Exploring Gender Gaps in Mathematics Achievement: The Case of Single-Sex Education in Saudi Arabia

Sadaf Basharat

Master's Programme

Assessment, Measurement and Evaluation

120 Credits

Centre for Educational Measurement (CEMO)
Faculty of Educational Sciences

13 May 2022

Popular Abstract

International surveys have revealed large international differences in gender gaps in mathematics achievement. While boys and girls perform almost equally in some countries, large gaps are observed in others. One reason that could explain why girls and boys score unequally in some countries could be that they are exposed to different learning conditions. This is especially the case if girls and boys do not attend the same schools. One example is Saudi Arabia where all boys go to boys' schools and all girls go to girls' schools. Using the case of Saudi Arabia, the present study has considered variations in *school climate*, the overall school experience that involves aspects of school emphasis on academic success, school safety, school community relationships, and the institutional and structural features of the school environment. In comparison to coeducational systems, if girls and boys attend separate schools, and if these schools have different *school climate* conditions, this might contribute to gender gaps in achievement outcomes. The study indeed found some first evidence of *school climate* differences between girls' and boys' schools at primary and secondary school level in Saudi Arabia. The study findings have reaffirmed the importance of school climate factors and implementation of a safe and positive school climate for girls' and boys' schools alike, where students can learn and thrive.

Abstract

Results from the Trends in International Mathematics and Science Study (TIMSS) 2019 indicate large international differences in gender gaps in mathematics achievement. Interestingly, countries with a high proportion of single-sex schools show unexpectedly large mean advantages for girls. One reason that could explain why girls and boys score unequally in these countries could be that they are exposed to different learning conditions. Saudi Arabia particularly stands out in this regard as it has consistently exhibited extreme gender gaps in mathematics achievement in favour of girls. It is also the country that has implemented 100% gender segregation in its education system, such that all girls and boys in the country go to separate single-sex schools at primary and secondary school level. By considering the case of Saudi Arabia, this study has sought to examine whether variations in *school climate* dimensions between girls' and boys' schools may help us understand why large unexpected girls' advantages in mathematics achievement exist. In school-level and two-level regression models using TIMSS 2019 data from fourth and eighth grade, this study indeed found some evidence of *school climate* differences between girls' and boys' schools at primary and secondary school level in Saudi Arabia. However, research results were not found to be robust across grades and after including control variables, especially in the models where all *school climate* dimensions were simultaneously included. The present study is the first attempt in understanding some of the endogenous forces operating at the school level that can explain variation in student achievement outcomes in single-sex education systems like those implemented in Saudi Arabia. The study limitations, implications and areas of future research are discussed.

Keywords: Mathematics gender gap, student achievement, school climate, single-sex schooling, gender-segregated schools, international large-scale assessment.

Acknowledgements

I am eternally grateful to my thesis supervisor, Dr. Isa Steinmann, for her unwavering support, constant guidance and wisdom throughout the thesis writing process. No words of appreciation can do justice to the amount of effort and patience she has shown me throughout the supervision process. Without her thoughtful comments, constant feedback and support with Mplus, this thesis would not have been possible.

My deepest gratitude to my co-supervisor, Dr. Sigrid Blömeke, for her insightful comments and feedback on my thesis. I have greatly benefited from her immense knowledge and wisdom.

I would also like to express my sincere gratitude towards Dr. Ronny Scherer, who inspired me with his knowledge, wit and teaching style throughout my two-year stay at CEMO. His course on multi-level modelling was particularly helpful for me for running my analyses.

Lastly and most importantly, my heartfelt gratitude to my husband, my rock and my pillar of strength, Ummar, for his patience and moral support. And to my parents and siblings, who always believed in me and pushed me towards excellence.

Exploring Gender Gaps in Mathematics Achievement: The Case of Single-Sex Education in Saudi Arabia

Gender gaps in mathematics achievement of students have become an important policy concern for countries around the world. The disparity of educational outcomes in mathematics between girls and boys raises concerns of equity in national education systems designed around the concept of *fairness* and *inclusion*. The fairness dimension of equity in education entails that students' personal and social circumstances such as gender, ethnic origin and socio-economic status should not hinder educational success, while the inclusion dimension implies ensuring a minimum standard of education for all (OECD, 2007). The Education 2030 Framework for Action, designed to advance progress on Sustainable Development Goal (SDG) 4 on reducing *educational inequalities*, has called for reducing inequities "related to access, participation, and learning processes and outcomes" (UNESCO, 2017, p. 12) within national education systems. Through adopting the slogan of "every learner matters and matters equally" (UNESCO, 2017, p. 12), UNESCO (United Nations Educational, Scientific and Cultural Organization) has also emphasized the importance of adopting inclusive and equitable educational practices to foster student learning and outcomes.

Given the salience of achieving equitable educational outcomes for girls and boys, international comparative research has increasingly paid attention to understanding the extent of gender gaps in mathematics achievement as well as analysing the factors and antecedents that may explain variation in girls' and boys' performance. International Large-Scale Assessments (ILSAs) such as Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA) have particularly facilitated the cross-national comparability of girls' and boys' performance across learning domains such as mathematics, science and reading. For instance, recent data on mathematics achievement from the TIMSS 2019 cycle showed that in fourth grade, boys had higher average achievement than girls in nearly half of the 58 participating countries, while girls had higher average achievement than boys in four countries (Mullis et al., 2020). In eighth grade, girls outperformed boys in seven countries, boys outperformed girls in six countries, while boys and girls performed almost equally on average in 26 countries (Mullis et al., 2020). For 15-year old students living in OECD countries and some non-OECD countries, PISA 2018 results showed that on average boys significantly outperformed girls in mathematics in 32 of the 79 participating countries, while girls outperformed boys in 14 countries and economies (OECD, 2019). Results from both TIMSS 2019 and PISA 2018 indicate that the pervasiveness and magnitude of gender gaps in mathematics achievement among students varies considerably across countries.

Explaining Gender Gaps in Mathematics Achievement

A number of theories have been put forward to explain gender differences in mathematics achievement (see Hyde (2014) for an extensive overview of these theories). One strand of theories focuses on biological causes to explain the difference in boys' and girls' mathematics performance in schools. These theories argue that in particular genetic differences in spatial ability, higher order thinking, and brain development between boys and girls produce a gap in mathematics achievement (Dickerson et al., 2015; Fryer & Levitt, 2010; Penner, 2008). However, if gender differences in mathematics merely result from genetic or biological factors, international variation in the size and direction of gender gaps should be limited (Meinck & Brese, 2019; Penner, 2008; Reilly et al., 2019) and countries should display similar patterns of gender gaps. In fact, cross-national studies of gender gaps in mathematics achievement show that there is "no clear advantage to either gender when viewed globally, but important differences are present at the national level" (Reilly et al., 2019, p. 8), thus biological explanations cannot be deemed suitable to explain between-country variation in mathematics performance alone.

A second explanation that could potentially explain international variation in gender gaps, relies on social or structural theories. These explanations tend to attribute gender gaps in mathematics as an outcome of social stratification and gender socialization processes operating through stereotype threat and expectancy-value mechanisms (Eccles et al., 1990; Hyde, 2014) in which families, schools, labour markets and national governments affect the varying opportunity structure for girls and boys (Dickerson et al., 2015; Tsai et al., 2018). Baker and Jones's (1993) gender stratification hypothesis is noteworthy in this regard. It argues that as gender stratification of opportunity decreases in countries such that females gain more access to higher education and labour market opportunities, gender differences in mathematics performance also decrease.

The empirical tests of the gender stratification hypothesis offer mixed results. Using data from 40 countries participating in the PISA 2003 study, the widely cited paper by Guiso et al. (2008) has lent support to the gender stratification hypothesis by finding a positive correlation between indicators of gender equality (World Economic Forum’s Gender Gap Index) and gender gaps in mathematics. They found that in more gender-equal countries, such as Norway and Sweden, the mathematics gender gap disappears. Else-Quest et al. (2010) also tested the gender stratification hypothesis by testing composite indices of gender equity (e.g., Gender Empowerment Measure, Gender Equality Index, Standardized Index of Gender Equality, Gender Gap Index) as moderators of gender differences in mathematics achievement, attitudes, and affect. They found that gender equity did not significantly predict gender differences in TIMSS 2003 math achievement, however, each of the four gender equity composite indices significantly and negatively predicted gender differences in PISA 2003 mathematics achievement.

In contrast to the aforementioned findings, Stoet and Geary (2015) by using data from PISA 2000-2009 studies, concluded that there was no reliable relation between gender equality measures and gender differences in mathematics achievement. They argued that the positive findings for the PISA 2003 data were primarily driven because of the inclusion of a small number of Nordic countries with high levels of equality. Stoet and Geary's (2015) findings were confirmed by Reilly et al. (2019) who reported similar results for TIMSS 2011 mathematics achievement data.

An interesting and rather contradictory finding with regards to the gender stratification hypothesis appears in the case of Middle-eastern countries out of which one is the focus of this study. Despite having the lowest proportion of female representation in politics and labour markets as measured by the Global Gender Gap Index (World Economic Forum, 2021), these countries either show a significant girl advantage or no significant gender differences in mathematics achievement (Mullis et al., 2020). Given this paradoxical pattern, Ayalon and Livneh (2013) have concluded

> the association between women’s political and economic participation and the gender gap in mathematics is not universal. The link between these indicators of gender stratification and educational outcomes, which is expected in Western culture, does not necessarily hold for other cultures.
>
> (Ayalon & Livneh, 2013, pp. 440–441)

Given the paradoxical pattern of gender stratification hypothesis in the Middle-Eastern context, Fryer and Levitt (2010) have tentatively attributed the presence of single-sex schooling as a possible explanation for a significant girl advantage in mathematics. In education systems that exhibit high levels of gender segregation in terms of single-sex schools, girls may have less pressure to comply with gender role expectations and may be encouraged to take gender-atypical courses like mathematics and science (Spielhofer et al., 2004). Additionally, same-gender teachers in single-sex schools may prove to be effective role models for girls, thereby encouraging them to take interest in mathematics and science that are otherwise considered male subjects (Riordan, 2015; Thompson, 2003).

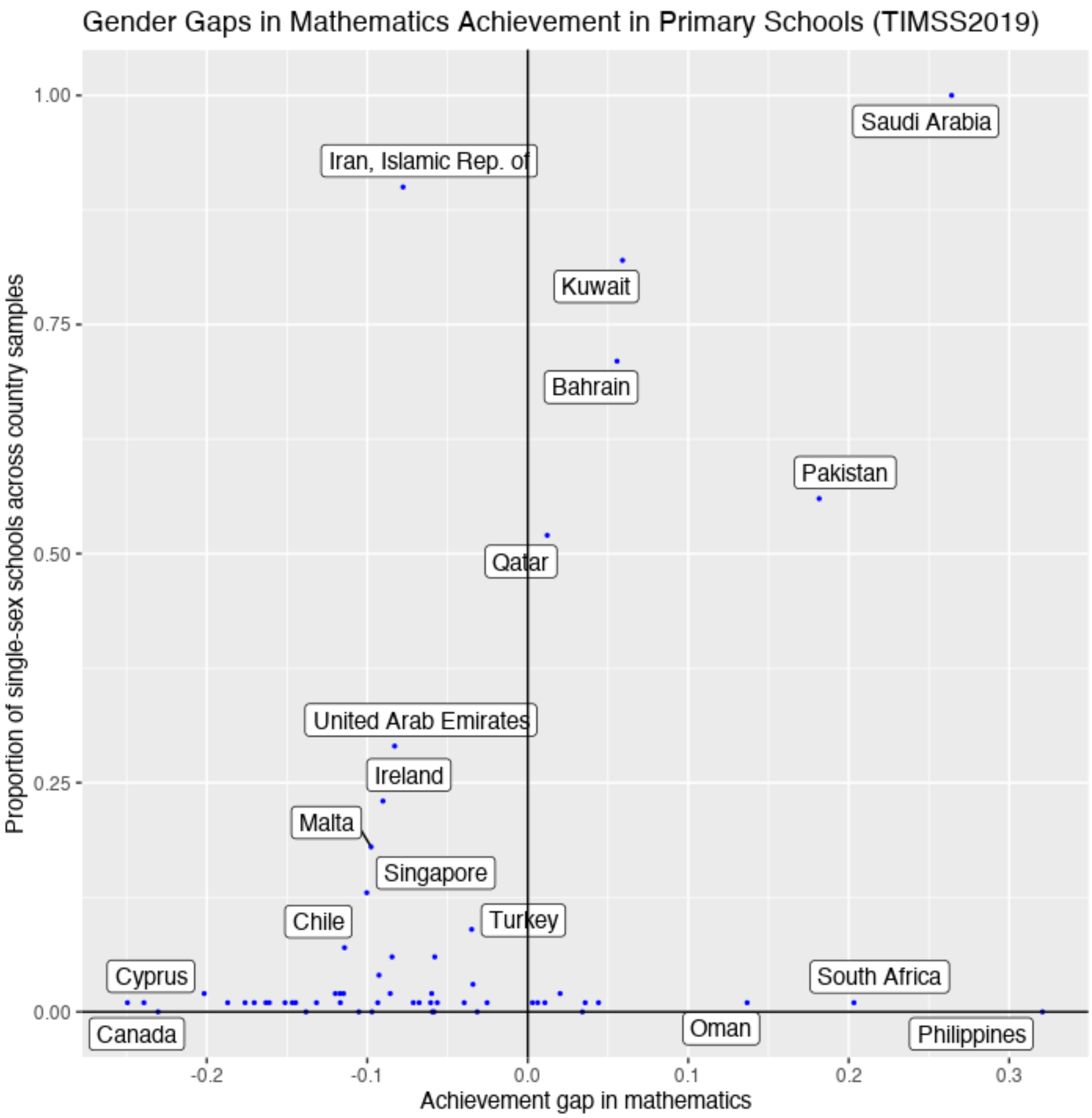
Single-Sex Education and Mathematics Achievement

When looking at TIMSS 2019 mathematics achievement results at the primary and secondary school level, it is interesting to note that among the countries with the largest gender gaps in favour of girls, countries with single-sex education systems seem to be overrepresented. Single-sex education refers to “the practice of educating female and male students within separate learning spaces” (Robinson et al., 2021, p. 3). It can take place within separate schools where there are separate all-girls schools or all-boys schools. Alternatively, it can take place within separate single-sex classes that are organized within coeducational school settings where students, selectively, attend classes reserved for members of the same sex (Robinson et al., 2021).

Figure 1 and Figure 2 visualize student achievement gap in mathematics across all participating countries excluding benchmarking participants (specific country regions, additional grades or additional language groups) in TIMSS 2019 for fourth and eighth grade respectively. Both figures illustrate that more extreme gender gaps in mathematics performance are observed in countries that have higher proportion of single-sex schools in their education systems compared to countries with lower or zero proportion of single-sex schools. For instance, at the secondary school level, Saudi Arabia, Jordan and Iran, which have 100%, 97% and 96% of single-sex schools respectively, have higher gender gaps in mathematics in favour of girls. Countries such as Norway, Cyprus, Finland with 0% single-sex schools show close to zero gender gaps in mathematics achievement (Appendix A).

Figure 1

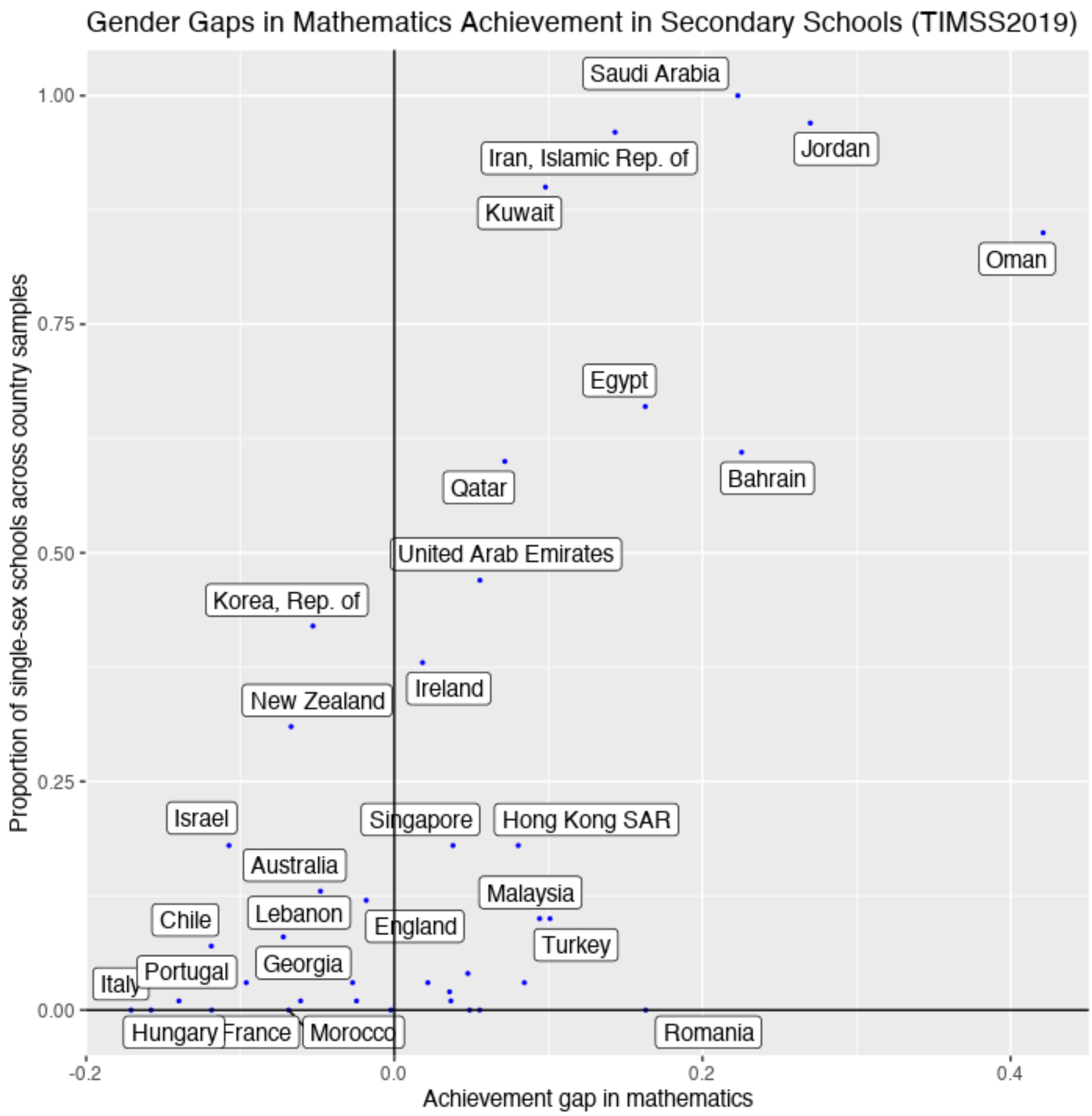
Proportion of Single-Sex Schools and Gender Gaps in Mathematics Achievement (Grade 4)



Note. Effect sizes *d* (x-axis) representing average gender gaps in mathematics achievement in Grade 4 for all TIMSS participating countries (represented by blue dots) excluding benchmarking participants. Values above zero indicate higher scores of girls in contrast to boys. Proportion of single-sex schools compared with all schools in country sample (y-axis) across TIMSS participating countries. A value of zero shows that there are no single-sex schools in a country sample, while a value of 1 shows that all schools in a country sample are single-sex schools. The figure shows labels for only a few countries to overcome overplotting issues. *Source*: Own analysis.

Figure 2

Proportion of Single-Sex Schools and Gender Gaps in Mathematics Achievement (Grade 8)



Note. Effect sizes *d* (x-axis) representing average gender gaps in mathematics achievement in Grade 8 for all TIMSS participating countries (represented by blue dots) excluding benchmarking participants. Values above zero indicate higher scores of girls in contrast to boys. Proportion of single-sex schools compared with all schools in country sample (y-axis) across TIMSS participating countries. A value of zero shows that there are no single-sex schools in a country sample, while a value of 1 shows that all schools in a country sample are single-sex schools. The figure shows labels for only a few countries to overcome overplotting issues. *Source*: Own analysis.

With the exception of South Africa and Philippines at the primary school level and Romania at the secondary school level, we can see a visible cluster of Middle-Eastern countries that exhibit high gender gaps in mathematics achievement in favour of girls. Although these countries differ on the basis of economic development, institutional and political maturity, a common feature shared by these countries is the presence of single-sex education systems, the design of which has been influenced by a common religious and cultural context. The potential role of single-sex education systems in explaining gender gaps in mathematics achievement is explained below.

Academic research on the effects of single-sex education on educational outcomes in mathematics has a long tradition, with much of the research focusing on Western contexts like the USA and the UK where coeducational and single-sex schools coexist. The dominant line of inquiry for single-sex education literature focuses on whether this type of schooling enhances students' mathematics performance as compared to mixed or coeducational schools (Doris et al., 2013; Eisenkopf et al., 2015; Halpern et al., 2011; Pahlke et al., 2014; Park et al., 2018). In a large meta-analysis, Robinson et al. (2021) have noted that there have been mixed and equivocal research results with some studies reporting benefits of single-sex education, others reporting disadvantages and some reporting no effects on student's mathematics achievement.

The inconsistency of research findings in single-sex education literature can, at least partly, be attributed to some "serious methodological weaknesses" (Hayes et al., 2011, p. 695) in research studies. While analysing the effect of single-sex schooling on achievement, most studies have used samples from Western countries where single-sex education is a specialized and unique feature of education systems in which students or their families freely choose single-sex schools over coeducational schools (Park et al., 2018). Hayes et al. (2011) have identified two types of selections effects in literature that can potentially bias research results. These are student-driven selection effects and school-driven selection effects. *Student-driven selection effects* occur when students who choose to attend single-sex schools may differ systematically from those students who do not choose to attend single-sex schools. For instance, students choosing to attend single-sex schools can have higher levels of prior ability or motivation. Furthermore, since the majority of single-sex schools in Western countries tend to be private institutions (Signorella et al., 2013), they tend to attract students with higher socio-economic status compared to those who choose to attend free public coeducational schools. In contrast, *school-driven selection effects* occur when schools adopt selective admission processes in the form of setting specific admissions criteria (e.g., prior achievement scores, religiosity) for students attending single-sex schools. The presence of student-driven selection effects and school-driven selection effects raise questions about the validity of previous studies that have reported significant achievement differences between single-sex and coeducational schools (Hayes et al., 2011).

A few studies have tried to overcome the issue of selection effects by either exploiting data from natural experiments or carefully controlling for selection effects. In the South Korean context, Kim and Law (2012) examined the effects of single-sex schools and coeducational schools in 15-year-old students' mathematics achievement using data from PISA 2006 study. South Korea offers a unique opportunity to study the effects of single-sex schooling as students are randomly assigned into single-sex or coeducational schools within residential school districts as a result of an equalization policy implemented in the 1970s (Pahlke et al., 2013). Through using hierarchical linear modelling (HLM) to account for the nested nature of the data, Kim and Law (2012) found that girls as well as boys in single-sex schools on average had higher mathematics performance compared to their counterparts in coeducational schools. In contrast, Pahlke et al. (2013), using eighth grade data from TIMSS

2003 and 2007 studies found contradictory results for South Korean students. After testing for random assignment of students to the specific school type through controlling for a range of background variables, Pahlke et al. (2013) found that in both TIMSS 2003 and 2007 waves, the gender composition of the schools was not related to either boys' or girls' performance in mathematics. In other words, they concluded that girls in single-sex schools performed as well as girls in coeducational schools, and boys at boys-school performed as well as boys in coeducational schools.

While the above studies focused on middle school education, Park et al. (2018) used data from high school seniors' national college entrance exam scores on mathematics in Seoul, South Korea to examine the effects of single-sex schools on mathematics performance. They found that boys in all-boys schools in Seoul perform better on a general-mathematics test compared to their counterparts in coeducational schools, however girls attending all-girls schools do not show significantly better performance on the mathematics test compared to girls attending coeducational schools. Park et al. (2018) acknowledge that despite the random assignment of students in single-sex and coeducational schools in Seoul, most single-sex schools tend to be private institutions in contrast to coeducational schools. These private schools tend to have significant differences in teacher recruitment and teacher tenure policies as compared to public coeducational schools, which can lead to spurious benefits to attending single-sex schools.

In the Swiss context, Eisenkopf et al. (2015) exploited school performance mathematics data from a high school where girls were assigned to single-sex and coeducational classes on the basis of school management's decision. They found that students in all-girls classes obtain better grades in mathematics as compared to their female fellow students in coeducational classes. They suggested that girls perform better in single-sex classrooms due to the absence of stereotype threat in mathematics and by engaging in competitive behaviour in the absence of boys in the classroom. Since Eisenkopf et al. (2015)

primarily utilized data from a female-dominated student body, they do not offer comparable results for the impact of single-sex classes on the academic performance of boys.

Through exploiting data from the Irish education system where about a quarter of primary school children are educated in single-sex, state-funded schools, Doris et al. (2013) investigated the extent to which single-sex schooling affects gender gap in mathematics at the top of the test distribution in nine years-old students. They used scores that tested a range of skills in problem solving, reasoning, and mathematical understanding. Through estimating a number of probit models, they concluded that gender gap is wider among primary school students attending single-sex schools compared to those studying in co-educational schools. They have argued that single-sex education exacerbates educational inequalities among boys and girls, rather than reducing them. However, they have not explained or investigated possible mechanisms through which single-sex education perpetuates gender gaps in mathematics.

Using data from TIMSS 2003, Wiseman (2008) found that countries (predominantly Middle-eastern) with a significantly high percentage of single-sex schools either had no significant gender differences in mathematics score or had a significant girl advantage. In an extension to their gender gaps analysis, Wiseman (2008) looked at different indicators of student's opportunity to learn in single-sex schooling and coeducational schools by focusing on factors related to teacher characteristics, teaching methods, and curriculum coverage. He found that some of the countries with high levels of gender segregation in their national education systems showed considerable variation in opportunities to learn for students educated in girls' schools, boys' schools and coeducational schools. For example, he found that teachers at boys-only schools in Saudi Arabia and Bahrain had significantly more experience (in years) than teachers at girls-only schools. Similarly, teachers at boys-only schools in the Palestinian National Authority had significantly more training (in years) than teachers at coeducational schools. However, these results are counter-intuitive in relation to the actual mathematics achievement scores of boys and girls. If teachers in boys' schools are more experienced and more trained compared to those in girls' schools, then we should see a boy advantage in mathematics achievement rather than a girl advantage in these countries.

Contrary to Wiseman's finding about Saudi teachers' education and experience, Haroun et al. (2016) report that in Saudi Arabia, female teachers in girls' schools scored significantly higher than their male counterparts in both Content Knowledge (CK) and Knowledge of Content and Student (KCS) in mathematics. Combined together, CK and KCS represent teachers' subject matter knowledge in mathematics as well as their pedagogical content knowledge. If female teachers in Saudi all-girls' schools have significantly more mathematics knowledge compared to their male counterparts, then this could possibly explain why girls outperform boys in mathematics in Saudi Arabia, a country with 100% single-sex schools.

In the light of limited and often contradictory results with respect to mathematics achievement in the context of single-sex education, the present study has taken advantage of the Middle-Eastern context of single-sex schooling to analyse gender differences in mathematics achievement. By focusing on an educational context where single-sex schooling is the default school type, this study has sought to overcome the issue of selection effects and school-driven selection effects that are present in extant literature (Hayes et al., 2011).

Features of Single-Sex Education Systems that could Explain Gender Gaps in Mathematics Achievement

In their 20-year (2001-2020) scoping review, Robinson et al. (2021) note that much of the research on single-sex schooling focuses on analysing *exogenous forces* that contribute to achievement differences, such as race, class, gender, socioeconomic status and national curricula. However, very little research focuses on *endogenous forces* operating at the classroom and school level and more proximal to students' actual learning processes that could explain variation in student achievement in single-sex education systems. In order to fill this research gap, the present study examines variation in such more proximal variables that may account for gender gaps in mathematics achievement of boys and girls. More specifically, this study considers variations in *school climate* that can potentially account for gender gaps in mathematics achievement. In comparison to coeducational systems, if boys and girls attend separate schools, and if these schools have different *school climate* conditions, this might contribute to gender gaps in achievement outcomes.

Within the broader school effectiveness research, *school climate* has been recognized as an important factor linked with student outcomes across academic, behavioural, and psychosocial domains (Bryk & Driscoll, 1988; Scheerens et al., 2013; Wang & Degol, 2016). School climate refers to the overall school experience that involves aspects of quality of teaching and learning, school organization, school community relationships, and the institutional and structural features of the school environment (Thapa et al., 2013; Wang & Degol, 2016).

While recognizing the multidimensionality of school climate in research literature, Wang and Degol (2016) proposed four dimensions of school climate: academic, community, safety, and institutional environment. *Academic climate* relates to the quality of the academic atmosphere, including curricula, instruction, teacher training, and professional development. *Community* relates to the quality of interpersonal relationships within the school. *Safety* indicates the degree of physical and emotional security provided by the school. Lastly *institutional environment* reflects the organizational or structural features of the school environment such as psychical infrastructure, availability of resources, class size and school size.

With regards to the academic climate of schools, empirical research has found that schools that put greater emphasis on academic success and reinforce high standards for academic performance experience higher student achievement (Bodovski et al., 2013; Hoy et al., 2006; Scheerens et al., 2013). Similarly, students who attend schools characterized by high-quality interpersonal relationships, sense of belongingness and cohesiveness tend to have significantly higher academic achievement (Stewart, 2008). School environments that provide a safe learning environment for students are also known to contribute positively to academic achievement (Gietz & McIntosh, 2014; Thapa et al., 2013). In terms of the institutional characteristics of schools such as size, type, location and structural features of learning environments, research suggests that these structural characteristics may not directly affect student achievement, but may in fact alter classroom processes that can indirectly affect student achievement levels (Wang & Degol, 2016).

In a multilevel meta-analysis of 38 studies published between 2000 and 2020, Demirtas-Zorbaz et al. (2021) found that school climate was positively correlated with academic achievement, albeit the effect size for the relationship between school climate and academic achievement was small ($r = 0.178$, $p < 0.01$). They found that the academic dimension of school climate yielded larger effect sizes compared to other domains, followed by safety, community, and institutional environment. This is in contrast to Ma and Klinger's (2000) finding who found school's disciplinary climate (rule clarity, perceived fairness of rules, consistency of application and disruptive behaviour) to be the most important predictor of academic achievement among two other school climate variables (academic press and parental involvement).

Given past research results on the positive association between school climate and academic achievement, it is worth investigating whether differences in school climate contribute to differences in academic achievement between boys and girls who attend separate schools in single-sex education systems.

Single-Sex Education System in Saudi Arabia

The specific country that is the focus of this study is Saudi Arabia, a Muslim-majority country located in the Middle East. Given Saudi Arabia's pervasive gender segregation policies in its education sector, it provides a unique opportunity to analyse the association between school-level characteristics and gender gaps in mathematics achievement. By taking advantage of the Saudi context of single-sex schooling where students or parents do not have the choice to opt in or opt out of this form of schooling, we can overcome the issue of selection bias present in extant literature on single-sex education (Hayes et al., 2011).

Since the 1960s, there has been a strict segregation between female and male workplaces and socializing spaces in Saudi Arabia reflecting the country's strict adherence to religious values (Renard, 2008). Gender segregation is considered 'the norm' or 'default' in Saudi society (van Geel, 2016) that is also reflected in the country's education system where co-education is prohibited beyond kindergartens (El-Sanabary, 1994). In Saudi educational institutions, all students, teachers and personnel are from the same sex, working in educational institutions designed for their specific gender (Barry, 2019). Furthermore, teacher training provided by Ministry of Education is conducted separately for male and female staff (Al-Zarah, 2008).

The Saudi education system is divided into pre-primary (3-5 years), primary (6-11 years), secondary (12-17 years) and tertiary (18-22 years) education (UNESCO Institute of Statistics, 2022). According to World Economic Forum's Global Gender Gap Report 2021, Saudi Arabia has achieved gender parity in primary school enrolment with near equal enrolment between girls (94.4%) and boys (94.7%). At secondary school level, there exists some gender disparity in enrolment rates with 94.1% of girls enrolled versus 98.5% of boys enrolled (UNESCO Institute of Statistics, 2022).

In terms of gender equity, Saudi Arabia ranked 147 out of 156 in World Economic Forum's 2021 Global Gender Gap Index. Women's labour force participation was reported at only 23.3% compared to 80.7% for men (World Economic Forum, 2021). Despite the limited opportunity structure for women and girls in the labour market, the country has consistently reported a significant girl advantage in average mathematics scores in TIMSS 2011, 2015 and 2019 at both fourth and eighth grade in contrast to their male counterparts (Mullis et al., 2020)

– a finding that is in stark contrast with Baker and Jones's (1993) gender stratification hypothesis.

The Present Study

In light of the above, the present study seeks to answer the following questions:

RQ1: To what extent does the reported level of *school climate* vary between girls' schools and boys' schools at the primary and secondary school level in Saudi Arabia?

RQ2: To what extent is the reported level of *school climate* associated with students' mathematics achievement at the primary and secondary school level in Saudi Arabia?

Based on the state of research, the hypothesis is that *school climate* and students' mathematics achievement are positively associated (RQ2). In addition, if within Saudi Arabia, girls' schools experience higher levels of perceived *school climate* (RQ1), then these two pieces of evidence could tentatively suggest why girls have such a large advantage in mathematics achievement in Saudi Arabia (Figure 1 and Figure 2).

This study diverges from other existing gender gap studies in several ways. *First*, it is the first attempt to analyse gender gaps in mathematics achievement in non-Western contexts where single-sex education is a not a specialised or unique feature of education systems, thereby overcoming issues of student or school selection effects. *Second*, by considering school climate, this study seeks to provide useful insights into antecedents that could potentially help us understand why large unexpected girls' advantages in mathematics achievement exist in Saudi Arabia. *Third*, it uses the most recent TIMSS data (2019) to examine gender differences in mathematics achievement, thereby providing latest evidence in the extant body of literature.

Methods and Analyses

Data

This study is based on fourth and eighth grade data for Saudi Arabia from TIMSS 2019 study. Conducted on a four-year assessment cycle for students in the fourth and eighth grades, TIMSS focuses on assessing student achievement on common school curriculum across participating countries and the test items typically reflect what students have learned in the classroom. Through providing contextual background information from participating students and their homes, teachers, and schools, TIMSS data provides a unique opportunity to link student outcomes with instructional characteristics as they collect data from intact classrooms.

TIMSS uses a stratified two-stage cluster design for choosing participants within a country (LaRoche et al., 2020). In the first stage, schools are sampled with probability proportional to their sizes, whereby larger schools have higher probability of being sampled (LaRoche et al., 2020). In this stage, the schools within the sampling frame may be stratified according to some important demographic variables. In the case of Saudi dataset, explicit stratification was done on the basis of school type (public, private, international/foreign) and school gender (boys' and girls' schools). In the second sampling stage, one or more intact classroom from the target grade of each participating school is randomly selected such that all students in the selected classes are then assessed (LaRoche et al., 2020). This sampling design enables to achieve a nationally representative sample of schools and students across the TIMSS participating countries.

Within the national target population, TIMSS sometimes excludes certain types of schools and students within countries. School-level exclusions mostly consist of schools for students with disabilities and remote schools, while student-level, or within-school exclusions generally comprise of students with disabilities or students who are non-native language speakers (LaRoche & Foy, 2020). According to TIMSS guidelines, the overall percentage of excluded students (combining school and within-school levels) should be 5% or less. In the present study, it is important to note that the Saudi sample had an overall exclusion rate of 10%, which is above the 5% recommended threshold. 9.1% were school-level exclusions (very small schools, special needs schools, schools using different language other than Arabic or English), while 0.9% were within-sample exclusions (students with intellectual disabilities, students with functional disabilities, and non-native language speakers).

The fourth-grade student dataset used in this study contained 5453 students nested in 220 schools (Table 1). The average age of the students in fourth grade at the time of testing was 9.9 years (Mullis et al., 2020). With an exception of one school that was an international school, all schools in the dataset were single-sex schools. After excluding the one coeducational school from the dataset, the effective school sample size contained 219 schools and 5432 students (50.42 % girls). There were 111 girls' schools, while 108 schools were boys' schools. Majority of schools ($N = 217$) had one classroom per school. The fourth-grade teacher dataset, nested in 219 schools, contained observations from 436 teachers (48.3% female), that taught students both mathematics and science. The majority of schools contained data from 2 or more teachers per school. Almost all schools had same-gender teachers, with an exception of one or two teachers being of the opposite gender.

The eighth-grade student dataset contained 5680 (50.77 % girls) students nested in 209 schools (Table 1). The average age of the students in eight grade at the time of testing was 13.9 years (Mullis et al., 2020). All schools in the dataset were single-sex schools**,** with 106 girls' schools and 103 boys' schools. Majority of schools ($N = 196$) had one classroom per school. The teacher dataset contained 449 students' mathematics teachers, with majority of teachers being of the same-gender as well.

Measures

To assess the relations among school gender, school climate and students' mathematics achievement, existing TIMSS 2019 fourth and eighth grade student, teacher and principal scales were used. Given school level is the focus of RQ1 and RQ2, student and teacher reported data for school climate variables was aggregated at the school level.
The measures and scales used in the present study are briefly described below (cf. Yin & Fishbein, 2020) and the precise item wording for the scales is given in Appendix B:

Table 1

Descriptive Statistics for Girls' Schools and Boys' Schools for Grade 4 and Grade 8

	Grade 4			**Grade 8**		
	Girls' Schools (*N* = 111)	**Boys' Schools (*N* = 108)**	**All Schools (*N* = 219)**	**Girls' Schools (*N* = 106)**	**Boys' Schools (*N* = 103)**	**All Schools (*N* = 209)**
Number of Students	2739	2693	5432	2884	2796	5680
Number of teachers	221	215	436	229	220	449
Students' Gender						
Female	2739	0	2739	2884	0	2884
Male	0	2693	2693	0	2796	2796
Missing	-	-	-	-	-	-
School Teachers' Gender						
Female	209	2	211	102	2	104
Male	1	207	208	0	102	102
Missing	11	6	17	127	116	243
School Location						
Urban	87	93	180	89	84	173
Rural	15	12	27	13	16	29
Missing	9	3	12	4	3	7

Note. Students' and teachers' gender based on self-reported data.

School Gender

Students' self-reports on their gender was used to derive the school gender, where schools with 100% female student population were categorized as girls' schools, while schools with 100% male population represented boys' schools. Boys' schools were coded as 0 and girls' school were coded as 1. There was no missing data on students' self-reports on gender.

School Climate Measures

In line with the theoretical framework proposed by Wang and Degol (2016), four dimensions of school climate were considered in this study to represent school climate, namely; *academic, safety, institutional environment, and community* dimensions. Consistent with past empirical research (Ker, 2016; Lee & Chen, 2019; Scherer & Nilsen, 2016), four TIMSS scales were used as proxies to represent the different dimensions of the *school climate* construct*: School Emphasis on Academic Success (SEAS), Safe and Orderly School (SOS), Instruction Affected by Mathematics Resource Shortages (SHORTAGE) and Students' Sense of School Belonging (SSB).*

School Emphasis on Academic Success (SEAS): The SEAS scale, created by TIMSS using the Item Response Theory (IRT) partial credit model, was used to measure the *academic dimension* of school climate. Based on teachers' responses to 12 items (Appendix B), the scale covered teachers' expectations on successful curriculum implementation and student achievement, parental support for student achievement, and the students' desire to achieve. A higher score on the scale indicates higher level of school emphasis on academic success. Cronbach's alpha reliability coefficient for this scale for Saudi Arabia was 0.90 and 0.89 for fourth and eighth grade respectively. Since two or more teachers reported on SEAS for the same school, teacher ratings were aggregated at the school level. At the teacher level, missing data on the scale score was 4.8% and 53.6% for fourth and eighth grade respectively. At the school level, the missing data was 1.8% and 4.7% for fourth and eighth grade respectively.

Safe and Orderly School (SOS): Constructed by TIMSS using the IRT partial credit model, the SOS scale based on teachers' responses to eight items, was chosen to represent the *safety* dimension of school climate. The scale included aspects of school's safety posture, students' (mis-)behaviour, and the school's disciplinary rules and procedures (Appendix B). A higher score on the scale indicates a higher safety and orderly environment in the school. For Saudi sample, Cronbach's alpha reliability coefficient for this scale was 0.85 and 0.86 for fourth and eighth grade respectively (Mullis et al., 2020). Since two or more teachers reported on SOS for the same school, teacher ratings were aggregated at the school level. At the teacher level, missing data on the scale score was 10.3% and 55.5% for fourth and eighth grade respectively. At the school level, the rate of missing data on this measure was 3.6% and 8.1% for fourth and eighth grade respectively.

Instruction Affected by Mathematics Resource Shortages (SHORTAGE): This scale was used to represent the *institutional environment* dimension of school climate. TIMSS constructed this scale based on principals' responses to 13 items (Appendix B) using the IRT partial credit model. This scale measured the extent to which school's instruction was affected by shortage of general school resources (such as teaching materials, supplies, school buildings and grounds, heating/cooling and lighting systems) and specific mathematic resources (specialized mathematic teachers, library and computer resources). A higher score indicates that instruction is not affected by resource shortages. Cronbach's alpha reliability coefficient for this scale for Saudi Arabia was 0.90 and 0.92 for fourth and eighth grade respectively (Mullis et al., 2020). Missing data on this variable was negligible at 1.8% and 0.95% for fourth and eighth grade respectively.

Students' Sense of School Belonging (SSB): The SSB scale created by TIMSS using IRT scaling methods was used to represent the *community* dimension of school climate. Based on students' responses to five items, the scale measured whether students feel safe at school, enjoy school, and have good relationships with teachers and classmates (Appendix B). A higher score indicates a higher sense of belonging. Cronbach's alpha reliability coefficient for this scale for Saudi Arabia was 0.75 and 0.78 for fourth and eighth grade respectively. Given school level is the focus of this study, student ratings were aggregated at the school level. At the student level, missing data for this measure amounted to 4.8% and 2.5% for fourth and eighth grade respectively. At the school level, there was no missing data on this measure.

Math Achievement

TIMSS measures students' achievement with a mathematics test and estimates it via a measurement model that produces a set of five plausible values that represent the posterior distribution of student proficiency given achievement and contextual data (Foy et al., 2020). Analyses were conducted by using all five plausible values and were conducted for each set of these values. Resultant model parameters were combined using Rubin's (1987) rules. The scores were scaled to an international mean of 500 with a standard deviation of 100. The plausible values contained no missing data.

School Socio-Economic Status (SES):

To control for the effects of socio-economic status, the Home Educational Resources scale (derived from students' ratings of the number of books at home, their parents' highest education and home study supports such as students having their own room and internet connection) provided by TIMSS was used. Since this scale was measured on the student level, it was aggregated at the school level to represent schools' socioeconomic status. An aggregated measure of SES based on student ratings was chosen because the manifest school SES variable for TIMSS Saudi data had high levels of missing data (34.5% for fourth grade and 28.2% for eighth grade). A higher score on the scale indicates a higher level of socio-economic status. At student level, the missing rate on this measure for fourth and eighth grade was 12.5% and 1.7% respectively. At school level, there was no missing data on this measure.

School Location

In order to control for the effects of school location, a recoded school location variable was used where TIMSS' five-category variable was recoded to a binary variable (rural = 0, urban = 1). School locations coded as 'urban-densely populated', 'suburban-on fringe or outskirts of urban area', and 'medium size city or large town' were recoded as 'urban'. School locations coded as 'small-town or village' and 'remote rural' were recoded as 'rural'. Percentage of missing data (5.5% for fourth grade and 3.34% for eighth grade) on this measure was relatively low.

Statistical Models

RQ1: To what extent does the reported level of school climate vary between girls' schools and boys' schools at the primary and secondary school level in Saudi Arabia?

In order to model the differences in perceived levels of school climate between girls' and boys' school (RQ1), four separate models were considered for the four different dimensions of school climate: School Emphasis on Academic Success (SEAS), Safe and Orderly Schools (SOS), Instruction Affected by Mathematics Resource Shortages (SHORTAGE) and Students' Sense of School Belonging (SSB).

As a first step, school climate variables were regressed on school gender type (girls' or boys' schools) at the school level using the following equation:

$$Y_j = \beta_{0j} + \beta_{1j}X_j + \mu_{0j}$$

Where Y_j is the outcome variable (school climate dimension) in school j and X_j is a dichotomous variable that indicates school gender (0 = boys' schools, 1 = girls' schools).

As a second step, school SES and school location were added in the models to control for differences in school climate variables depending on school SES and location. Robustness analyses were also run to make sure that the results were not driven by violations of linear regression assumptions.

RQ2: To what extent is the reported level of school climate associated with students' mathematics achievement at the primary and secondary school level in Saudi Arabia?

Four different models were fitted to estimate the association between different dimensions of school climate and students' mathematics achievement. The intra-class correlation (ICC) for students' mathematics achievement at fourth grade was 31.2% and 37.7% for eighth grade and the corresponding design effect was 8.42 and 10.84 for fourth grade and eighth grade respectively. Since the design effect was above the recommended threshold of 2 (Muthen & Satorra, 1995), multi-level modelling was considered (Snijders & Bosker, 2012). The following equation can be used to specify this type of model:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + r_{ij}$$
$$r_{ij} \sim \mathrm{N}\,(0, \sigma^2 \mathrm{r})$$

$$\text{Level 2: } \beta_{0j} = \gamma_{oo} + \gamma_{01}.W_j + \mu_{0j}$$
$$\mu_{0j} \sim \mathrm{N}\,(0, \sigma^2 \mu_0)$$

$$\text{Total equation: } Y_{ij} = \gamma_{oo} + \gamma_{01}.W_j + \mu_{0j} + r_{ij}$$

The student level (L1) equation contains only the simple decomposition of an observed mathematics score into the cluster mean (β_{0j}) plus an individual deviation from the cluster mean (r_{ij}). At level 2 (L2), differences in the cluster means of math achievement scores are explained through the L2 predictor W_j (SEAS, SOS, SHORTAGE or SSB). The parameters γ_{oo} and γ_{01} are fixed L2 regression coefficients for the regression of the cluster means on the L2 predictor. The term μ_{0j} reflects that part of the cluster mean β_{0j} that cannot be predicted by W_j.

As a first step, four different multi-level models with outcome variable at L1 and school climate predictors at L2 were run. As a second step, school SES and school location were added as control variables at L2 in these models. As a third step, another model was run where all the different predictors were included simultaneously. Subsequently, control variables were added in this combined model. In all these models, school climate variables were grand-mean centred while the outcome variable, school SES and location remained unaltered.

Robustness analyses were also run to make sure that the results were not driven by violations of multilevel regression assumptions.

Statistical Analyses

The data was prepared and merged using the EdSurvey package (Bailey et al., 2021) in statistical software (R Core Team, 2021), while statistical analyses were carried out in Mplus 8.6 (Muthén & Muthén, 1998-2017). All models used maximum likelihood estimation with robust standard errors. The full-information-maximum-likelihood procedure was used to handle the missing data in the sample. The school level missing data ranged between 0% to 8.1%. In order to account for sampling design applied in TIMSS 2019, school sampling weights (SCHWGT) were included in models specified to answer RQ1. Since student level data was used in RQ2 models, therefore total student sampling weights (TOTWGT) were included in these models.

Before all models were run, measurement invariance (MI) across school gender type was tested for the latent constructs. MI is important to establish in order to check if the latent variable of interest is understood and measured equivalently across girls' and boys' schools (Rutkowski & Svetina, 2017). MI was investigated by conducting multiple-group confirmatory factor analysis. Firstly, *configural* invariance was investigated, which means that in each school group, the same items had to be associated with the same latent factors. Secondly, *metric invariance* was tested, by studying whether the factor loadings were invariant across school types and thirdly, *scalar invariance* was tested, by studying whether the intercepts were invariant across school types. The results given in Appendix C indicate that it was not possible to establish metric invariance for some latent constructs. The implications of MI testing results are discussed in the discussion section.

Results

Descriptive Results

Descriptive statistics for girls' schools, boys' school and all schools in the sample for fourth grade are presented in Table 2. Consistent with Figure 1, girls' schools at fourth grade level had higher average mathematics achievement (*M* = 411.76, *SD* = 91.31) compared to boys' schools (*M* = 384.99, *SD* = 108.17). When looking at the school climate variables, girls' schools had slightly higher reported levels of average SEAS, SOS, SHORTAGE and SSB, compared to boys' schools (see Appendix D.1 for the boxplot). Similarly, average school SES in girls' school (*M* = 9.61, *SD* = 0.74) was reported to be slightly more in contrast to boys' schools (*M* = 9.38, *SD* = 0.82). When looking at the correlations between key study variables for fourth grade (Appendix D.2), we see that none of the school climate variables were significantly correlated with mathematics achievement in girls' schools. On the contrary, SEAS and SOS had positive and significant correlation with mathematics achievement in boys' schools. When looking at the combined school sample, SEAS (r = 0.224), SOS (r = 0.247) and SSB (r = 0.147) were found to be positively and significantly correlated with mathematics achievement.

For eighth grade (Table 3), girls' schools also had higher average mathematics achievement (*M* = 402.52, *SD* = 74.49) compared to boys' schools (*M* = 385.25, *SD* = 80.29). Similar to fourth grade, girls' schools had higher reported levels of average SEAS, SOS, SHORTAGE and SSB, compared to boys' schools in eighth grade (see Appendix D.1 for the boxplot). Similarly, average school SES in girls' school (*M* = 9.73, *SD* = 0.79) was reported to be slightly more in contrast to boys' schools (*M* = 9.58, *SD* = 0.76). Inspection of the correlations (Appendix D.3) reveals that SEAS and SOS had positive and significant correlation with mathematics achievement in girls' schools. On the contrary, none of the school climate variables were significantly correlated with mathematics achievement in boys' schools. When looking at the combined school sample, SOS was found to be positively and significantly correlated with mathematics achievement, albeit the correlation coefficient was found to be small (r = 0.157).

Table 2

School Level Descriptive Statistics for Study Variables for Grade 4

	Girls' Schools (*N* = 111)					**Boys' Schools (*N* = 108)**					**All Schools (*N* = 219)**				
	M	*SD*	Min	Max	Missing	*M*	*SD*	Min	Max	Missing	*M*	*SD*	Min	Max	Missing
MATH	411.76	91.31	101.04	27.01	-	384.99	108.17	47.61	728.11	-	397.89	101.28	47.62	739.53	-
SEAS	11.28	2.15	7.41	17.128	2	10.53	2.00	6.27	15.83	2	10.89	5.218	6.28	17.13	4
SOS	11.79	1.63	6.51	13.352	3	10.71	1.63	6.70	13.35	5	11.24	4.875	6.51	13.35	8
SHORTAGE	8.65	2.13	2.27	15.434	2	8.51	2.19	2.27	15.77	2	8.57	3.140	2.27	8.45	4
SSB	11.10	0.77	8.68	12.572	-	9.94	0.88	7.82	12.54	-	397.89	101.28	47.62	739.53	-
SES	9.61	0.74	8.31	11.560	-	9.38	0.82	7.87	11.42	-	9.496	0.794	7.875	11.560	-

Note. Descriptive statistics based on original scale scores. MATH: Student mathematics’ achievement, SEAS: School Emphasis on Academic Success, SOS: Safe and Orderly Schools, SHORTAGE: Instruction Affected by Mathematics Resource Shortages; SSB: Students’ Sense of School Belonging, SES: School socio-economic status.

Table 3

School Level Descriptive Statistics for Study Variables for Grade 8

	Girls' Schools (*N* = 106)					**Boys' Schools (*N* = 103)**					**All Schools (*N* = 209)**				
	M	*SD*	Min	Max	Missing	*M*	*SD*	Min	Max	Missing	*M*	*SD*	Min	Max	Missing
MATH	402.52	74.49	145.44	693.70	-	385.25	80.29	135.25	709.97	-	393.77	77.99	124.67	713.9	-
SEAS	11.55	2.09	7.309	17.78	8	10.68	2.32	4.88	17.78	2	11.102	2.26	4.879	17.78	10
SOS	11.88	1.89	6.288	13.87	11	11.13	1.64	7.26	13.87	6	11.496	1.96	6.29	13.87	17
SHORTAGE	8.9	1.91	4.235	14.25	2	8.66	1.79	4.23	14.25	-	8.779	1.85	4.23	14.25	2
SSB	10.29	0.85	7.973	10.88	-	10.15	0.78	7.97	12.83	-	10.225	0.82	7.970	12.83	-
SES	9.73	0.79	8.311	11.56	-	9.58	0.76	7.87	11.42	-	9.66	0.77	7.86	11.56	-

Note. Descriptive statistics based on original scale scores. MATH: Student mathematics' achievement, SEAS: School Emphasis on Academic Success, SOS: Safe and Orderly Schools, SHORTAGE: Instruction Affected by Mathematics Resource Shortages; SSB: Students' Sense of School Belonging, SES: School socio-economic status.

Main Findings

RQ1: To what extent does the reported level of school climate vary between girls' schools and boys' schools at the primary and secondary school level in Saudi Arabia?

Different models were fitted separately to investigate the extent to which school climate dimensions vary between girls' schools and boys' schools at the primary school and secondary school level in Saudi Arabia.

Results from fourth grade (Table 4) indicate that there were significant mean differences between girls' and boys' schools for the *safety* (SOS) and *community* (SSB) dimension of school climate in Saudi Arabia, after controlling for the effects of school SES and location. On average, students' in girls' schools reported higher on the SOS ($b = 0.74$, $p < .05$) and SSB ($b = 1.128$, $p < .05$) scales compared to their male counterparts, after controlling for the effects of school SES and location. When looking at the *academic* climate (SEAS) for fourth grade, significant differences between girls' and boys' school were also found, however, these differences were not found to be robust when controls for school SES and location were added. Not in line with expectations, no statistically significant differences between girls' and boys' schools were found for the *institutional environment* (SHORTAGE) dimension for school climate, also after controlling for school SES and school location.

Results from eighth grade (Table 5) show that there were significant differences between girls' and boys' schools only for the *safety* (SOS) dimension of school climate, after controlling for the effects of school SES and location. No significant differences between girls' and boys' schools were found for the *community* (SSB) and *institutional environment* (SHORTAGE) dimension, with and without control variables. Similar to fourth grade, significant differences between the two school types were found for the *academic* dimension, but these results were not found to be robust after inclusion of control variables.

RQ2: To what extent is the reported level of school climate associated with students' mathematics achievement at the primary and secondary school level in Saudi Arabia?

In order to answer RQ2 that sought to examine the extent to which reported levels of *school climate* is associated with students' mathematics achievement, a number of multi-level models were fitted. In these analyses, the extent to which the variance in average mathematics achievement score (the intercept) at L1 is explained by L2 predictors was examined.

Table 6 presents the results for different multi-level models for fourth grade. Results indicate that in line with expectations, the *academic* (SEAS)*, safety* (SOS) and *community* (SSB) dimensions of school climate were positively and significantly associated with students' mathematics achievement, after controlling for the effects of school SES and location. However, the *institutional environment* (SHORTAGE) dimension was not found to be significantly associated with students' mathematics achievement in models with and without control variables. In Model 9 in Table 6, all school climate dimensions were added simultaneously to check their association with students' mathematics achievement. Only SEAS was found to be positively and significantly associated with students' mathematics achievement, albeit this association became statistically insignificant after controls for school SES and location were added in the model (Model 10 in Table 6).

When looking at eighth grade results given in Table 7, we observe that the *safety* (SOS) dimension is positively and significantly associated with students' mathematics achievement, also after controlling for the effects of school SES and location. In line with expectations, the *institutional environment* (SHORTAGE) dimension was also found to be negatively and significantly associated with students' mathematics achievement, after controlling for the effects of school SES and location. The *academic (*SEAS) and *community* (SSB*)* dimensions were not found to be significantly associated with students' mathematics achievement, after controlling for the effects of school SES and location. However, in Models 9 and 10 (Table 7) where all school climate dimensions were added simultaneously, none of

Table 4

Grade 4 Results of Regressing School Climate Dimensions on School Gender Type

	School Emphasis on Academic Success (SEAS)		Safe and Orderly Schools (SOS)		Instruction Affected by Mathematics Resource Shortages (SHORTAGE)		Sense of School Belonging (SSB)	
	M1a	M1b	M2a	M2b	M3a	M3b	M4a	M4b
Regression parameter	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)
Intercept	10.496* (0.298)	2.616 (2.621)	10.781* (0.22)	6.432* (2.345)	8.691* (0.207)	11.491* (2.18)	10.06* (0.108)	13.288* (1.024)
Girls' Schools	0.834* (0.375)	0.527 (0.372)	0.968* (0.289)	0.74* (0.286)	-0.054 (0.303)	0.079 (0.301)	1.002* (0.138)	1.128* (0.133)
School SES		0.973* (0.302)		0.566* (0.255)		-0.363 (0.266)		-0.366* (0.111)
School Location (Urban)		-1.242* (0.633)		-1.034 * (0.367)		0.691 (0.492)		0.151 (0.235)
Variances	4.353 (0.575)	3.993 (0.543)	4.348 (0.57)	0.732 (0.08)	4.348 (0.57)	4.039 (0.513)	0.732 (0.08)	0.658 (0.066)

Note. The outcome variables are the four dimensions of school climate. Main predictor variable is school gender type (0 = Boys' Schools, 1 = Girls' Schools). Control variables are school SES (positive values indicate higher SES) and school location (coded as 0 = Rural and 1 = Urban). * $p < .050$

Table 5

Grade 8 Results of Regressing School Climate Dimensions on School Gender Type

	School Emphasis on Academic Success (SEAS)		Safe and Orderly Schools (SOS)		Instruction Affected by Mathematics Resource Shortages (SHORTAGE)		Sense of School Belonging (SSB)	
	M1a	M1b	M2a	M2b	M3a	M3b	M4a	M4b
Regression parameter	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)
Intercept	10.635*	8.843*	11.088*	10.549*	8.763*	7.756*	10.199*	11.89*
	(0.295)	(2.578)	(0.221)	(2.309)	(0.182)	(2.105)	(0.108)	(1.078)
Girls' Schools	0.784*	0.728	0.997*	0.987*	0.223	0.188	0.204	0.252
	(0.369)	(0.376)	(0.33)	(0.338)	(0.281)	(0.285)	(0.159)	(0.149)
School SES		0.178		0.074		0.131		-0.148
		(0.279)		(0.257)		(0.227)		(0.123)
School Location (Urban)		0.241		-0.227		-0.29		-0.45*
		(0.502)		(0.464)		(0.273)		(0.223)
Variances	4.464	4.52	3.39	3.416	3.273	3.338	0.747	0.678
	(0.641)	(0.661)	(0.291)	(0.297)	(0.439)	(0.458)	(0.105)	(0.1)

Note. The outcome variables are the four dimensions of school climate. Main predictor variable is school gender type (0 = Boys' Schools, 1 = Girls' Schools). Control variables are School SES (positive values indicate higher SES) and school location (coded as 0 = Rural and 1 = Urban). * $p < .050$

Table 6

Modelling the Effects of School Climate Dimensions (L2) on Students' Mathematics Achievement (L1) for Grade 4

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
	b (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)
Fixed effects										
Intercept γ_{oo}	406.823* (4.007)	78.789 (60.428)	407.702* (4.088)	74.21 (59.988)	407.903* (4.196)	43.506 (58.218)	407.175* (4.158)	16.895* (56.883)	408.021* (4.059)	51.402 (62.431)
School SEAS	8.18* (1.888)	3.426* (1.714)							5.12* (2.427)	0.733 (2.311)
School SOS			8.828* (2.41)	5.521* (2.249)					5.222 (3.169)	3.672 (3.106)
School Shortage					-1.73 (1.675)	0.74 (1.438)			-1.831 (1.627)	0.539 (1.479)
School SSB							4.005 (3.914)	9.259* (3.613)	0.192 (3.888)	6.171 (3.796)
School SES		37.665* (6.433)		37.99* (6.387)		42.131* (6.251)		44.76* (6.217)		40.472* (6.727)
School Location (Urban)		-34.145* (16.121)		-31.804* (16.018)		-41.439* (16.176)		-39.482* (16.275)		-32.649* (16.407)
Random effects										
Within schools ($\sigma^2 r$)	7181.678 (216.273)	7244.568 (220.735)	7157.01 (197.534)	7246.305 (228.287)	7166.777 (190.993)	7218.13 (220.714)	7176.713 (205.869)	7248.004 (226.728)	7182.681 (225.039)	7248.813 (249.478)
Between schools ($\sigma^2 \mu_0$)	2955.267 (333.33)	2360.143 (306.852)	3065.967 (342.4)	2351.402 (294.905)	3336.173 (375.212)	2433.503 (307.813)	3320.451 (367.915)	2321.362 (294.303)	2945.639 (334.06)	2273.805 (284.384)

Note. The outcome variable is students' mathematics achievement at the student level (L1). The predictor variables (SEAS, SOS, SHORTAGE, SSB, SES, LOCATION) are all located at the school level (L2). * $p < .050$

Table 7

Modelling the Effects of School Climate Dimensions (L2) on Students' Mathematics Achievement (L1) for Grade 8

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
	b (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)
Fixed effects										
Intercept γ_{oo}	407.908* (3.904)	-75.975 (41.348)	408.795* (4.136)	-96.477* (42.262)	409.141* (3.93)	-90.283* (40.559)	408.594* (3.923)	-95.928* (41.13)	408.144* (4.024)	-90.785* (41.676)
School SEAS	5.724* (1.896)	1.876 (1.195)							4.425 (2.663)	0.604 (1.417)
School SOS			5.881* (1.958)	3.055* (1.414)					2.557 (2.708)	2.054 (1.651)
School Shortage					-3.213 (2.204)	-3.028* (1.485)			-1.892 (2.196)	-2.431 (1.55)
School SSB							-3.82 (5.513)	3.121 (4.373)	-5.891 (5.336)	1.758 (4.264)
School SES		50.684* (4.664)		52.462* (4.806)		52.313* (4.561)		52.742* (4.611)		51.834* (4.703)
School Location (Urban)		-6.779 (8.446)		-2.612* (9.19)		-7.317 (8.145)		-5.583 (8.766)		-2.888 (9.438)
Random effects										
Within schools ($\sigma^2 r$)	4698.263 (119.384)	4736.046* (118.559)	4709.91* (120.28)	4742.099 (123.167)	4711.769 (112.355)	4742.887 (115.28)	4718.544 (111.688)	4743.013 (114.968)	4714.345 (120.738)	4743.098 (123.371)
Between schools ($\sigma^2 \mu_0$)	2583.794 (357.39)	1281.912 (168.208)	2849.411 (396.272)	1367.312 (194.34)	2731.78 (374.139)	1278.417 (170.383)	2733.523 (358.908)	1302.151 (175.377)	2627.192 (364.913)	1280.542 (185.573)

Note. The outcome variable is students' mathematics achievement at the student level (L1). The predictor variables (SEAS, SOS, SHORTAGE, SSB, SES, LOCATION) are all located at the school level (L2). * $p < .050$

the school climate dimensions were found to be significantly associated with students' mathematics achievement for eighth grade, also after controlling for the effects of school SES and location.

When looking at the control variables, school SES was found to be a significant predictor of students' mathematics achievement in all models for both fourth and eighth grades. At fourth grade level, schools in urban areas showed lower academic achievement. However, school location was not found to be a significant predictor of students' mathematics achievement at the eighth grade level.

Findings of Robustness Checks

The perquisites for linear and multilevel regression models were evaluated**.** The findings for linear regression models and multilevel regression models are summarized in Appendix F and E respectively. Findings from Appendix F support linear regression models overall, with some deviations from the normality assumption. Findings from Appendix E also support running multilevel regression models for fourth grade overall. For eighth grade, we observe slight deviation from normality in the normal probability plot of the standardized L2 residuals. We also observe some deviation from linearity when looking at the relationship between L1 residuals and the outcome variable.

In addition, the school level and two-level models were rerun after removing influential observations. As shown in Appendix E.9 and E.10, the results for robustness checks replicated the main findings for school level models with one exception. For eighth grade, mean differences between girls' and boys' schools for the *academic* dimension became statistically significant, after controlling for the effects of school SES and location. For two-level models (Appendix F.11), the results for robustness checks replicated the main findings for fourth grade only, with slight changes in the size of coefficients. For eighth grade (Appendix F.12), the *safety* dimension was not found to be significantly associated with students' mathematics achievement, after controlling for the effects of school SES and location.

Discussion

Results from TIMSS 2019 indicate that gender gaps in average mathematics achievement in favour of girls are more pronounced in countries that have high proportion of single-sex schools in their education systems compared to countries with coeducational systems (Figure 1 and Figure 2). Saudi Arabia particularly stands out in this regard as it has consistently exhibited extreme gender gaps in mathematics achievement in favour of girls (Mullis et al., 2020). Out of all the TIMSS participating countries, it is also the only country that has implemented 100% gender segregation in its education system, such that girls and boys in the country go to separate single-sex schools at primary and secondary school level (El-Sanabary, 1994). By considering the case of Saudi Arabia, this study sought to examine whether variations in *school climate* dimensions between girls' and boys' schools may help us understand why such extreme gender gaps in students' mathematics performance exist.

Using data from fourth and eighth grade from TIMSS 2019, this study investigated the extent to which the reported level of school climate varies between girls' and boys' schools at the primary and secondary school level in Saudi Arabia. It further examined the extent to which reported levels of school climate were associated with students' mathematics achievement at the primary and secondary school level. If within Saudi Arabia, girls' schools experience higher levels of perceived *school climate* and if *school climate* is associated positively with mathematics achievement, then these two pieces of evidence would point to a possible explanation for the large girls' advantages in mathematics achievement in Saudi Arabia. Recognizing the complexity and multidimensionality of the *school climate* construct as illustrated by past research, this study considered four different dimensions of school climate, namely a*cademic*, *safety, institutional environment* and *community* dimensions (Wang & Degol, 2016) that were measured by TIMSS scales for School Emphasis on

Academic Success (SEAS), Safe and Orderly Schools (SOS), Instruction Affected by Mathematics Resource Shortages (SHORTAGE) and Students' Sense of School Belonging (SSB).

Results from the study show that at both fourth and eighth grade level, there were small but significant differences in the *safety* dimension of school climate between girls' and boys' schools in Saudi Arabia, after controlling for the effects of school SES and location and when subjected to robustness analyses that accounted for influential observations. These results suggest that girls' schools were perceived to have a better safety posture, experienced fewer behavioural problems, had clearer rules for enforcing school safety and more fairness as compared to boys' schools. These results are consistent with past research that has shown that Saudi boys are more likely to experience behavioural problems at school such as engaging in school fights and facing disciplinary action in contrast to girls (AlMakadma & Ramisetty-Mikler, 2015). Furthermore, the *safety* dimension measured by the SOS scale was found to be positively and significantly associated with students' mathematics achievement at fourth grade level, after controlling for the effects of school SES and location and when subjected to robustness analyses that accounted for influential schools. Adopting clear rules for safety and orderliness, observing consistency and fairness in rule application can result in fewer disruptions and behavioural problems at schools, which in turn can promote a stable learning environment. This might partly explain the fact that schools with greater levels of reported safety and orderliness had higher levels of mathematics achievement.

When considering the results for the *community* dimension of school climate measured by SSB, we see that fourth-grade students in girls' schools reported to have a stronger sense of school belonging or school connectedness compared to students in boys' schools, after controlling for the effects of school SES and location. Past research has suggested that girls' in single-sex schools as compared to boys in similar settings tend to have a higher sense of belonging as same-gender settings may provide and reinforce a sense of social solidarity, thereby increasing their sense of school belonging (Brutsaert & Van Houtte, 2002). Students who feel a greater sense of belonging to school are more motivated to attend school, have a positive orientation towards school, classwork and teachers and invest more in the learning process (Osterman, 2000). This in turn could improve their achievement outcomes, which is corroborated by this study's findings where SSB was found to be positively and significantly associated with students' mathematics achievement, after controlling for school SES and location. However, similar findings were not found for eighth grade, where neither significant difference in the sense of school belongingness was found between girls' and boys' schools, nor was the sense of belonging found to be significantly associated with students' mathematics achievement. This might indicate that the effect of school belongingness on mathematics achievement may decline with higher grade levels.

The *academic* dimension of school climate was found to be significantly different between girls' and boys' schools at both fourth and eighth grade level. For eighth grade only, the robustness analyses suggest that difference in reported levels of SEAS across school gender type is significantly different from zero when controlling for the effects of school SES and location. This dimension was also found to be positively and significantly associated with students' mathematics achievement for fourth grade, even after controlling for SES and location. This finding is consistent with past research that showed that schools that put greater emphasis on academic success and reinforce high standards for academic performance experience higher student achievement (Bodovski et al., 2013; Hoy et al., 2006; Scheerens et al., 2013).

Contrary to expectations, the *institutional environment* dimension measured by instruction affected by mathematics shortages was not found to be significantly different between girls' and boys' schools at both fourth and eighth grade. Unlike eighth grade, it was not found to be significantly associated with students' mathematics achievement for fourth grade. The latter finding is in line with Wang and Degol's (2016) observation who note that institutional features of learning environments may not directly affect student achievement levels, but may indirectly affect them by altering classroom processes.

Using data from fourth and eighth grade from TIMSS 2019, the study found some evidence supporting that there might be some *school climate* differences between girls' and boys' schools at primary and secondary school level in Saudi Arabia. In contrast to boys' schools, girls' schools had higher levels of perceived *school climate*, which could tentatively explain why achievement gaps between girls and boys exist. However, research results were not found to be robust across grades and after including control variables, especially in the models where all *school climate* dimensions were simultaneously included. Consistent with past research, school SES was found to be a significant predictor of students' mathematics achievement for both grades and this finding was remarkably robust across models and when subjected to robustness analyses that accounted for influential schools. Students in higher SES schools tend to be better prepared for school work and face stronger parental pressure for academics, which could explain their better academic performance (Gustafsson et al., 2018).

Limitations

Given the exploratory nature of this study, it must be emphasised that this study focused on only on a subset of school-level factors that can potentially explain gender gaps in mathematics achievement in Saudi Arabia. Variation in other educational inputs can also potentially elucidate achievement gaps within single-sex education systems. As students in girls' and boys' schools are taught by different, often same-gender teachers in single-sex education systems, it is possible that teacher quality differs considerably between girls' and boys' schools. Furthermore, male and female teachers often receive professional training in segregated institutions in Saudi Arabia, that could also create variations in teacher quality (Al-Zarah, 2008). Therefore, differences in teacher quality factors, such as characteristics of teachers' educational background, amount of teaching experience, teachers' participation in professional development activities and teachers' self-efficacy, can also potentially explain gender differences in students' mathematics achievement.

Using the framework of instructional quality to understand classroom practices can also be useful, as it can help us understand if girls and boys receive different levels of instructional quality in single-sex educational contexts. However, existing TIMSS scale on instructional quality can be considered deficient, in terms of construct under representation, as it does not adequately cover all three dimensions (*supportive climate, cognitive activation and classroom management)* of instructional quality that have been identified in extant literature (Fauth et al., 2014; Nilsen & Gustafsson, 2016). Future research can use data from independent observers to investigate classroom practices and teaching strategies within single-sex educational contexts, however this type of data collection may prove to be time-consuming and costly.

In addition to the above, it is important to note a few more central limitations pertaining to the analysis conducted in this study. *Firstly*, the present study has used cross-sectional data and therefore, precludes inference regarding cause and effect. While cross-sectional data may be useful for conducting preliminary exploratory analyses, longitudinal data is better suited for establishing causality. Therefore, the findings of this study must be interpreted with caution.

Secondly, even though the sample used in the study was large, it is country-specific and may not be representative of other target populations and countries. However, the present study can offer useful insights for conducting further analysis in countries with similar single-sex education systems that have also experienced extreme gender achievement gaps.

Thirdly, as the present study has used TIMSS ready-made scales to assess differences in *school climate* dimensions between girls' and boys' schools, it has assumed measurement invariance of latent constructs across school gender type, even though it was not possible to establish measurement invariance for some latent constructs. Failure to establish measurement invariance across girls' and boys' schools point to fundamental differences between the two school types in Saudi Arabia, that warrants further research.

Fourthly, this study has only focused on mathematics achievement in Saudi Arabia, even though the country has experienced extreme gender gaps in science and reading achievement as well (Mullis et al., 2020). Future research can extend the present analysis by exploring potential antecedents of gender gaps in other learning domains.

Conclusion

The present research is a step forward in understanding some of the *endogenous forces* operating at the school level that can explain variation in student achievement outcomes in single-sex education systems like those implemented in Saudi Arabia. By focusing on variations in school climate, the study results suggest that there might be some *school climate* differences between girls' and boys' schools at primary and secondary school level in Saudi Arabia, which could tentatively explain why achievement gaps between girls and boys exist. Education practitioners must be cognizant of the importance of school climate factors and must strive to build and maintain a safe and positive school climate for girls' and boys' schools alike, where students can learn and thrive. School management should take into consideration possible perceived deficiencies in school climate of boys' schools, and provide behavioural and learning support that can facilitate the creation of a conducive learning environment for Saudi boys.

While the present study focused on school level factors that can potentially elucidate gender achievement gaps, future research can focus on student level affective factors, such as student engagement and academic self-concept. One possible reason for the relative girls' advantage in average mathematics achievement in Saudi Arabia could be that girls, in contrast to boys, are more engaged in the learning process, exert additional effort in class activities and show greater interest and motivation to learn. Another reason could be that girls have a higher academic self-concept, compared to boys, which could explain girls' relative performance advantage. Future research can empirically test these explanations to further enhance our understanding of why extreme gender mathematics achievement gaps exist in single-sex education systems.

References

AlMakadma, A. S., & Ramisetty-Mikler, S. (2015). Student, school, parent connectedness, and school risk behaviors of adolescents in Saudi Arabia. *International Journal of Pediatrics and Adolescent Medicine*, *2*(3–4), 128–135. https://doi.org/10.1016/j.ijpam.2015.09.004

Al-Zarah, L. N. (2008). *An investigative study of the female initial teacher preparation programme in Saudi Arabia in an Islamic Context: Perceptions of the key participants* [Thesis (Doctoral), Durham University]. http://etheses.dur.ac.uk/1893/

Ayalon, H., & Livneh, I. (2013). Educational standardization and gender differences in mathematics achievement: A comparative study. *Social Science Research*, *42*(2), 432–445. https://doi.org/10.1016/j.ssresearch.2012.10.001

Bailey, P., Emad, A., Huo, H., Lee, M., Liao, Y., Lishinski, A., Nguyen, T., Xie, Q., Yu, J., Zhang, T., Buehler, E., Bundsgaard, J., C'deBaca, R., & Christensen, A. A. (2021). *EdSurvey: Analysis of NCES Education Survey and Assessment Data* (2.7.1) [Computer software]. https://CRAN.R-project.org/package=EdSurvey

Baker, D. P., & Jones, D. P. (1993). Creating Gender Equality: Cross-national Gender Stratification and Mathematical Performance. *Sociology of Education*, *66*(2), 91. https://doi.org/10.2307/2112795

Barry, A. (2019). Gender Differences in Academic Achievement in Saudi Arabia: A Wake-Up Call to Educational Leaders. *International Journal of Education Policy and Leadership*, *15*(15). https://doi.org/10.22230/ijepl.2019v15n15a890

Bodovski, K., Nahum-Shani, I., & Walsh, R. (2013). School Climate and Students' Early Mathematics Learning: Another Search for Contextual Effects. *American Journal of Education*, *119*(2), 209–234. https://doi.org/10.1086/667227

Brutsaert, H., & Van Houtte, M. (2002). Girls' and Boys' Sense of Belonging in Single-Sex versus Co-Educational Schools. *Research in Education*, *68*(1), 48–56. https://doi.org/10.7227/RIE.68.5

Bryk, A. S., & Driscoll, M. E. (1988). *The High School as Community: Contextual Influences and Consequences for Students and Teachers*. https://eric.ed.gov/?id=ED302539

Demirtas-Zorbaz, S., Akin-Arikan, C., & Terzi, R. (2021). Does school climate that includes students' views deliver academic achievement? A multilevel meta-analysis. *School Effectiveness and School Improvement*, *32*(4), 543–563. https://doi.org/10.1080/09243453.2021.1920432

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. https://doi.org/10.1016/j.econedurev.2015.02.005

Doris, A., O'Neill, D., & Sweetman, O. (2013). Gender, single-sex schooling and maths achievement. *Economics of Education Review*, *35*, 104–119. https://doi.org/10.1016/j.econedurev.2013.04.001

Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender Role Stereotypes, Expectancy Effects, and Parents' Socialization of Gender Differences. *Journal of Social Issues*, *46*(2), 183–201. https://doi.org/10.1111/j.1540-4560.1990.tb01929.x

Eisenkopf, G., Hessami, Z., Fischbacher, U., & Ursprung, H. W. (2015). Academic performance and single-sex schooling: Evidence from a natural experiment in Switzerland. *Journal of Economic Behavior & Organization*, *115*, 123–143. https://doi.org/10.1016/j.jebo.2014.08.004

El-Sanabary, N. (1994). Female Education in Saudi Arabia and the Reproduction of Gender Division. *Gender and Education*, *6*(2), 141–150. https://doi.org/10.1080/0954025940060204

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*(1), 103–127. https://doi.org/10.1037/a0018053

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, *29*, 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Fryer, R. G., & Levitt, S. D. (2010). An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics*, *2*(2), 210–240. https://doi.org/10.1257/app.2.2.210

Gietz, C., & McIntosh, K. (2014). Relations Between Student Perceptions of Their School Environment and Academic Achievement. *Canadian Journal of School Psychology*, *29*(3), 161–176. https://doi.org/10.1177/0829573514540415

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). DIVERSITY: Culture, Gender, and Math. *Science*, *320*(5880), 1164–1165. https://doi.org/10.1126/science.1154094

Gustafsson, J.-E., Nilsen, T., & Hansen, K. Y. (2018). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation*, *57*, 16–30. https://doi.org/10.1016/j.stueduc.2016.09.004

Halpern, D. F., Eliot, L., Bigler, R. S., Fabes, R. A., Hanish, L. D., Hyde, J., Liben, L. S., & Martin, C. L. (2011). The Pseudoscience of Single-Sex Schooling. *Science*, *333*(6050), 1706–1707. https://doi.org/10.1126/science.1205031

Haroun, R. F., Ng, D., Abdelfattah, F. A., & AlSalouli, M. S. (2016). Gender Difference in Teachers' Mathematical Knowledge for Teaching in the Context of Single-Sex Classrooms. *International Journal of Science and Mathematics Education*, *14*(S2), 383–396. https://doi.org/10.1007/s10763-015-9631-8

Hayes, A. R., Pahlke, E. E., & Bigler, R. S. (2011). The Efficacy of Single-Sex Education: Testing for Selection and Peer Quality Effects. *Sex Roles*, *65*(9–10), 693–703. https://doi.org/10.1007/s11199-010-9903-2

Hoy, W. K., Tarter, C. J., & Hoy, A. W. (2006). Academic Optimism of Schools: A Force for Student Achievement. *American Educational Research Journal*, *43*(3), 425–446. https://doi.org/10.3102/00028312043003425

Hyde, J. S. (2014). Gender Similarities and Differences. *Annual Review of Psychology*, *65*(1), 373–398. https://doi.org/10.1146/annurev-psych-010213-115057

Ker, H. W. (2016). The impacts of student-, teacher- and school-level factors on mathematics achievement: An exploratory comparative investigation of Singaporean students and the USA students. *Educational Psychology*, *36*(2), 254–276. https://doi.org/10.1080/01443410.2015.1026801

Kim, D. H., & Law, H. (2012). Gender gap in maths test scores in South Korea and Hong Kong: Role of family background and single-sex schooling. *International Journal of Educational Development*, *32*(1), 92–103. https://doi.org/10.1016/j.ijedudev.2011.02.009

LaRoche, S., & Foy, P. (2020). Sample Implementation in TIMSS 2019. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report*. TIMSS & PIRLS International Study Center.

LaRoche, S., Foy, P., & Joncas, M. (2020). Sample Design in TIMSS 2019. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report*. TIMSS & PIRLS International Study Center.

Lee, J., & Chen, M. (2019). Cross-Country Predictive Validities of Non-cognitive Variables for Mathematics Achievement: Evidence based on TIMSS 2015. *EURASIA Journal of Mathematics, Science and Technology Education*, *15*(8). https://doi.org/10.29333/ejmste/106230

Ma, X., & Klinger, D. A. (2000). Hierarchical Linear Modelling of Student and School Effects on Academic Achievement. *Canadian Journal of Education / Revue Canadienne de l'éducation*, *25*(1), 41. https://doi.org/10.2307/1585867

Meinck, S., & Brese, F. (2019). Trends in gender gaps: Using 20 years of evidence from TIMSS. *Large-Scale Assessments in Education*, *7*(1), 8. https://doi.org/10.1186/s40536-019-0076-3

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College and International Association for the Evaluation of Educational Achievement (IEA). https://timssandpirls.bc.edu/timss2019/index.html

Muthén, B., & Muthén, L. (1998). *Mplus Version 8.6*. Muthén & Muthén.

Muthen, B. O., & Satorra, A. (1995). Complex Sample Data in Structural Equation Modeling. *Sociological Methodology*, *25*, 267. https://doi.org/10.2307/271070

Nilsen, T., & Gustafsson, J.-E. (Eds.). (2016). *Teacher Quality, Instructional Quality and Student Outcomes: Relationships Across Countries, Cohorts and Time* (Vol. 2). Springer International Publishing. https://doi.org/10.1007/978-3-319-41252-8

OECD. (2007). *No More Failures: Ten Steps to Equity in Education*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/education/no-more-failures_9789264032606-en

OECD. (2019). *PISA 2018 Results (Volume II): Where All Students Can Succeed*. OECD. https://doi.org/10.1787/b5fd1b8f-en

Osterman, K. F. (2000). Students' Need for Belonging in the School Community. *Review of Educational Research*, *70*(3), 323–367. https://doi.org/10.3102/00346543070003323

Pahlke, E., Hyde, J. S., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychological Bulletin*, *140*(4), 1042–1072. https://doi.org/10.1037/a0035740

Pahlke, E., Hyde, J. S., & Mertz, J. E. (2013). The effects of single-sex compared with coeducational schooling on mathematics and science achievement: Data from Korea. *Journal of Educational Psychology*, *105*(2), 444–452. https://doi.org/10.1037/a0031857

Park, H., Behrman, J. R., & Choi, J. (2018). Do single-sex schools enhance students' STEM (science, technology, engineering, and mathematics) outcomes? *Economics of Education Review*, *62*, 35–47. https://doi.org/10.1016/j.econedurev.2017.10.007

Penner, A. M. (2008). Gender Differences in Extreme Mathematical Achievement: An International Perspective on Biological and Social Factors. *American Journal of Sociology*, *114*(S1), S138–S170. https://doi.org/10.1086/589252

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Reilly, D., Neumann, D. L., & Andrews, G. (2019). Investigating Gender Differences in Mathematics and Science: Results from the 2011 Trends in Mathematics and Science Survey. *Research in Science Education*, *49*(1), 25–50. https://doi.org/10.1007/s11165-017-9630-6

Renard, A. L. (2008). 'Only for Women:' Women, the State, and Reform in Saudi Arabia. *Middle East Journal*, *62*(4), 610–629.

Riordan, C. (2015). *Single-Sex Schools: A Place to Learn*. Rowman & Littlefield.

Robinson, D. B., Mitton, J., Hadley, G., & Kettley, M. (2021). Single-sex education in the 21st century: A 20-year scoping review of the literature. *Teaching and Teacher Education*, *106*, 103462. https://doi.org/10.1016/j.tate.2021.103462

Rubin, D. B. (Ed.). (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470316696

Rutkowski, L., & Svetina, D. (2017). Measurement Invariance in International Surveys: Categorical Indicators and Fit Measure Performance. *Applied Measurement in Education*, *30*(1), 39–51. https://doi.org/10.1080/08957347.2016.1243540

Scheerens, J., Witziers, B., & Steen, R. (2013). A meta-analysis of school effectiveness studies. *Revista de Educación*, *361*. https://doi.org/10.4438/1988-592X-RE-2013-361-235

Scherer, R., & Nilsen, T. (2016). The Relations Among School Climate, Instructional Quality, and Achievement Motivation in Mathematics. In T. Nilsen & J.-E. Gustafsson (Eds.), *Teacher Quality, Instructional Quality and Student Outcomes: Relationships Across Countries, Cohorts and Time* (Vol. 2). Springer International Publishing. https://doi.org/10.1007/978-3-319-41252-8

Signorella, M. L., Hayes, A. R., & Li, Y. (2013). A Meta-Analytic Critique of Mael et al.'s (2005) Review of Single-Sex Schooling. *Sex Roles*, *69*(7–8), 423–441. https://doi.org/10.1007/s11199-013-0288-x

Snijders, T., & Bosker, R. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd ed.). Sage Publications Ltd.

Spielhofer, T., Benton, T., & Schagen, S. (2004). A study of the effects of school size and single-sex education in English schools. *Research Papers in Education*, *19*(2), 133–159. https://doi.org/10.1080/02671520410001695407

Stewart, E. B. (2008). School Structural Characteristics, Student Effort, Peer Associations, and Parental Involvement: The Influence of School- and Individual-Level Factors on Academic Achievement. *Education and Urban Society*, *40*(2), 179–204. https://doi.org/10.1177/0013124507304167

Stoet, G., & Geary, D. C. (2015). Sex differences in academic achievement are not related to political, economic, or social equality. *Intelligence*, *48*, 137–151. https://doi.org/10.1016/j.intell.2014.11.006

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A Review of School Climate Research. *Review of Educational Research*, *83*(3), 357–385. https://doi.org/10.3102/0034654313483907

Thompson, J. S. (2003). The Effect of Single-Sex Secondary Schooling on Women's Choice of College Major. *Sociological Perspectives*, *46*(2), 257–278. https://doi.org/10.1525/sop.2003.46.2.257

Tsai, S.-L., Smith, M. L., & Hauser, R. M. (2018). Gender Gaps in Student Academic Achievement and Inequality. In H. Park & G. Kao (Eds.), *Research in the Sociology of Education* (Vol. 20, pp. 181–218). Emerald Publishing Limited. https://doi.org/10.1108/S1479-353920180000020008

UNESCO. (2017). *A Guide for Ensuring Inclusion and Equity in Education; 2017*. https://unesdoc.unesco.org/ark:/48223/pf0000248254

UNESCO Institute of Statistics. (2022). *Saudi Arabia*. Education and Literacy. http://uis.unesco.org/en/country/sa

van Geel, A. (2016). Separate or together? Women-only public spaces and participation of Saudi women in the public domain in Saudi Arabia. *Contemporary Islam*, *10*(3), 357–378. https://doi.org/10.1007/s11562-015-0350-2

Wang, M.-T., & Degol, J. L. (2016). School Climate: A Review of the Construct, Measurement, and Impact on Student Outcomes. *Educational Psychology Review*, *28*(2), 315–352. https://doi.org/10.1007/s10648-015-9319-1

Wiseman, A. W. (2008). A Culture of (in)Equality?: A Cross-National Study of Gender Parity and Gender Segregation in National School Systems. *Research in Comparative and International Education*, *3*(2), 179–201. https://doi.org/10.2304/rcie.2008.3.2.179

World Economic Forum. (2021). *Global Gender Gap Report*. World Economic Forum. https://www3.weforum.org/docs/WEF_GGGR_2021.pdf

Yin, L., & Fishbein, B. (2020). Creating and Interpreting the TIMSS 2019 Context Questionnaire Scales. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report*. TIMSS & PIRLS International Study Center.

Appendix I

GDPR Documentation

The present study did not require obtaining any GDPR (General Data Protection Regulation) documentation as no personal data were collected or processed.

NOTIFICATION FORM (ENGLISH TRANSLATION) – NSD

Which personal data will be processed?

N/A. No personal data will be collected or processed.

What are personal data?

Personal data consist of any data relating to an identified or identifiable person. Collected data that can be linked directly or indirectly to individual persons are considered personal data. Answer 'Yes' if you are processing personal data, including if there exists a link between the collected data and personally identifiable information (e.g. name, identity number, contact details etc).

Name (also with signature/written consent)

N/A.

National ID number or other personal identification number

N/A.

Date of birth

N/A.

Address or telephone number

N/A.

Email address, IP address or other online identifier

N/A.

Photographs or video recordings of people

N/A.

Sound recordings of people

N/A.

GPS data or other geolocation data (electronic communications)

N/A.

Background data that can identify a person

N/A.

Genetic data

N/A.

Biometric data

N/A.

Other data that can identify a person

N/A.

Appendix II

Data Management and Analysis Code

The R and Mplus code for selected models are given below for eighth grade data only, since the code used for fourth grade data was identical.

Data Preparation of TIMSS 2019 data using the Edsurvey Package in R

```
*******************************************************
library(tidyverse)
library("EdSurvey")
library(lavaan)
library(psych)
library(ggplot2)
library (MplusAutomation)

rm(list=ls(all =TRUE))

# Reading in the data for Saudi Sample of Grade 8
TIMSS19_SAU8<- readTIMSS(path = "T19_G8_SPSS Data/",
                countries = c("SAU"), gradeLvl = 8,
                forceReread = FALSE, verbose = TRUE)

View(showCodebook(TIMSS19_SAU8))

        ##############  Subsetting student data ##########

TIMSS19_S <- getData(data = TIMSS19_SAU8, varnames = c("idschool","idclass","itsex",
                     "mmat","totwgt","bcdgsbc","bcbg05b","bsbgher","schwgt","
                     bcbgmrs","bsbgssb","bcbgmrs"), omittedLevels = FALSE,
                     addAttributes = TRUE)

summary2(TIMSS19_S, variable = c("itsex"), weightVar = c("totwgt"))
summary2(TIMSS19_S, variable = c("mmat"), weightVar = c("totwgt"))
summary2(TIMSS19_S, variable = c("bsbgher"), weightVar = c("totwgt"))
summary2(TIMSS19_S, variable = c("bsbgssb"))


#Descriptive Stats for Boys' and Girls' Schools
Girls_8 <-  TIMSS19_S [ which(TIMSS19_S$itsex=='FEMALE'),]
summary2(Girls_8, variable = c("mmat"), weightVar = c("totwgt"))

Boys_8 <-  TIMSS19_S [ which(TIMSS19_S$itsex=='MALE'),]
summary2(Boys_8, variable = c("mmat"), weightVar = c("totwgt"))

#Number of dinstinct schools
Schools<- TIMSS19_S %>%
        group_by(idschool,itsex,bcbg05b) %>%
        summarise(n_distinct(idclass))
```

```
colnames(Schools) <- c("idschool", "itsex", "location", "classes")
Schools$classes <- as.factor (Schools$classes)
summary (Schools)

Schools$location<-
        ifelse(Schools$location==c("URBAN-DENSELY POPULATED"), "URBAN",
        ifelse(Schools$location==c("SUBURBAN-ON FRINGE OR OUTSKIRTS OF
        URBAN AREA"), "URBAN",
        ifelse(Schools$location==c("MEDIUM SIZE CITY OR LARGE TOWN"),
        "URBAN",
        ifelse(Schools$location==c("SMALL TOWN OR VILLAGE"), "RURAL",
        ifelse(Schools$location==c("REMOTE RURAL"), "RURAL",
                                        NA)))))
Schools$location <- as.factor(Schools$location)
summary (Schools)

#Mean SES
SSES <- aggregate(bsbgher ~ idschool, data = TIMSS19_S, mean)
colnames(SSES) <- c("idschool", "SchSES")
Schools <- merge(Schools,
            SSES,
            all.x = TRUE)

summary (Schools)
sd(Schools$SchSES)

                    ##############  Subsetting teacher data ##########

TIMSS19_t8  <- getData(data = TIMSS19_SAU8, varnames = c("idschool",
"idclass","idteach","btbg02.math","btbgsos.math",
                                          "btbgeas.math",
                                          "btbglsn.math", "schwgt")
                                          ,omittedLevels = FALSE, addAttributes = TRUE)

#Number of unique teachers in the dataset
Teachers <- TIMSS19_t8 %>%
  group_by(idschool) %>%
  summarise(n_distinct(idteach))

str (Teachers)
colnames(Teachers) <- c("idschool", "teachers")
Teachers$teachers <- as.factor (Teachers$teachers)
summary (Teachers)

#Keeping data for distinct teachers
T_data <- TIMSS19_t8 %>% distinct(idteach, .keep_all = TRUE)

summary2(T_data , variable = c("btbgeas.math"))
summary2(T_data , variable = c("btbgsos.math"))

#School gender vs teacher gender
```

```
T_data <- left_join (T_data, Schools)
( with (T_data, tapply (btbg02.math,itsex,summary)) )

summary (T_data)

#Aggregate score of SOS
tempSOS_Sc <- aggregate(btbgsos.math ~ idschool, data = T_data, mean,na.rm= T)
colnames(tempSOS_Sc) <- c("idschool", "PooledSOS")

T_data <- merge(T_data,
                tempSOS_Sc,
                all.x = TRUE)

#Aggregate score of SEAS
tempSEAS_Sc <- aggregate(btbgeas.math ~ idschool, data = T_data, mean,na.rm= T)

colnames(tempSEAS_Sc) <- c("idschool", "PooledSEAS")

T_data <- merge(T_data,
            tempSEAS_Sc,
            all.x = TRUE)

#Keeping variables of interests
T_data <- T_data[,-(2:7),drop=FALSE]

str (T_data)

#Keeping data for distinct schools
TSchool <- T_data %>% distinct(idschool, .keep_all = TRUE)

#Merging with school data
TS_data <- merge(Schools,
             TSchool,
             all.x = TRUE)
colnames (TS_data)

#Merging teacher data with student data
TS_data <- merge(TS_data ,
             TIMSS19_S,
             all.x = TRUE)
colnames (TS_data)

#Aggregate score of SSB
tempSSB <- aggregate(bsbgssb ~ idschool, data = TS_data, mean,na.rm= T)

colnames(tempSSB) <- c("idschool", "PooledSSB")
str (tempSSB)

TS_data <- merge(TS_data,
             tempSSB,
            all.x = TRUE)
```

```
#Preparing data for Mplus
g8_teach <- subset(as.data.frame(TS_data),
                   select = c(idschool,idclass, itsex, totwgt,schwgt,
                             PooledSOS, PooledSEAS,bcbgmrs,PooledSSB,
                             SchSES,
                             location,
                             bsmmat01, bsmmat02, bsmmat03, bsmmat04, bsmmat05))
str (g8_teach)
summary (g8_teach)

#Prepare Mplus Data with MplusAutomation package
prepareMplusData(list(subset(g8_teach, select = -c(bsmmat02, bsmmat03, bsmmat04,
bsmmat05)),
                   subset(g8_teach, select = -c(bsmmat01, bsmmat03, bsmmat04, bsmmat05)),
                   subset(g8_teach, select = -c(bsmmat01, bsmmat02, bsmmat04, bsmmat05)),
                   subset(g8_teach, select = -c(bsmmat01, bsmmat02, bsmmat03, bsmmat05)),
                   subset(g8_teach, select = -c(bsmmat01, bsmmat02, bsmmat03, bsmmat04))),
                   "g8_teach.dat", imputed = TRUE)

                          ##############  Subsetting principal data ##########

#Instruction Affected by Math Resources
TIMSS19_Sch <-  getData(data = TIMSS19_SAU8, varnames = c("idschool",
                        "schwgt","bcbg05b","bcbgmrs","bcdgsbc"),   omittedLevels = FALSE,
                        addAttributes = TRUE)

str (TIMSS19_Sch)

#Keeping data for distinct schools
TIMSS19_Sch <- TIMSS19_Sch %>%
  distinct(idschool, .keep_all = TRUE)

#School location
Schloc <-summary2(TIMSS19_Sch, variable = c("bcbg05b"))
sum (Schloc[["summary"]][["N"]])

#School SES
ScSES <- summary2(TIMSS19_Sch, variable = c("bcdgsbc"))

summary2(TIMSS19_Sch, variable = c("bcbgmrs"))

#Merging teacher school data with principal data
TSP_data <- merge(TIMSS19_Sch,
                TS_data,
                all.x = TRUE)
```

Data Preparation for Measurement Invariance Testing of Measures using the Edsurvey Package in R

```
***********************************************

library(tidyverse)
library("EdSurvey")
library (MplusAutomation)

# Reading in the data
TIMSS19_SAU8<- readTIMSS(path = "T19_G8_SPSS Data/",
                  countries = c("SAU"), gradeLvl = 8,
                  forceReread = FALSE, verbose = TRUE)

                                ########     SSB scale ########
#Subsetting SSB scale items froms student data
TIMSS19_SS <- getData(data = TIMSS19_SAU8, varnames =
c("idschool","idclass","idstud","itsex",
                                    "totwgt","schwgt",
                                    "bsbg13a","bsbg13b","bsbg13c","bsbg13d","bsbg13e"),
omittedLevels = FALSE, addAttributes = TRUE)

str (TIMSS19_SS)
TIMSS19_SSr <-TIMSS19_SS
colnames (TIMSS19_SSr)
TIMSS19_SSr[6:10] <- ifelse(TIMSS19_SSr[6:10] == "AGREE A LOT", 4,
                    ifelse(TIMSS19_SSr[6:10] == "AGREE A LITTLE", 3,
                        ifelse(TIMSS19_SSr[6:10] == "DISAGREE A LITTLE", 2,
                            ifelse(TIMSS19_SSr[6:10] == "DISAGREE A LOT", 1,
                                NA))))
#Recoding itsex
TIMSS19_SSr$itsex <- ifelse(TIMSS19_SSr$itsex==c("MALE"), 0,
                    ifelse(TIMSS19_SSr$itsex==c("FEMALE"), 1, NA))

#Preparing data for Mplus
g8_SSB <- subset(as.data.frame(TIMSS19_SSr),
              select = c(idschool,idclass,itsex,totwgt,schwgt,
                    bsbg13a, bsbg13b,bsbg13c,bsbg13d,bsbg13e))

str (g8_SSB)

#Prepare Mplus Data with MplusAutomation package
prepareMplusData(g8_SSB, "g8_SSB.dat")

                           ########### SEAS and SOS scales ########

#Subsetting SEAS and SOS scale items from teacher data
TIMSS19_t8  <- getData(data = TIMSS19_SAU8, varnames = c("idschool",
                  "idclass","idteach","btbg06a.math","btbg06b.math","btbg06c.math","btbg06d
                  .math","btbg06e.math","btbg06f.math","btbg06g.math","btbg06h.math","btbg
                  06i.math","btbg06j.math","btbg06k.math","btbg06l.math","btbg07a.math","bt
                  bg07b.math","btbg07c.math","btbg07d.math","btbg07e.math","btbg07f.math"
```

```
                    ,"btbg07g.math","btbg07h.math","matwgt","schwgt","itsex"),omittedLevels =
                    FALSE, addAttributes = TRUE)

#Recoding the SEAS items
TIMSS19_t8r <-TIMSS19_t8
colnames (TIMSS19_t8)
summary2(TIMSS19_t8r, variable = c("btbg07c.math"))

#SEAS items
TIMSS19_t8r[4:15] <- ifelse(TIMSS19_t8r[4:15] == "VERY HIGH", 5,
                      ifelse(TIMSS19_t8r[4:15] == "HIGH", 4,
                           ifelse(TIMSS19_t8r[4:15] == "MEDIUM", 3,
                                ifelse(TIMSS19_t8r[4:15] == "LOW", 2,
                                     ifelse(TIMSS19_t8r[4:15] == "VERY LOW", 1,
                                         NA)))))
#SOS items
TIMSS19_t8r[16:23] <- ifelse(TIMSS19_t8r[16:23] == "AGREE A LOT", 4,
                      ifelse(TIMSS19_t8r[16:23] == "AGREE A LITTLE", 3,
                           ifelse(TIMSS19_t8r[16:23] == "DISAGREE A LITTLE", 2,
                                ifelse(TIMSS19_t8r[16:23] == "DISAGREE A LOT", 1,
                                     NA))))

#Recoding itsex
TIMSS19_t8r$itsex <- ifelse(TIMSS19_t8r$itsex==c("MALE"), 0,
                      ifelse(TIMSS19_t8r$itsex==c("FEMALE"), 1, NA))

TIMSS19_t8r <- TIMSS19_t8r %>% distinct(idteach, .keep_all = TRUE)

#Preparing data for Mplus
g8_SEAS_SOS2 <- subset(as.data.frame(TIMSS19_t8r),
             select = c(idschool,idclass,itsex,matwgt,schwgt,
                      btbg06a.math,btbg06b.math,btbg06c.math,btbg06d.math,btbg06e.math,
                      btbg06f.math,btbg06g.math,btbg06h.math,btbg06i.math,btbg06j.math,
                      btbg06k.math,btbg06l.math,
                      btbg07a.math,btbg07b.math,btbg07c.math,btbg07d.math,btbg07e.math,btbg
                      07f.math,btbg07g.math,btbg07h.math,idteach))

#Prepare Mplus Data with MplusAutomation package
prepareMplusData(g8_SEAS_SOS2, "g8_SEAS_SOS2.dat")

                    ###################SHORTAGE scale ################

#Subsetting Shortage scale items from the principal data
TIMSS19_p8  <- getData(data = TIMSS19_SAU8, varnames = c("idschool",
"itsex","schwgt","bcbg13aa","bcbg13ab","bcbg13ac","bcbg13ad","bcbg13ae","bcbg13af","bc
bg13ag","bcbg13ah", "bcbg13ba","bcbg13bb","bcbg13bc","bcbg13bd","bcbg13be")
,omittedLevels = FALSE, addAttributes = TRUE)

TIMSS19_p8r <-TIMSS19_p8
colnames (TIMSS19_p8)
summary2(TIMSS19_p8r, variable = c("bcbg13aa"))
```

```
# Recoding Shortage items
TIMSS19_p8r[4:16] <- ifelse(TIMSS19_p8r[4:16] == "NOT AT ALL", 4,
                         ifelse(TIMSS19_p8r[4:16] == " A LITTLE", 3,
                            ifelse(TIMSS19_p8r[4:16] == "SOME", 2,
                               ifelse(TIMSS19_p8r[4:16] == "A LOT", 1,
                                  NA))))
#Recoding itsex
TIMSS19_p8r$itsex <- ifelse(TIMSS19_p8r$itsex==c("MALE"), 0,
                   ifelse(TIMSS19_p8r$itsex==c("FEMALE"), 1, NA))

TIMSS19_p8r <- TIMSS19_p8r %>% distinct(idschool, .keep_all = TRUE)

#Preparing data for Mplus
g8_Shortage <- subset(as.data.frame(TIMSS19_p8r),  select = c(idschool,itsex,schwgt,
              bcbg13aa,bcbg13ab,bcbg13ac,bcbg13ad,bcbg13ae,bcbg13af,bcbg13ag,bcbg13a
              h, bcbg13ba,bcbg13bb,bcbg13bc,bcbg13bd,bcbg13be))
str (g8_Shortage)

#Prepare Mplus Data with MplusAutomation package
prepareMplusData(g8_Shortage, "g8_Shortage.dat")
```

Mplus Code for selected models

****

```
TITLE:
   Master Thesis Sadaf;  !SEAS model

DATA: FILE = "g8_teach.dat";
   TYPE = IMPUTATION;
VARIABLE:
   NAMES = idschool idclass itsex totwgt schwgt SOS SEAS bcbgmrs SchSES
   location bsmmat;
   USEVAR = SEAS itsex;
   CLUSTER= idschool;
   MISSING=.;
   WEIGHT= schwgt;

ANALYSIS:
   TYPE = COMPLEX;
MODEL:
   SEAS on itsex;
OUTPUT:
   sampstat standardized;
***
TITLE:
   Master Thesis Sadaf; !SEAS model with controls
DATA: FILE = "g8_teach.dat";
   TYPE = IMPUTATION;
VARIABLE:
   NAMES =  idschool idclass itsex totwgt schwgt SOS SEAS bcbgmrs SchSES
```

```
    location bsmmat;
   USEVAR = SEAS itsex SchSES location;
   CLUSTER= idschool;
   MISSING=.;
   WEIGHT= schwgt;
ANALYSIS:
   TYPE = COMPLEX;
MODEL:
   SEAS on itsex SchSES location;
OUTPUT:
   sampstat standardized;

***
TITLE:
   Master Thesis Sadaf; !SEAS and achievement model
DATA: FILE = "g8_teach.dat";
   TYPE = IMPUTATION;
VARIABLE:
   NAMES = idschool idclass itsex totwgt schwgt SOS SEAS bcbgmrs SchSES
    location bsmmat;
   USEVAR = SEAS bsmmat;
   BETWEEN= SEAS;
   CLUSTER= idschool;
   MISSING=.;
   WEIGHT= totwgt;

DEFINE:
 CENTER SEAS(GRANDMEAN);

ANALYSIS:
   TYPE = TWOLEVEL;
MODEL:
   %between%
   bsmmat on SEAS;
OUTPUT:
   sampstat standardized;

***
TITLE:
   Master Thesis Sadaf;  !SEAS and achievement model with controls
DATA: FILE = "g8_teach.dat";
   TYPE = IMPUTATION;
VARIABLE:
   NAMES =  idschool idclass itsex totwgt schwgt SOS SEAS bcbgmrs SchSES
    location bsmmat;
   USEVAR = SEAS itsex SchSES location;
   CLUSTER= idschool;
   MISSING=.;
   WEIGHT= schwgt;

ANALYSIS:
```

```
   TYPE = COMPLEX;
MODEL:
   SEAS on itsex SchSES location;
OUTPUT:
   sampstat standardized;
***
TITLE:
   Master Thesis Sadaf;       !Model with all school climate dimensions
DATA: FILE = "g8_teach.dat";
   TYPE = IMPUTATION;
VARIABLE:
   NAMES = idschool idclass itsex totwgt schwgt SOS SEAS bcbgmrs SSB
    SchSES location bsmmat;
   USEVAR = SOS SEAS bcbgmrs SSB bsmmat;
   BETWEEN= SOS SEAS bcbgmrs SSB;
   CLUSTER= idschool;
   MISSING=.;
   WEIGHT= totwgt;

DEFINE:
 CENTER SOS(GRANDMEAN);
  CENTER SEAS(GRANDMEAN);
   CENTER bcbgmrs(GRANDMEAN);
   CENTER SSB(GRANDMEAN);

ANALYSIS:
   TYPE = TWOLEVEL;
MODEL:

   %between%
   bsmmat on SEAS SOS bcbgmrs SSB;
OUTPUT:
   sampstat standardized;
***
TITLE:
   Master Thesis Sadaf;       !Model with all school climate dimensions and controls
DATA: FILE = "g8_teach.dat";
   TYPE = IMPUTATION;
VARIABLE:
   NAMES = idschool idclass itsex totwgt schwgt SOS SEAS bcbgmrs SSB
    SchSES location bsmmat;
   USEVAR = SOS SEAS bcbgmrs SSB bsmmat SchSES location;
   BETWEEN= SOS SEAS bcbgmrs SSB SchSES location;
   CLUSTER= idschool;
   MISSING=.;
   WEIGHT= totwgt;

DEFINE:
 CENTER SOS(GRANDMEAN);
  CENTER SEAS(GRANDMEAN);
   CENTER bcbgmrs(GRANDMEAN);
```

```
   CENTER SSB(GRANDMEAN);

ANALYSIS:
   TYPE = TWOLEVEL;
MODEL:

   %between%
   bsmmat on SEAS SOS bcbgmrs SSB SchSES location;
OUTPUT:
   sampstat standardized;
***

TITLE:
   Master Thesis Sadaf           !Measurement invariance testing for the SSB scale
   Two-level CFA model
   Level 1 students: saturated model
   Level 2 classrooms: not an explicit level
   Level 3 schools: factor model
   Configural invariance across gender
DATA:
    FILE = g8_SSB.dat;

VARIABLE:
   NAMES =  idschool idclass itsex totwgt schwgt bsbg13a bsbg13b bsbg13c bsbg13d
    bsbg13e;   !itsex is student gender (0= Male, 1= Female)
   USEVAR =   bsbg13a bsbg13b bsbg13c bsbg13d bsbg13e;
                          !Items on the Student Sense of Belonging Scale
   MISSING=.;
   WEIGHT = totwgt;

   CLUSTER = idschool;

   GROUPING IS itsex (0=Male 1=Female);           ! Gender as a grouping variable

ANALYSIS:
   TYPE = TWOLEVEL;
   ESTIMATOR = MLR;
   PROCESSORS = 4;                   ! To speed things up
   H1ITERATIONS = 10000;! To make convergence reasonable

MODEL:
         %WITHIN%
         ! Student level

         ! Saturated model (only variances and covariances)
         bsbg13a-bsbg13e WITH bsbg13a-bsbg13e;

         %BETWEEN%
         ! School level

         ! Factor model
```

```
        SSB by bsbg13a*
                bsbg13b
                bsbg13c
                bsbg13d
                bsbg13e;

        SSB@1;          ! To set the scale of the latent variable

        ! Intercepts of items exist only at this level
        ! they are freely estimated
        [bsbg13a-bsbg13e*];

        ! Factor mean
        ! fixed to zero to identify the mean structure
        ! because intercepts are freely estimated
        [SSB@0];


MODEL Female:
        ! Specify if the group of female students has any different
        ! or equal parameters in the factor model at the school level

        %BETWEEN%
        ! School level

        ! Factor model
        SSB by          bsbg13a*
                bsbg13b
      bsbg13c
      bsbg13d
      bsbg13e;

        SSB@1;          ! To set the scale of the latent variable

        ! Intercepts of items exist only at this level
        ! they are freely estimated
        [bsbg13a-bsbg13e*];

OUTPUT:
   sampstat;
   standardized;
***
TITLE:
   Master Thesis Sadaf;        ! Measurement invariance testing for SEAS scale

DATA:
   FILE = g8_SEAS_SOS2.dat;

VARIABLE:
NAMES = idschool idclass itsex matwgt schwgt btbg06a btbg06b btbg06c
    btbg06d btbg06e btbg06f btbg06g btbg06h btbg06i
```

```
    btbg06j btbg06k btbg06l btbg07a btbg07b btbg07c
    btbg07d btbg07e btbg07f btbg07g btbg07h idteach;

    USEVAR =  btbg06a btbg06b btbg06c btbg06d btbg06e btbg06f btbg06g
    btbg06h btbg06i btbg06j btbg06k btbg06l;

   MISSING=.;
   WEIGHT = schwgt;

  GROUPING = itsex (0=Male 1=Female);

ANALYSIS:

   ESTIMATOR = MLR;
   MODEL= configural metric scalar;

MODEL:

  SEAS by btbg06a btbg06b btbg06c btbg06d btbg06e btbg06f btbg06g
    btbg06h btbg06i btbg06j btbg06k btbg06l;

OUTPUT:
   sampstat;
   standardized;
```

Running robustness checks for multilevel regression models for selected models

****

```
library(lme4)
library ("HLMdiag")

rm(list=ls(all =TRUE))

#Boxplots for the school climate dimensions for girls' and boys' schools
p1<-ggplot(TS_data, aes(x=as.factor(itsex), y=PooledSEAS)) +
  geom_boxplot(fill="slateblue", alpha=0.2) +
  labs( x ="School Type", y = "School SEAS")+
  scale_x_discrete(labels = c('Girls Schools','Boys Schools'))+
  labs(title = "                School SEAS") +
  ylim(4,16)


p2<-ggplot(TS_data, aes(x=as.factor(itsex), y=PooledSOS)) +
  geom_boxplot(fill="slateblue", alpha=0.2) +
  labs( x ="School Type", y = "School SOS")+
  scale_x_discrete(labels = c('Girls Schools','Boys Schools'))+
  labs(title = "                School SOS") +
  ylim(4,16)


p3<- ggplot(TS_data, aes(x=as.factor(itsex), y=bcbgmrs)) +
```

```
  geom_boxplot(fill="slateblue", alpha=0.2) +
  labs( x ="School Type", y = "School Shortage")+
  scale_x_discrete(labels = c('Girls Schools','Boys Schools'))+
  labs(title = "                    School SHORTAGE") +
  ylim(4,16)

p4<- ggplot(TS_data, aes(x=as.factor(itsex), y=PooledSSB)) +
  geom_boxplot(fill="slateblue", alpha=0.2) +
  labs( x ="School Type", y = "School SSB")+
  scale_x_discrete(labels = c('Girls Schools','Boys Schools'))+
  labs(title = "                    School SSB") +
  ylim(4,16)

grid.arrange(p1,p2,p3,p4)

                                ############  SEAS scale  ##########

mlm1 <- lmer(bsmmat01 ~ 1 + PooledSEAS + (1 | idschool),
             data = TS_data,
             REML = FALSE)

summary (mlm1)

#Level 1 residuals (rij) are independent and normally distributed.
#This can be tested with a histogram or normality test.
#Q-Q Plot of the L1 residuals
qqnorm(resid(mlm1), main="Plot 1
     Normal Q-Q Plot of L1 Residuals")
qqline(resid(mlm1), col = "red")

resid1_fm1 <- hlm_resid(mlm1, level = 1, standardize = TRUE)

#With standardized residuals
ssresids <- (resid1_fm1$.std.ls.resid)
qqnorm(ssresids, ylab="Standardized Residuals", main="Plot 1
     Normal Q-Q Plot of L1 Residuals")
qqline(ssresids, col = "red")


#Level-2 residuals
# Q-Q Plot of the random effect of the intercept (Standardized residuals)
level2 <- hlm_resid(mlm1, level = "idschool", standardize = TRUE)

qqnorm(level2$.std.ranef.intercept,ylab="Standardized Residuals", main="Plot 2
Normal Q-Q Plot of L2 Residuals
(Random Effect of the Intercept)")
qqline(level2$.std.ranef.intercept, col = "red")


#Level 1 predictors (Xij) are independent of Level 1 residuals (rij).
#Do a scatter plot between the two or test the correlation.
```

```
qplot(x = bsmmat01, y = .ls.resid, data = resid1_mlm1,
      geom = c("point", "smooth")) +
  ylab("LS Level-1 residuals")+
  labs(title = "                    Plot 3
L1 residuals against bsmmat01 values")

qplot(x = PooledSEAS, y = .ls.resid, data = resid1_mlm1,
      geom = c("point", "smooth")) +
  ylab("LS Level-1 residuals")+
  xlab ("SEAS")+
  labs(title = "                    Plot 4
L1 residuals against SEAS values")

#We check the LS level-1 residuals for fm1, plotting them against the LS fitted values
ggplot(data = resid1_mlm1, aes(x = .ls.fitted, y = .ls.resid)) +
  geom_point(alpha = 0.2) +
  geom_smooth(method = "loess", se = FALSE) +
  labs(x = "LS fitted values",
       y = "LS level-1 residuals", title = "                    Plot 5
LS residuals against LS fitted values")

#Identify influential observations at student level
infl <- hlm_influence(mlm1, level = 1)

tb1 <- infl %>%
  arrange(desc(cooksd))

dotplot_diag(infl$cooksd, name = "cooks.distance", cutoff = "internal")

#Identify influential groups at school level
infl.schools <- hlm_influence(mlm1, level = "idschool")

infl.schools %>%
  arrange(desc(cooksd))

dotplot_diag(infl.schools$cooksd, name = "cooks.distance", cutoff = "internal",modify =
FALSE)

hlm_influence(mlm1, level = "idschool", delete = c("5011", "5023", "5180","5096", "5125"))

#Removing influential observations
TS_SEAS <- subset(TS_data, idschool!="5011" & idschool!="5023" & idschool!="5180"
&idschool!="5096" & idschool!="5125")
str (TS_SEAS)
```

Appendix III

Appendix A: Gender Achievement Gaps in Mathematics and Percentage of Single-Sex Schools in TIMSS 2019

Appendix A.1

Percentage of single-sex schools in Grade 4 across TIMSS 2019 participating countries excluding benchmarking participants

Country	Total number of schools	Number of Single Sex Schools	Percentage of Single-Sex Schools
1. Albania	167	0	0
2. Armenia	150	2	0.01
3. Australia	287	5	0.02
4. Austria	193	3	0.02
5. Azerbaijan	194	1	0.01
6. Bahrain	185	132	0.71
7. Belgium (Flemish)	147	1	0.01
8. Bosnia and Herzegovina	178	2	0.01
9. Bulgaria	151	1	0.01
10. Canada	704	8	0.01
11. Chile	169	12	0.07
12. Chinese Taipei	162	0	0
13. Croatia	153	1	0.01
14. Cyprus	151	2	0.01
15. Czech Republic	152	1	0.01
16. Denmark	166	1	0.01
17. England	139	3	0.02
18. Finland	158	4	0.03
19. France	155	2	0.01
20. Georgia	154	6	0.04
21. Germany	203	3	0.01
22. Hong Kong SAR	139	9	0.06
23. Hungary	149	1	0.01
24. Iran. Islamic Rep. of	224	202	0.9
25. Ireland	150	34	0.23

26. Italy	162	1	0.01
27. Japan	147	1	0.01
28. Kazakhstan	168	1	0.01
29. Korea. Rep. of	151	2	0.01
30. Kosovo	145	1	0.01
31. Kuwait	164	135	0.82
32. Latvia	154	2	0.01
33. Lithuania	207	3	0.01
34. Macedonia. Rep. of	150	2	0.01
35. Malta	98	18	0.18
36. Montenegro	140	8	0.06
37. Morocco	264	0	0
38. Netherlands	112	0	0
39. New Zealand	160	3	0.02
40. Northern Ireland	134	1	0.01
41. Norway	150	0	0
42. Oman	228	3	0.01
43. Pakistan	139	78	0.56
44. Philippines	180	0	0
45. Poland	149	0	0
46. Portugal	181	0	0
47. Qatar	242	126	0.52
48. Russian Federation	200	4	0.02
49. Saudi Arabia	220	219	1
50. Serbia	165	3	0.02
51. Singapore	187	25	0.13
52. Slovak Republic	157	1	0.01
53. South Africa	297	4	0.01
54. Spain	501	10	0.02
55. Sweden	145	0	0
56. Turkey	180	16	0.09
57. United Arab Emirates	688	200	0.29
58. United States	287	3	0.01

Appendix A.2

Absolute scores and effect size d of gender differences in mathematics achievement for Grade 4 students

Country	Gender difference in achievement score (Absolute value)	Standard error of gender difference in achievement score	p-value	Effect size *(d)*
1. Albania	-2.70	3.66	0.46	-0.03
2. Armenia	2.44	3.19	0.45	0.04
3. Australia	-9.96	2.82	0.00	-0.11
4. Austria	-7.55	2.86	0.01	-0.12
5. Azerbaijan	3.71	2.97	0.21	0.04
6. Bahrain	4.84	4.25	0.26	0.06
7. Belgium (Flemish)	-10.84	3.18	0.00	-0.16
8. Bosnia and Herzegovina	-8.76	2.87	0.00	-0.12
9. Bulgaria	-2.18	3.67	0.55	-0.03
10. Canada	-18.87	2.29	0.00	-0.25
11. Chile	-8.54	3.67	0.02	-0.11
12. Chinese Taipei	-3.84	2.74	0.16	-0.06
13. Croatia	-11.82	3.09	0.00	-0.18
14. Cyprus	-18.70	3.33	0.00	-0.24
15. Czech Republic	-10.86	3.04	0.00	-0.15
16. Denmark	-6.85	2.92	0.02	-0.09
17. England	-7.34	4.03	0.07	-0.09
18. Finland	-2.61	3.13	0.41	-0.03
19. France	-13.56	3.08	0.00	-0.17
20. Georgia	-7.50	3.21	0.02	-0.09
21. Germany	-10.49	2.77	0.00	-0.15
22. Hong Kong SAR	-5.81	3.25	0.08	-0.08
23. Hungary	-11.19	2.85	0.00	-0.14
24. Iran. Islamic Rep. of	-7.41	8.58	0.39	-0.08
25. Ireland	-6.83	3.84	0.08	-0.09
26. Italy	-12.29	3.35	0.00	-0.19
27. Japan	0.75	2.25	0.74	0.01

28. Kazakhstan	0.42	2.49	0.87	0.01
29. Korea. Rep. of	-5.03	2.61	0.06	-0.07
30. Kosovo	-4.52	3.63	0.22	-0.06
31. Kuwait	6.54	8.95	0.47	0.06
32. Latvia	-4.61	2.53	0.07	-0.07
33. Lithuania	-4.58	3.82	0.23	-0.06
34. Macedonia. Rep. of	0.29	3.88	0.94	0.00
35. Malta	-7.45	2.67	0.01	-0.10
36. Montenegro	-4.95	3.10	0.11	-0.06
37. Morocco	3.49	3.13	0.27	0.03
38. Netherlands	-8.51	2.91	0.00	-0.14
39. New Zealand	-5.41	4.65	0.26	-0.06
40. Northern Ireland	-3.47	4.20	0.41	-0.04
41. Norway	-4.40	3.36	0.19	-0.06
42. Oman	13.97	2.91	0.00	0.14
43. Pakistan	18.88	15.67	0.24	0.18
44. Philippines	34.90	3.53	0.00	0.32
45. Poland	-8.10	2.81	0.00	-0.11
46. Portugal	-17.32	2.51	0.00	-0.23
47. Qatar	1.11	5.25	0.83	0.01
48. Russian Federation	-8.17	2.61	0.00	-0.12
49. Saudi Arabia	26.48	8.13	0.00	0.26
50. Serbia	1.70	3.90	0.66	0.02
51. Singapore	-7.91	2.83	0.01	-0.10
52. Slovak Republic	-12.47	3.66	0.00	-0.16
53. South Africa	20.26	2.98	0.00	0.20
54. Spain	-14.57	2.91	0.00	-0.20
55. Sweden	-7.11	2.67	0.01	-0.10
56. Turkey	-3.49	5.02	0.49	-0.03
57. United Arab Emirates	-8.17	3.37	0.02	-0.08
58. United States	-11.13	2.82	0.00	-0.13

Appendix A.3

Percentage of single-sex schools in Grade 8 across TIMSS 2019 participating countries excluding benchmarking participants

Country	Total number of schools	Number of Single Sex Schools	Percentage of Single-Sex Schools
1. Australia	284	38	0.13
2. Bahrain	112	68	0.61
3. Chile	164	12	0.07
4. Chinese Taipei	203	7	0.03
5. Cyprus	98	0	0
6. Egypt	169	112	0.66
7. England	136	17	0.12
8. Finland	154	0	0
9. France	150	0	0
10. Georgia	145	4	0.03
11. Hong Kong	136	25	0.18
12. Hungary	154	0	0
13. Iran	220	211	0.96
14. Ireland	149	57	0.38
15. Israel	157	29	0.18
16. Italy	158	0	0
17. Japan	142	4	0.03
18. Jordan	235	227	0.97
19. Kazakhstan	168	7	0.04
20. Korea. Rep. of	168	70	0.42
21. Kuwait	171	154	0.9
22. Lebanon	204	17	0.08
23. Lithuania	194	2	0.01
24. Malaysia	177	17	0.1
25. Morocco	251	0	0
26. New Zealand	134	42	0.31
27. Norway	157	0	0
28. Oman	228	193	0.85
29. Portugal	156	1	0.01

30. Qatar	152	91	0.6
31. Romania	198	0	0
32. Russia	204	2	0.01
33. Saudi Arabia	209	209	1
34. Singapore	153	27	0.18
35. South Africa	519	13	0.03
36. Sweden	150	1	0.01
37. Turkey	181	18	0.1
38. United Arab Emirates	623	294	0.47
39. United States	273	5	0.02

Appendix A.4

Absolute scores and effect size d of gender differences in mathematics achievement for Grade 8 students

Country	Gender difference in achievement score (Absolute value)	Standard error of gender difference in achievement score	p-value	Effect size *(d)*
1. Australia	-4.28	5.44	0.43	-0.05
2. Bahrain	21.34	3.02	3.55e-11	0.23
3. Chile	-8.94	4.51	0.059	-0.12
4. Chinese Taipei	2.13	3.26	0.51	0.02
5. Cyprus	3.98	3.06	0.19	0.05
6. Egypt	15.57	8.01	0.05	0.16
7. England	-1.63	7.20	0.82	-0.02
8. Finland	4.06	2.76	0.14	0.06
9. France	-8.12	2.59	0.00	-0.12
10. Georgia	-8.44	4.18	0.04	-0.10
11. Hong Kong	7.26	6.49	0.27	0.08
12. Hungary	-14.08	3.62	0.00	-0.16
13. Iran	13.22	7.47	0.08	0.14
14. Ireland	1.34	3.41	0.69	0.02
15. Israel	-10.66	4.57	0.02	-0.11

16. Italy	-12.14	3.00	7.97e-05	-0.17
17. Japan	-2.28	2.85	0.42	-0.03
18. Jordan	23.15	6.51	0.00	0.27
19. Kazakhstan	3.91	3.92	0.32	0.05
20. Korea. Rep. of	-5.11	3.39	0.14	-0.05
21. Kuwait	8.63	8.52	0.31	0.10
22. Lebanon	-5.22	3.60	0.15	-0.07
23. Lithuania	-2.00	2.82	0.47	-0.02
24. Malaysia	8.55	3.47	0.016	0.09
25. Morocco	-4.91	2.30	0.03	-0.07
26. New Zealand	-5.99	5.05	0.24	-0.07
27. Norway	-0.18	3.12	0.95	0.00
28. Oman	40.71	4.68	1.33e-14	0.42
29. Portugal	-10.33	3.41	0.00	-0.14
30. Qatar	6.82	6.62	0.30	0.07
31. Romania	15.88	3.47	9.03e-06	0.16
32. Russia	-4.93	3.23	0.13	-0.06
33. Saudi Arabia	17.27	4.35	0.00	0.22
34. Singapore	3.37	4.17	0.42	0.04
35. South Africa	6.49	2.12	0.00	0.08
36. Sweden	2.81	3.08	0.36	0.04
37. Turkey	11.05	5.72	0.06	0.10
38. United Arab Emirates	5.71	5.55	0.31	0.06
39. United States	3.55	3.74	0.36	0.04

Appendix B: Item wording for TIMSS 2019 Scales used in the Study

School Emphasis on Academic Success (SEAS Scale)
Teachers' Reports – Grade 4 and Grade 8

How would you characterize each of the following within your school?
Response categories: 5=*Very High, 4= High, 3= Medium, 2= Low, 1=Very Low*

1. Teachers' understanding of the school's curricular goals
2. Teachers' degree of success in implementing the school's curriculum
3. Teachers' expectations for student achievement
4. Teachers' ability to inspire students
5. Parental involvement in school activities
6. Parental commitment to ensure that students are ready to learn
7. Parental expectations for student achievement
8. Parental support for student achievement
9. Students' desire to do well in school
10. Students' ability to reach school's academic goals
11. Students' respect for classmates who excel academically
12. Collaboration between school leadership (including master teachers) and teachers to plan instruction

Safe and Orderly School (SOS Scale)
Teachers' Reports – Grade 4 and Grade 8

Thinking about your current school, indicate the extent to which you agree or disagree with each of the following statements.
Response categories: 4= *Agree a lot, 3=Agree a little, 2= Disagree a little, 1= Disagree a lot*

1. This school is located in a safe neighborhood
2. I feel safe at this school
3. This school's security policies and practices are sufficient
4. The students behave in an orderly manner
5. The students are respectful of the teachers
6. The students respect school property
7. This school has clear rules about student conduct
8. This school's rules are enforced in a fair and consistent manner

Instruction Affected by Mathematics Resource Shortages (SHORTAGE Scale)
Principals' Reports – Grade 4 and Grade 8

How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following:
Response categories: 4= *Not at all, 3= A little, 2= Some, 1= A lot*

A. General School Resources
1. Instructional materials (e.g., textbooks)
2. Supplies (e.g., papers, pencils, materials)
3. School buildings and grounds
4. Heating/cooling and lighting systems
5. Library resources relevant to mathematics
6. Technologically competent staff
7. Audio-visual resources for delivery of instruction (e.g., interactive white boards, digital projectors)
8. Computer technology for teaching and learning (e.g., computers or tablets for student use)

B. Resources for Mathematics Instruction
1. Teachers with a specialization in mathematics
2. Computer software/applications for mathematics instruction
3. Library resources relevant to mathematics instruction
4. Calculators for mathematics instruction
5. Concrete objects or materials to help students understand quantities or procedures

Sense of School Belonging (SSB Scale)
Students' Reports – Grade 4 and Grade 8

What do you think about your school? Tell how much you agree with these statements.
Response categories: 4= *Agree a lot, 3=Agree a little, 2= Disagree a little, 1= Disagree a lot*

1. I like being in school
2. I feel safe when I am at school
3. I feel like I belong at this school
4. Teachers at my school are fair to me
5. I am proud to go to this school

Appendix C: Results of Measurement Invariance Testing

Appendix C.1

Results of Measurement Invariance Testing of Measures for Grade 4

Invariance model	Chi-square	df	CFI	TLI	RMSEA	SRMR	SRMRw	SRMRb	ΔCFA	ΔTLI	ΔRMSEA	ΔSRMR(b)
School Emphasis on Academic Success (12 items)												
Configural	437.431*	108	0.767	0.715	0.120	0.099	-	-	-	-	-	-
Metric	446.118*	119	0.769	0.743	0.114	0.123	-	-	0.002	0.028	-0.006	0.024
Scalar	471.279*	130	0.759	0.755	0.112	0.112	-	-	-0.008	0.04	-0.008	0.013
Safe and Orderly Schools (8 items)												
Configural	537.632*	40	0.348	0.087	0.250	0.135	-	-	-	-	-	-
Metric	340.507*	47	0.615	0.542	0.177	0.221	-	-	0.267	0.455	-0.073	0.086
Scalar	354.00*	54	0.607	0.592	0.167	0.218	-	-	-0.008	0.505	-0.083	0.083
Students' Mathematics Achievement on Instruction Affected by Mathematics Resource Shortages (13 items)												
Configural	565.473*	130	0.54	0.449	0.176	0.199	-	-	-	-	-	-
Metric	580.180*	142	0.538	0.492	0.169	0.211	-	-	-0.002	0.043	-0.007	0.012
Scalar	586.602*	154	0.543	0.538	0.161	0.215	-	-	0.003	0.089	-0.015	0.016
Students' Sense of School Belonging (5 items)												
Configural	12.943*	10	0.999	0.995	0.011	-	0001	005	-	-	-	-
Metric	28.239*	15	0.994	0.985	0.018	-	0002	0164	-0005	-001	0007	0114
Scalar	40.357*	19	0.991	0.981	0.021	-	0002	0198	-0008	-0014	001	0148

Note. For model fit comparisons, the configural models served as references. df = degrees of freedom. SRMRw: within level SRMR. SRMRb = between level SRMR. *p < 0.05

Appendix C.2

Results of Measurement Invariance Testing of Measures for Grade 8

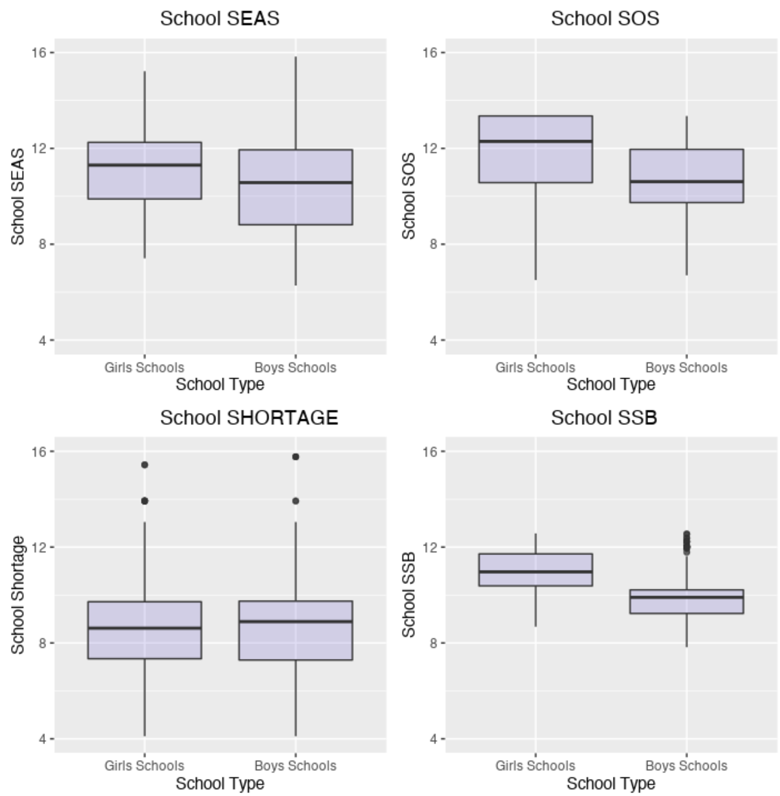
Invariance model	Chi-square	df	CFI	TLI	RMSEA	SRMR	SRMRw	SRMRb	ΔCFA	ΔTLI	ΔRMSEA	ΔSRMR(b)
School Emphasis on Academic Success (12 items)												
Configural	*The model could not be identified*											
Metric	308.802*	119	0.765	0.74	0.124	0.154	-	-	*NA*	*NA*	*NA*	*NA*
Scalar	341.711*	130	0.738	0.734	0.125	0.143	-	-	*NA*	*NA*	*NA*	*NA*
Safe and Orderly Schools (8 items)												
Configural	179.140*	40	0.569	0.396	0.185	0.133	-	-	-	-	-	-
Metric	141.743*	47	0.706	0.65	0.141	0.144	-	-	0.137	0.254	-0.044	0.011
Scalar	146.237*	54	0.714	0.704	0.13	0.13	-	-	0.145	0.308	-0.055	-0.003
Students' Mathematics Achievement on Instruction Affected by Mathematics Resource Shortages (13 items)												
Configural	738.569*	130	0.455	0.347	0.213	0.234	-	-	-	-	-	-
Metric	743.188*	142	0.462	0.409	0.202	0.232	-	-	0.007	0.062	-0.011	-0.002
Scalar	728.670*	154	0.486	0.479	0.19	0.235	-	-	0.031	0.132	-0.023	0.001
Students' Sense of School Belonging (5 items)												
Configural	26.623*	10	0.997	0.987	0.024		0.001	0.025	-	-	-	-
Metric	24.722*	15	0.998	0.995	0.015		0.001	0.055	0.001	0.008	-0.009	0.03
Scalar	52.373*	19	0.994	0.986	0.025		0.001	0.059	-0.003	-0.001	0.001	0.034

Note. For model fit comparisons, the configural models served as references. df = degrees of freedom. SRMRw: within level SRMR. SRMRb = between level SRMR. *p < 0.05

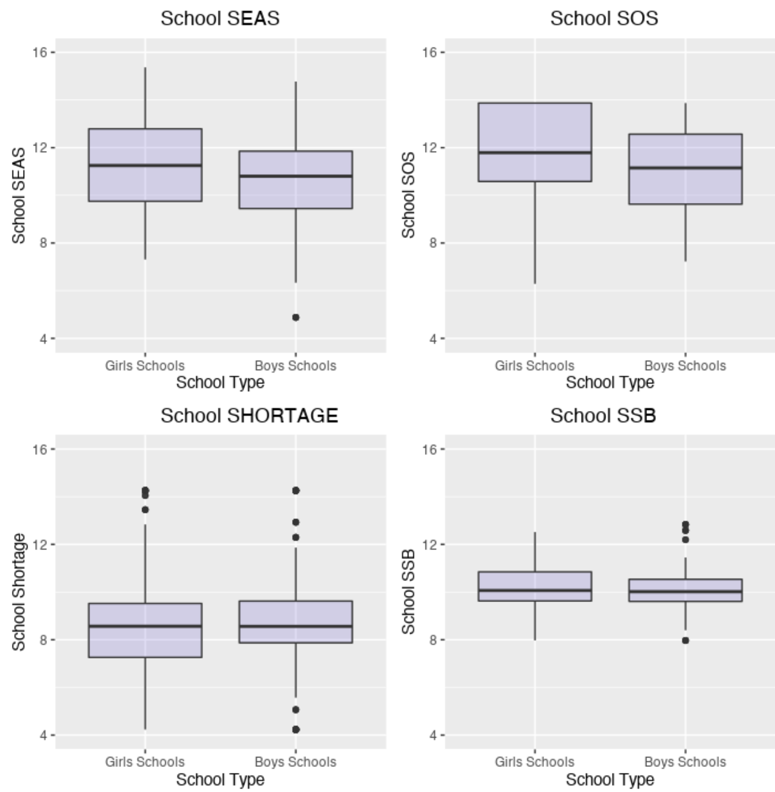
Appendix D: Selected Descriptives for Grade 4 and 8

Appendix D.1

Box and Whisker plot for School Climate Dimensions for Girls' and Boys' Schools (Grade 4)



Box and Whisker plot for School Climate Dimensions for Girls' and Boys' Schools (Grade 8)



Note: SEAS: School Emphasis on Academic Success, SOS: Safe and Orderly Schools, SHORTAGE: Instruction Affected by Mathematics Resource Shortages; SSB: Students' Sense of School Belonging.

Appendix D.2

School Level Bivariate Correlations between Key Study Variables for Grade 4

	Girls' Schools (*N* = 111)					**Boys' Schools (*N* = 108)**					**All Schools (*N* = 219)**				
	MATH	SEAS	SOS	SHORTAGE	SSB	MATH	SEAS	SOS	SHORTAGE	SSB	MATH	SEAS	SOS	SHORTAGE	SSB
MATH	1					1					1				
SEAS	0.139	1				0.236*	1				0.224*	1			
SOS	-0.082	0.517*	1			0.392*	0.756*	1			0.247*	0.645*	1		
SHORTAGE	-0.083	-0.017	0.17	1		-0.090	-0.108	0.061	1		-0.077	0.050	0.268*	1	
SSB	0.043	0.024	0.087	0.037	1	0.003	0.126	0.072	0.051	1	0.147*	0.081	0.082	0.054	1

Note. Correlations based on original scale scores aggregated at the school level. MATH: Student mathematics' achievement, SEAS: School Emphasis on Academic Success, SOS: Safe and Orderly Schools, SHORTAGE: Instruction Affected by Mathematics Resource Shortages; SSB: Students' Sense of School Belonging, SES: School socio-economic status. * $p < .050$

Appendix D.3

School Level Bivariate Correlations between Key Study Variables for Grade 8

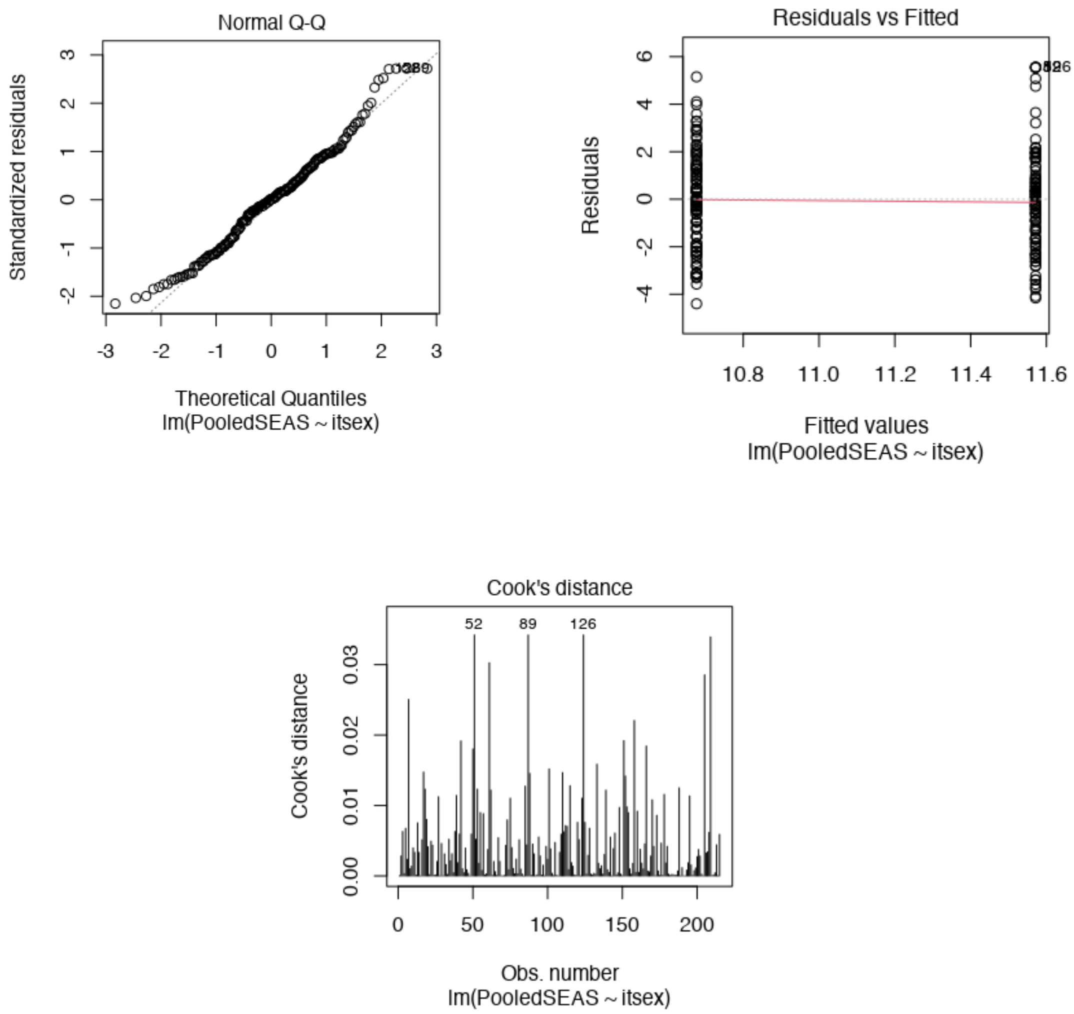
	Girls' Schools (*N* = 111)					**Boys' Schools (*N* = 108)**					**All Schools (*N* = 219)**				
	MATH	SEAS	SOS	SHORTAGE	SSB	MATH	SEAS	SOS	SHORTAGE	SSB	MATH	SEAS	SOS	SHORTAGE	SSB
MATH	1					1					1				
SEAS	0.266*	1				-0.023	1				0.130	1			
SOS	0.188*	0.668*	1			0.070	0.592*	1			0.157*	0.638*	1		
SHORTAGE	-0.069	-0.021	0.033	1		-0.128	-0.023	-0.047	1		-0.083	-0.008	0.005	1	
SSB	0.083	0.126	0.255*	0.338*	1	-0.197	0.022	0.163	-0.058	1	-0.042	0.088	0.222*	0.156*	1

Note. Correlations based on original scale scores aggregated at the school level. MATH: Student mathematics' achievement, SEAS: School Emphasis on Academic Success, SOS: Safe and Orderly Schools, SHORTAGE: Instruction Affected by Mathematics Resource Shortages; SSB: Students' Sense of School Belonging, SES: School socio-economic status. * $p < .050$

Appendix E: Linear Regression Prerequisite Analyses for Grade 4 and 8

Appendix E.1

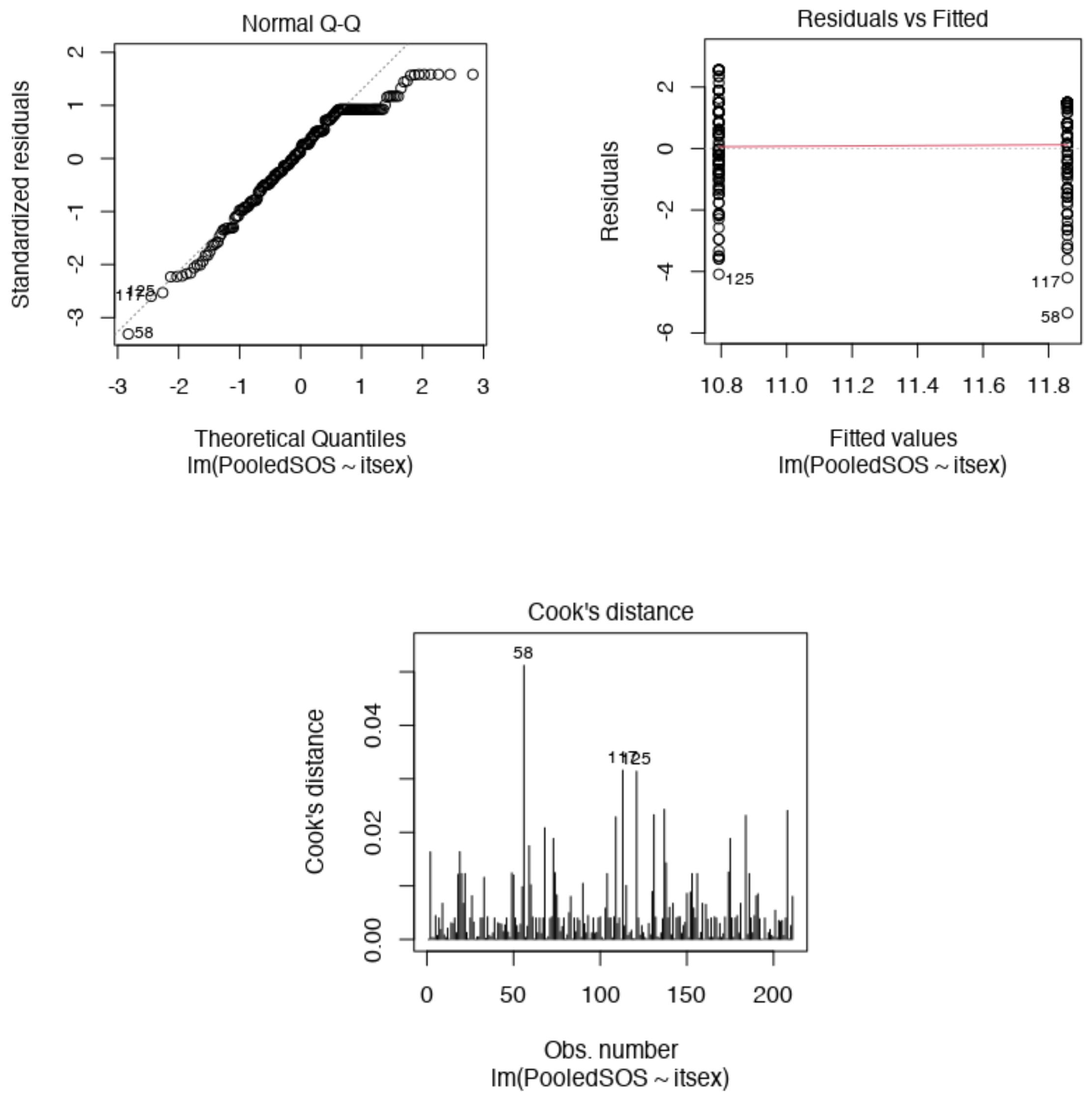
Results of Linear Regression Assumption Checks for Regressing School Emphasis on Academic Success (SEAS) on School Gender for Grade 4



Note. Quantile-quantile plot (upper left) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Homoscedasticity plot (upper right) displays residual (y-axis) by fitted values (x-axis). Plot on the bottom shows flagged influential observations using Cook's distance. The plots have been generated using base R functions.

Appendix E.2

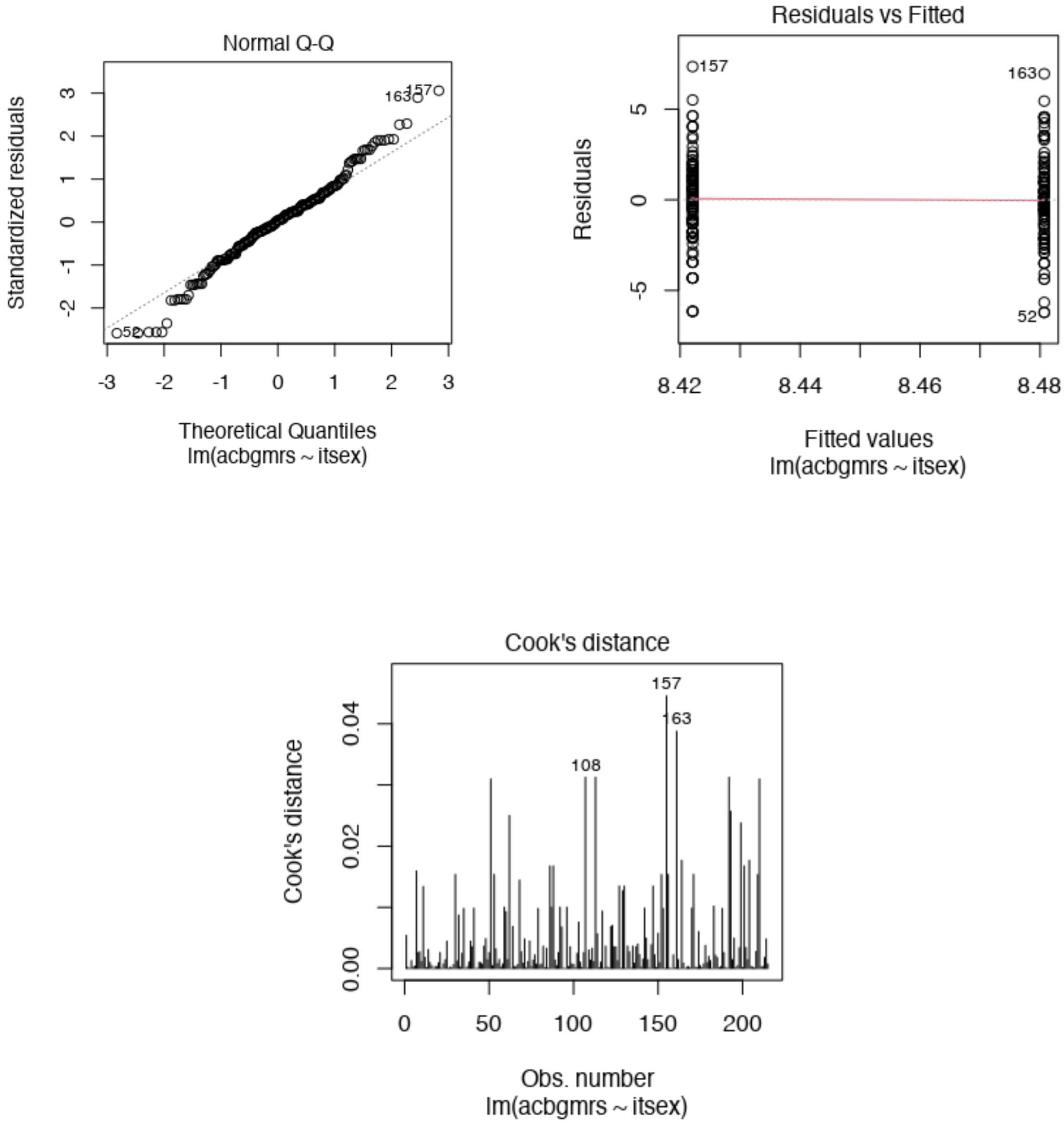
Results of Linear Regression Assumption Checks for Regressing Safe and Orderly Schools (SOS) on School Gender for Grade 4



Note. Quantile-quantile plot (upper left) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Homoscedasticity plot (upper right) displays residual (y-axis) by fitted values (x-axis). Plot on the bottom shows flagged influential observations using Cook's distance. The plots have been generated using base R functions.

Appendix E.3

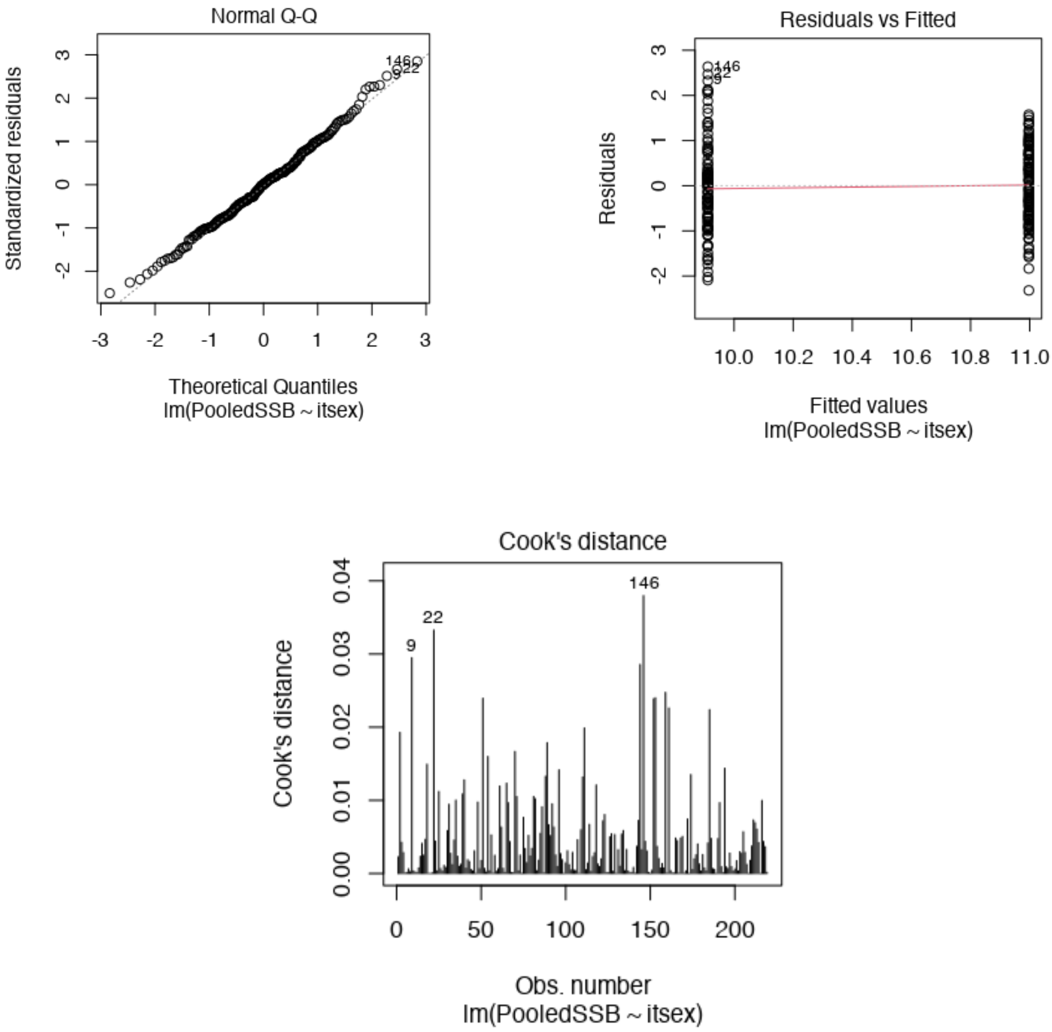
Results of Linear Regression Assumption Checks for Regressing Instruction Affected by Mathematics Resource Shortages (SHORTAGE) on School Gender for Grade 4



Note. Quantile-quantile plot (upper left) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Homoscedasticity plot (upper right) displays residual (y-axis) by fitted values (x-axis). Plot on the bottom shows flagged influential observations using Cook's distance. The plots have been generated using base R functions.

Appendix E.4

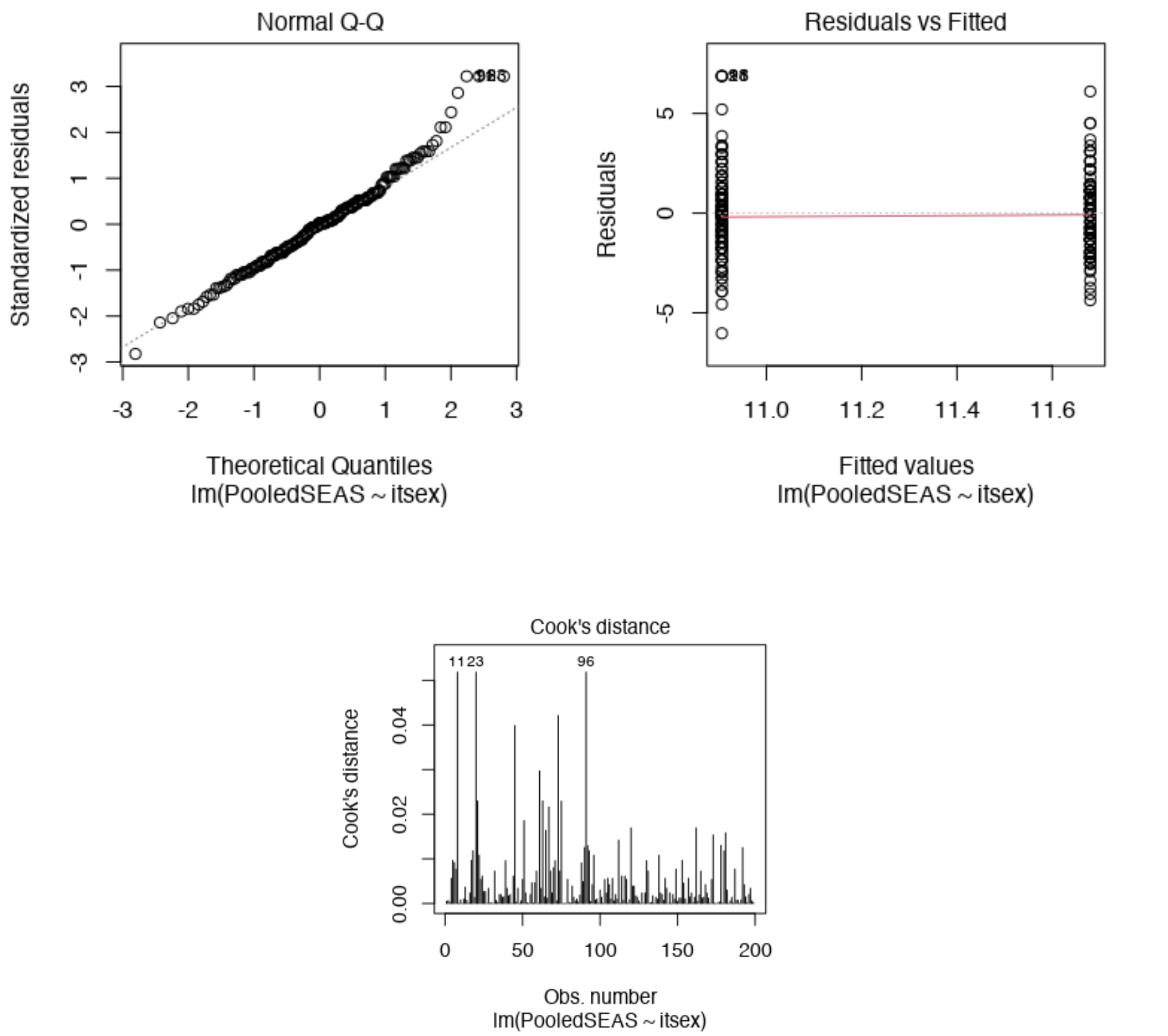
Results of Linear Regression Assumption Checks for Regressing Students' Sense of School Belonging (SSB) on School Gender for Grade 4



Note. Quantile-quantile plot (upper left) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Homoscedasticity plot (upper right) displays residual (y-axis) by fitted values (x-axis). Plot on the bottom shows flagged influential observations using Cook's distance. The plots have been generated using base R functions.

Appendix E.5

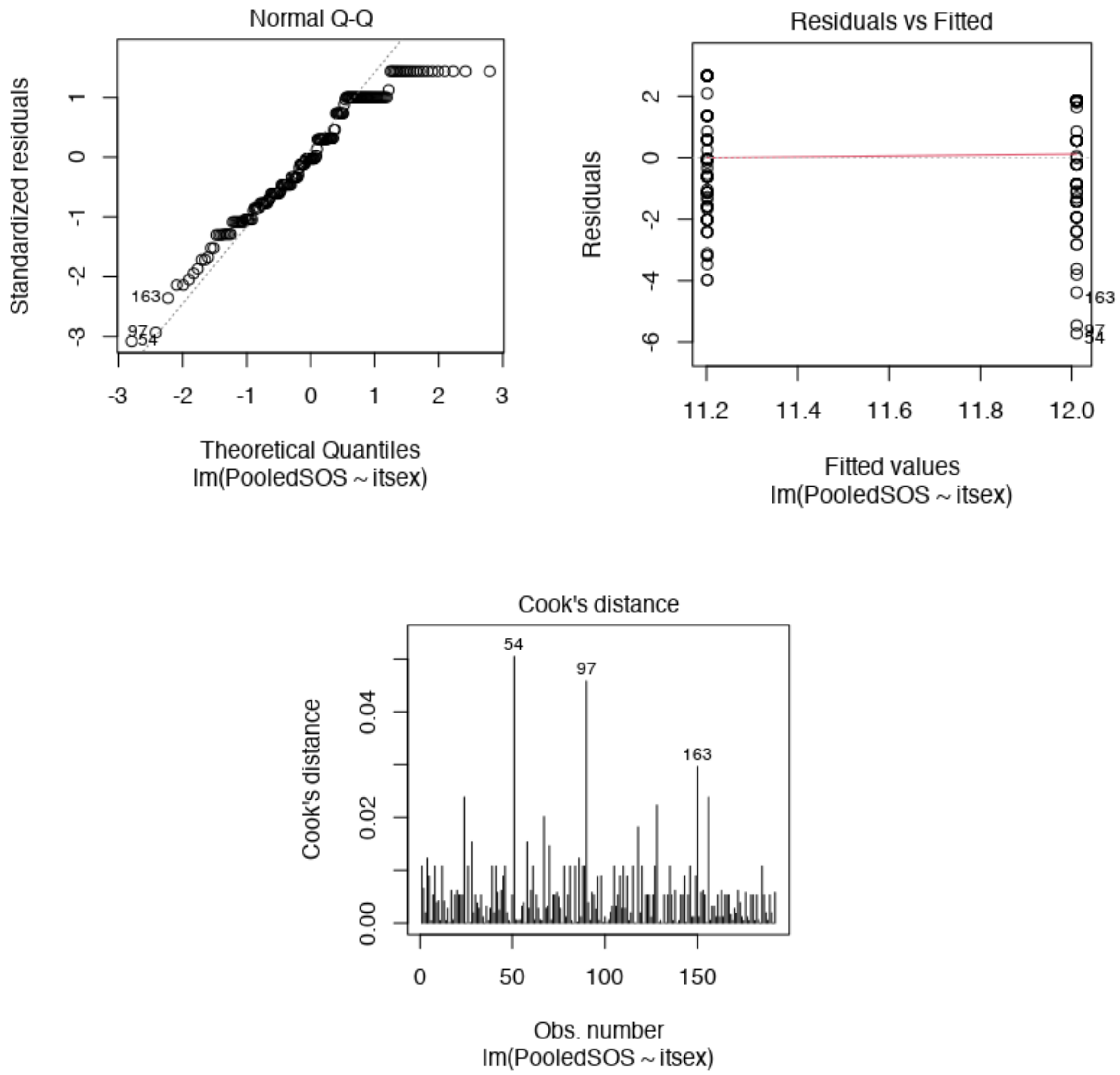
Results of Linear Regression Assumption Checks for Regressing School Emphasis on Academic Success (SEAS) on School Gender for Grade 8



Note. Quantile-quantile plot (upper left) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Homoscedasticity plot (upper right) displays residual (y-axis) by fitted values (x-axis). Plot on the bottom shows flagged influential observations using Cook's distance. The plots have been generated using base R functions.

Appendix E.6

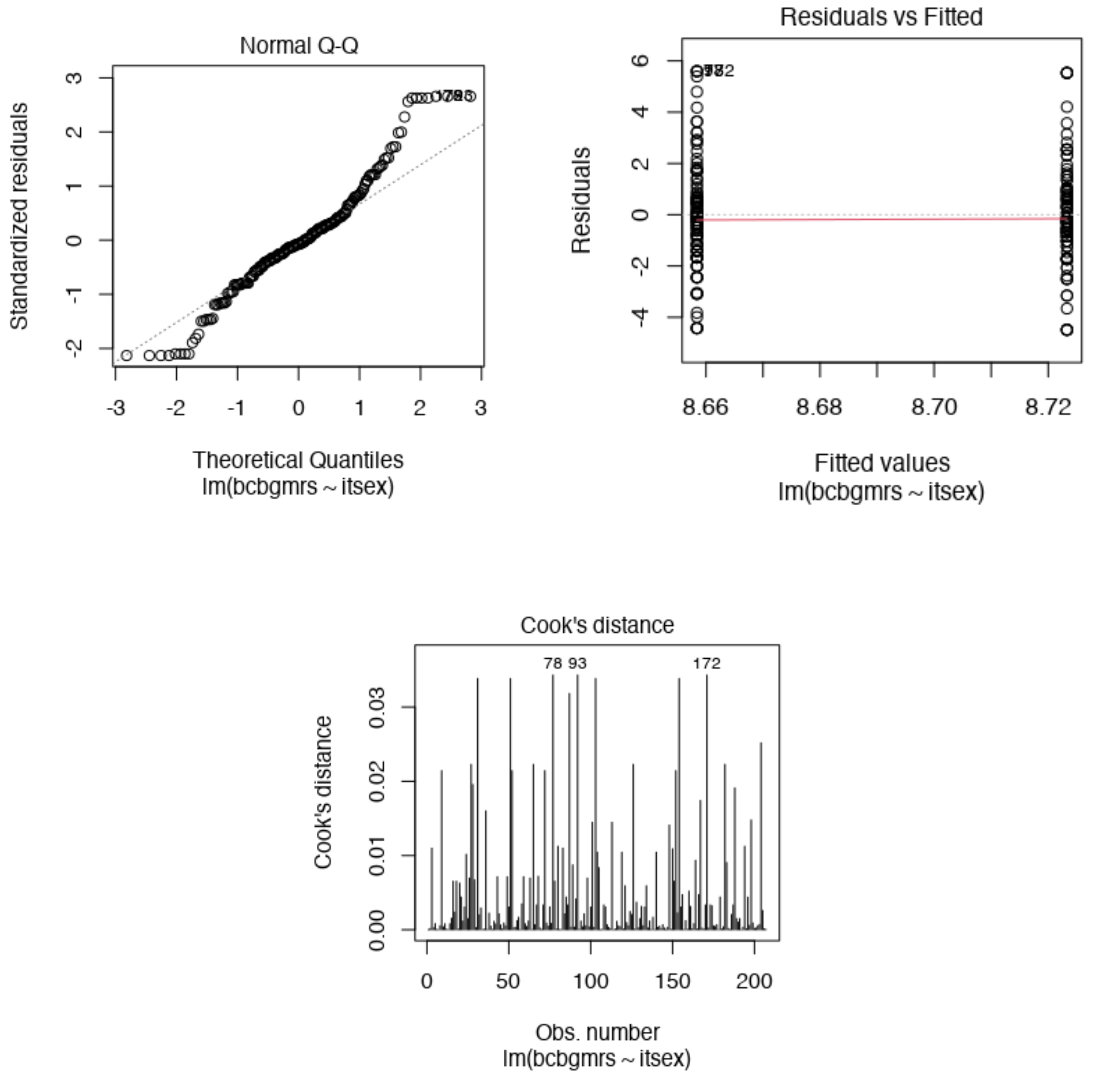
Results of Linear Regression Assumption Checks for Regressing Safe and Orderly Schools (SOS) on School Gender for Grade 8



Note. Quantile-quantile plot (upper left) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Homoscedasticity plot (upper right) displays residual (y-axis) by fitted values (x-axis). Plot on the bottom shows flagged influential observations using Cook's distance. The plots have been generated using base R functions.

Appendix E.7

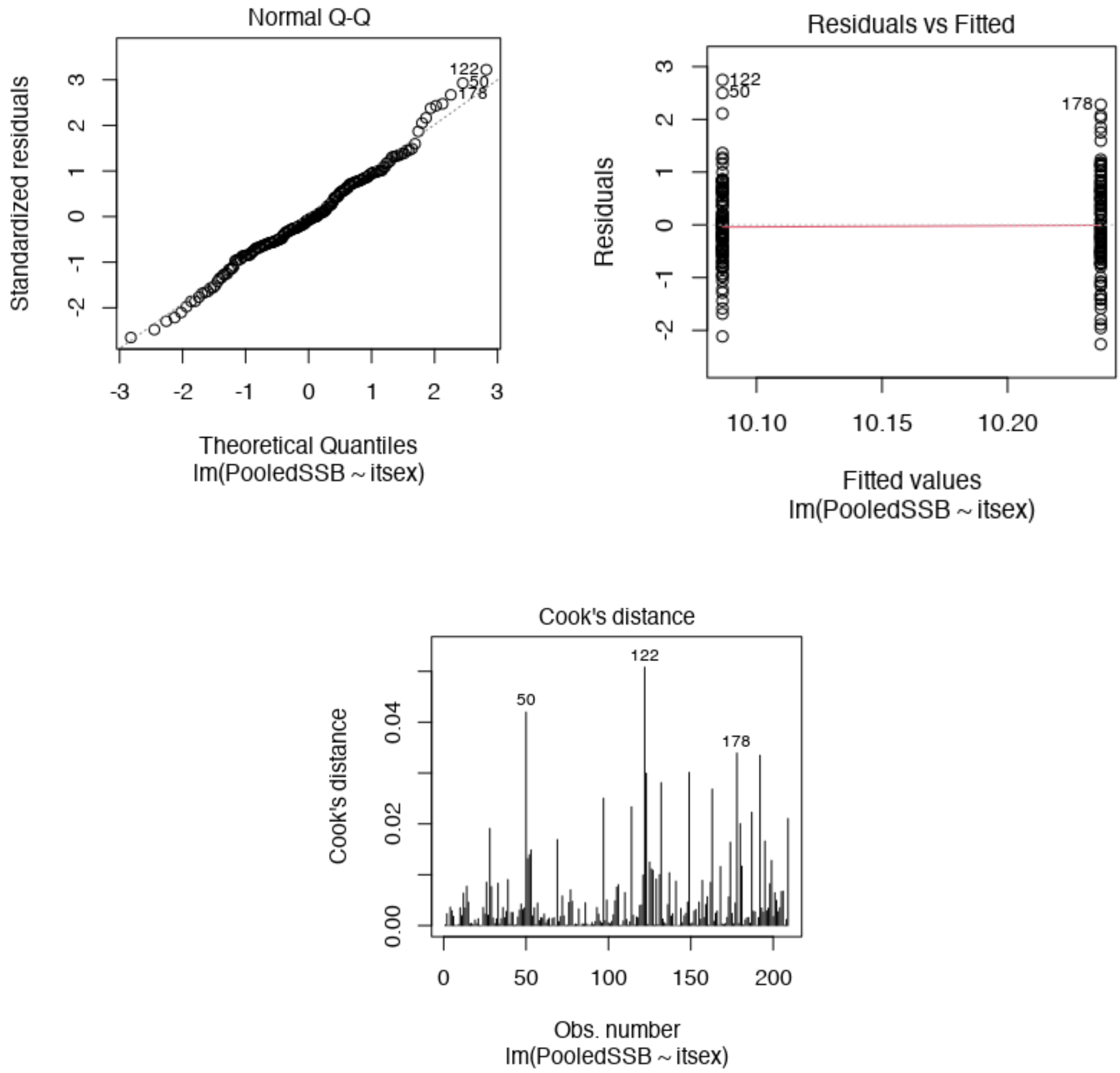
Results of Linear Regression Assumption Checks for Regressing Instruction Affected by Mathematics Resource Shortages (SHORTAGE) on School Gender for Grade 8



Note. Quantile-quantile plot (upper left) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Homoscedasticity plot (upper right) displays residual (y-axis) by fitted values (x-axis). Plot on the bottom shows flagged influential observations using Cook's distance. The plots have been generated using base R functions.

Appendix E.8

Results of Linear Regression Assumption Checks for Regressing Students' Sense of School Belonging (SSB) on School Gender for Grade 8



Note. Quantile-quantile plot (upper left) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Homoscedasticity plot (upper right) displays residual (y-axis) by fitted values (x-axis). Plot on the bottom shows flagged influential observations using Cook's distance. The plots have been generated using base R functions.

Appendix E.9

Results of Robustness Checks for Regressing School Climate Dimensions on School Gender Type (Grade 4)

	School Emphasis on Academic Success (SEAS)		Safe and Orderly Schools (SOS)		Instruction Affected by Mathematics Resource Shortages (SHORTAGE)		Sense of School Belonging (SSB)	
	M1a	M1b	M2a	M2b	M3a	M3b	M4a	M4b
Regression parameter	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)
Intercept	10.514* (0.319)	3.353 (3.120)	10.942* (0.242)	9.098* (2.264)	8.742* (0.245)	13.126* (2.803)	10.036* (0.122)	11.725* (1.101)
Girls' Schools	0.720 (0.390)	0.342 (0.392)	0.955* (0.304)	0.780* (0.311)	-0.252 (0.364)	0.031 (0.310)	0.994* (0.160)	1.077* (0.171)
School SES		0.853* (2.384)		0.260 (0.250)		-0.574 (0.341)		-0.189 (0.120)
School Location (Urban)		-0.720 (0.568)		-0.702 (0.398)		1.183 (0.604)		0.033 (0.244)
Variances	3.994 (0.512)	0.886 (0.084)	2.385 (0.284)	2.340 (0.311)	4.377 (0.755)	3.944 (0.574)	0.729 (0.086)	

Note. The outcome variables are the four dimensions of school climate. Main predictor variable is school gender type (0 = Boys' Schools, 1 = Girls' Schools). Control variables are school SES (positive values indicate higher SES) and school location (coded as 0 = Rural and 1 = Urban). * $p < .050$. Influential observations having high Cook's distance were diagnosed using base R functions. These were removed and the models was rerun in Mplus.

Appendix E.10

Results of Robustness Checks for Regressing School Climate Dimensions on School Gender Type (Grade 8)

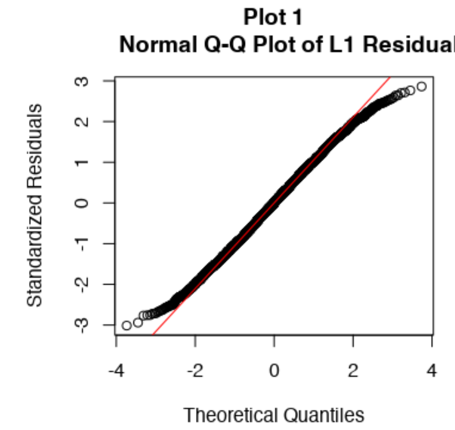
	School Emphasis on Academic Success (SEAS)		Safe and Orderly Schools (SOS)		Instruction Affected by Mathematics Resource Shortages (SHORTAGE)		Sense of School Belonging (SSB)	
	M1a	M1b	M2a	M2b	M3a	M3b	M4a	M4b
Regression parameter	*b* *(SE)*	*b* *(SE)*	*b* *(SE)*	*b* *(SE)*	*b* *(SE)*	*b* *(SE)*	*b* *(SE)*	*b* *(SE)*
Intercept	10.407* (0.369)	8.911* (2.240)	11.099* (0.258)	12.778* (2.554)	8.774* (0.191)	8.123* (0.309)	10.158* (0.127)	11.462* (1.261)
Girls' Schools	1.178* (0.369)	1.140* (0.386)	1.294* (0.353)	1.321* (0.350)	0.014 (0.306)	-0.014 (0.309)	0.228 (0.177)	0.274 (0.176)
School SES		0.160 (0.252)		-0.177 (0.278)		0.093 (0.251)		-0.120 (0.142)
School Location (Urban)		0.058 (0.117)		-0.046 (0.457)		-0.321 (0.304)		-0.309 (0.243)
Variances	3.486 (0.374)	3.526 (0.377)	3.044 (0.265)	3.045 (0.265)	3.157 (0.460)	3.195 (0.479)	0.694 (0.106)	0.649 (0.099)

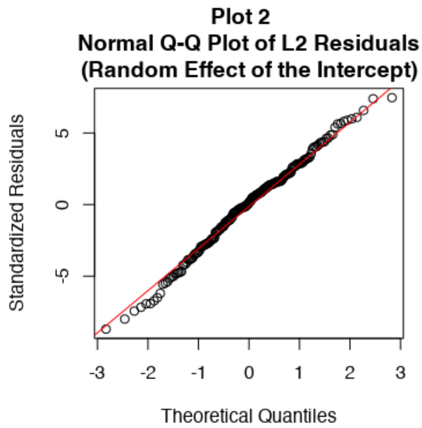
Note. The outcome variables are the four dimensions of school climate. Main predictor variable is school gender type (0 = Boys' Schools, 1 = Girls' Schools). Control variables are school SES (positive values indicate higher SES) and school location (coded as 0 = Rural and 1 = Urban). * $p < .050$. Influential observations having high Cook's distance were diagnosed using base R functions. These were removed and the models was rerun in Mplus.

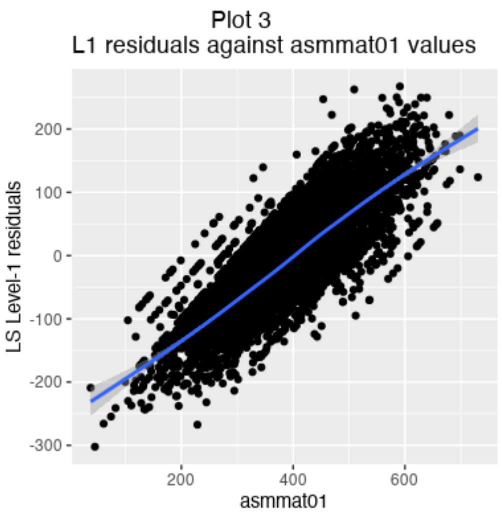
Appendix F: Multi-level Regression Prerequisite Analyses for Grade 4 and 8

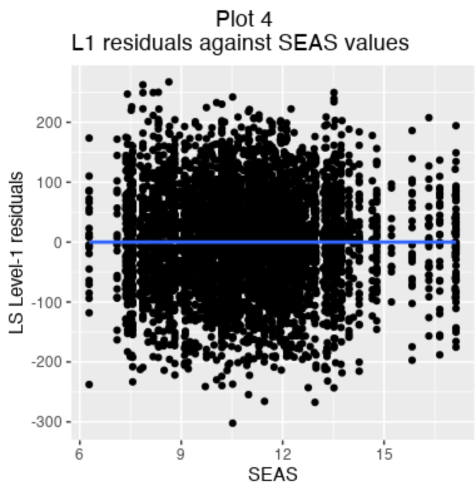
Appendix F.1

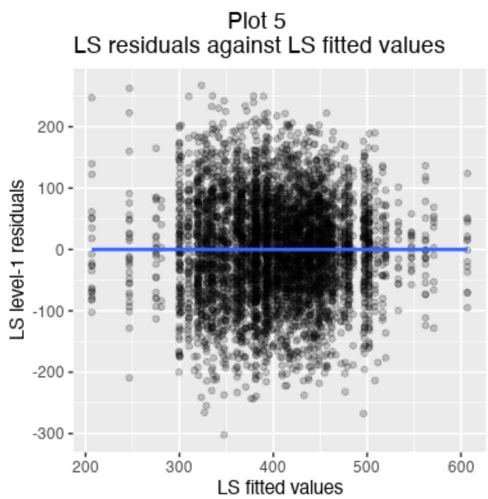
Results of Multilevel Regression Assumption Checks for Regressing Students' Mathematics Achievement on School Emphasis on Academic Success (SEAS) for Grade 4







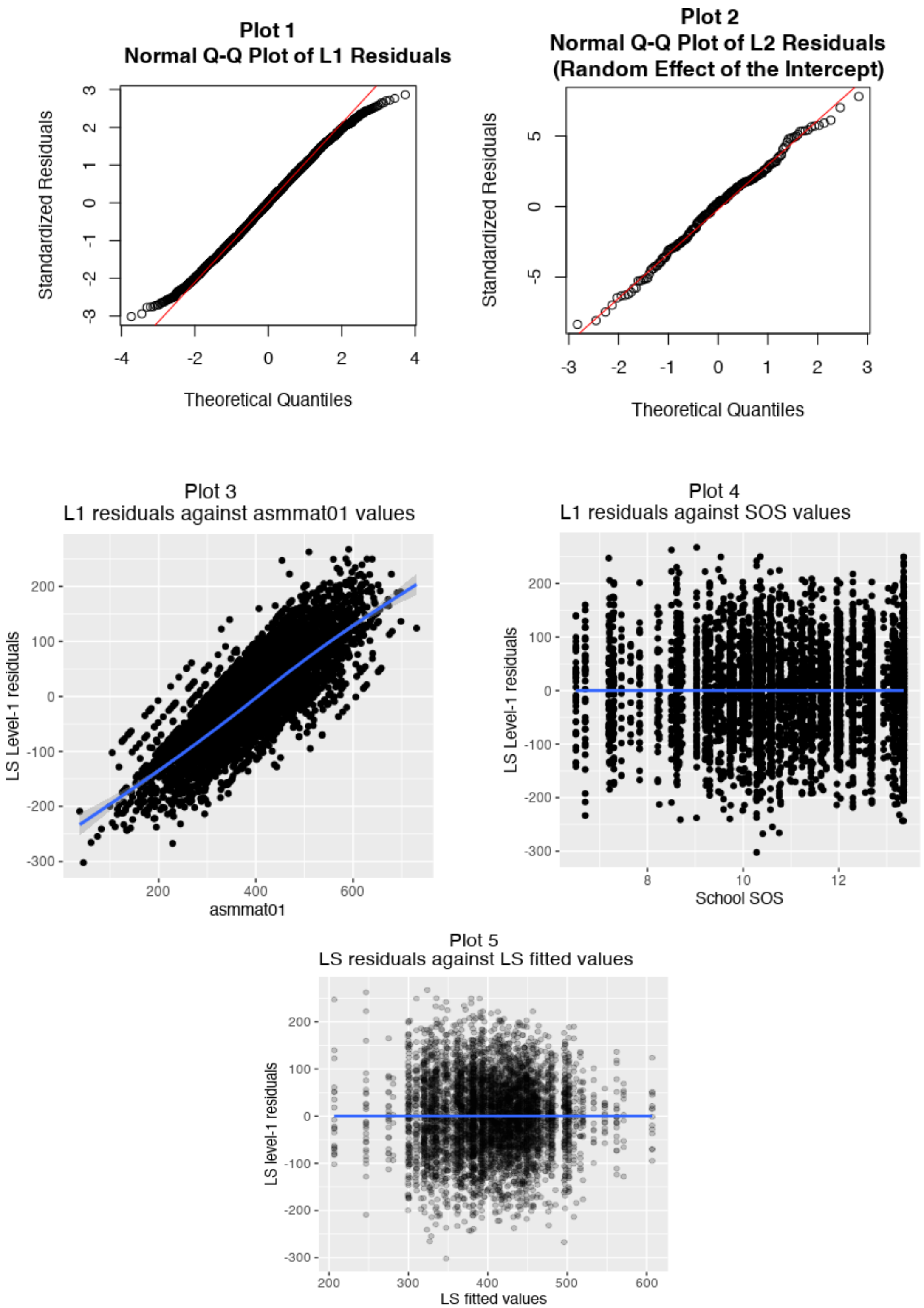




Note: Quantile-quantile plot (plot 1) for level 1 residuals displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Quantile-quantile plot (plot 2) for level 2 residuals (random effect of the intercept) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Scatter plot (plot 3) shows the relationship between Least Squared (LS) level 1 residuals (y-axis) and the first plausible value for Mathematics Achievement (asmmat01). Scatter plot (plot 4) shows the relationship between LS level 1 residuals (y-axis) and the SEAS predictor variable. Homoscedasticity plot (plot 5) displays LS level 1 residuals (y-axis) by LS fitted values (x-axis). The plots have been generated using the R package "HLMdiag" and multilevel models have been fitted using the R package "lme4".

Appendix F.2

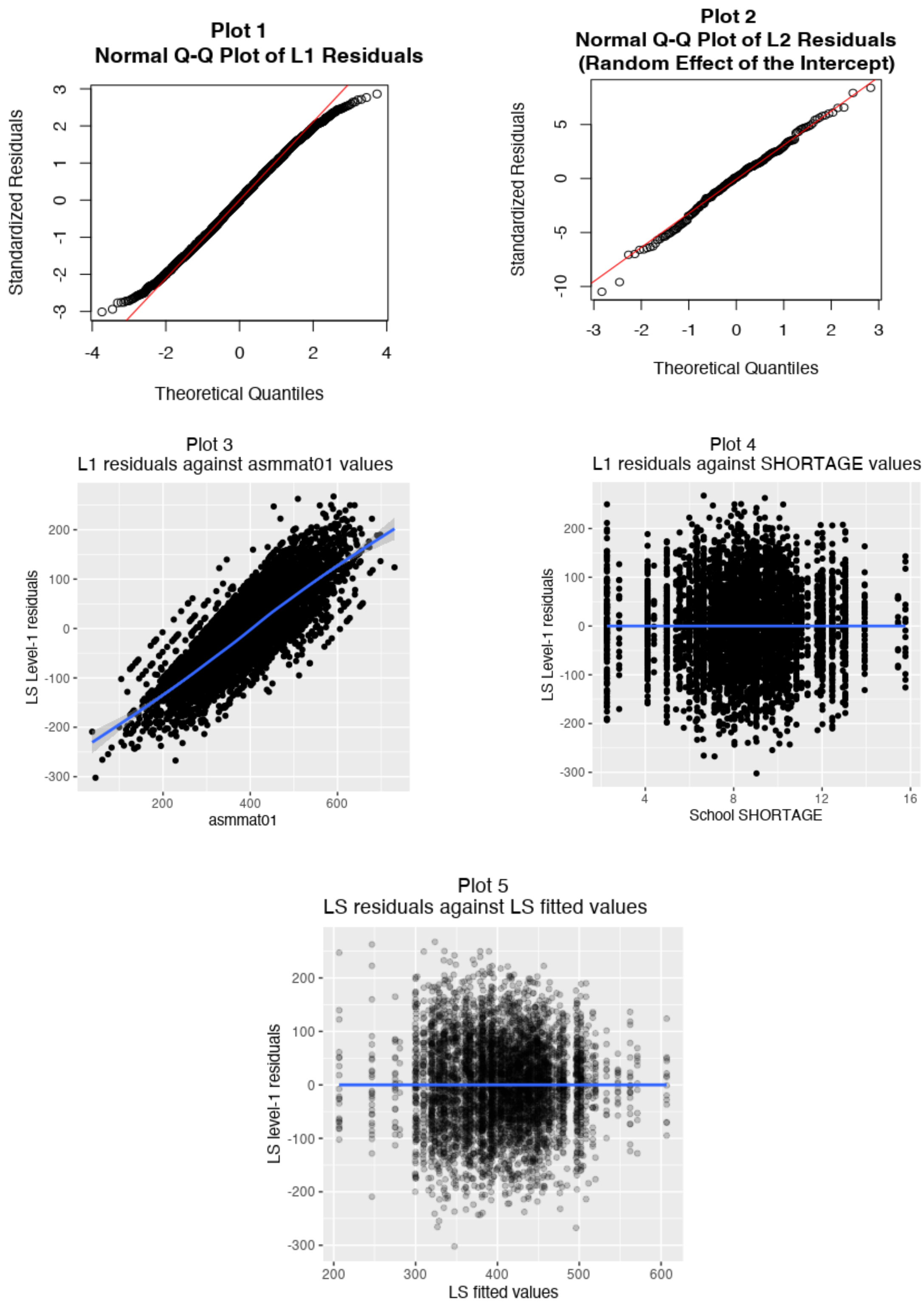
Results of Multilevel Regression Assumption Checks for Regressing Students' Mathematics Achievement on Safe and Orderly Schools (SOS) for Grade 4



Note: Quantile-quantile plot (plot 1) for level 1 residuals displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Quantile-quantile plot (plot 2) for level 2 residuals (random effect of the intercept) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Scatter plot (plot 3) shows the relationship between Least Squared (LS) level 1 residuals (y-axis) and the first plausible value for Mathematics Achievement (asmmat01). Scatter plot (plot 4) shows the relationship between LS level 1 residuals (y-axis) and the SOS predictor variable. Homoscedasticity plot (plot 5) displays LS level 1 residuals (y-axis) by LS fitted values (x-axis). The plots have been generated using the R package "HLMdiag" and multilevel models have been fitted using the R package "lme4".

Appendix F.3

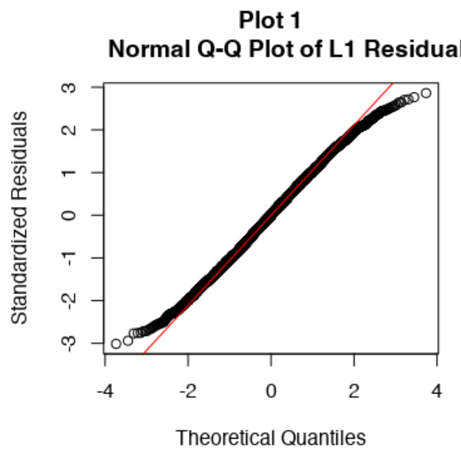
Results of Multilevel Regression Assumption Checks for Regressing Students' Mathematics Achievement on Instruction Affected by Mathematics Resource Shortages (SHORTAGE) for Grade 4

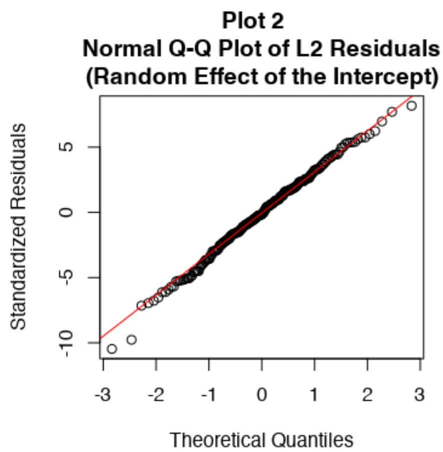


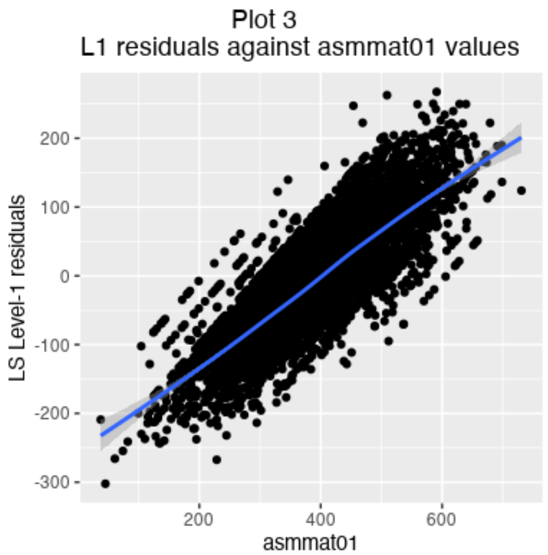
Note: Quantile-quantile plot (plot 1) for level 1 residuals displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Quantile-quantile plot (plot 2) for level 2 residuals (random effect of the intercept) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Scatter plot (plot 3) shows the relationship between Least Squared (LS) level 1 residuals (y-axis) and the first plausible value for Mathematics Achievement (asmmat01). Scatter plot (plot 4) shows the relationship between LS level 1 residuals (y-axis) and the SHORTAGE predictor variable. Homoscedasticity plot (plot 5) displays LS level 1 residuals (y-axis) by LS fitted values (x-axis). The plots have been generated using the R package "HLMdiag" and multilevel models have been fitted using the R package "lme4".

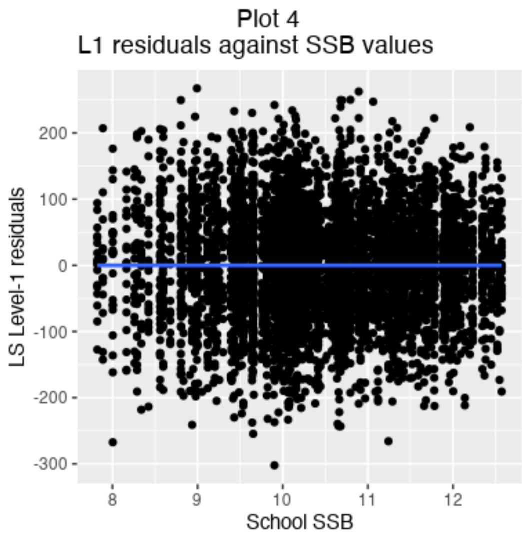
Appendix F.4

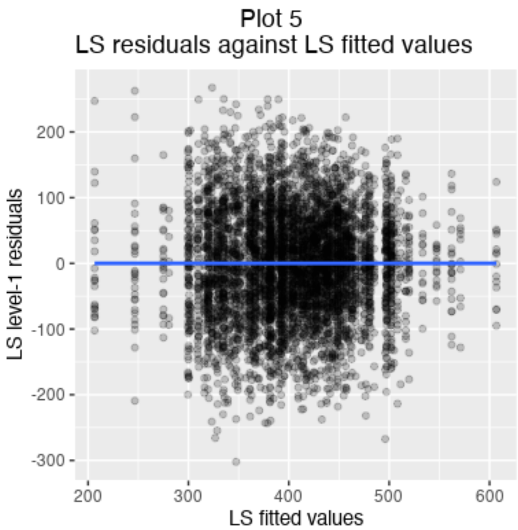
Results of Multilevel Regression Assumption Checks for Regressing Students' Mathematics Achievement on Students' Sense of School Belonging (SSB) for Grade 4







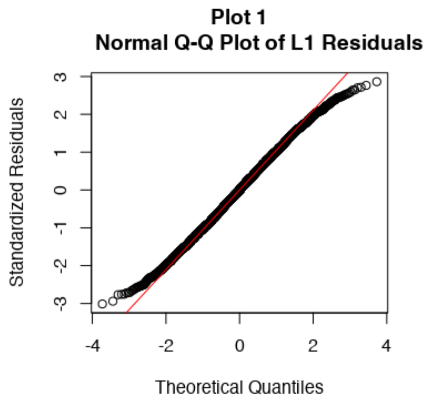


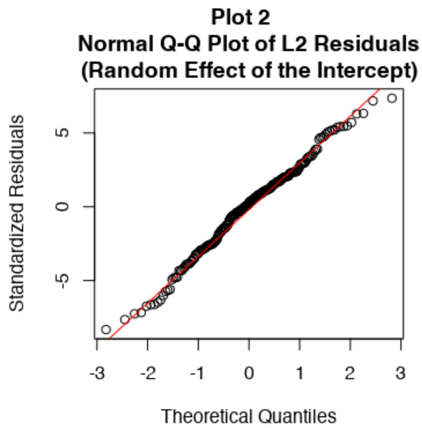


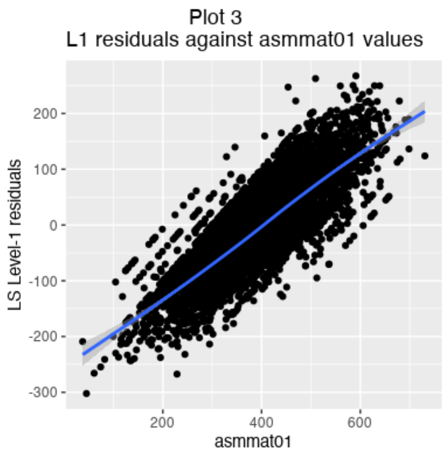
Note: Quantile-quantile plot (plot 1) for level 1 residuals displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Quantile-quantile plot (plot 2) for level 2 residuals (random effect of the intercept) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Scatter plot (plot 3) shows the relationship between Least Squared (LS) level 1 residuals (y-axis) and the first plausible value for Mathematics Achievement (asmmat01). Scatter plot (plot 4) shows the relationship between LS level 1 residuals (y-axis) and the SSB predictor variable. Homoscedasticity plot (plot 5) displays LS level 1 residuals (y-axis) by LS fitted values (x-axis). The plots have been generated using the R package "HLMdiag" and multilevel models have been fitted using the R package "lme4".

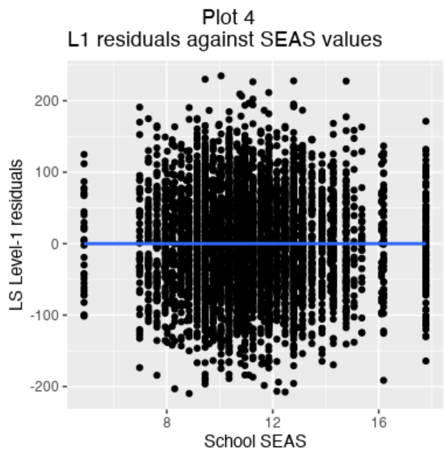
Appendix F.5

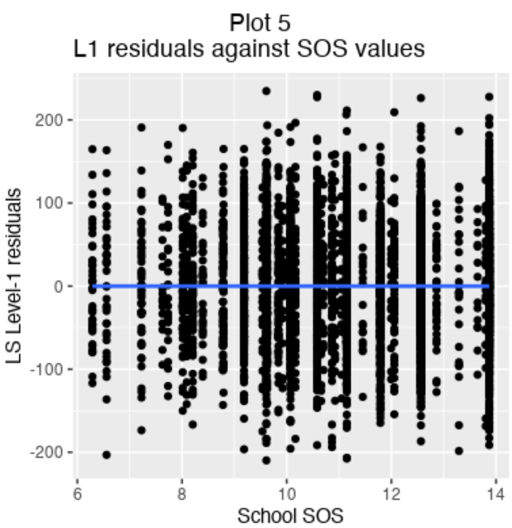
Results of Multilevel Regression Assumption Checks for Regressing Students' Mathematics Achievement on SEAS, SOS, SHORTAGE and SSB for Grade 4

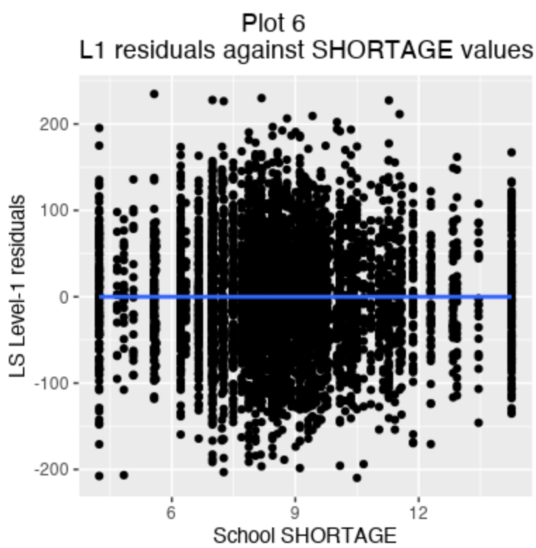


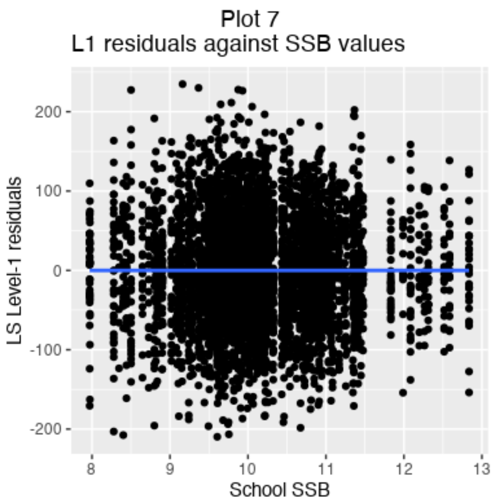


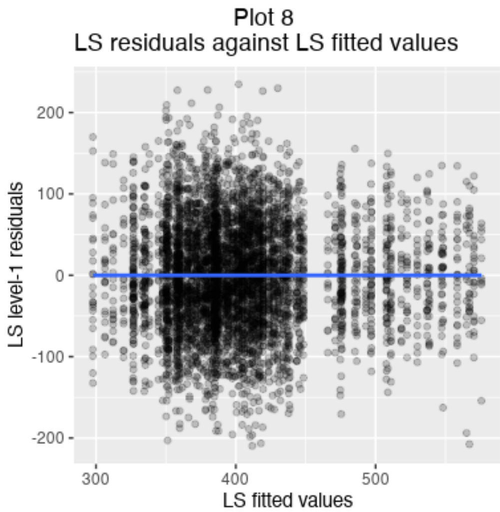








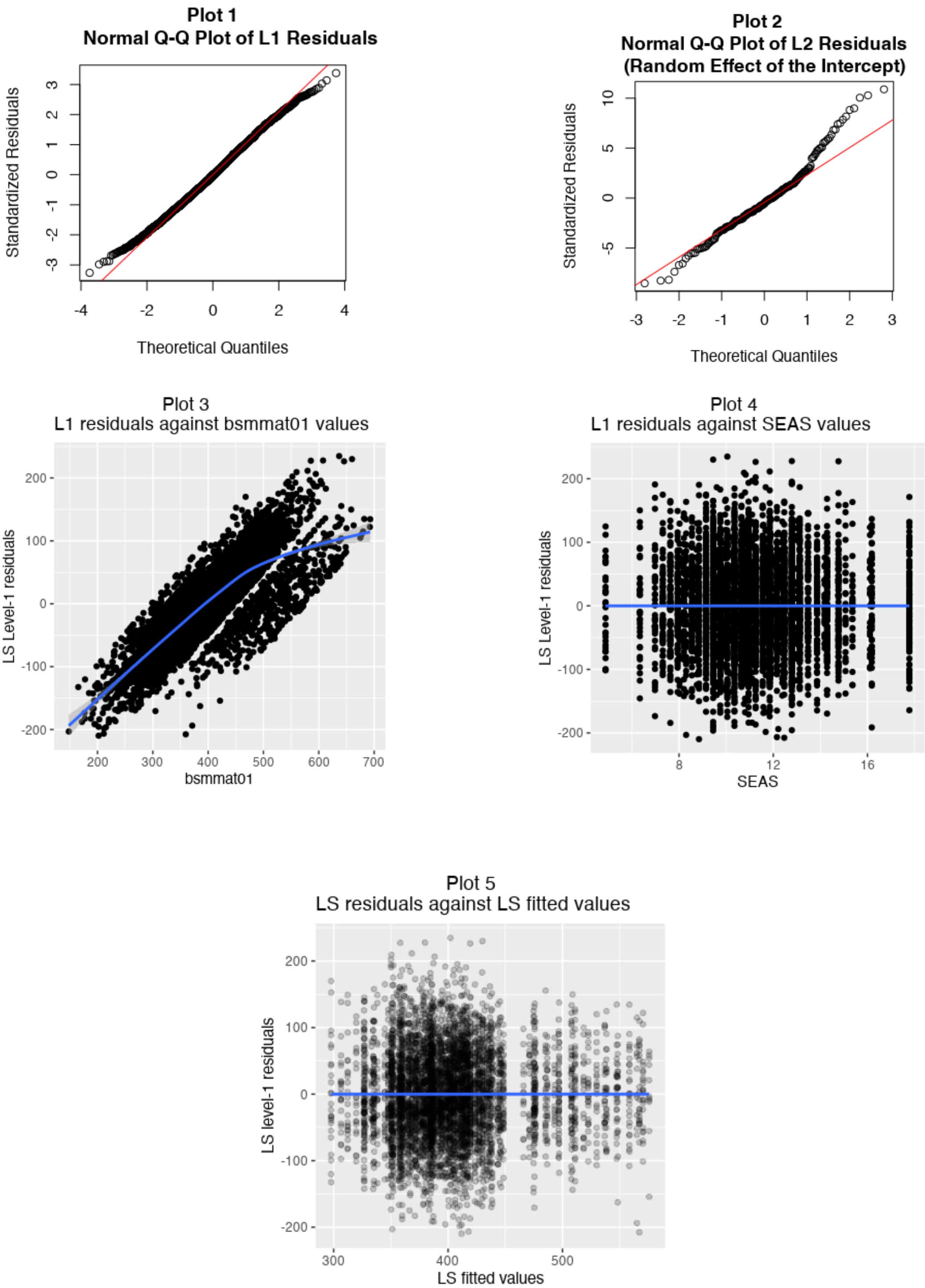




Note: Quantile-quantile plot (plot 1) for level 1 residuals displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Quantile-quantile plot (plot 2) for level 2 residuals (random effect of the intercept) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Scatter plot (plot 3) shows the relationship between Least Squared (LS) level 1 residuals (y-axis) and the first plausible value for Mathematics Achievement (asmmat01). Scatter plots (plot 4-7) show the relationship between LS level 1 residuals (y-axis) and the four predictor variables. Homoscedasticity plot (plot 5) displays LS level 1 residuals (y-axis) by LS fitted values (x-axis). The plots have been generated using the R package "HLMdiag" and multilevel models have been fitted using the R package "lme4".

Appendix F.6

Results of Multilevel Regression Assumption Checks for Regressing Students' Mathematics Achievement on School Emphasis on Academic Success (SEAS) for Grade 8

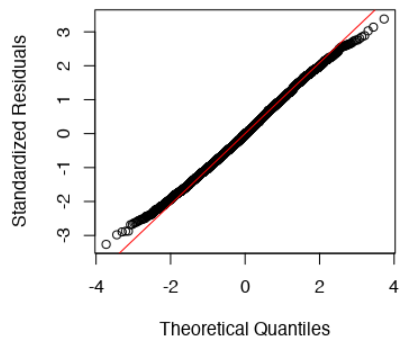


Note: Quantile-quantile plot (plot 1) for level 1 residuals displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Quantile-quantile plot (plot 2) for level 2 residuals (random effect of the intercept) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Scatter plot (plot 3) shows the relationship between Least Squared (LS) level 1 residuals (y-axis) and the first plausible value for Mathematics Achievement (bsmmat01). Scatter plot (plot 4) shows the relationship between LS level 1 residuals (y-axis) and the SEAS predictor variable. Homoscedasticity plot (plot 5) displays LS level 1 residuals (y-axis) by LS fitted values (x-axis). The plots have been generated using the R package "HLMdiag" and multilevel models have been fitted using the R package "lme4".

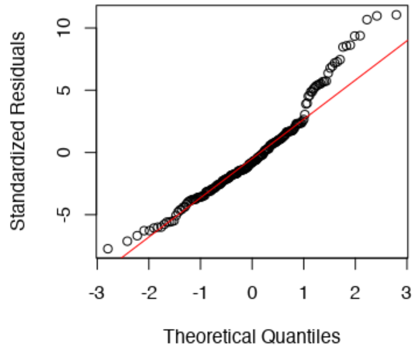
Appendix F.7

Results of Multilevel Regression Assumption Checks for Regressing Students' Mathematics Achievement on Safe and Orderly Schools (SOS) for Grade 8

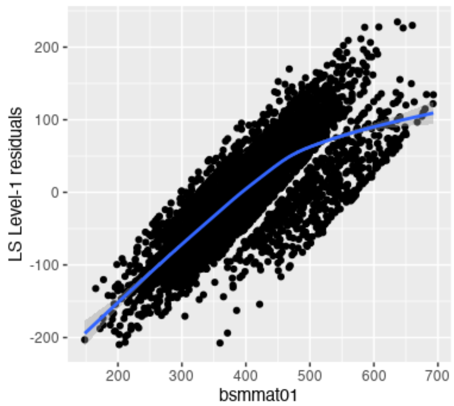
Plot 1
Normal Q-Q Plot of L1 Residuals



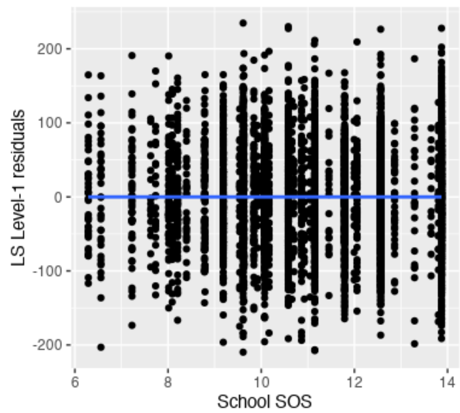
Plot 2
Normal Q-Q Plot of L2 Residuals
(Random Effect of the Intercept)



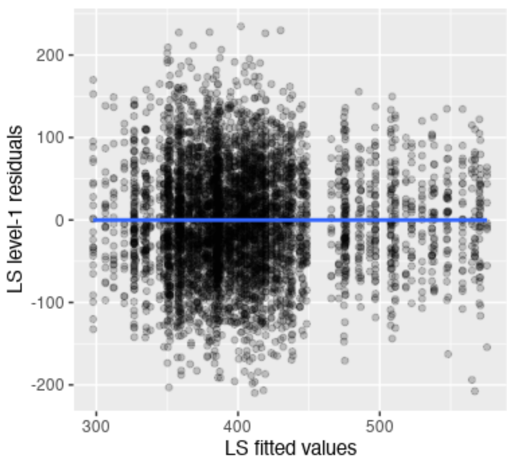
Plot 3
L1 residuals against bsmmat01 values



Plot 4
L1 residuals against SOS values



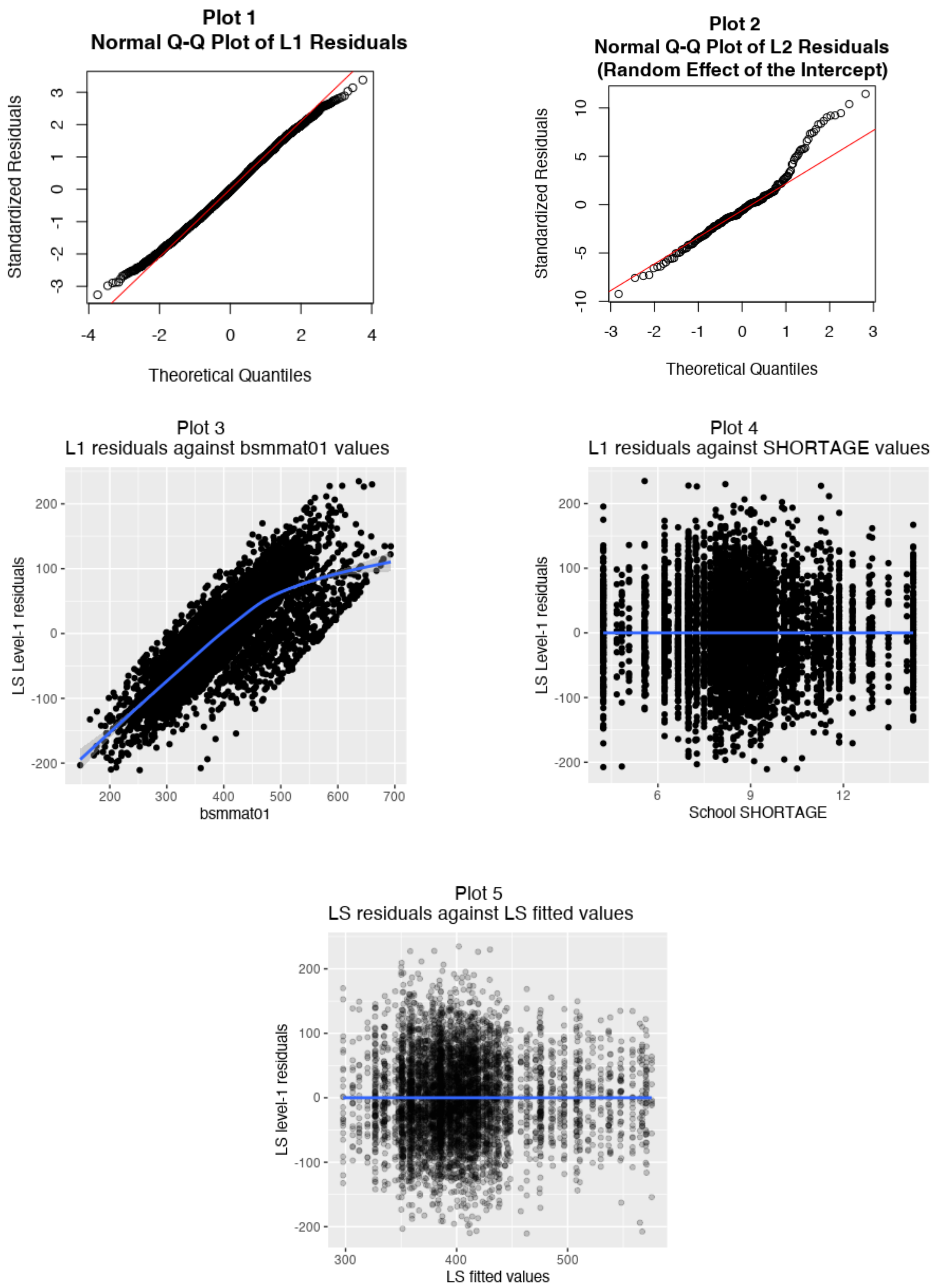
Plot 5
LS residuals against LS fitted values



Note: Quantile-quantile plot (plot 1) for level 1 residuals displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Quantile-quantile plot (plot 2) for level 2 residuals (random effect of the intercept) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Scatter plot (plot 3) shows the relationship between Least Squared (LS) level 1 residuals (y-axis) and the first plausible value for Mathematics Achievement (bsmmat01). Scatter plot (plot 4) shows the relationship between LS level 1 residuals (y-axis) and the SOS predictor variable. Homoscedasticity plot (plot 5) displays LS level 1 residuals (y-axis) by LS fitted values (x-axis). The plots have been generated using the R package "HLMdiag" and multilevel models have been fitted using the R package "lme4".

Appendix F.8

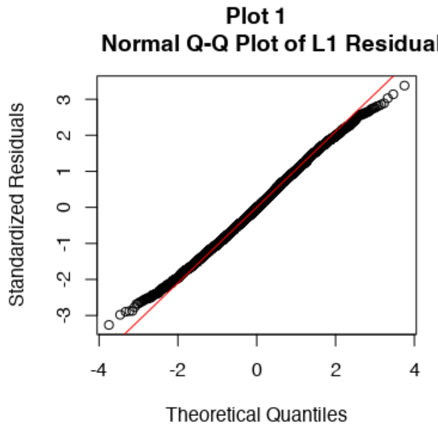
Results of Multilevel Regression Assumption Checks for Regressing Students' Mathematics Achievement on Instruction Affected by Mathematics Resource Shortages (SHORTAGE) for Grade 8

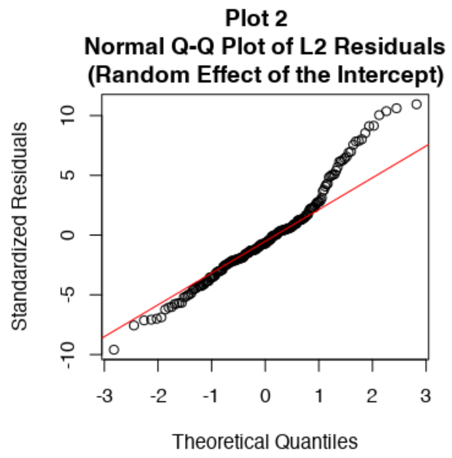


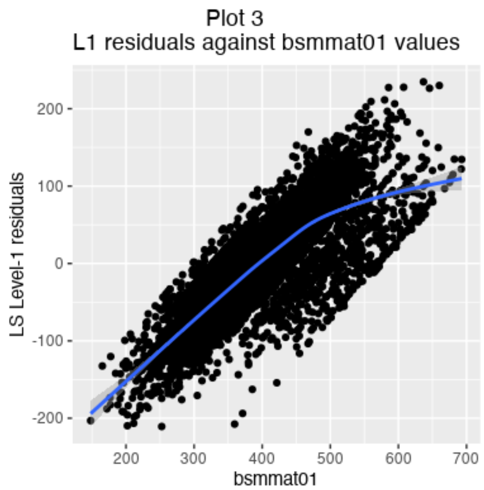
Note: Quantile-quantile plot (plot 1) for level 1 residuals displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Quantile-quantile plot (plot 2) for level 2 residuals (random effect of the intercept) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Scatter plot (plot 3) shows the relationship between Least Squared (LS) level 1 residuals (y-axis) and the first plausible value for Mathematics Achievement (bsmmat01). Scatter plot (plot 4) shows the relationship between LS level 1 residuals (y-axis) and the SHORTAGE predictor variable. Homoscedasticity plot (plot 5) displays LS level 1 residuals (y-axis) by LS fitted values (x-axis). The plots have been generated using the R package "HLMdiag" and multilevel models have been fitted using the R package "lme4".

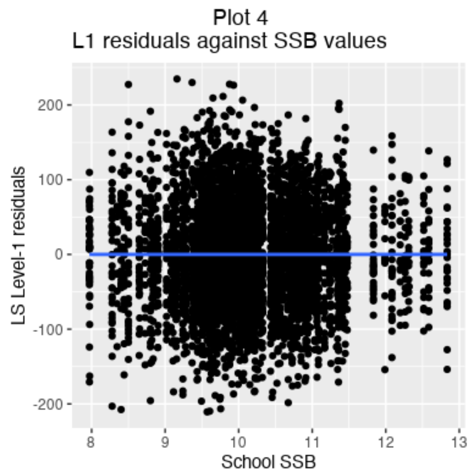
Appendix F.9

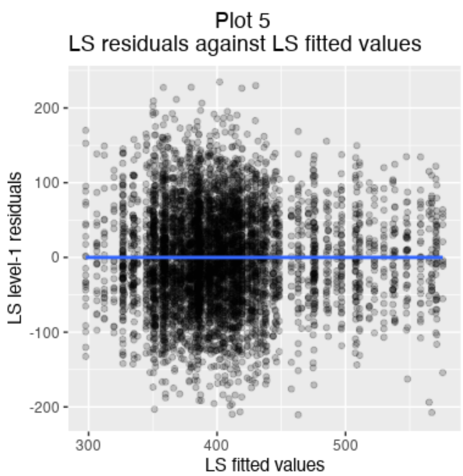
Results of Multilevel Regression Assumption Checks for Regressing Students' Mathematics Achievement on Students' Sense of School Belonging (SSB) for Grade 8







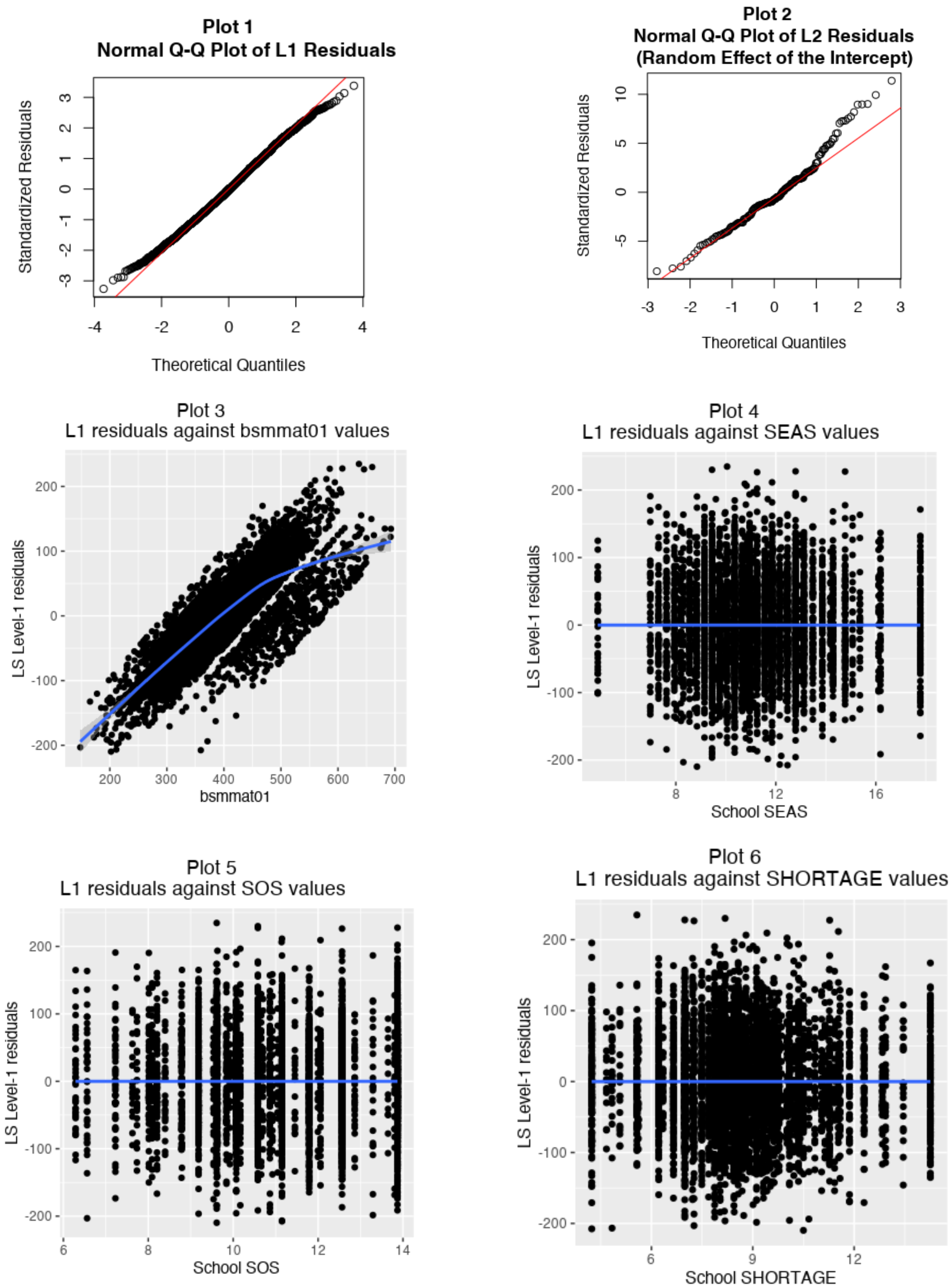


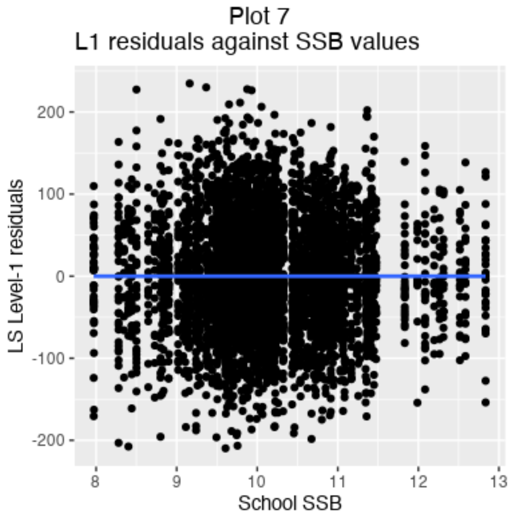


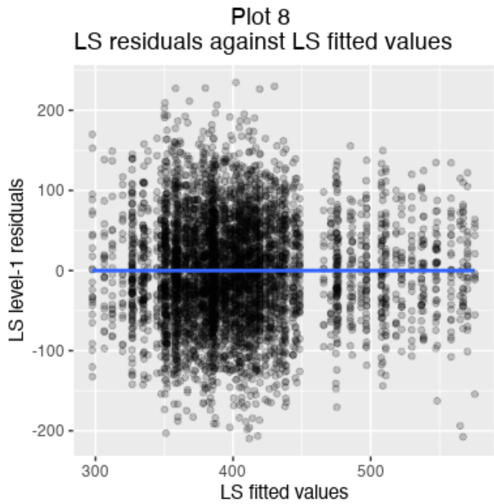
Note: Quantile-quantile plot (plot 1) for level 1 residuals displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Quantile-quantile plot (plot 2) for level 2 residuals (random effect of the intercept) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Scatter plot (plot 3) shows the relationship between Least Squared (LS) level 1 residuals (y-axis) and the first plausible value for Mathematics Achievement (bsmmat01). Scatter plot (plot 4) shows the relationship between LS level 1 residuals (y-axis) and the SSB predictor variable. Homoscedasticity plot (plot 5) displays LS level 1 residuals (y-axis) by LS fitted values (x-axis). The plots have been generated using the R package "HLMdiag" and multilevel models have been fitted using the R package "lme4".

Appendix F.10

Results of Multilevel Regression Assumption Checks for Regressing Students' Mathematics Achievement on SEAS, SOS, SHORTAGE and SSB for Grade 8







Note: Quantile-quantile plot (plot 1) for level 1 residuals displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Quantile-quantile plot (plot 2) for level 2 residuals (random effect of the intercept) displays observed quantiles (y-axis) by theoretical quantiles for normal distributions (x-axis). Scatter plot (plot 3) shows the relationship between Least Squared (LS) level 1 residuals (y-axis) and the first plausible value for Mathematics Achievement (bsmmat01). Scatter plots (plot 4-7) show the relationship between LS level 1 residuals (y-axis) and the four predictor variables. Homoscedasticity plot (plot 5) displays LS level 1 residuals (y-axis) by LS fitted values (x-axis). The plots have been generated using the R package "HLMdiag" and multilevel models have been fitted using the R package "lme4".

Appendix F.11

Results of Robustness Checks for the Effects of School Climate Dimensions (L2) on Students' Mathematics Achievement (L1) for Grade 4

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
	b (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)
Fixed effects										
Intercept γ_{oo}	408.039* (3.833)	120.028* (55.738)	408.996* (3.924)	109.247* (55.559)	407.695* (4.068)	73.2 (55.478)	405.686* (3.975)		407.02* (3.817)	79.815 (57.552)
School SEAS	7.865* (1.616)	3.673* (1.523)							6.113* (2.14)	2.23 (2.038)
School SOS			7.352* (2.17)	4.638* (2.088)					2.802 (2.819)	1.425 (2.71)
School Shortage					-2.956 (1.577)	-0.17 (1.443)			-1.998 (1.446)	0.259 (1.38)
School SSB							4.882 (3.67)	9.601* (3.501)	-1.329 (3.662)	4.429 (3.651)
School SES		33.502* (5.979)		34.588* (5.94)		38.585* (5.998)		42.903* (5.923)		37.007* (6.27)
School Location (Urban)		-34.636* (15.819)		-33.362* (15.807)		-36.923* (16.613)		-31.788* (15.236)		-28.646 (15.44)
Random effects										
Within schools ($\sigma^2 r$)	7146.493 (221.257)	7231.083 (234.818)	7123.988 (222.17)	7208.572 (255.19)	7173.139 (197.052)	7246.575 (238.867)	7157.481 (202.406)	7236.083 (208.811)	7193.855 (238.858)	
Between schools ($\sigma^2 \mu_0$)	2587.871 (279.956)	2084.93 (267.116)	2703.035 (284.488)	2083.676 (259.157)	3049.273 (328.272)	2303.298 (290.198)	3107.494 353.335	2024.614 (238.516)	2510.161 (266.898)	

Note. The outcome variable is students' mathematics achievement at the student level (L1). The predictor variables (SEAS, SOS, SHORTAGE, SSB, SES, LOCATION) are all located at the school level (L2). Influential schools having high Cook's distance were diagnosed using dotplot_diag() and hlm_influence () functions in HLMdiag R package. These were removed and the analysis was rerun in Mplus. * $p < .05$

Appendix F.12

Results of Robustness Checks for the Effects of School Climate Dimensions (L2) on Students' Mathematics Achievement (L1) for Grade 8

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
	b (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)
Fixed effects										
Intercept γ_{oo}	408.552* (3.784)	-52.282 (42.071)	406.069* (3.933)	-58.341 (41.74)	405.585* (3.674)	-51.474 (40.009)	406.181* (3.716)	-61.415 (40.44)	406.566* (3.794)	-59.842 (41.389)
School SEAS	7.309* (1.644)	2.259 (1.287)							6.475* (2.295)	0.822 (1.574)
School SOS			4.312* (1.769)	2.34 (1.341)					0.934 (2.437)	2.162 (1.756)
School Shortage					-4.165* (1.702)	-3.185* (1.19)			-2.046 (1.946)	-2.473 (1.331)
School SSB							-2.362 (4.299)	4.3 (3.963)	-4.832 (4.926)	2.037 (4.039)
School SES		48.144* (4.73)		47.713* (4.649)		47.99* (4.518)		48.558* 4.46		48.425* (4.648)
School Location (Urban)		-5.778 (8.586)		4.244 (7.705)		-6.351 (8.124)		-0.82 7.583		-2.209 (9.781)
Random effects										
Within schools (σ^2r)	4708.658 (122.093)	4737.685 (124.984)	4726.982 (125.514)	4757.378 (128.496)	4715.078 (113.473)	4752.905 (116.795)	4720.195 (117.101)	4756.901 (116.831)	4711.425 (122.28)	4744.845 (125.31)
Between schools ($\sigma^2\mu_0$)	2346.003 (329.725)	1238.797 (164.122)	2431.637 (356.363)	1205.561 (177.161)	2234.225 (310.98)	1091.184 (154.345)	2454.108 (345.559)	1141.715 (153.159)	2193.936 (301.797)	1117.461 (161.203)

Note. The outcome variable is students' mathematics achievement at the student level (L1). The predictor variables (SEAS, SOS, SHORTAGE, SSB, SES, LOCATION) are all located at the school level (L2). Influential schools having high Cook's distance were diagnosed using dotplot_diag() and hlm_influence () functions in HLMdiag R package. These were removed and the models was rerun in Mplus. * $p < .05$