patterns already contributes to a pro-girl advantage in Grade 4 and 5 suggest that girls' academic advantage is already present in the Foundation Phase (Grades R-3). This may indicate that girls already have an advantage at the start of formal schooling, however more research is required that explicitly investigates whether this is the case.

A more minor limitation of the results present here is that the TIMSS assessment is conducted at the grade level, and not by age (unlike, for example, the Programme for International Student Assessment (PISA), which tests all 15-year-olds regardless of what grade they are in). As a result, the TIMSS data does not allow for analysis of all members of the same cohort. Limiting the sample to students who are on-track in terms of age-for-grade allows one to get closer to the 'true' cohort that started school together in Grade 1. However, the sample of students remains a pseudo cohort, since it includes students from a previous cohort (students who started school early and repeated a grade), and excludes students from the true cohort (students who started school late and never repeated). This limitation points to the need for international educational assessments to include items in the student background questionnaire that would provide a more detailed picture of students' past educational trajectories, such as asking students whether they have ever repeated a grade.

6.3. Concluding remarks

While it is important to understand the reasons behind gendered learning outcomes, the results presented here (in combination with my earlier results regarding the reasons for the country's pro-girl gap at the primary school level) point to a more general feature of South Africa's schooling system, namely that the system seems to offer few opportunities for remediation of initial learning deficits. Weaker students fall behind in the early grades, and their disadvantage seems to accumulate over the course of their schooling careers. This result is supported by the findings of Spaull and Kotze (2015) and Van der Berg (2015), who show that South African students acquire learning deficits early on in their schooling careers and, in the absence of proper remediation, these early deficits are translated into underperformance in later years. The results presented here add to these findings by showing that there is a strong gender dimension to this underperformance at the Grade 9 level, which can be traced back to gender differences in grade progression patterns in the years leading up to Grade 9. Together, these results point to the importance of improving opportunities for remediation in the country's schooling system, particularly in the Foundation Phase.

9. Conclusion

This paper contributes to the literature documenting South Africa's pro-girl advantage in educational outcomes. The analysis focussed on the role of gender differences in prior grade progression in explaining the country's pro-girl gap in Grade 9 mathematics achievement in the 2019 round of TIMSS. The analysis showed that South Africa's Grade 9 TIMSS sample is unbalanced in terms of gender, which the results suggest is at least in part due to boys being more likely to be held back than girls, all else being equal. That is, weaker boys are held back and even drop out of the system entirely, while weaker girls move through the system without interruption. When accounting for these gender differences in grade repetition and dropout, a pro-girl gap in mathematics achievement is observable in the bottom three school quintiles, suggesting that girls' advantage in mathematics observable in the primary school grades extends to secondary school. Thus there is little evidence of true catch-up among boys in mathematics achievement between Grade 5 and Grade 9. This points to a more general feature of the South African education system, whereby early learning deficits are not remediated effectively, and thus extend to higher grades. We need to do more to understand why boys fall behind and why they are more likely to be held back, with research that focusses on gender gaps in the early grades likely constituting a particularly fruitful avenue for future research.

References

Autor, D. *et al.* (2016) 'School quality and the gender gap in educational achievement', *American Economic Review*, 106(5), pp. 289–295. doi: 10.1257/aer.p20161074.

Batista Gomez-Neto, J. and Hanushek, E. A. (1994) 'Causes and Consequences of Grade Repetition: Evidence from Brazil', *Economic Development and Cultural Change*, 43(1), pp. 117–148.

Bertrand, M. and Pan, J. (2011) *The Trouble With Boys: Social Influences and the Gender Gap in Disruptive Behaviour*. 17541. Cambridge, MA. doi: 10.16194/j.cnki.31-1059/g4.2011.07.016.

Branson, N., Hofmeyr, C. and Lam, D. (2014) 'Progress through school and the determinants of school dropout in South Africa Progress through school and the determinants of school dropout in South Africa', *Development Southern Africa*, 31(September), pp. 106–126. doi: 10.1080/0376835X.2013.853610.

Branson, N., Hofmeyr, C. and Lam, D. (2015) 'The impact of the no-fee school policy on enrolment and school performance: Evidence from NIDS Waves 1-3', (Psppd Ii), pp. 1–27.

Buhl-Wiggers, J., Jones, S. and Thornton, R. (2021) 'Boys lagging behind: Unpacking gender differences in academic achievement across East Africa', *International Journal of Educational Development*. Elsevier Ltd, 83(March), p. 102382. doi: 10.1016/j.ijedudev.2021.102382.

Burger, R. (2011) 'School effectiveness in Zambia: The origins of differences between rural and urban outcomes', *Development Southern Africa*, 28(2), pp. 157–176. doi: 10.1080/0376835X.2011.570064.

Cobb-Clark, D. A. and Moschion, J. (2017) 'Gender gaps in early educational achievement', *Journal of Population Economics*. Journal of Population Economics, 30(4), pp. 1093–1134. doi: 10.1007/s00148-017-0638-z.

Crouch, L. *et al.* (2020) 'Déjà vu all over again? Recent evidence on early childhood and early grade repetition in developing countries', *Prospects*. Springer Netherlands, (0123456789). doi: 10.1007/s11125-020-09473-2.

Department of Basic Education (2013) 'National Policy Pertaining To the Programme and Promotion Requirements of the National Curriculum Statement grades R - 12', *Government Gazette*, p. 74.

De Wet, N. and Mkwananzi, S. (2014) 'School Dropout among African Adolescents in South Africa: Is there any gender differential?', *Gender & Behaviour*, 12(4), pp. 1–23.

Entwisle, D. R., Alexander, K. L. and Olson, L. S. (2007) 'Early schooling: The handicap of being poor and male', *Sociology of Education*, 80(2), pp. 114–138. doi: 10.1177/003804070708000202.

Evans, D. K. and Yuan, F. (2019) *Equivalent Years of Schooling A Metric to Communicate Learning Gains in Concrete Terms, World Bank Policy Research Working Paper*. Available at: www.worldbank.org/research.

Fleisch, B. and Shindler, J. (2009) 'Gender repetition: School access, transitions and equity in the "Birth-to-Twenty" cohort panel study in urban South Africa', *Comparative Education*, 45(2), pp. 265–279. doi: 10.1080/03050060902920955.

Grant, M. J. and Behrman, J. R. (2010) 'Gender gaps in educational attainment in less developed countries', *Population and Development Review*, 36(1), pp. 71–89. doi: 10.1111/j.1728-4457.2010.00318.x.

Hofmeyr, H. (2019) 'Performance Beyond Expectations: Academic Resilience in South Africa', *Stellenbosch Economics Working Papers (19/2019)*, (December).

Hofmeyr, H. (forthcoming). 'Why do girls do better? Unpacking South Africa's gender gap in PIRLS and TIMSS'. *International Journal of Educational Development.*

Jann, B. (2008) 'The Blinder-Oaxaca decomposition for linear regression models', *Stata Journal*, 8(4), pp. 453–479. doi: 10.1177/1536867x0800800401.

Kingdon, G. G. (2002) 'The gender gap in educational attainment in India: How much can be explained?', *Journal of Development Studies*, 39(2), pp. 25–53. doi: 10.1080/00220380412331322741.

Lam, D., Ardington, C. and Leibbrandt, M. (2011) 'Schooling as a lottery: Racial differences in school advancement in urban South Africa', *Journal of Development Economics*. Elsevier B.V., 95(2), pp. 121–136. doi: 10.1016/j.jdeveco.2010.05.005.

Mullis, I. V. S. *et al.* (2016) *TIMSS 2015 International Results in Mathematics*, *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. Paris. doi: 10.4135/9781506326139.n704.

Mullis, I. V. S. *et al.* (2017) *PIRLS 2016 International Results in Reading*. Available at: http://timssandpirls.bc.edu/pirls2016/international-results/.

Mullis, I. V. S. *et al.* (2020) *TIMSS 2019 International Results in Mathematics and Science.*

OECD (2018) *PISA in Focus*. Paris.

Reddy, V. *et al.* (2020) *TIMSS 2019: Highlights of South African Grade 9 Results in Mathematics and Science*. Cape Town: HSRC.

Spaull, N. and Kotze, J. (2015) 'Starting behind and staying behind in South Africa: The case of insurmountable learning deficits in mafile:///Users/heleenhofmeyer/Downloads/10.1.1.487.9364.pdfthematics', *International Journal of Educational Development*. Elsevier Ltd, 41, pp. 13–24. doi: 10.1016/j.ijedudev.2015.01.002.

Spaull, N. and Makaluza, N. (2019a) 'Girls Do Better: The pro-female gender gap in learning outcomes in South Africa', *Agenda*. Taylor & Francis, 33(4), pp. 207–224. doi: 10.1080/10130950.2019.1672568.

Spaull, N. and Makaluza, N. (2019b) 'Girls Do Better: The pro-female gender gap in learning outcomes in South Africa 1995-2018', *Agenda*, 33(4), p. 207=224.

Van der Berg, S., Wills, G., *et al.* (2019) 'The cost of repetition in South Africa', *Stellenbosch Economics Working Papers*, pp. 117–121. doi: 10.1016/0360-5442(87)90115-0.

Van der Berg, S., Van Wyk, C., *et al.* (2019) 'The promise of SA-SAMS & DDD data for tracking progression, repetition, and drop-out', *Stellenbosch Economic Working Papers*, (December).

Van Der Berg, S. (2015) 'What the Annual National Assessments can tell us about career year learning deficits over the education system and the school', *Stellenbosch Economics Working Papers*

Von Davier, M. (2020) 'TIMSS 2019 scaling methodology: Item response theory, population models, and linking across modes', *Methods and procedures: TIMSS 2019 technical report*, pp. 11.1-11.25. Available at: https://timssandpirls.bc.edu/timss2019/methods/chapter-11.html.

Walton, E. (2018) 'Teacher beliefs about grade repetition: An exploratory South African study Elizabeth Walton', *Citizenship Teaching & Learning*, pp. 1–22.

Zuze, T. L. and Beku, U. (2019) 'Gender Inequalities in South African Schools: New Complexities', in Spaull, N. and Jansen, J. D. (eds) *South African Schooling: The Enigma of Inequality*. Cape Town: Springer, pp. 225–241.

Zuze, T. L. and Reddy, V. (2014) 'School resources and the gender reading literacy gap in South African schools', *International Journal of Educational Development*. Elsevier Ltd, 36, pp. 100–107. doi:

10.1016/j.ijedudev.2013.10.002.