FACULTY OF EDUCATION
DEPARTMENT OF EDUCATION AND SPECIAL EDUCATION

EDUCATIONAL QUALITY AND EQUITY IN SOUTH AFRICA: EVIDENCE FROM TIMSS 2015

Ernest Mensah

Master's thesis:	30credits
Programme/course:	L2EUR (IMER) PDA184
Level:	Second cycle (Advanced)
Term/year:	Spring 2020
Supervisor:	Kajsa Yang-Hansen
Examiner:	xx
Report nr:	xx (Supplied by supervisor)

Abstract

This study contributes to educational quality and equity research in South Africa. In grade 9 (TIMSS) 2015, a sample of 334 mathematics teachers, and 12514 students in 292 schools in South Africa was used. Applying a two-level structural equation modelling technique, the relationship between aspects of teacher characteristics, instructional quality, students' SES background, and mathematics achievement were examined. The results revealed that teacher qualification and characteristics, and instructional quality do not affect student mathematics achievement, once the student's family SES and classroom SES composition were taken into account. The classroom SES composition explained almost 80 % of the cross-classroom differences in mathematics achievement differences in South Africa, indicating a high level of socio-economic segregation between classrooms in mathematics achievement. A tentative explanation might be that qualified and experienced teachers are more likely to self-selected to schools and classes where the best students are. Since student's achievement level is related to their socio-economic background. High achieving schools and classrooms very often are also with students of a higher level of SES. Thus, in South Africa, the teacher effects are confounded with SES composition effect. These results are discussed, and policy implications and practice recommendations of the findings are suggested.

Master's thesis:	30 credits
Programme/Course:	L2EUR (IMER) PDA184
Level:	Second cycle (Advanced)
Term/year:	Spring 2020
Supervisor:	Kajsa Yang-Hansen
Examiner:	xx
Report nr:	xx (Supplied by supervisor)
Keywords:	Educational quality, educational equity, teacher quality, instructional quality, teacher qualification, teacher confidence, socioeconomic status, two-level structural equation modelling, TIMSS, South Africa.

Aim: This study aims to investigate the relationship between teacher qualification and characteristics, teacher instructional quality, students' family socioeconomic background, and student mathematics achievement with the South Africa data from TIMSS 2015.

Theory: The dynamic model of educational effectiveness, proposed by Creemers and Kyriakides (2008) in understanding variables within each level and across different levels related (such as student-level and classroom-level), and Input-Process-Outcome (IPO) model, proposed by Goe (2007) was used as the theoretical framework which leads the selection of variables and was operationalized with the achievement and contextual data available in South Africa TIMSS 2015 data.

Method: Two-level structural equation models were estimated at student and classroom-levels using Statistic Software Program SPSS Version 25 for proper data management to make the variables appropriate to be analysed in Mplus Version 8.3.

Results: No significant relationship between teacher qualification and characteristics, teacher instructional quality, and student average mathematics achievement was found. Teacher characteristics and teacher instructional quality had no relationship to the classroom average mathematics achievement. However, teacher characteristics were

significantly related to teacher instructional quality, indicating teachers with a higher level of confidence offered a higher quality of instruction, as required by their students. Teachers with experience, and level of formal education, and teachers who focus on either mathematics or mathematics education had no association with the classroom mean mathematics achievement. The context of classroom SES composition was found to be the strongest classroom-level factor, strongly associated with increased variations (inequity) in classroom average mathematics achievement scores of South African students. Student-level SES significantly affect mathematics achievement, but the effects were not strong at the student-level.

Foreword

Writing this thesis has been both challenging and rewarding. I am grateful to enrol in the IMER program, which gives me the opportunities to step out into the research world, which is different from my previous encounter with the social world.

I would like to thank my supervisor, Professor Kajsa Yang-Hansen, for her great support and diligent guidance in the planning and development of the research described in this thesis, and for her perceptive suggestions throughout the writing of the thesis. I am also grateful for her forbearance, wit, encouragement, and endless patience during the research process, especially when I struggled with Mplus and the complexity of two-level modelling.

I would like to gratefully acknowledge the help and support I have received from all the lecturers of IMER program, especially, Dr. Adrianna Nizinska, Dr. Aimee Haley, Dr. Dawn Sanders, Dr. Susanne Garvis, and Dr. Ernst Thoutenhoofd—who helped me out with practical issues and keeping a close eye on my progress.

Finally, I am grateful for my family members and friends, both near and far, for their financial support and kindness, which has guided me throughout my growing passion for academic work.

Ernest Mensah

June 2020, Gothenburg

Table of contents

List of Tables and Figures

List of Abbreviations

$\chi 2$	Chi-square
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
α	Cronbach's alpha
EER	Educational Effectiveness Research
ESCS	Index of Economic, Social, and Cultural Status
HER	Home Educational Resources
HSS	Home study support
IEA	International Association for the Evaluation of Educational Achievement
ICC	Intraclass correlation coefficient
INQUA	Instructional Quality
IPO	Input-Process-Outcome Theoretical and Analytical Model
MSEM	Multi-level structural equation modelling
N	Sample size
OECD	Organization for Economic Co-operation and Development
PEDU	Parent education level
RMSEA	Root Mean Square Error of Approximation
RQ	Research Question
SEM	Structural Equation Modeling
SES	Socio-economic status
SPSS	Statistical Package for the Social Sciences
SRMR	Standardized Root Mean Square Residual
TCM	Teacher Confidence in Mathematics
TIMSS	Trends in International Mathematics and Science Study
TLI	Tucker-Lewis Index

1 Introduction

1.1 Background

Education is of important benefits for the development of both individuals and society as a whole. It associates with qualified workforce and economic growth, social mobility, higher adult numeracy and literacy levels, and better health and wellbeing. Therefore, improving educational quality and providing education for all is of growing interests worldwide, and have become the most desirable goals stated in their policy documents in all educational systems (Van Damme & Bellens, 2017). They also are deemed relevant as two dimensions of effectiveness in education (Creemers & Kyriakides, 2010; Nachbauer & Kyriakides, 2019). This is evident by, for example, the Strategic Framework for Education and Training of the European Union (European Union Commission, 2016), and the law of No Child Left Behind Act (NCLB) of 2001 in the USA, and the National Development Plan 2030 for South Africa.

Quality and equity, as outcomes of an education system, imply that schools or education system of a country (e.g., South Africa) should evenly distribute the benefits of education among all students. There should not be any barriers to access and participation in education. Moreover, the expected outcomes of students from education should not be affected by their socioeconomic status (SES) or other backgrounds, such as gender, ethnicity or religion (e.g., Takyi et al., 2019).

Unfortunately, ensuring all learners to have the equitable opportunity and quality in education remains a challenge in many educational systems (Gorard & Smith, 2004; Van Damme & Bellens, 2017). It was observed in the Trends in International Mathematics and Science Study (TIMSS 2015) that the mathematics score for grade 8 students improved in many countries. Still, the achievement gaps concerning socioeconomic and ethnic background also have increased significantly (Mullis, Martin, & Loveless, 2016). South Africa, for example, although the government promotes a solid legal policy for the right and accessibility to education for everyone, irrespective of gender or ethnicity, a proper mechanism for the effective fulfilment of the quality and equity in education is still lacking. Inequity in achievement has increased in South African schools in recent years (Department of Education [DoE] 2003, 3; Frempong, Reddy, & Kanjee, 2011; Spaull, 2019). The South African learners performed significantly worse in TIMSS achievement, compared to all other developing countries in the study (Howie & Pietersen, 2001; Reddy et al., 2019). The average mathematics achievement in TIMSS 2015 in South Africa is 372, far below the international mean of 500 points. This indicates great challenges in education quality provision (Alex & Juan, 2017; Arends, Winnaar, & Mosimege, 2017; Frempong et al., 2011; Sayed & Ahmed, 2011; Visser, Juan, & Feza, 2015).

To achieve their educational goals of high performance and high equity, South African will need high-quality teaching for every student. Teacher and teaching quality have consistently been found to be substantial for student's academic achievement and well-being (Canales & Maldonado, 2018; Slater et al., 2012; Wayne & Youngs, 2003). However, despite the substantial evidence that teachers matter in student learning, there remains no agreement as to which teacher quality is most consequential for student achievement (Canales & Maldonado, 2018; Gustafsson, 2003; Rivkin, Hanushek, & Kain, 2005; Scheerens & Blömeke, 2016).

Reviews of the studies investigating the effects of teacher education, certification and years of formal teaching experience on student outcomes have been conducted (Darling-Hammond, 2000, 2014; Goe, 2007; Wayne & Youngs, 2003). In essence, several teacher characteristics seem to be consistently cited as important teacher input that may contribute to explaining variation in teacher 'effects' affecting student achievement. Much research has focused on investigating the influence of a particular aspect of teacher quality, for example, teacher education or certification on student outcomes. However, very little evidence exists on the relationship between instructional quality and student's mathematics achievement conditioned on student's family socioeconomic background and teacher's qualification.

Moreover, most of the research in this area has centred on individual countries, such as United States (Goe, 2007); Nordic region (Blömeke, Olsen, & Suhl, 2016; Nilsen & Gustafsson, 2016) mostly with TIMSS 2007 and 2011 dataset; and Germany (Atlay, Tieben, Hillmert, & Fauth, 2019). Only a few came from developing countries (e.g., South Africa included) (Frempong et al., 2011; Sayed & Ahmed, 2011; Visser et al., 2015). There is no previous attempt observed to investigate these attributes in TIMSS 2015 in South Africa. Therefore, Sayed and Ahmed (2011) highlighted that quantitative evidence is not enough in South Africa. Probably, this may limit our understanding of the impact of teacher qualification and characteristics, teacher instructional quality on student learning outcomes, controlling for student's SES, and classroom context. This topic thus deserves further research, especially in developing countries like South Africa.

1.2 Aim and relevance of the study

This study aims to investigate the relationship between teacher qualification and characteristics, teacher instructional quality, students' family socioeconomic background, and student mathematics achievement with the South Africa data from TIMSS 2015. It can be hypothesized that teacher qualification (indicated by teacher experience, formal education, and teacher major or specialization), teacher instructional quality, and classroom SES composition have significant effects on 9th graders' mathematics achievement in South African, after taking into account students' SES. The study may promote a precise explanation for variation in teacher effectiveness and student achievement, in terms of teacher characteristics and their classroom practises, conditioning on student's family SES background. Given this, I anticipate that this study would lead to findings and conclusions that can provide policy implications and practice recommendations that are useful for improving mathematics education in South Africa.

1.3 Organization of the rest of the study

The organization of the rest of this thesis is structured as follows. Section two addresses the theoretical framework. Section three undertakes a review of the relevant literature. Section four presents the research questions. Section five depicts the empirical approach of this study: methodology and data analysis, which contains detailed data sources, sample and sampling strategy, and variables of interest, reliability, and validity issues, analytical method-MSEM introduction, and modelling process. Modelling description and the findings from the model are presented and discussed in sections six and seven. Section eight to eleven is dedicated to the main findings and discussion of the results. Section twelve concludes with ethical consideration and concerns, policy implications, and practice recommendations of the findings, limitations, and further research, as well as the contribution of the study.

2 Theoretical Framework

With regards to the research problems, the current thesis tries to underpin teacher quality in South Africa via relating teacher qualifications and characteristics, instructional quality and socioeconomic background to student achievement with TIMSS 2015 data. The dynamic model of educational effectiveness (Creemers and Kyriakides, 2008) is adopted as the theoretical framework in understanding the mechanism of teacher quality within the South Africa context. The dynamic model explains how various factors situated at different organizational levels of education (e.g., student, classroom, school, and system-level) are hypothesized to intertwine and explain variations in student outcomes. Furthermore, the conceptualization of teacher and teaching quality, i.e., the phenomenon of teacher inputs, processes, and outcomes (IPO model; e.g., Goe, 2007) is used as the analytical guidance to my research, because it provides a suitable theoretical perspective for investigating the South Africa context.

2.1 Educational effectiveness research (EER)

According to Reynolds et al. (2014), the origins of EER stem from the reactions to the findings in the famous Coleman et al. report (1966). In that report, the differences in student outcomes are almost entirely explained by student's background characteristics, together with class and school composition concerning these characteristics. This leads to the conclusion that only a small proportion of the variation in student achievement could be attributed to schools or educational factors.

Creemers and Kyriakides (2008) present a summary of the most important characteristics of educational effectiveness research (EER) as follows:

> The main research question of educational effectiveness research EER is what factors in teaching, curriculum, and learning environment at different levels such as the classroom, the school, and the above-school levels can directly or indirectly explain the differences in the outcomes of students, taking into account background characteristics, such as ability, SES, and prior attainment. (p.12)

Basically, the main goal in EER is to investigate the possibility of educational factors to explain a diverse range of student outcomes, and these educational factors are situated at multiple levels of the education system (e.g., student-level factors, classroom-level factors, school-level factors, and system-level factors). The basic EER model was further developed into the dynamic model of educational effectiveness to study the abovementioned factors (Creemers & Kyriakides, 2008).

2.2 Dynamic model of educational effectiveness

The theoretical basis for this study draws essentially from the dynamic model of the educational effectiveness of Creemers and Kyriakides (2008, see Figure 1). Because there are factors in education that need to be understood to explain variations in student outcomes, the dynamic model takes into account the complexity of an important process of various factors in education situated at a different level to influence student outcomes. These processes have a nested structure, meaning that students are nested in classrooms, classrooms are nested in schools and schools in a system context. The dynamic model is a framework of the mechanisms about the educational factors within each level and across different levels related. The dynamic model thus provides the current thesis with a multilevel theoretical perspective that facilitates statistical modelling of the EER related research questions. It also highlights the importance of investigating both direct or indirect effects among the theoretical factors and student outcomes.

The validity of the dynamic model at the classroom-level has been found to be more significant for student results than the school level (Kyriakides et al., 2000). In addition, studies within this research tradition confirm that the dynamic model is also applicable to deal with policy and practice in teacher education, for instance, about "what works" at school and classrooms (Reynolds et al., 2014). According to Reynolds et al. (2014), the dynamic model places emphasis on the teacher and teaching quality and uses factors operating at different levels in defining effective teaching by focusing on factors related to student outcomes.

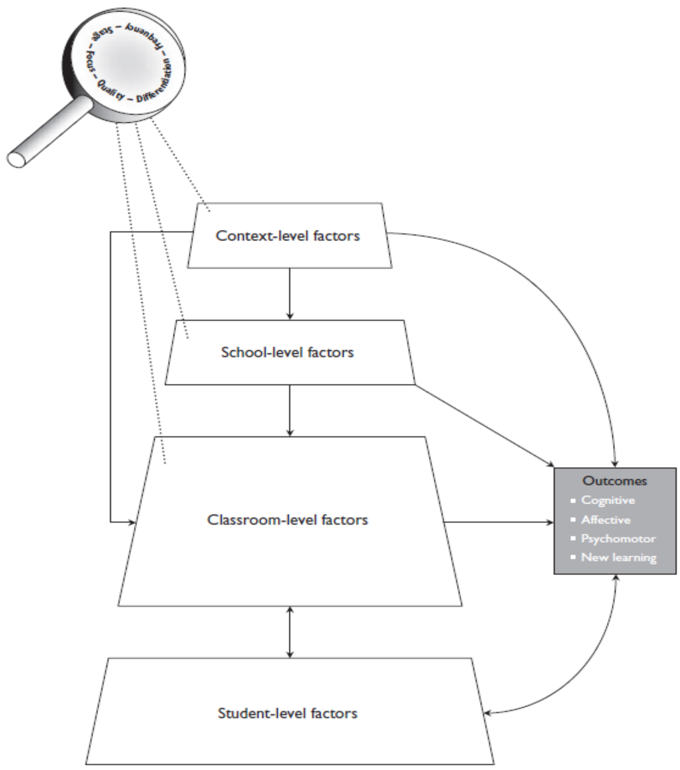

Figure 1. The main structure of the dynamic model of educational effectiveness (Creemers & Kyriakides, 2008).

Furthermore, to understand and study educational quality and equity issues in South Africa context, classroom-level factors such as indicators of teacher quality via teacher qualifications (indicated by teacher experience, teacher education, and teacher major or specialization) and teacher characteristics, teacher instructional quality, and student-level factors such as student's background variables (e.g., SES; the amount of books at home, level of parental education and number of home study supports) have been considered. It was observed that the complexities of studying the degree of which possible inputs affect an outcome involve variables that relate to one or more of the levels in the education system (Blömeke et al., 2016, p. 24). Fortunately, Trends in International Mathematics and Science Study (TIMSS) starts to fill these complexities by providing data at both the student and classroom or teacher level. TIMSS contextual questionnaire relates satisfactorily with the factors identified in the Dynamic Model, which are context-related factors, school-related factors, classroom-related factors, and student-related factors (Creemers & Kyriakides, 2008). This model allowed for the investigation of teacher qualification and characteristics, teacher instructional quality, classroom SES composition, and their relationship with mathematics achievement. The cognitive outcome factor highlights the importance of student-level factor, and is also applied in the study and such outcomes in TIMSS assessment is related to student mathematics achievement. Consequently, the theoretical model which leads the selection of variables in the educational context would be further discussed in the Input-Process-Outcome framework section below.

2.3 Conceptualizations of teacher and teaching quality (IPO Model)

Goe's (2007) reviewed a myriad of EER studies and conceptualized teacher and teaching quality into the framework of the so-called inputs, processes, and outcomes (IPO) model. The IPO model is widely applied in educational effectiveness research (EER), which effectively leads the selection of variables and specification of statistical models in educational context (Grabau & Ma, 2017). The model is feasible to be used and to combine with the EER dynamic model, and can serve to identify educational factors contributing to variation in student outcomes.

According to Grabau and Ma (2017), the IPO model is seen as a technique of multilevel modelling that deals with data with a hierarchical structure such as students nested within the classroom in turn nested within schools. In the IPO model, the components of the framework are presented (see Figure 2). Thus, teacher inputs consist of teacher qualifications such as education, certification status, and experience as well as teacher characteristics such as attitudes and self-efficacy. Next to inputs, processes include teacher practices such as instructional delivery, classroom management and interactions with students and outcomes as a result of teacher effectiveness, empirically defined as value-added measures. Goe (2007) noted that teacher effectiveness is considered as growth in student learning outcomes and usually measured by achievement scores, particularly the standardized achievement test scores. Within this framework, the two "inputs" (teacher qualifications and teacher characteristics), and the process measure (teacher practices) are used to define teacher quality and 'exist independently of student achievement'. Whereas the outcome measure (teacher effectiveness) is entirely dependent on student achievement (p. 9). This means that teacher effectiveness cannot be determined without student outcomes (Goe, 2007).

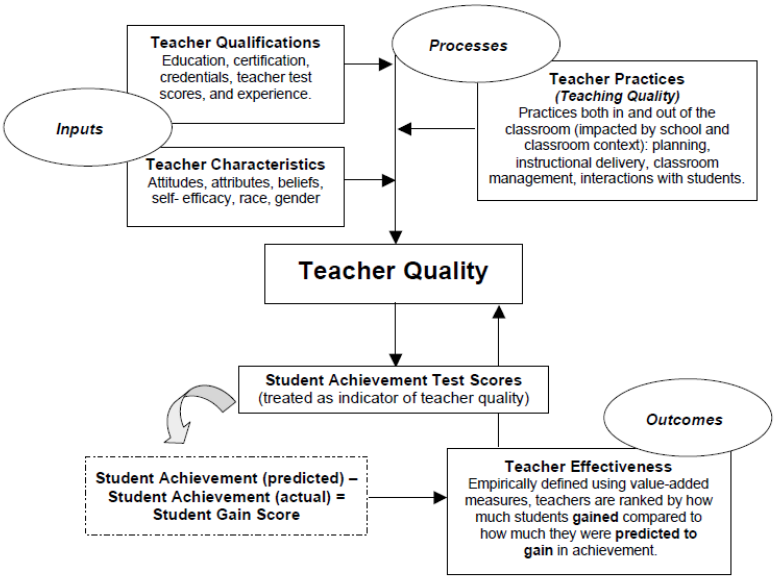

Figure 2. Goe's (2007) framework for teacher quality; IPO model.

Consequently, the two interdisciplinary models—the dynamic model of educational effectiveness and the IPO model provide a suitable theoretical basis for this 2015 TIMSS secondary data analysis for the South Africa perspective. They reflect not only support to identify the variables at different levels, i.e., factors situated at the student, classroom, and school-level factors contributing to variation in student outcomes but also attempted to investigate and explain complex interrelation between these factors at multiples levels. Both interdisciplinary models allow for a statistical technique of multilevel modelling that looks at data with hierarchical or nested structure. They also make a significant contribution informing the subsequent research design, for example, to determine which variables to be included in the hypothetical model based on the research questions and the aim of the study.

3 Literature review

This section critically reviews existing research on educational quality and equity in South Africa and around the world. Firstly, the concepts of *educational quality* and *equity* will be defined. Secondly, previous research on the relationship between teacher qualification, instructional quality and student outcomes will be reviewed. Next, this section highlights and discusses studies on teacher quality, emphasizing teaching experience, teacher education, teacher characteristics, and their instructional practices perspectives. The need to consider students' family socioeconomic background and student achievement also is discussed in a South Africa context.

3.1 Educational quality

3.1.1 Definition

Educational quality is described as a multi-faceted and multidimensional construct. It can be measured and viewed from different perspectives, and it appears to challenge all attempts at reaching any consensus as to its meaning. So how do we handle such concepts? How should we explain the fact that it is so difficult for us to find a universal, agreed-upon definition? However, some considerable agreement on the basic dimensions of quality has been reached. According to Adams (1993), the analysis of quality should consider the degree to which schools are able to meet specific objectives and achieve the desired level of accomplishment (p. 7). The term *effectiveness* has often been used synonymously due to the complexity and multi-faceted nature of quality (Adams, 1993). In some instances, the terms, efficiency, and equity have been used synonymously.

Increasingly, however, at different stages of a country's economic development, priority is given to different types of indicators. Thus, to some extent, quality is commonly viewed as a "self-evident goal for countries, based on research, stating the positive effects of education on different domains" (Van Damme & Bellens, 2017, p. 128). This definition implies that quality is committed to the normative settings of society. Hanushek and Woessmann (2008), for example, studied the role of cognitive skills in promoting economic well-being, with a particular focus on the role of school quality. They established that among the numerous indicators of quality, (e.g., student outcomes, years of schooling, and dropout rate and so forth), the cognitive achievement is the most important predictor of a diverse set of desirable educational outcomes (Hanushek & Woessmann, 2008).

When it comes to evaluating countries' effort towards quality, most researchers focus merely on the average level of achievement. It was observed that most of the educational effectiveness studies so far have examined the quality dimension, and hence school and teacher effectiveness is measured by looking at average student achievement (Nachbauer & Kyriakides, 2019).

In the present study, educational quality, which will be henceforth referred to as teacher *effects* to student learning outcomes, entails examining the degree to which the teacher's qualification and teacher instructional quality achieve educational objectives through the level of educational attainment of the student. Basically, in the absence of a generally accepted definition of educational quality, this thesis regards the educational quality as teacher quality that comprises teacher qualifications and characteristics (inputs) that influence teachers' instruction practices or quality (process) and student outcomes (e.g., achievement and motivation) (Goe, 2007). Accordingly, one can look for an acceptable and workable definition of educational quality in terms of teacher quality, and student achievement. Further, educational quality is inextricably linked to equity and is most frequently viewed as trade-offs (Adams et al., 2012).

3.2 Educational equity

3.2.1 Definition

The term *educational equity* is a multidimensional concept (Nachbauer & Kyriakides, 2019), and offers the most important explanations for student performance differences, and thus finding consensus or a universal definition seems impossible. In these premises, only some of those many definitions will shortly be presented in this section. Educational equity has been connected within the issue of social justice, asking the question "how can we contribute to the creation of a more equitable, respectful, and just society for everyone?" and is defined as the absence of a relationship between student achievement and background characteristics (Van Damme & Bellens, 2017). This notion implies that there is a need to make a distinction between the terms *equality* and *equity* (Nachbauer & Kyriakides, 2019); where an individual should preferably be treated based on their needs (Takyi et al., 2019). Since the concepts of equality and equity carry different assumptions about the goal of education, educational equality is taken to mean that schools should offer the same access education to all students. In contrast, equity considers the different needs of individual or learning opportunities for students of different backgrounds.

Building on this assumption, the global concern related to education is shifting away from equality and instead focusing on equity. Moreover, equity in education has been differentiated and identified as two dimensions: mainly equity as fairness and equity as inclusion (Field, Kuczera, & Pont, 2007; Kyriakides, Charalambous, Creemers, & Dimosthenous, 2019; Nachbauer & Kyriakides, 2019). Therefore, equity as fairness implies that personal and social factors that are unlikely to change, such as gender, socio-economic status or ethnic origin, should not be an obstacle to achieving educational potential. Also, the inclusion perspective has implications for ensuring a basic minimum standard of education for all, and in such a way everyone would be able to read, write and do simple arithmetic (Field et al., 2007).[1]

3.3 Reviews of the previous research

Studies about that the relationship between student achievement and teacher and teaching quality have consistently shown that teacher and teaching quality are substantial for student's academic achievement (Canales & Maldonado, 2018; Luque et al., 2020; Slater, Davies &

[1] It should be noted that since the concepts of educational quality and equity are multidimensional and do not have an appropriate and comprehensive definition, a specific definition was used in this study and may, consequently, differ from other definitions that can be observed in scholarly literature.
Caution must be taken because there is no absolute consensus on the definitions and appropriate meaning of educational quality and equity, and thus finding a comprehensive definition is unlikely to be found. One possible reason for this difficulty may due to differences in countries' expectations.

Burgess, 2012; Wayne & Youngs, 2003), especially for disadvantaged and low-performance students (Darling-Hammond, 2000; Goe, 2007; Hattie, 2003; Kyriakides, Christoforou, & Charalambous, 2013; Nye, Konstantopoulos, & Hedges, 2004). However, there has been no clear consensus on which aspects of teacher and teaching quality matter most for student outcomes (Canales & Maldonado, 2018; Gustafsson, 2003; Rivkin et al., 2005; Scheerens & Blömeke, 2016).

Following Booth et al. (2016) and Jesson et al. (2011) and Petticrew & Roberts (2008), the literature review was systematically undertaken in line with features that make a review systematic (see Appendix 1 for the description of reviewing process).

3.3.1 The Relation of Teacher Quality and Student Achievement

A number of studies have found a relation between an aspect of teacher qualification, teacher instructional quality and student outcomes. Blömeke et al. (2016) analyzed the grade 4 data in TIMSS 2011 using a multilevel structural equation modeling technique (MSEM) and noted that teacher qualification was significantly related to instructional quality and student achievement, while student achievement was not always predicted by instructional quality. They also found an average effect size, estimated around 0.20, which is a notable effect in the study. The significance of teacher qualifications varies across countries around the world, suggesting that the relationship between teacher qualification, instructional quality, and student achievement needs to be further investigated. Similarly, Blömeke and Olsen (2019) investigated the effects of teacher quality on the fourth and eighth-grade student's mathematics and science achievement in 5 countries using TIMSS 2011 data. They too found an average effect size estimated around 0.12 for the relationship between teacher qualification, instructional quality and student achievement. These authors mentioned that little consistent patterns within countries existed regarding instructional quality as predictors. These studies have found little evidence for the significant effect of instructional quality on achievement. This may be due to the varying operationalization of instructional quality (Akiba et al., 2007; Luschei & Chudgar, 2011). Also, Scherer and Nilsen (2016) and Blömeke et al. (2016) who examined the consistency of relations between teacher qualification, instructional quality, and student achievement warned against quick generalization.

3.3.1.1 Teacher Qualifications

As observed by Goe (2007), the evidence for teacher qualifications indicated by teacher experience, teacher education, certification, and test scores are significantly related to student academic achievement. In these premises, the study will only focus on two aspects of teacher qualifications, namely: teacher experience and education. These would further be discussed below.

3.3.1.1.1 Teacher experience

Teacher experience is usually measured by the number of years of service. In the studies examining the effects of teacher experience on student's learning outcomes (e.g., Nye et al., 2004), a positive impact is often found, i.e., the longer the in-service years of teachers, the better their students' achievement is (e.g., Wayne & Youngs, 2003). Using 50-state data from the National Assessment of Educational Progress (NAEP) mathematics and reading test scores, it was found that the effects of teacher experience on student achievement differ, depending on the degree of teacher effectiveness (Darling-Hammond, 2000).

Further, the evidence reported indicates that teacher experience matter, but still there remains some disagreement on how teacher experience is being measured. Even though in-service years as a teacher is very often used to measure teacher experience, other proxies of teacher experience are also used, for example, subject specificity, and this may cause the estimate of teacher experience effect to differ across countries and studies. This is why some studies claim that the effects of teacher experience are likely to vary based on grade level and subject-specific (Goe, 2007; Wayne & Youngs, 2003). Similarly, the experience effects of teaching may reflect the impact of labour market conditions for teachers (Wayne & Youngs, 2003). Besides, Blömeke et al. (2016), Nilsen & Gustafsson (2016), Hanushek & Luque (2003), and Luschei & Chudgar (2011), reported that the relationship between student achievement and teacher experience is not statistically significant, and thus inconsistent.

Meanwhile, researchers such as Akiba et al. (2007) corroborated the importance of teacher experience relating to grade four and eight mathematics teachers across different TIMSS cycle. Interestingly, some studies have contended that experience matters most in the first year (Rivkin et al., 2005), some have said after first few years (Nye et al., 2004), some pointed out three years (Akiba et al., 2007), others have claimed 4 or 5 years (Goe, 2007), and still, others have mentioned ten years (Papay & Kraft, 2015; Wiswall, 2013). The evidence that experience effect seems to level off after five years, has found support in previous research (e.g., Darling-Hammond, 2000). Still, some researchers have put forward that teachers reach a certain peak in their careers, particularly 19 years after which their contribution to student achievement decline (Toropova et al., 2019). The Trends in International Mathematics and Science Study (TIMSS), for example, have shown that teachers of mathematics in grade 8 who teach mathematics have on average 15.77 years of experience.

3.3.1.1.2 Teacher Education

Research in education commonly employs teacher degree and the main academic disciplines as indicators of teacher's education. Toropova et al. (2019) stress that teacher educational level, certification status, and coursework are indicators of teacher knowledge. Moreover, teacher knowledge may analytically be differentiated into both teacher education and experience. Also, Wayne and Youngs (2003) highlighted that measures of teacher knowledge are subject-and grade-specific. This suggests that domain-specific (e.g., subject-and grade-specific) may be more critical for successful teaching, thereby contributing to student outcomes. In the case of mathematics instruction, teachers with higher degrees and who majored in mathematics during their level of education seem to be positively related to their student's mathematics achievement (Goe, 2007), subsequent studies have supported these conclusions (e.g., Baumert et al., 2010). Further, a consensus across domain specificity (subject and grade-specific), has emerged concerning the relationship between teacher education and student achievement (Goe, 2007).

It has also been suggested that more detailed and precise measures of teachers' education tend to be better predictors (Johansson & Myrberg, 2019). While other studies generally confirmed the relevance of teachers for student achievement, some research on the effects of teacher education is still conflicting or inconclusive. For instance, some studies revealed almost no significant relationship exists between teacher education, (e.g., teacher degree), and student achievement (Blömeke et al., 2016; Luschei & Chudgar, 2011). The effect size from meta-analysis research on the relationship between teacher education and student achievement was estimated at approximately 0.10 (Hattie, 2008). This effect size is pretty negligible. These differences in opinion by these researchers may be because teacher education training may vary across countries, and in some instances within countries (Blömeke et al., 2016; Wayne & Youngs, 2003).

3.3.1.2 Teacher confidence

Confidence is defined as one's perception of self-regarding achievement in school (Reyes, 1984, p. 559). For example, it reflects a teachers' sense of personal ability in successfully completing a specific activity. It is worth noting that the belief that an individual can organize and execute a particular activity is often discussed under the heading of self-confidence or self-efficacy (Bandura, 1997). To avoid confusion, based on my theoretical framework, the concepts *self-efficacy* will not be used; instead, *teacher self-confidence* is used.

Teacher's confidence, as one of the important characteristics, play a crucial role in student achievement. It has commonly been stated that the influence of teachers' confidence levels contributes to student performance. It was observed that teachers' self-confidence in their teaching skills is not only related to their professional behaviour, but also with students' performance and motivation (Bandura, 1997; Henson, 2002). Other researchers (Klassen et al., 2011; Klassen & Tze, 2014; Tschannen-Moran et al., 1998) have concluded that the teachers' self-confidence is a vital aspect of teacher competence that influence teachers' instructional practices. These findings were also supported by Beswick (2007). Moreover, there has been a focus on the relationship between teacher self-confidence, (e.g., teacher self-efficacy) and student achievement at the classroom level (Wyatt, 2014; Zee & Koomen, 2016).

In TIMSS 2015, teachers are asked to indicate how confident they feel about teaching mathematics to the TIMSS class. For example, "Inspiring students to learn mathematics", "Showing students a variety of problem-solving strategies", and "Improving the understanding of struggling students" (See **Sub-Section 6.4.3.2** for more detail).

It was found that a construct that is closely related to teacher confidence is that of teacher's instructional practices which play a role in student learning outcomes (Tschannen-Moran et al., 1998; Zee & Koomen, 2016). From these contributions from various researchers prove the fact that teacher confidence appears to have particular relevance to teacher instructional practices.

3.3.1.3 Instructional Quality (InQua)

Apart from teacher qualifications such as teacher experience and education that have proven to be paramount to student's achievement in a range of studies, quality of teaching also matters for student educational outcomes. Instructional quality is a construct that reflects the features of teachers' teaching practices well known to be positively associated with student outcomes, both cognitive and affective ones (p. 5, Nilsen et al., 2016). The definition focuses on the measures of the classroom process concerning student outcomes. Some studies reported that constructs about teacher instructional practices had shown more significant relationships with teacher self-efficacy, especially at the primary school level (Toropova et al., 2019). Further, instructional quality is presumed to be an important aspect of an individual teacher's characteristics (Bellens, Van Damme, Van Den Noortgate, Wendt, & Nilsen, 2019).

Also, majority of the scholarly literature reviewed revealed that instructional quality is multi-dimensional, including three main aspects: classroom management (CM), cognitive activation (CA) and supportive climate (SC) (Atlay, Tieben, Hillmert, & Fauth, 2019; Bellens et al., 2019; Fauth, Decristan, Rieser, Klieme, & Büttner, 2014; Kunter & Baumert, 2006; Nilsen et al., 2016). As stated by Fauth et al. (2014), the first dimension, can be thought about in terms of classroom management, and which "focuses on classroom rules and procedures, coping with disruptions, and smooth transitions" (p. 2). This dimension focuses on using strategies to solve disrupting behaviour (Atlay et al., 2019), and serves as a prerequisite for learning (Bellens et al., 2019). SC is the second dimension of instructional quality which is "constituted by characteristics of teacher-student relations, feedback by the teacher, mutual respect and a proactive attitude towards student mistakes and misunderstandings" (Atlay et al., 2019, p. 3). This construct is probably related to features of social interactions in the classroom. A third dimension is CA, "addresses features of the instruction, which facilitate students' conceptual understanding" (Atlay et al., 2019, p. 2). Also, it determines the "types of problems selected and how they are implemented" (Kunter & Baumert, 2006, p. 235).

While there is an agreement of researchers on the three global dimensions of instructional quality, there remains some concern from other studies. Surprisingly, researchers within the same strand of research found clarity of instruction (CI) as the fourth dimension of instructional quality (Nilsen et al., 2016). Also, there is a similar pattern of contradictory research which places emphasis on four dimensions of instructional quality (Blömeke & Olsen, 2019). One could argue that the field of measuring instructional quality is disintegrated or perhaps fragmented (Bellens et al., 2019).

Instructional quality has been observed as best indicators of examined teaching practices (Blömeke & Olsen, 2019), and traditionally assessed through students, teachers, researchers, and or external observers (Toropova et al., 2019). Student-assessed instructional quality has been applied, for example, as a predictor of student mathematics achievement (Wagner et al., 2016). The student perceived instructional quality scale was developed by Fauth et al. (2014) and was one of the most commonly known scales measuring the three major sub-dimensions of the student questionnaire (Bellens et al., 2019). Similarly, the teacher-rated instructional quality in the Trends in International Mathematics and Science Study (TIMSS) also measure the three main aspects of the instructional quality construct in the teacher questionnaire. Examples of the InQua indicators are, "use questioning to elicit reasons and explanation", and "praise students for good effort". However, Fauth et al. (2014) emphasized that student-rated InQua is a reliable source in assessing teacher's instructional quality in primary school levels.

However, Bellens et al. (2019) found different factor structure of InQua in Germany, Norway and Belgium. They pinpointed that other existing instructional quality sub-dimensions existed across different educational systems (e.g., Kunter & Baumert, 2006). Likewise, PISA[2] has different indicators to measure the three dimensions of instructional quality (OECD, 2013). In TIMSS, students were endorsed according to their degree of agreement on the students' view on engaging teaching in mathematics lessons which reflects some aspects of instructional quality (Mullis, Martin, Foy, & Hooper, 2016). Meanwhile, the student-rated InQua was criticized for being biased by teacher popularity; for example, a simple operationalization is the item "I like my teacher"(Fauth et al., 2014). Given this, it was noted that teacher popularity is assumed to contaminate student ratings of instructional quality. Further, the presence of a researcher (e.g., an external observer), might be affected or contaminate teachers' behaviour (Bellens et al., 2019). Also, teaching strategies are equally to be affected too. Therefore, teacher popularity in student ratings should be concerned.

Atlay et al. (2019) did a study in this respect by examining the relationship between the three major dimensions of instructional quality and student background using German panel data (longitudinal data from PISA-I-Plus). These authors reported three major findings: First, they found positive relations between CM and student achievement. Second, students with high socioeconomic background seem to have a strong relation between CA and SC compared to students with middle and low SES background. Thus these two dimensions of instructional quality positively moderated the relationship between SES background and achievement levels, and as a result leading to a larger achievement gap. Similarly, Bellens et al. (2019) analysed TIMSS 2015 grade four mathematics data across three countries to study the dimensionality of InQua. They too found that CM is related to student achievement but CA and SC did not. These authors suggested that effort is needed to solve the issues at stake to capture SC and CA dimensions. This little consistent finding indicates that the relations between each of these major constructs and student outcomes might not be established in most studies. Indeed, instructional quality has also been investigated as a mediating variable between teacher qualifications like teacher knowledge (e.g., teacher experience and education) and student achievement (Toropova et al., 2019).

3.3.1.4 Student background and their academic achievement in an international perspective

Educational equity is measured by, for instance, the effect of students' socioeconomic background (SES) on their achievement and the gap in outcomes between low-and high-SES schools and students. Therefore, one frequently used component is the strength of the SES factor when predicting achievement (Akiba et al., 2007; Yang-Hansen & Gustafsson, 2004).

Admittedly, SES is commonly conceptualized as the relative position of an individual or family within a hierarchical social structure, based on their access to, or control over wealth, prestige, and power (Gustafsson, Nilsen, & Yang-Hansen, 2018; Mueller & Parcel, 1981). Research shows that SES is typically measured by parental education levels, parental occupation prestige, and family income (Caro & Cortés, 2012; Yang-Hansen & Gustafsson, 2004). Yang-Hansen (2003) identified various aspects of the SES, and they related differently to student learning outcomes.

[2] Programme for International Student Assessment, conducted by OECD.

However, there is no consensus on how to measure SES. Moreover, different explanations and operationalization procedures exist depending on what one wants to measure. For example, some studies used the number of books at home as a single SES indicator. PISA studies compute the index ESCS to measure student's family economic, social, culture status, and is a combination of the highest level of parent education, highest parental occupational position, cultural possessions and home possessions. TIMSS studies also created a home educational resources (HER) index, which is measured by different educational aids from student questionnaire. As part of the discourse, SES operationalization procedures may relate to data availability. However, a number of studies indicated that the number of books at home is a useful indicator of SES (Yang-Hansen et al., 2014; Sirin, 2005), and has made valuable contributions to educational research (Caro & Cortés, 2012). However, it was also showed that books at home mainly captures the cultural capital component of SES (Sirin, 2005; Yang-Hansen, 2003; Yang-Hansen & Gustafsson, 2004). Further, Yang-Hansen (2003) revealed that the cultural capital component of SES is the most important component relating to student outcomes.

It has frequently been asserted that a perfectly equitable education system would reflect students' achievement unrelated to their socioeconomic status (SES). The bottom line of concern is that educational inequities related to socioeconomic status (SES) are unfair and that they can be measured through the relationship between students' family socioeconomic background and student achievement. Previous research has confirmed that student socioeconomic status (SES) significantly affects student achievement (Yang-Hansen & Gustafsson, 2016). Further, the effect from four meta-analyses which is based on 499 studies (957 effects) on socioeconomic status (SES) found the effect size ($d = .57$), which is a notable influence on student achievement (Hattie, 2008). It is worth noting that Hattie (2008) defined effect size as follows: small, $d = 0.2$; medium, $d = 0.4$; and large, $d = 0.6$. This implies that what is really necessary when judging educational outcomes is the *size of the effects*. Similarly, White's (1982) meta-analysis revealed that the aggregated effect of socioeconomic status (SES) at the school level was $d = 0.73$, whereas the effect was $d = 0.55$ at the individual student level. As Sirin pointed out in 2005, the relationship between SES and mathematics achievement was estimated at $d = 0.70$. A study by Gustafsson (1998) found that in most countries, the effect of individual background factors on student achievement was relatively stable over time. However, in the case of South Africa, other studies indicate that students are segregated by SES in South Africa schools. Thus, the effect of SES on students' achievement level has been observed in high achieving schools (Frempong et al., 2011). This indicates that student background factors, for example, SES still determines academic achievement to some extent, and the issue of SES differences in outcomes could vary across countries.

Some research from South Africa using TIMSS 2011 data has indicated that student from high-SES backgrounds, who speak the language of test at home performed better in mathematics. They also found that the effect of SES as measured by school infrastructure such as school building has a positive effect on student achievement (Visser et al., 2015). These researchers also point out that the SES of a school has a significant effect on student performance (Spaull, 2013a; van der Berg, 2008).

Moreover, it is well-established that a large variety of variables are used as indicators of SES (White, 1982), and regarded as the most frequent used contextual variable in educational research (Sirin, 2005). Further, Sirin (2005), drawing on several conceptual frameworks to capture students' socioeconomic background suggest that conceptualization of SES have to consider (a) the unit of analysis for SES data (findings are likely to be contaminated when aggregated data or high-level data are used to make explanations at the student level or individual level effect), (b) the type of SES measure (influence the relationship between SES and student achievement), (c) the range of the SES variable (studies that used dichotomous SES variables are less likely to produce strong correlations), and (d) the source of SES data (factors such as student family background, age, and performance level are more likely to produce accurate reports).

It is reported by Sirin (2005) that SES has less predictive effect for minority students. Similarly, students with a lower level of SES or socioeconomically disadvantaged background are likely to perform worse over those students who come from affluent families (Caro & Cortés, 2012). Therefore, it can be observed that the gap between high and low SES students tends to widen as students get older, and hence, closing the gaps between and across students from varying socioeconomic backgrounds has important implications for student wellbeing.

3.3.2 Quality, equity and teacher situation in South Africa

South African's educational reforms envision educational policies. The country has seen a proliferation of educational policies in the post-apartheid South Africa education system (e.g., Sayed & Ahmed, 2011), and triggered significant adoption of 'inclusive' education (Frempong et al., 2011). The proliferation of these policies was intended to enhance quality education for all students regardless of their background characteristics. Nevertheless, the quality of South African's educational system has remained low and characterized mainly by inequality of educational opportunities (Spaull, 2013b). Two schooling systems can be seen in the South Africa education system. According to Spaull (2013b), the smaller wealthy population, which is better performing and provides students with required skills, and another for a larger and poor population, which is low-performing and poorly equipped to provide students with the knowledge they should be acquiring at school. Besides, in many countries of which South Africa is of no exception, where educational achievement remains strongly linked to family background, education may be attributable to reinforcing inequalities. South Africa is almost experiencing a highly unequal school system (Spaull, 2019). Therefore, learning that takes place in schools may be unequal due to student background characteristics, and this may as well limit students' abilities.

Moreover, it has become increasingly clear that inefficiencies and teacher quality in South Africa have been noted as a determining factor that is key to underachievement in education quality. The most striking example of this is that South Africa's primary school level mathematics teachers are least-knowledgeable in Sub-Saharan Africa. Therefore, many of these mathematics teachers, who serve poor and rural communities, have below basic levels of content knowledge. In several cases, these teachers cannot answer questions their students are required to answer based on the curriculum (Spaull, 2013b, p. 8). This issues needed policy attention to improving teacher quality, and also, as the main route to addressing instructional quality in South Africa.

3.3.3 Reviewed Gap

Across the region of the world, it was observed that the relation of teacher qualification, instructional quality on student achievement had been studied, it is mainly with regard to primary school students' mathematics achievement (Blömeke et al., 2016). Still, there exist very few concerning secondary school students' mathematics. Also, no evidence exists in investigating the relationship between teacher qualification (indicated by teacher experience, teacher education, and teacher major or specialization) and teacher characteristics, teacher instructional quality, students' family socioeconomic background, and student mathematics achievement with the South Africa data from TIMSS 2015. Besides, most of the research in this area focused on individual countries, such as United States (Goe, 2007); Nordic region (Blömeke et al., 2016; Nilsen & Gustafsson, 2016) mostly with TIMSS 2007 and 2011 dataset; and Germany (Atlay et al., 2019), whiles very little ones came from developing countries (e.g., South Africa included) using TIMSS 2011 (Frempong et al., 2011; Visser et al., 2015). Moreover, it also was observed that a study that employs a quantitative method in South Africa is not enough (Sayed & Ahmed, 2011). Therefore, there will be a need for closing the gaps above by using a quantitative research method, employing TIMSS 2015 data to promote a precise explanation to the variation in teacher effectiveness and student achievement, in terms of teacher characteristics and their classroom practises, conditioning on student's family socioeconomic background. Also, undertaken this study would lead to findings and conclusions that can provide policy implications and practice recommendations that are useful for improving mathematics education for South Africa. Furthermore, one of the biggest gaps has been the gap of Goe (2007) theoretical framework used in the study. None of the South Africa-based research literature has ever used this framework.

3.3.4 Summary of previous studies

A plethora of studies is conducted in the United States and Europe, within which most studies also have been undertaken in the Nordic region (e.g., Sweden and Norway) and Germany. In sum, the relationship between teacher qualification, instructional quality and student achievement, including equity dimensions have been primarily addressed in the Nordic region, particularly, in the Swedish and Norwegian context (Blömeke et al., 2016; Nilsen & Gustafsson, 2016). However, in Germany and the United States, researchers have investigated the subject domain and grade level, specifically at the secondary school level mathematics (Baumert et al., 2010; Goe, 2007). In general, the research pointed to the positive influence of teacher qualification and instructional quality on student's achievement after controlling for student's family SES. However, such positive impact often disappears or reduced once the contextual factor classroom or school SES composition being accounted for. Variation in the findings also is observed. As in the case of an education system like South Africa, studies on the indicators of educational quality or teacher quality and educational equity have been more infrequent. For example, few studies investigating into the quality of teachers (e.g., least-knowledge content) (Spaull, 2013a), and the dimensions between school quality and socioeconomic disadvantage (Frempong et al., 2011; Spaull, 2013a), and balancing equity, diversity, rights, and participation (Sayed & Ahmed, 2011) have been conducted.

4 Research Questions

Using the data from TIMSS 2015, this study investigates the following research questions in terms of teacher qualification and characteristics and their classroom practises, conditioning on student's family socioeconomic background to fulfil the aim.

1) Is there any relationship between teacher qualification, teacher characteristics, instructional quality, and classroom SES composition, and student mathematics achievement for South African nine graders?
2) To what extent do the teacher qualification and characteristics, instructional quality, and classroom SES composition affect 9th graders' mathematics achievement in South Africa?
3) Is SES the strongest predictor of South Africa grade nine student's achievement?

5 Method

This section presents the data sources, sample and sampling strategy that were applied in the thesis. It also discusses teacher data, matrix sampling, variables and measures, as well as the validity and reliability issues. The section concludes with the analytical method that has been taken into account the two-level structural equation modelling and analytical procedure.

Data is managed in a proper manner using SPSS V.25 to make the variables appropriate to be analyzed in Mplus V.8.3 (Muthén & Muthén, 1998-2017) for modelling and to answer the research questions.

5.1 Data and Samples

5.1.1 Data Source

The Trends in International Mathematics and Science Study (TIMSS) assessment is a large international study conducted by the International Association for the Evaluation of Educational Achievement (IEA). The project was initiated in 1995 by IEA and runs in a four-year recurrence. The assessment seeks to continue to track trends in mathematics and science at two target populations: the fourth grade and eighth grade. In TIMSS, students are selected based on the grade they attend (Reimer et al., 2018). According to IEA, the main goal of TIMSS is to evaluate mathematics and science knowledge in line with curricula in participating countries. TIMSS provides comparative information about educational achievement across participating countries to improve educational policies related to mathematics and science teaching and learning. The assessment also aimed at enables participating countries to examine and monitor their educational system by taking into account whether the student achievement scores are improving or not. The data for this present study was drawn from the 2015 TIMSS from South Africa, as this country is among those participating most frequently in the assessment cycles. South Africa participated in 1995, 1999, 2003, 2011 and 2015 cycles.[3] TIMSS 2015 is the sixth in a series of international

[3] In TIMSS 2007, South Africa did not participate.

assessments undertaken by IEA. The assessment data contains an ambitious contextual questionnaire (e.g., students, teachers, schools and home), and achievement tests both completed by students. Teachers of the assessed classes are required to complete a questionnaire. School principals also are asked to respond to a questionnaire. These various aspects of the contextual features may be relevant to the learning process and its educational outcomes. The assessment covers four content domains in mathematics, namely: number, algebra, geometry, data, and chance. The assessment also requires three behavioural levels or cognitive domain: knowing, applying and reasoning. These domains are to assess students' mastery of basic knowledge in mathematics.

5.1.2 Sample and Sampling Strategy in TIMSS 2015

TIMSS employs a two-stage stratified sample design or cluster sampling design for students in eighth grade. First, samples of schools are drawn as a first stage, which may be stratified. The second stage consisted of intact classrooms of students selected from each of the sampled schools (Joncas & Foy, 2010). This means that students are not sampled individually at the second stage of sampling but rather TIMSS select intact classrooms, as TIMSS has a focus on curriculum-based assessment. Importantly, the operational advantage of the selection of intact classrooms in TIMSS is that it provides the opportunity for researchers to study the relationship of classroom-level factors with student outcomes. Before continuing to describe the cluster sampling strategy in 2015 TIMSS for South Africa, it is worth noting that, TIMSS data is collected in such a way that the actual standard error is underestimated because of a hierarchical observational structure with students being nested in classrooms or schools. Additionally, the assumptions for estimating standard errors based on a simple random sample of independency of observations are violated, requiring special analytical methods to estimate the uncertainty of sampling accuracy (Rutkowski et al., 2010). Therefore, a multilevel structural equation modelling technique (MSEM) was applied to correct the standard error estimation, since individuals belonging to the same classroom or school are much similar, compared to students from different classrooms or schools. Also, with the help of multilevel SEM, it is possible to use the two-level modelling simultaneously to account for dependencies between individual observations, as well as to analyse individual-level data and the classroom-level data. According to Hox (2002), an assumption of independency of observations would produce standard errors which could lead to invalid significant results. The abovementioned description further attests to the appropriateness of the choice of the analytical method.

In the case of South Africa, the same sampling procedure was applied to ensure the sample representative of its population. First, the sampling process entailed sampling a list of schools stratified by school type (public and independent), province and language of instruction, followed by randomly selecting a class within the sampled school, after which intact classes participated in the survey. Further, participating countries are met with the minimum requirement of sampling 150 schools and a student sample of at least 4000. However, in a situation where a country's class size is smaller or not more than 30 students per school than expected, countries were advised to sample more schools or more classrooms per school to align with the minimum target of sample size 4000. Along with these, to ensure an acceptable response, experts from the National Research Coordinators put in place a minimum student response or participation rates of 85%, while 50 % or less are deemed to be not desirable and could lead to exclusion from the study. Approximately, South Africa TIMSS sample consisted of 334 mathematics teachers and 12514 students in 292 schools. Consequently, these samples were nationally representative of all grade 9 students in South Africa, with an average age of 15.7 years (LaRoche & Foy, 2016).

There is a point that needs clarifying. TIMSS is tested at the fourth and eighth grade internationally. However, the earlier assessment data for South African indicated that international grade 8 test was too difficult for south African learners (Spaull, 2013a). Consequently, a voluminous number of learners did not attempt many of the test items. This seems to be the biggest challenge when estimating achievement scores (Reddy et al., 2016). These authors highlighted that in TIMSS 2003, South Africa assessed grade 8 and 9 learners, but in TIMSS 2011 and 2015, grade 9 learners were assessed for better estimates of achievement scores.

5.2 Teacher Data

The Trends in International Mathematics and Science Study (TIMSS) sample teachers, as teachers are connected to students, and therefore teachers should be student representative. Rutkowski et al. (2010) point out that using teacher data is considered unbiased if it is connected to implications at the student level due to the sampling structure of TIMSS. Students must form the lowest level of analysis when using teacher-level data in analyses from TIMSS assessment (Rutkowski et al., 2010). In such a sense, a concept at the teacher level (classroom-level factors) may impact concept at the lowest level (student-level factors). Another critical issue for the choice of teacher-level data is the issue of the theoretical framework. As it was mentioned above, Goe's (2007) theoretical framework contains variables such as teacher qualifications (indicated by teacher experience and education), teacher characteristics (teacher self-confidence) and teacher instructional practices, situated at different levels (Creemers & Kyriakides, 2008). Thus, as the focus of the study is to investigate teacher effects at the classroom level, as well as students at the individual level, this description is one step further leading to the choice of the teacher level data. According to Gustafsson (2013), some mechanisms which operate strongly at student and classroom levels do not operate at other levels, such as the country level. However, for this thesis, Goe's theoretical framework is guided by the study to build up a hypothetical model and to be operationalized by the data.

5.3 Matrix Sampling

Several challenges stem from the sampling procedure of educational assessment data like IEA TIMSS assessment. TIMSS 2015 employs a matrix sampling technique, where each student responded to only a subset or answered partially of the extensive pool of the same items, and not the entire assessment. Also, there is a little bit of overlapping in different booklet students have, that is why TIMSS does not have a single outcome that applies for every participant, in that the items in each of the booklet have a missing. Rutkowski et al. (2010) point out that large scale assessments such as TIMSS and possibly PISA, use variations on a specific sampling procedure in that not all participants receive all the items. One possible reason for this is because of a large number of items and time limitations for each student. Therefore, the measurement of student's achievement scores is achieved with a measurement error (von Davier et al., 2009). Strictly speaking, it is particularly important to minimize the measurement error associated with each student's estimate. To estimate the uncertainty of sampling accuracy of the measurement in the large scale assessment, multiple scores or imputations technique, called plausible values (PVs), are presented for each student (Laukaityte & Wiberg, 2017). In this way, plausible values estimate a missing data analysis for each student. In general, plausible values are based on student responses to a part of an extensive pool of items they receive, and other relevant information of responses in test scores and background variables (Mislevy, 1991). For instance, in the case of mathematics, TIMSS assessment employs multiple imputation or PV to estimate student' scores in mathematics by using student responses to the items in the eighth-grade mathematics assessment and adjusting these responses in terms of the background data available for that particular student (Laukaityte & Wiberg, 2017). Typically, 5 plausible values are generated for each student, taking into account scales and sub-scales in the international database. In TIMSS, a single precise estimate for a student is not possible. Therefore, the dependent variable, which is the overall mathematics achievement for each student is represented by five plausible values, BSMMAT01 to BSMMAT05. The plausible values are set to have a mean of 500 and a standard deviation of 100 (Mullis, Martin, Foy, & Arora, 2012).

However, plausible values cannot be recognized as individual test scores, instead of as a measure of population performance. For instance, individual scores are not reported in TIMSS, but the assessment only estimates population parameters. According to Wu (2005), plausible values are a representation of *"the range of abilities that an individual might reasonably have based on their responses to test items"* (p. 115). Plausible values have been successfully used to describe the performance of a population or to estimate population characteristics compared with simply point estimates of abilities (Wu, 2005).

5.4 Variables and Measures

The primary variables of interest were the student achievement scores in mathematics, student's family background variables, three teacher qualification variables, as well as variables indicating teacher characteristics (e.g., teacher confidence) and instructional quality from the 2015 TIMSS.

5.4.1 Student Outcome: Mathematics Achievement (BSMMAT01 to BSMMAT05)

Mathematics achievement measured by the five plausible values[4] (BSMMAT01 to BSMMAT05) is used to describe students' mathematics competence level in South Africa in TIMSS 2015. The five plausible values were used as the endogenous variables (dependent variables) in this thesis. With regards to TIMSS 2015 international report, a set of four international benchmarks was provided for countries with considerable descriptive of what students know based on their mathematics scores (Mullis et al., 2016). Students with score

[4] The plausible values were standardized with a mean of 500 and a standard deviation of 100 internationally.

points between 400 and 475 are classified as low proficiency level, score points between 475 and 550 are intermediate achievement level, and score points from 550 to 625 are regarded as high proficiency and scores above 625 points as proficiency at an advanced level. As presented in the descriptive statistics table of mathematics achievement variable, *Table 1* shows that the valid sample size of N = 12514. It should be pointed out that the average students' performance for South Africa is below the international benchmark. The average mathematics performance is 372 points, which is far below the international mean (set at 500 points). It also can be seen that there is some variability in their abilities in different achievement scores presented in Table 1 when taken into account the mean and standard deviation of achievement scores. Since the five plausible values of mathematics achievement were estimations of population values, given all available information of achievement test as well as from student contextual questionnaires, the mean for these PVs are not exactly the same, and variation can be observed in both mean and SD values. Therefore, the current analyses use all five plausible values as recommended, which will provide more precise estimations of the population parameters.

The model estimations were run five times separately for each PV five mathematics achievement scores, and the results of the five analysis will generate a single value by averaging the estimated parameter or statistics. By doing so, the measurement error can be correctly estimated to ensure valid statistical inferences (Rutkowski et al., 2010).

Table 1. Descriptive statistics of the dependent variable "BSMMAT01 to BSMMAT05".

Plausible Values	*Variables Labels*	*N*	*Mean (M)*	*Standard Deviation (SD)*
BSMMAT01	1st Plausible Value Mathematics	12514	371.42	81.73
BSMMAT02	2nd Plausible Value Mathematics	12514	370.90	82.13
BSMMAT03	3rd Plausible Value Mathematics	12514	369.67	82.89
BSMMAT04	4th Plausible Value Mathematics	12514	369.24	82.56
BSMMAT05	5th Plausible Value Mathematics	12514	369.88	83.10

5.4.2 Student background variables

In this study, based on the available information in TIMSS 2015, the student's responded number of books in the home (BOOKS), number of home study support (HSS), and parents' highest education level (PEDU) are used as proxies for socioeconomic status (SES) background at the individual (within) level. The aggregated values of these individual proxies are used as indicators of classroom socioeconomic composition at the classroom (between) level.

The indicator of BOOKS contains the information on students' estimated number of books in the home (see Table 2), provided by students. This information is retrieved from the student questionnaire. The question contains five responses categories: (1) None or very few (0-10) books); (2) Enough to fill one shelf (11-25 books); (3) Enough to fill one bookcase (26-100 books); (4) Enough to fill two bookcases (101-200 books); (5) Enough to fill three or more bookcases (more than 200) (Martin et al., 2016). In table 2, the mean and standard deviation of these variables are presented. Mean value describes the central tendency, and standard deviation indicates a measure of spread or variation of a set of the data value around its mean value. The low standard deviation indicates that the data closely cluster to the mean or average (Bland & Altman, 1996).

Further, the indicator HSS was computed based on student responses concerning the availability of home resources study supports such as internet connection, own room or both.

Similarly, information concerning the parents' highest education level (PEDU) was derived from responses to the student questionnaire and was captured by five response categories: University or higher, post-secondary but not University, upper secondary, lower secondary, some primary, lower secondary or no school. The mean and standard deviation for the three indicators as a proxy for SES are presented in Table 2 below.

The choice of the derived indicators of SES stems from the lowest percentage of missing values in these indicators: 1.1% in students' estimated number of books at home (BOOKS), 3.4 % in the number of home study support (HSS), and 5.1 % in parents' highest education level (PEDU). Furthermore, the number of books at home is well established as a good measure of student background.

Table 2. Descriptive statistics for student estimated number of books in the home, the number of home study support, and parents' highest education level.

Indicator/ item	Question/ Variable labels	Mean	Standard Deviation
BOOK	About how many books are there in your home? (Do not count magazines, newspapers, or your school books)	1.89	.99
HSS	Number of Home Study Supports	1.22	.73
PEDU	Parents' Highest Education Level	3.18	1.69

5.4.3 Teacher background variables

5.4.3.1 Teacher experience, education and major

The variable of BTBG01 contains the information on teacher experience. Teacher experience is commonly used to represent the total number of years the teacher reported teaching (see Table 3). For teacher experience, "the number of years the teacher reported teaching" is used as a proxy. In other words, teacher experience is measured by the total number of years of service. This measure was preferred since it has a more explicit relation to my theoretical framework as proposed by Goe (2007). Teachers' experience is captured by the question: By the end of this school year, how many years will you have been teaching altogether? This is an open response question, and the years are reported in whole numbers. In an attempt to investigate the general impact of teacher experience, this variable is included in the study.

It should be noted that since the question for teacher experience was an open-ended item on a continuous scale, internal consistency of reliability is not necessary. This is because a single item cannot measure a latent construct, and cannot account for, and isolate measurement error.

For the level of teacher education, "The teachers' degree" is used as a proxy. In this way, to measure the level of formal education completed by teachers, the item, BTBG04 were used. This item, BTBG04 contains information about teachers' degrees from teacher education. It implies that teachers had to provide information about their degree. The indicator, BTBG04 is captured by the question, "What is the highest level of formal education you have completed?" Teachers' responses were coded in the ISCED system, ranging from "Did not complete ISCED level 3" to "Doctor or equivalent level-ISCED level 8". The indicator BTBG04 was included to investigate whether a large proportion of teachers with a bachelor's degree, master's degree or higher degree had a more positive effect on student achievement or not.

It should be noted that most South African teachers indicated having the highest level of education, and relates to University, this turns out to be a lack of variation in teachers' responses. Hence, no internal consistency of reliability is necessary in this regard.

Apart from the teachers' level of formal education, teachers also provided information about their major or main area(s) of study. Teachers were asked to rate their major or main area of study. The item, BTDM05 reflects whether teachers' main area of study was either mathematics or mathematics education.

The three indicators BTBG01, BTBG04 and BTDM05 to teacher qualification variables with their mean and standard deviation are represented in Table 3 below.

Table 3. Descriptive Statistics for number of years of teacher experience, teacher degree and teacher major (manifest variables)

(From IEA, TIMSS 2015 in the international database).

Indicator	Question/ Variable labels	Response categories	M	SD
BTBG01	Number of years of experience	Reported in whole number	13.73	9.77
BTBG04	Level of formal education completed/ Teacher degree	Did not complete upper secondary, upper secondary, post-secondary, non-tertiary, short-cycle tertiary, bachelor's or equivalent, master's or equivalent, Doctor or equivalent	4.67	.68
BTDM05	Major	Focus on either mathematics major OR mathematics education	3.84	1.04

5.4.3.2 Teacher Confidence in Mathematics

In recent cycles in TIMSS, the measure of teacher self-efficacy in mathematics included items in the "confidence in teaching mathematics scale", (hereafter referred to as TCM) that could be related to aspects of Bandura's theoretical foundation for the self-efficacy construct (Mullis et al., 2009). To investigate teachers' confidence in teaching mathematics to the TIMSS class, teachers of the students taking the eighth-grade TIMSS assessment were asked to indicate how confident they feel about doing each of the following: A = Inspiring students to learn mathematics, B = Showing students a variety of problem-solving strategies, C = Providing challenging tasks for the highest achieving students, D = Adapting my teaching to engage students' interest, E = Helping students appreciate the value of learning mathematics, F = Assessing student comprehension of mathematics, G = Improving the understanding of struggling students, H = Making mathematics relevant to students, and I = Developing students' higher-order thinking skills.

Responses were endorsed by teachers on a 4-point scale ranging from "Very high", "High", "Medium", and "low" with no midpoint. Teachers' responses to these items were recoded so that the high values reflect a higher assessment of teacher confidence in mathematics. This means that higher values indicate greater teacher confidence. The items in the teacher confidence in mathematics scale or index TCM scale are presented in Table 4. Teachers of mathematics in the South Africa sample report above-average levels of self-confidence in all aspects. Mean confidence levels appear to be nearly the same across the dimensions, except for the somewhat lower mean for teacher confidence for providing challenging tasks for the highest achieving students, and also for improving the understanding of struggling students.

Table 4. Descriptive Statistics for Teacher Confidence in Mathematics.

Note. _R Item was reverse coded.
(From IEA, TIMSS 2015 in the international database)

Indicator/ item	Question/ Variable labels	Response categories	M	SD
BTBM17A_R	Inspiring students to learn mathematics	Very high, high, medium, low	3.52	.596
BTBM17B_R	Showing students a variety of problem solving strategies		3.40	.596
BTBM17C_R	Providing challenging tasks for the highest achieving students		3.01	.804
BTBM17D_R	Adapting my teaching to engage students' interest		3.25	.656
BTBM17E_R	Helping students appreciate the value of learning mathematics		3.41	.649
BTBM17F_R	Assessing student comprehension of mathematics		3.20	.686
BTBM17G_R	Improving the understanding of struggling students		3.04	.722
BTBM17H_R	Making mathematics relevant to students		3.26	.641
BTBM17I_R	Developing students' higher-order thinking skills		3.10	.726

5.4.3.3 Instructional quality as assessed by teachers

TIMSS teacher background questionnaire included several aspects relating to instructional quality (Blömeke et al., 2016). These aspects relate to form a latent variable of instructional quality (hereafter referred to as InQua). The measure of InQua applied in this thesis is based on the teacher background questionnaire in TIMSS where teachers were asked to report "How often do you do the following in teaching this class?" The construct concerned the cognitive activation (CA), which was measured by the items ("Relate the lesson to students' daily lives", "Link new content to students' prior knowledge", and "Ask students to complete challenging exercises that require them to go beyond the instruction"), supportive climate (SC) measured by the items ("Encourage classroom discussions among students", "Encourage students to express their ideas in class"), and instructional clarity (CI) measured by the item ("Ask students to explain their answers", "Ask students to decide their problem-solving procedures"). These measures were used since it has a more explicit relation to the dimensions of high-quality instructions.

The response was given on a 4-point scale, ranging from "Every or almost every lesson", "About half the lessons", "Some lessons", and "Never". Teachers' responses to these items were recoded so that high values reflect a higher assessment of instructional quality. Mathematics teachers in the South Africa sample reported high levels of linking new content to students' prior knowledge. They also rated students completing a challenging exercise that require them to go beyond the instruction relatively low (see Table 5).

Table 5. Descriptive statistics for instructional quality as assessed by teachers.

Note. _R Item was reverse coded.
(From IEA, TIMSS 2015 in the international database).

Indicator/ item	Question/ Variable labels	Response categories	M	SD
BTBG14A_R	Relate the lesson to students' daily lives	Every or almost every lesson, about half the lessons, some lessons, never	3.12	.842
BTBG14B_R	Ask students to explain their answers		3.30	.800
BTBG14C_R	Ask students to complete challenging exercises that require them to go beyond the instruction		2.90	.864
BTBG14D_R	Encourage classroom discussions among students		3.14	.843
BTBG14E_R	Link new content to students' prior knowledge		3.62	.593
BTBG14F_R	Ask students to decide their own problem-solving procedures		2.97	.836
BTBG14G_R	Encourage students to express their ideas in class		3.35	.758

5.5 Validity and Reliability

Validity and reliability are the main principles for considering trustworthiness and accuracy in research. Validity refers to how effectively an instrument measures what it claims to measure, and reliability denotes results consistency and measurement precision (Field, 2009). As for the relationship between validity and reliability, Pedhazue and Schmelkin (1991) noted that "a measure cannot be valid if it is not reliable, but being reliable it is not necessarily valid for the purpose its author or user has in mind" (p.81). Reliability then is fundamental but not a sufficient condition for validity.

Evaluating construct validity of the latent constructs lies in the centre of educational measurement, and is of great priority in the present study. The extent to which inferences can be drawn from the findings of a study is necessary with regards to how a particular construct is conceptualized and operationalized. Many concepts in social and behavioural sciences cannot be measured directly in a physical sense. These unobservable error-free concepts often are called latent variables or factors. What we can measure are different indications or representations of the concept by employing different approaches that attempt to collect measurable information about the latent variable. This operationalization process captures the researcher's definitions of a construct, and measures what it sets out to measure, based on a particular theoretical framework (e.g., Cohen et al., 2011). And this is a crucial step for ensuring construct validity.

In accordance with my theoretical framework of input, processes, and outcome in education (e.g., see Goe, 2007), a number of theoretical constructs of interest were operationalized by the contextual questionnaires and achievement test in TIMSS 2015. For example, the latent construct *teacher qualification* is measured by teacher's responses to the questions about their level of formal education, certification and subject major of their teacher education, and years of teaching. Each of these questions covers a special aspect of teacher qualification, but none of them alone can capture *teacher qualification* as a whole. In other words, the latent construct *teacher qualification* explains the different amount of variance in these indicators, and it is the shared variances among these indicators attribute to the measurement of the latent variable. Thus, the higher the explained variance is, the more valid the indicator will be to the latent variable.

The fundamental assumption is that the indicators of a latent variable should be uni-dimensional, meaning that these indicators underlie only the construct that they supposed to measure. Testing the uni-dimensionality of the indicators is a way to evaluate construct validity, which can be done by setting up a single factor measurement model to regress the observed variables (i.e., indicators) on the factor (i.e., latent variable). The relationship between the observed variables and latent variables is captured by the regression weights (factor loadings). The higher the regression weights, the better indicator will be. Normally standardized factor loading greater than .30 and statistically significant is used as a criterion when assessing construct validity.

To evaluate the reliability of the indicators of a theoretical concept and to measure internal consistency of a composite scale, Cronbach's alpha (α) is used. A coefficient of 0.70 has often been regarded as an acceptable threshold for reliability; however, between 0.80 and 0.90 inclusive and above is preferred based on the psychometric quality of scales (Cronbach, 1951).

In this study, the α coefficient was between 0.80 and 0.91, indicating high reliability of all the composite scales (George & Mallery, 2003). Cronbach's alpha reliability coefficient is presented in Table 6 below. The index of TCM shows high internal consistency of the coefficient of teacher confidence (Cronbach's $\alpha = .91$). This means that about 91 % of the variance in the teacher confidence scores is reliable. Therefore, only 9% is due to error variance. This presupposes that the TCM is measured with good reliability and lower error. The same applies to the index of InQua, with Cronbach's $\alpha = .80$, indicating the reliability for the InQua is very high. The error variance is estimated at 20%. Typically, it should be noted that the measurement procedure used in TIMSS is in line with the variables of interest that are part of my theoretical framework (Goe, 2007), and which serves as a good reflection of my hypothetical model.

Table 6. The reliability coefficient for instructional quality (InQua) and teacher confidence in mathematics (TCM).

Scale	Indicators	Cronbach's alpha α
InQua	BTBG14A_R, BTBG14B_R, BTBG14C_R, BTBG14D_R, BTBG14E_R, BTBG14F_R, BTBG14G_R	.80
TCM	BTBM17A_R, BTBM17B_R, BTBM17C_R, BTBM17D_R, BTBM17E_R, BTBM17F_R, BTBM17G_R, BTBM17H_R, BTBM17I_R	.91

5.6 Analytical Method

5.6.1 Introduction to two-level structural equation modelling

Structural equation modelling can be viewed as a general statistical modelling technique to empirical data, such as the analysis of variance and covariance, multiple regression, factor analysis and path analysis (Bowen & Guo, 2011). Hu & Bentler (1999) proposed that Structural Equation Modelling is a statistical modelling technique that is used when investigating the plausibility of theoretical models that might explain the interrelations among a set of variables (p. 2). Structural equation models consist of measurement models representing the relationship between the latent variables and the observable indicators, and structural models representing the relationship between the latent variables of interest (Nachtigall et al., 2003). Within the framework of SEM, confirmatory factor analysis represents a measurement model, where a variety of indicators (observed) variable are used as measures of latent variables, which can be further related in structural models to investigate the relationship among latent variables (Brown, 2015). CFA is used to test a model that specifies the relationship between observed variables and latent variables, usually based on existing theory. A latent variable in SEM generally defines a hypothetical construct that is not directly observable through single indicators (Kline, 2015). Therefore, SEM is viewed from the perspective of multivariate statistical analysis technique, it takes on a confirmatory factor analysis, as well as examining direct, indirect and total effects between variables.

The present study uses two-level structural equation modelling (SEM) as the main analytical methods to answer the research questions. Since the TIMSS 2015 data has a hierarchical structure, therefore, employing the appropriate method for handling such a data structure is necessary. Multilevel analysis is a powerful tool to adjust for the cluster effect of the sample and correctly estimate the standard error based on the variation at different levels of observation. In such a way, type I error can be avoided (Hox et al., 2010). In addition, educational assessment data, such as TIMSS study, often have unobserved variables that are not directly measured. One possible example is the construct of teacher instructional quality.

There is no single, definitive measure of teacher instructional quality. Instead, a variety of observed indicators are used to assess facets of teacher instructional quality. This means that a latent variable is useful when hypothetical constructs are operationalized. Other latent variables feature in SEM corresponds to an explicit representation of measurement error, which can be related to factors specified as outcomes or observed variables (Kline, 2015). In addition to its greater emphasis to make corrections for errors of measurement, it follows that the unique variance of the observed indicators variable, the error variance (Brown, 2015), is separated from the unexplained part.

An example of teacher instructional quality is depicted to distinguish between a latent (unobserved) variable and observed (manifest) variables (see Figure 3). Indicators (observed) variables are bg14a_r, bg14b_r, bg14c_r, bg14d_r, bg14e_r, bg14f_r and bg14g_r indicated as rectangles (or squares), each capturing a response related to how often teachers perform various activities in the classroom. These seven observed variables are indicators of the underlying unobserved latent variable. The ellipse represents unobserved latent variables. The straight arrows represent direct effects.

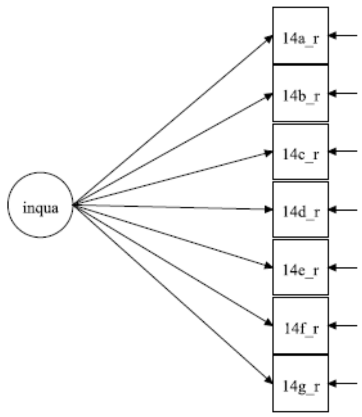

Figure 3. A measurement model of teacher instructional quality.

There are different test statistics to evaluate the model fit for a study's indicators. A model must be tested to investigate how well the model estimates fit the observed data. A chi-square ($\chi 2$) of statistically non-significant value is commonly used as a measurement to establish a good model fit. The statistically-based measure of fit indicates the difference between the sample variance-covariance matrix (observed data) and model variance-covariance matrix (predicted data) does not differ significantly. This, in turn, implies that a non-significant difference denotes that the hypothetical model tested captured correctly the structure of the data. Hence, the chi-square test, which is most often used, is criticized as being meaningless for large samples (Kline, 2011; Khine, 2013). To deal with this problem, several other measures of model-fit indices have been developed by researchers based on the chi-square test to determine whether to reject or accept the model. The given study uses the recommended model-fit indices (Brown, 2015), such as Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the Standardized Root Mean Square Residual (SRMR), performed using Mplus (Muthén, & Muthén, 2010).

Further, there are different recommendations regarding which cut off values should be taken into account to assess model fit. The RMSEA is regarded as "*one of the most informative fit indices*" (Diamantopoulos & Siguaw, 2000, p.85) as a result of its sensitivity to the number of estimated parameters in the model (Hooper, Coughlan & Mullen, 2008). It explains how well the model parameter estimates would fit the population's covariance matrix. A RMSEA value between 0.08 and 0.10 indicates a mediocre fit and below 0.08 reflects a good model fits. In addition, it has been found that if a RMSEA value is 0.07 or below, then it can be recommended or considered as an acceptable fit (Hu & Bentler, 1999). SRMR which is a measure of the residuals or the square root of the difference between the residuals computed for the sample covariance matrix and the hypothesized covariance model, values below 0.08 are regarded acceptable (Hu & Bentler, 1999). For CFI, which is a fixed index that assumes that all latent variables are uncorrelated and compares the sample covariance matrix with the independent model. CFI may also depend on the average size of the correlations in the data, and values of 0.95 and above are deemed as acceptable, while values close to 1.0 are recognized as indicative of good model fit. Values for TLI are not normed, it may fall below 0 or above 1, but a good fit model usually has values between 0.95 and 1.0. In addition, a factor loading of .30 or greater signifies a higher measure of their respective latent construct (Brown, 2015).

5.6.2 Intraclass Correlation Coefficient

In the previous section, it was shown that TIMSS data is obtained through a multi-stage cluster sampling design. Data achieved in such a way have commonly a hierarchically nested structure; for example, students nested within a classroom and classrooms within a school. The intraclass correlation coefficient (ICC), denoted as ρ, captures the extent to which individuals, say students are clustered. ICC simply expresses the amount of variation in a target variable that can be attributable to the fact that individuals belong to different groups or aggregated level units, such as classrooms or schools (e.g., Raudenbush & Bryk, 2002; Yang-Hansen et al., 2014). In other words, ICC indicates the proportion of the variance accounted for by the grouping structure of the population (p. 15, Hox, 2010). Whenever the average variation in the student's level of performance is small across different classrooms or groups, the ICC becomes small, and vice versa.

Given that students from the same classroom or schools are much similar compared to the students from different classrooms or schools, the variation among the total sampled students by cluster sampling design will be smaller than that of the same amount of students that are sampled with simple random sampling strategy. If we use the ordinary analytical method to handle clustered data, the standard error will be underestimated. In such a situation, one more than often conducts type I error, i.e., when one wrongly rejected the null hypothesis when actually the null hypothesis is true. The results for the ICC of this study are presented in *Section 7.3* below.

5.6.3 Analytical Process

As it was mentioned above, TIMSS comes to terms with hierarchically structured data with students nested within the classroom level. In the analysis using a multilevel approach, the weighting factor must be applied for each hierarchical level appropriately. In this study, because of the two-stage cluster sample design of TIMSS, weight for student and teacher effect at the classroom level was used in the analysis to account for bias in the parameter estimates or unequal probability of students and classroom being selected. As recommended by Rutkowski et al. (2010), weights at the student level are a product of student weight and student weight adjustment. At the classroom level, weights are a product of class weight, a class weight adjustment, a school weight, and a school weight adjustment.

The analysis was performed using Mplus V. 8.3. In a stepwise manner, first, path analysis using latent variables or unobserved measurement models employing CFA was formulated. Next, with the help of structural equation modelling, the relationship between variables of interest was specified accordingly. A latent measurement model of teacher instructional quality and perhaps, teacher confidence in mathematics was formulated at only the classroom level. This means that the teacher's effects were modelled at the teacher level since teacher variation is only at the teacher level. Also, the individual (student) level was modelled at both the individual (student) level and teacher level, which means that individuals' variation was decomposed at both levels. As the focus of this study was on the relations between teacher qualification and characteristics, teacher instructional quality, and student mathematics achievement, the researcher is interested in the aggregated level of students' family socioeconomic background. Students shared teaching and learning-related factors or common characteristics within the same classroom, and classroom level is necessary when considering such a group-level construct (Lüdtke et al., 2009).

The student background variables were created (PEDU, BOOKS, HSS), designed to reflect parents' highest education level, amount of books in student's home and the number of home study support, and teacher background variables (T_EXP, T_EDU, T_Major, InQua, TCM) to explain variation in achievement. Student's family SES and classroom SES composition were taken into account. Other variables are distinguished into simple and complex scale. The simple scale including items like teacher experience, teacher education, and teacher major or specialization. The complex scale is defined by observable (manifest) items and constructed using procedures that encompass a variety of factors. Items are recoded, including teacher instructional quality (BG14A_R, BG14B_R, BG14C_R, BG14D_R, BG14E_R, BG14F_R, BG14G_R) and teacher confidence in teaching mathematics (BM17A_R BM17B_R BM17C_R BM17D_R BM17E_R BM17F_R BM17G_R BM17H_R BM17I_R).

The study makes use of two-level model analysis. The final model was used to investigate the relationship between teacher qualification, teacher characteristics, instructional quality, and classroom SES composition, and student mathematics achievement for South African nine graders. To investigate the extent to which teacher qualification and characteristics, instructional quality, and classroom SES composition have a significant effect on 9th graders' mathematics achievement in South Africa, the same 'model' was estimated. Finally, the two-level model was estimated to investigate if SES is the strongest predictor of South Africa grade nine student's achievement.

6 Results

In the following, measurement models are formulated and evaluated. Results of structural models and parameter estimates are presented.

6.1 Measurement Model of Teacher Instructional Quality

Confirmatory factor analysis (CFA) was conducted to test the measurement property of teacher instructional quality (InQua) using Mplus. Since this construct was assessed by teachers, the measurement model of teacher instructional quality is at the aggregated level. Table 7 showed that all the model fit indices were at the acceptable level, which implied that the one-factor model fits the data very well. Chi-square value $\chi2$ (14)= 22.865 is not significant, with $p = 0.0625$, indicating that the model implied variance-covariance matrix is not significantly different from the observed variance-covariance matrix from the data. SRMR = .029. Furthermore, the Comparative Fit Index (CFI) is 0.980, and the Tucker-Lewis Index (TLI) is 0.970. Root Mean Square Error Approximation (RMSEA) of 0.044 satisfies the acceptable threshold of less than 0.08. Both CFI and TLI are greater than the cut-off value of 0.95, but in some cases, 0.90 is acceptable, since chi-square is too sensitive to sample size (Hair et al., 2006).

Figure 4 showed the standardized factor loadings and residuals of the measurement model of InQua. A standardized factor loading is between plus and minus one and can be interpreted as the regression coefficient of the indicator, say Q14a, on the latent variable, InQua. The factor loadings in Figure 4 range from .46 for indicator Q14a (Relate the lesson to students' daily lives) to .68 for indicator Q14b (Ask students to explain their answers), all being significant and positive. These model results confirmed that only a single dimension, i.e., teacher's instructional quality can be identified by the 7 indicators. In other words, the indicators properly capture the underlying construct of teacher instructional quality (InQua). Hence, this result exhibited good construct validity of InQua.

6.2 Measurement Model of Teacher Confidence in Mathematics

Further, confirmatory factor analysis (CFA) was performed to test the measurement property of teacher confidence in mathematics (TCM) using Mplus. Since this construct was assessed by teachers, the measurement model of teacher confidence in mathematics is at the aggregated level. Table 7 exhibited that all the model fit indices were at the acceptable level, which implied that the one-factor model fits the data very well. However, the chi-square value $\chi2$ (27)= 57.826 is significant, with p = .0005, since $\chi2$ is known to be too sensitive to the sample size (Hair et al., 2006). SRMR = .035. Moreover, the Comparative Fit Index (CFI) is .969, and the Tucker-Lewis Index (TLI) is .959. Root Mean Square Error Approximation (RMSEA) of .060 satisfies the recommended level of acceptable fit of less than .08. Both CFI and TLI are greater than the cut-off value of 0.95, but in some cases, .90 is acceptable, since chi-square is found to be too sensitive to the sample size (Hair et al., 2006).

Figure 5 showed the standardized factor loadings and residuals of the measurement model of TCM. A standardized factor loading is between plus and minus one and can be interpreted as the regression coefficient of the indicator, say Q17a, on the latent variable, TCM. The factor loadings in Figure 5 range from .60 for indicator Q17a (inspiring students to learn mathematics) to .82 for indicator Q17I (developing students' higher-order thinking skills), all being significant and positive. These model results confirmed that only a single dimension, i.e., teacher's confidence in mathematics can be identified by the nine indicators. In other words, the indicators properly capture the underlying construct of teacher confidence in mathematics (TCM). Hence, this result demonstrated good construct validity of TCM.

Table 7. The model fit indices of teacher instructional quality (InQua) and teacher confidence in teaching mathematics (TCM).

Parameters	*χ2*	*df*	*CFI*	*TLI*	*RMSEA*	*SRMR*
InQua	22.865	14	0.980	0.970	0.044	0.029
TCM	57.826	27	0.969	0.959	0.060	0.035

*RMSEA=Root Mean Square Error of Approximation
*CFI=Comparative fit index
*TLI=Tucker-Lewis Index
*SRMR=Standardized Root Mean Square Residual

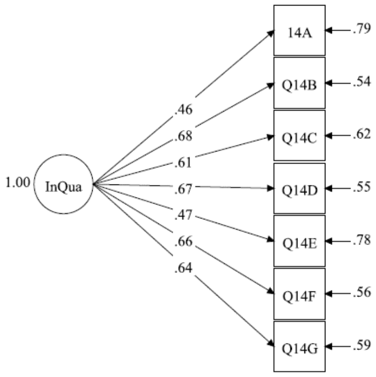

Figure 4. The measurement model of teacher instructional quality (InQua).

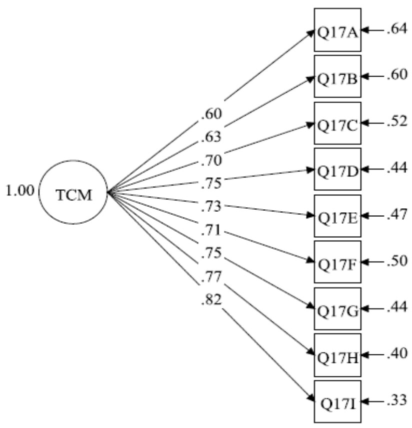

Figure 5. The measurement model of teacher confidence in mathematics (TCM).

7 Structural models

7.1 Relations between teacher qualification and characteristics, instructional quality, classroom composition, and student mathematics achievement

Considering the above latent variable models of InQua and TCM as classroom level construct which yielded substantially better fitting model and higher validity, the researcher can now have more confidence to test the hypothetical model. The interrelationship among these latent variables and student mathematics achievement were simultaneously examined at student and teacher levels in a two-level structural model, controlling for the socioeconomic status of students at both levels. The structural model at the teacher level was constructed so that teacher instructional quality (InQua), teacher confidence (TCM) and three teacher qualification (TQ) variables, namely, teacher experience, their formal education and teacher major or specialization, are interrelated and affecting mathematics achievement, after taking into account the variation of the classroom SES composition and student's SES at the respective levels. Student mathematics achievement scores (i.e., all five plausible values) was used as the dependent variables in the model. At the individual level, only the relationship between student's family socioeconomic status and their mathematics achievement was tested.

The model structure was portrayed in Figure 6.

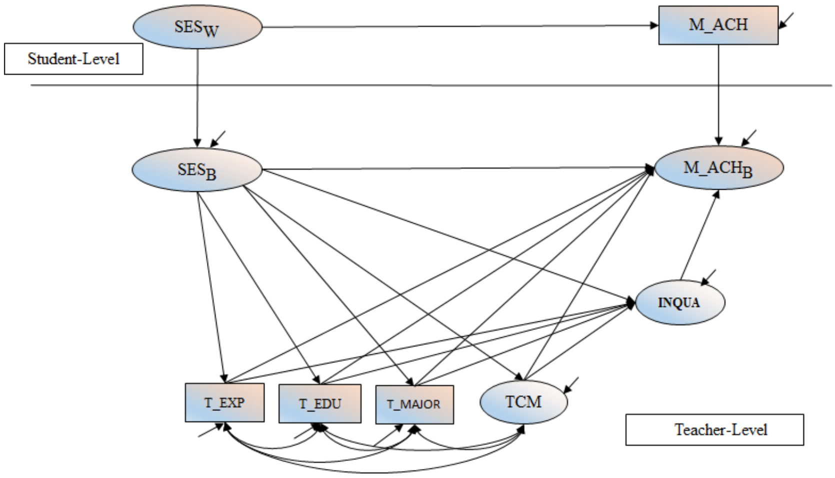

Figure 6. The hypothetical model with the relations between indicators of teacher quality (TQ) and characteristics (TCM), instructional quality (InQua), student and classroom SES composition, and mathematics achievement (M_ACH).

7.2 Model fit of the final structural model for South Africa

In a further step, a structural model for the South Africa sample was constructed to assess whether the parameter estimates to be analysed are valid and reliable. It is critical that the parameters to be analysed should be providing reliable estimates. To conduct a two-level analysis, Mplus was used to run the (same) model structure five times based on the five plausible values generated for each student. In this way, Mplus do calculation for the final parameter estimate of the standard error for the target parameters. The option for the evaluation in Mplus was TYPE = TWOLEVEL; ESTIMATOR = MLR functions in ANALYSIS command together with CLUSTER = IDTEACH; WEIGHT = HOUWGT (house weight); and BWEIGHT = MATWGT teacher weight). My reasoning here of incorporating house weight and teacher weight at individual and teacher levels is to properly adjust the standard error due to the clustered sampling strategy. Moreover, each classroom may not only have one teacher, and a teacher may not only teach one classroom, with the teacher weight, we also can adjust the teacher-classroom one to one match.

Results showed that the two-level model estimates had a satisfactorily good fit to the data. The chi-square test associated with the mean and standard deviation were $\chi2$ (215) = 347.576, SD = 6.295, and $\chi2$ is derived by the formula $\chi2$ = mean/ degree of freedom; $347.576/215 = 1.62 < 5$. When the value is less than 5, it is regarded as substantially good fit chi-square estimates. The formula holds if and only if the model is run five times, as mentioned before. Thus, the $\chi2$ satisfies the goodness-of-fit. Moreover, the measure of fit indices with regards to the mean and standard deviation are CFI (M = 0.930, SD = 0.003); RMSEA (M = 0.007, SD = 0.000), within-level SRMR (M = 0.013, SD = 0.003) and between-level SRMR (M = 0.058, SD = 0.000). These fit indices were found to exhibit an exceptionally good fit to the data. Further, it can be observed that the standard deviation for RMSEA is very small almost close to zero, which means that all the 5 times estimation or the number of successful computations from different plausible values provides an equally good fit. Similarly, SRMR value at the individual level is good, and there is almost no variation at the between level, indicating an excellent model fit. Hence, this model is valid and reliable for two-level modelling (Refer to Table 8 for more detail).

Table 8. Model fit indices for the final structure model (two-level model).

Fit Indices		*Mean*	*Standard deviation*
Chi-square ($\chi2$)		347.576	6.295
df	215	-	-
CFI		0.930	0.003
RMSEA		0.007	0.000
SRMR		0.013 (Within-level)	0.003 (Within-level)
		0.058 (Between-level)	0.000 (Between)

*RMSEA=Root Mean Square Error of Approximation
*CFI= Comparative Fit Index
*SRMR=Standardized Root Mean Square Residual

7.3 Reporting Results for ICC

The research questions in the current thesis are focused on how teacher-related factors affect student achievement. In the two-level structure equation model, only the student and teacher levels are simultaneously analysed in the analysis. Taken advantage of two-level analysis, the intraclass correlation coefficients were examined. To perform two-level modelling, an important aspect is to determine whether the between teacher-level effects were reliable for the purpose of the study's variables. Two-level modelling is usually required when the intraclass correlation coefficient is .05 or higher (Muthén, 1994). However, when the ICC is less than .05, two-level modelling will not be warranted. This is usually the case when the between-group variability is not present.

As is shown in Table 9, the intraclass correlations coefficient varies between .09 and .62. About 9.4% of the between-classroom variation in the parent's highest level of education is due to the fact that students are sorted into different classrooms. 15.8% of the between-classroom variation in the number of books at home and 16.2% of the between-classroom variation in the home study supports is due to students going to different classrooms. Concerning student mathematics achievement, different classroom belongingness explains 62.3% of between-classroom differences in mathematics achievement. In other words, 62.3% of the variance in mathematics test scores in TIMSS 2015 across different classrooms is explained by the classroom effects. The phenomenon of classroom effects could mean that different teachers teach different classrooms, or with different students' family background composition, classroom learning culture, etc. The proportion of variance of the mathematics achievement shows a measure of the extent to which the classrooms are segregated in terms of student achievement. The main argument is that students going to different classrooms should achieve equally, but the ICC coefficient for between-classroom differences were very high, indicating classroom segregation indicator in South Africa.

Table 9. Estimated Intraclass correlation coefficients of the latent variable indicators.

CLUSTER	AVERAGE CLUSTER SIZE	BSMMAT	PEDU	BOOKS	HSS
326	38.172	0.623	0.094	0.158	0.162

7.4 R-Square

Generally, the R-square of the estimates from the model provides information about the proportion of the variation in each of the dependent variables that are accounted for by the independent variables. Since SES significantly affects mathematics achievement, it is important to examine if differences in mathematics achievement are explained by this model. R-square, in other words, is the percentage of the explained variance in the outcome variable by the model. The final model (Figure 6) can explain almost 80% of the variation in mathematics achievement across different classrooms.

7.5 Parameter estimates

Corresponding to the research questions formulated in the previous section, the main results are presented. Table 10 and Table 11 show both the direct and indirect effects of the interplay among factors depicted in the constructed structural model respectively (see Figure 6).

Table 10. Standardized direct effects in the final model.

Parameter Estimate	*Factor loading*	*Standard Error* (S.E.)*	*P-value*
Within level			
SESW effect on BSMMAT	*0.073*	*0.030*	*0.014*
Between level			
Teacher qualification effects on mathematics achievement			
BTBG01 effect on BSMMAT	0.032	0.047	0.498
BTBG04 effect on BSMMAT	0.076	0.045	0.093
BTDM05 effect on BSMMAT	0.006	0.044	0.895
Teacher qualification effects on instructional quality			
BTBG01 effect on INQUA	*-0.148*	*0.070*	*0.034*
BTBG04 effect on INQUA	-0.026	0.064	0.690
BTDM05 effect on INQUA	0.018	0.072	0.802
Classroom student's SES composition effect on class-level factors			
SESB effect on BSMMAT	*0.874*	*0.032*	*0.000*
SESB effect on INQUA	0.014	0.079	0.862
SESB effect on TCM	0.022	0.079	0.778
SESB effect on BTBG01	0.007	0.085	0.935
SESB effect on BTBG04	0.098	0.059	0.096
SESB effect on BTDM05	-0.019	0.080	0.809
Interrelationship among teacher-related factors and mathematics achievement			
INQUA effect on BSMMAT	-0.085	0.077	0.267
TCM effect on BSMMAT	0.018	0.081	0.825
TCM effect on INQUA	*0.713*	*0.051*	*0.000*
Correlation among teacher-related factors			
BTBG01 correlates with TCM	0.014	0.086	0.871
BTBG04 correlates with TCM	-0.067	0.061	0.274
BTDM05 correlates with TCM	0.051	0.077	0.508
BTBG04 correlates with BTBG01	*-0.158*	*0.080*	*0.048*
BTDM05 correlates with BTBG01	0.038	0.099	0.701
BTDM05 correlates with BTBG04	*0.220*	*0.108*	*0.041*

*Based on the recommendations of Muthén and Muthén (2002), standard error bias was taken into account in the analysis as follows:
For any parameter estimate in the model, standard error bias should be less than 0.10 (10%). However, one S.E. was at 0.108 (10.8) but can be considered small.
Parameter estimates in Italic in the table are statistically significant at $p < .05$ level.

8 Results RQ 1

Is there any relationship between teacher qualification, teacher characteristics, instructional quality, and classroom SES composition, and student mathematics achievement for South African nine graders?

No significant direct relationship between the teacher background indicators (experience, education, and major), and teacher confidence, teacher instructional quality, and classroom average mathematics achievement was found. As seen in Table 10, the p-values of the effects of almost all teacher-related factors are greater than .05. Further, student-level SES has a statistically significant relationship with students' mathematics achievement with the path coefficient of 0.073 ($p = .014 < .05$). And the classroom SES composition has a very strong statistical relationship with average mathematics achievement with a regression coefficient of 0.874, $p < .05$. While in the correlation part, teacher confidence was significantly positively related to teacher instructional quality, $p < .05$, meaning that teachers with higher levels of confidence offered a higher quality of instruction, as needed by their students. Teacher's years of working affected teacher's instructional quality negatively (-.15, $p = .034$). No significant indirect effects were found of SESB on classroom average mathematics achievement, as presented in Table 11.

9 Results RQ 2.

To what extent do the teacher qualification and characteristics, instructional quality, and classroom SES composition affect 9th graders' mathematics achievement in South Africa?

With regards to the findings of RQ 1, as mentioned before, Figure 6 (see Table 10) graphically shown the model at the classroom level chosen to answer RQ2. Since the analyses of research RQ 1, reveal no significant association between teacher qualification and characteristics, teacher instructional quality, and mathematics achievement, it follows that the effects of the three teacher qualification variables on student average achievement were not significant, and estimated at $\beta_{exp} = 0.032$, ($p = 0.498 > .05$; $\beta_{edu} = 0.076$, ($p = 0.093 > .05$) and $\beta_{major} = 0.006$, ($p = 0.895 > .05$) (i.e., with extremely very small effect, not differ from zero). In terms of teacher confidence, a non-statistically significant effect on mathematics exist, estimated at $\beta_{TCM} = 0.018$, ($p = 0.825 > .05$), indicating that teacher confidence does not impact achievement. For instructional quality, the results exhibited regression weights of $\beta_{INQUA} = -0.085$ ($p = 0.267 > .05$), which means that the regression weights of teacher instructional quality have a negative impact on mathematics achievement. Also, results reveal the regression weight of $SES_B = 0.874$, $p < .05$, and a significant effect is due to SES composition in terms of student achievement. Taking into account the RQ 2, the size of the effects of the relationship between teacher qualification, and characteristics, teacher instructional quality, and classroom SES composition do not influence 9th graders' mathematics achievement in South Africa. However, the classroom SES composition is the only variable significantly attributable to mathematics achievement variation in South Africa.

Table 11. Standardized total and specific indirect effects estimated in the structural model.

Relations	*Total Indirect effects (Z-effects)**	***P-value*
Total indirect effect of SESB on mathematics		
BSMMAT - SESB	0.006 (0.587)	0.557
The specific indirect effect of SESB on mathematics via teacher-related factors		
BSMMAT - BTBG01 - SESB	0.000 (0.003)	0.935
BSMMAT - BTBG04 - SESB	0.007 (0.006)	0.192
BSMMAT - BTDM05 - SESB	0.000 (0.001)	0.918
BSMMAT – INQUA - SESB	-0.001 (0.007)	0.870
BSMMAT – TCM - SESB	0.000 (0.002)	0.859
BSMMAT – INQUA - BTBG01 - SESB	0.000 (0.001)	0.936
BSMMAT – INQUA - BTBG04 - SESB	0.000 (0.001)	0.731
BSMMAT – INQUA - BTDM05 - SESB	0.000 (0.000)	0.851
BSMMAT - INQUA – TCM - SESB	-0.001 (0.005)	0.786

*Z-effect is the ratio of the estimate or regression coefficient and standard error, (i.e., Z_{value} = estimate/ standard error) and when is > 1.96, then the effect is significant. The estimates reflect the regression weight or the strength of the relationship between variables.
** P-value is significant at <.05.

10 Results RQ 3

Is SES the strongest predictor of South Africa grade nine student's achievement?

Socioeconomic status is a significant predictor at the individual student level with the effect size (0.07, p = .014), but SES is a statistically significantly strong predictor at the classroom level (0.879, p < .05). Returning to the intra-class correlation coefficient estimated at 62.3% in Table 9 above, it can be seen that classroom composition in terms of SES is very high taking into account this estimate. Almost 80 % of the classroom-level variance of the mathematics achievement was explained by the classroom SES composition. High variation across classrooms in South Africa exists, which indicates that if a student has a better condition of SES, the student in question tends to have a higher level of mathematics achievement or go to a better classroom. Therefore, SES composition is a strong predictor of South Africa grade nine student's achievement.

11 Discussion

South Africa is a country of higher classroom segregation. South African students show a huge achievement variation in mathematics in the ninth grade in TIMSS 2015. South African's highly achievement inequitable environment provides unique insights into understanding the relationship among factors that are known to contribute to achievement differences in mathematics. This study developed a hypothetical model to examine the relationship between teacher qualification and characteristics, teacher instructional quality, students' family socioeconomic background, and student mathematics achievement with the South Africa data from TIMSS 2015. This study examines whether it can be hypothesized that aspects of teacher quality and classroom SES composition have a significant effect on 9th graders' mathematics achievement in South African. Also, whether SES is the strongest predictor of South Africa grade nine student's achievement.

In this context, the results of the study revealed no significant relationship between teacher qualification and characteristics, teacher instructional quality, and student's mean mathematics achievement. Teachers with experience, a degree, and who majored in mathematics had no significant relationship with student mathematics achievement. In terms of teacher's experience, the results of the study are concordant with previous research which points out that the relationship between student achievement and teacher experience is non-statistically significant (Nilsen & Gustafsson, 2016). The study results suggest that the effect of teacher experience on student achievement is not deterministic in South Africa. No relationship was found between teacher experience, teacher education, and teacher major or specialization. Besides, when examining the relationship between teacher experience and teacher education, it was found that the teacher's experience was negatively correlated with teacher education, and the effects were not significant. Undeniably, the idea that teachers with experience cease to contribute to student mathematics achievement is not in agreement with other research (Akiba et al., 2007).

With regards to teacher confidence in teaching mathematics, the results of the study suggest, however, is not deterministic with student achievement in mathematics. Still, when examining the effects of the teacher confidence in teaching mathematics, no significant effect from SES is possible. As demonstrated by the study results, teacher confidence in teaching mathematics does not show any significant direct or indirect effects on classroom average mathematics achievement. These findings also contradict those found by Bandura (1997) and Henson (2002), who found the relationship between teacher self-confidence and mathematics ability. When it comes to the effects of teacher confidence on instructional quality, it appears significantly strong, as revealed by the study results, which means that teachers with higher levels of confidence offered a higher quality of instruction. This result corresponds with previous studies (Klassen et al., 2011; Klassen & Tze, 2014; Tschannen-Moran et al., 1998) which have shown that teacher confidence influences teacher's instructional practices.

On the other hand, the absence of a relationship between teacher instructional quality and average student mathematics achievement was less anticipated, as previous research has indicated that teacher instructional quality is positively related to academic achievement (Nilsen et al., 2016). However, in the South Africa context, the relationship between teacher instructional quality and classroom average mathematics achievement is not significant. Somewhat surprisingly, the results also indicated that the effects of teacher instructional practices on student's average mathematics achievement were rather weak. In other words, teacher instructional practices cannot be determined as a result of the student's average mathematics achievement.

Indeed, findings of the study demonstrated only a direct significant effect from classroom SES composition to student average mathematics achievement. This finding also points out that individual-level SES significantly affects student mathematics achievement. The strength of the effects was not strong at the student level, but at the classroom level, it is really strong. These results are in agreement with previous research that found significant effects of SES on student achievement (Yang-Hansen & Gustafsson, 2016). Taking the analysis one step further, what is really affecting achievement is because of classroom segregation, which means that student mathematics achievement differences depend on classroom SES composition. In South Africa, the dominant contribution to achievement differences is student SES and classroom SES composition, other factors like teacher qualification, teacher confidence and teacher instructional quality do not matter very much. As mentioned above, apart from student SES and classroom SES explaining variations in mathematics achievement, other mechanisms like aspects of teacher quality, are not deterministic in terms of achievement differences. The findings indicated more than half of the total variation in South African students' mathematics achievement can be attributed to classroom differences, of which the differences in classroom SES composition was found the major explanatory factor. The between classroom differences in mathematics achievement may be further explained by other classroom and/or school level factors, such as uneven distributions of the physical and human resources of different kinds, where public classrooms or schools have a less resources compare to their private counterpart. This is an important equity issue to be examined in the further study of South Africa data.

In view of the much stronger influence of student SES and classroom SES composition in South Africa, the education system is relatively ineffective because the country is highly segregated in terms of the mechanisms of SES and reduces equity, and is being reflected in classrooms and schools. These mechanisms responsible for the student SES and classroom SES composition is alarming and deserve the attention of further research. Essentially, student SES and classroom SES composition may not only attributable to variations in achievement and student SES characteristics of the classroom's student intake, but also to physical resources-related characteristics (e.g., school building conditions, library materials, access to computers and flushing toilet, etc.) that vary across classrooms are captured by the classroom SES effect.

The question ought to be how we can best ameliorate the impact of students' SES background characteristics on mathematics achievement, rather than eradicate it. Deliberation of targeted provisions of physical resources may, therefore, be a better alternative to classrooms in respect of improving access to good equitable and quality education in the South Africa context.

12 Conclusion

The current study aimed to investigate the relationship between aspects of teacher quality, classroom SES composition, and student mathematics achievement. The study findings are as follows: First, results demonstrate no significant relationship between teacher qualification and characteristics, teacher instructional quality, and student average mathematics achievement. Second, findings indicate that individual student level SES significantly affect mathematics achievement. However, classroom level SES composition strongly affects classroom average mathematics achievement. Third, teacher characteristics (teacher confidence) were positively related to teacher instructional quality. But, teacher characteristics and teacher instructional quality had no relationship to the classroom average mathematics achievement. Fourth, teacher experience, teacher level of formal education, and teachers who focus on either mathematics or mathematics education had no association with the classroom mean mathematics achievement.

SES strongly appears to be a dominant contribution to achievement differences between classrooms in South Africa. SES may not only capture the socio-economic conditions of the classroom, but also variations in physical resources across classrooms with different SES effects.

The analysis from South Africa TIMSS 2015 data, highlights the view that the quality and equity of education can be achievable through an increase in school physical resources.

12.1 Ethical consideration and concerns

There are always ethical considerations and concerns to a greater extent with regard to human subjects in educational research. This thesis has followed the issues of ethical concern regarding some of the principles of research integrity. First, these principles are informed by the European Science Foundation (ESF, 2011), which stated that:

> Reliability in ensuring the quality of research reflected in the design, the methodology, the analysis and the use of resources. Honesty in developing, undertaking, reviewing, reporting and communicating research in a transparent, fair, full and unbiased way. Respect for colleagues, research participants, society, ecosystems, cultural heritage and the environment. Accountability for the research from idea to publication, for its management and organization, for training, supervision and mentoring, and its wider impacts. (p. 4)

As data practices and management is essential in quantitative research, for example, researchers, research institutions and organizations are responsible for handling the data "as open as possible, as closed as necessary, and where appropriate in line with the FAIR principles (Findable, Accessible, Interoperable and Re-usable) for data management" (ESF, 2011, P. 6).

This thesis works with earlier collected secondary data of the IEA's mathematics and science assessments. Therefore, the ethical issues relating to TIMSS assessment data has already been thoroughly dealt with by the national research coordinator in each participating country. Since

TIMSS data is secondary data, no students, schools, and teachers can be identified, which implies that TIMSS assessment does not contain confidential information on students, teachers or schools. This is because the identities of the students, schools, and teachers have been assigned with unique identification numbers (ID numbers). As mentioned before, TIMSS highlights the view that each student only answers a part of the total item. In this way, each student testing time is reasonably not mandatory, while it is possible that a student can withdraw from the survey or not responding to some of the assessment items. To this end, TIMSS data is freely and publically available to the researcher in the international TIMSS homepage, for which authorization will not be requested by the researcher from the IEA. However, permission may be sought to use restricted information from IEA. Hence, informed consent is not needed.

Further, based on the abovementioned principle, the TIMSS data will be stored for the intended purpose following the duration of this research.

12.2 Implications for policymakers, students, teachers and educational researchers

The findings of this study have influential implications, in particular for policymakers, students, teachers, and educational researchers. Interventions that aim to improve educational quality and equity in terms of mathematics achievement cannot be solely aimed at the human resource-related factors (e.g., teacher qualification, teacher instructional quality, and teacher confidence). The findings point to the greatest need to consider physical resources as an enriching tool for improving educational quality and equity rather than just any of the human resource-related factors. This can be achieved through equitable distribution of physical resources within the educational system. Students must be made aware of the effects of these physical resource-related factors on their achievement to compensate for any deficiencies.

Also, the findings generally point to the importance of reducing socioeconomic inequity conditions in societies for the sustainability of educational and economic outcomes. Moreover, efforts to reduce socioeconomic inequities in societies will be benefited given its positive and significant effect on mathematics achievement.

South Africa is one of the growing economies on the African continent. Therefore, greater efforts rest on the national Department of Basic Education and the Provincial Education Departments to design and implement cost-effective interventions to target students and classrooms in SES disadvantaged. To this end, South Africa can achieve the Sustainable Development Goals (SDGs) as documented in the National Development Plan of 2030.

The evidence from these findings suggests that increasing classroom physical resources in the education sector is one way to enhance student learning and achievement in South Africa. A greater collaboration of school physical resources should be pursuing by the Department of Basic Education and should be considered as a better school condition to improve student learning and achievement.

12.3 Practice recommendations

On the basis of these findings, some far-reaching practice recommendations can be attributable to teachers of mathematics and school principals regarding their daily work. The findings seem to reveal that the classroom SES composition is the strongest factor for measuring inequity in educational outcomes. High classroom SES effect on achievement is a segregation measure, and would not be recommended. However, if schools can work against school segregation, and mathematics teachers and school principals can establish networks with other schools and try to help each other to reduce the segregation, that will further be helpful to reduce educational inequality.

12.4 Limitation of the study

My study may have some limitations that need to be discussed. First of all, one possible limitation is the use of the cross-sectional design of the South Africa TIMSS study which makes it possible for a correlational conclusion to be made rather than for example, drawing causal relations between classroom SES composition and classroom average mathematics achievement. In this study, in light of the cross-sectional design, we cannot claim causal inferences between classroom SES composition and classroom mean mathematics achievement. These issues of investigating quality in education such that achievement together with teacher qualifications (indicated by teacher experience, teacher education, and teacher major or specialization), teacher characteristics and teacher instructional practices as the main and only indicators of quality in this thesis also point to the limitation of this study. Other indicators of quality were left out of scope, for example, students' motivation, well-being, etc.

Whereas equity was taken into account as background variable (SES) such as the number of books at home, the number of home study supports, and parent education level, other indicators of SES were not used, for example, immigrant status and language spoken at home. Usually, the amount of books at home reflects the best possible measurement available in TIMSS.

South African students have experienced varying teacher qualifications and characteristics, instructional practices during schooling, therefore teacher effects may not matter. Moreover, the correlation among certain teacher related-factors may be reciprocal. However, cautionary notes should be taken into consideration when generalizing the results of the findings, since the findings are merely limited to the South Africa educational system being investigated.

12.5 Further Research

It has been established that the teacher qualification and characteristics, and teacher instructional quality had no significant relation to the class mean mathematics achievement. The individual student level SES and classroom SES composition can cause differences in mathematics achievement.

In an educational system like South Africa, there are great diversities in educational quality and equity across schools. Varying educational resources and students' intake characteristics across private and public schools, and between urban and rural areas also assumed to attribute to student's achievement differences. Therefore, these variables need to take into consideration in further research. Moreover, variations in economic-related resources, such as physical resources conditions, may serve as constraints to effective teaching and thus deserve special attention in further investigation.

It is interesting and important for researchers to carefully investigate the contribution of teacher qualification, instructional quality, classroom SES composition on mathematics achievement between urban and rural area as well as between private and public schools. Since the mechanism among these attributes on mathematics achievement may be different, or even of opposite strengths, when studying together, Simpson's paradox may be conducted. We may also validate the current study with the TIMSS 2019 data of South Africa.

Moreover, we have reason to believe that different schools in South Africa may have different compensatory effects on student's SES backgrounds. The student's SES effect on their mathematics achievement may differ across different schools. It may thus be interesting and important to estimate such a random slope model to investigate the compensatory effect.

12.6 Contribution of the study

My original contribution of knowledge of this study is that the researcher employed TIMSS 2015 data of South Africa to contribute to educational quality and equity research in South Africa settings. The study investigated aspects of teacher qualification (indicated by teacher experience, teacher level of formal education, and teacher who focuses on either mathematics or mathematics education), teacher characteristics, teacher instructional quality, and classroom composition concerning students' family socioeconomic background in relation to mathematics achievement for South Africa 9 graders. The findings confirmed partly in line with some results of the previous research, and also fill the knowledge gaps and point out some new findings.

Results enhance our understanding or give some insights into demonstrating no significant relationship between teacher qualification and characteristics, teacher instructional quality, and student average mathematics achievement. It has also been found that classroom SES composition has a greater impact on classroom mean mathematics achievement than individual student SES factors. Teacher characteristics (teacher confidence) were positively related to teacher instructional quality. But, teacher characteristics and teacher instructional quality had no relationship to the classroom average mathematics achievement. Teacher experience, teacher level of formal education, and teachers who focus on either mathematics or mathematics education had no association with the classroom mean mathematics achievement. I argue that my analysis of the evidence of these findings is the original contribution, as no prior research investigates these teacher's attributes to classroom composition and mathematics achievement in the context of South Africa.

This thesis also adds to prior research by showing how the theoretical framework, particularly Goe's (2007) theoretical model has been useful to this study since the selection of the variables was based on this theoretical framework of the study.

13 References

Adams, D. (1993). *Defining educational quality*. Arlington, VA: Institute for International Research.

Adams, D., Acedo, C., & Popa, S. (2012). Quality Education. In C. Acedo, D. Adams, & S. Popa (Eds.), *Quality and Qualities: Tensions in Education Reforms* (pp. 197–208). Sense Publishers. https://doi.org/10.1007/978-94-6091-951-0_11

Alex, J. K., & Juan, A. (2017). Quality education for sustainable development: Are we on the right track? Evidence from the TIMSS 2015 study in South Africa. *Perspectives in Education; Bloemfontein*, *35*(2), 1–15. http://dx.doi.org.ezproxy.ub.gu.se/10.18820/2519593X/pie.v35i2.1

Arends, F., Winnaar, L., & Mosimege, M. (2017). Teacher classroom practices and mathematics performance in South African schools: A reflection on TIMSS 2011. *South African Journal of Education, 37*(3).

Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher Quality, Opportunity Gap, and National Achievement in 46 Countries. *Educational Researcher*, *36*(7), 369–387.

Atlay, C., Tieben, N., Hillmert, S., & Fauth, B. (2019). Instructional quality and achievement inequality: How effective is teaching in closing the social achievement gap? *Learning and Instruction*, *63*, 101211. https://doi.org/10.1016/j.learninstruc.2019.05.008

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., … Tsai, Y.-M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, *47*(1), 133–180.

Bellens, K., Van Damme, J., Van Den Noortgate, W., Wendt, H., & Nilsen, T. (2019). Instructional quality: Catalyst or pitfall in educational systems' aim for high achievement and equity? An answer based on multilevel SEM analyses of TIMSS 2015 data in Flanders (Belgium), Germany, and Norway. *Large-Scale Assessments in Education*, *7*(1), 1. https://doi.org/10.1186/s40536-019-0069-2

Beswick, K. (2007). Teachers' beliefs that matter in secondary mathematics classrooms. *Educational Studies in Mathematics, 65,* 95–120.

Bland, J. M., & Altman, D. G. (1996). *Statistics notes: measurement error*. Bmj, 313(7059), 744. doi:10.1136/bmj.312.7047.1654

Blömeke, S., & Olsen, R. V. (2019). Consistency of results regarding teacher effects across subjects, school levels, outcomes and countries. *Teaching and Teacher Education*, *77*, 170–182. https://doi.org/10.1016/j.tate.2018.09.018

Blömeke, S., Olsen, R. V., & Suhl, U. (2016). Relation of student achievement to the quality of their teachers and instructional quality. In T. Nilsen & J.-E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes*. IEA Research for Education 2 (pp. 21–50). Cham: Springer. doi:10.1007/978-3-319-41252-8

Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful literature review*. SAGE.

Bowen, N., & Guo, S. (2011). *Structural Equation Modeling*. Retrieved from https://www-oxfordscholarship-com.ezproxy.ub.gu.se/view/10.1093/acprof:oso/9780195367621.001.0001/acprof-9780195367621-chapter-0001

Brown, T. A. (2015). Methodology in the social sciences. *Confirmatory factor analysis for applied research* (2nd ed.). New York, USA: The Guilford Press.

Canales, A., & Maldonado, L. (2018). Teacher quality and student achievement in Chile: Linking teachers' contribution and observable characteristics. *International Journal of Educational Development*, *60*, 33–50. https://doi.org/10.1016/j.ijedudev.2017.09.009

Caro, D. H. & Cortés, D. (2012). Measuring family socioeconomic status: An illustration using data from PIRLS 2006. *IERI Monograph Series: Issues and Methodologies in Large Scale Assessments, 5,* 9-33.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research Methods in Education* (7th ed.). London: Routledge.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools.* London, New York: Routledge.

Creemers, B. P. M., & Kyriakides, L. (2010). Using the Dynamic Model to develop an evidence-based and theory-driven approach to school improvement. *Irish Educational Studies, 29*(1), 5–23. https://doi.org/10.1080/03323310903522669

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1), 1–44.

Darling-Hammond, L. (2014). Strengthening clinical preparation: The holy grail of teacher education. *Peabody Journal of Education, 89*(4), 547–561.

Diamantopoulos, A., & Siguaw, J. A. (2000). Introducing LISREL: A guide for the uninitiated. Sage.

DoE. (2003). Review of the financing, resourcing and cost of education in public schools. *Government Gazette*.

European Commission. (2016). *Strategic framework for education and training*. Retrieved from http://ec.europa.eu/education/policy/strategic-framework/index_en.htm

European Science Foundation (ESF). (2011). *The European code of conduct for research integrity. Revised version*. Berlin: ALLEA - All European Academies.

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

Field, S., Kuczera, M., & Pont, B. (2007). *No more failures: Ten steps to equity in education*. Paris, France: OECD Publishing. https://doi.org/10.1787/19901496

Field, A. (2009). *Discovering statistics using IBM SPSS statistics*. Sage.

Frempong, G., Reddy, V., & Kanjee, A. (2011). Exploring equity and quality education in South Africa using multilevel models. *Compare: A Journal of Comparative and International Education*, *41*(6), 819–835.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference 11.0 update* (4th ed.). Boston: Allyn & Bacon, 231.

Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis*. NCCTQ Report. National Comprehensive Center for Teacher Quality, Washington, DC, USA.

Gorard, S., & Smith, E. (2004). An international comparison of equity in education systems. *Comparative Education*, *40*(1), 15–28. https://doi.org/10.1080/0305006042000184863

Grabau, L. J., & Ma, X. (2017). Science engagement and science achievement in the context of science instruction: A multilevel analysis of U.S. students and schools.

International Journal of Science Education, *39*(8), 1045–1068. https://doi.org/10.1080/09500693.2017.1313468
Gustafsson, J. E. (1998). Social background and teaching factors as determinants of reading achievement at class and individual levels. *Journal of Nordic Educational Research, 18*(4), 241–250.
Gustafsson, J.-E. (2003). What do we know about effects of school resources on educational results? *Swedish Economic Policy Review, 10,* 77–110.
Gustafsson, J.E. (2013). Causal inference in educational effectiveness research: a comparison of three methods to investigate effects of homework on student achievement. *School Effectiveness and School Improvement, 24,* 275-295.
Gustafsson, J.-E. (2016). Lasting effects of quality of schooling: Evidence from PISA and PIAAC. *Intelligence*, *57*, 66–72. https://doi.org/10.1016/j.intell.2016.05.004
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
Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Englewood Cliffs, NJ: Prentice Hill.
Hamel, R. (2007). The dominance of English in the international scientific periodical literature and the future of language use in science. *AILA Review. 20*. 53-71. 10.1075/aila.20.06ham.
Hanushek, E. A., & Luque, J. A. (2003). Efficiency and equity in schools around the world. *Economics of Education Review*, *22*(5), 481–502.
Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, *46*(3), 607–668. https://doi.org/10.1257/jel.46.3.607
Hanushek, E. A., & Woessmann, L. (2015). *Universal basic skills: What countries stand to gain*. Paris: OECD Publishing.
Hattie, J. (2003). Teachers make a difference: what is the research evidence? Paper presented at the Australian Council for Educational Research Annual Conference on Building Teacher Quality, Melbourne.
Hattie, J. (2008). Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement. Routledge: Taylor & Francis Group.
Henson, R. (2002). From adolescent Angst to Adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist, 37* (2), 137–150.
Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Articles,* 2.
Howie, S. J., & Pietersen, J. J. (2001). Mathematics literacy of final year students: South African realities. *Studies in Educational Evaluation, 27*(1), 7–25.
Hox, J. (2002). *Quantitative methodology series. Multilevel analysis techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
Hox, J. (2010). *Quantitative methodology series. Multilevel analysis: Techniques and applications* (2nd ed.). Publisher: Routledge.
Hox, J. J., Moerbeek, M., & Schoot, R. van de. (2010). *Multilevel Analysis: Techniques and Applications,* Second Edition. Routledge.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

Jesson, J., Matheson, L., & Lacey, F. M. (2011). *Doing your literature review: Traditional and systematic techniques*. SAGE.

Johansson, S., & Myrberg, E. (2019). Teacher specialization and student perceived instructional quality: What are the relationships to student reading achievement? *Educational Assessment, Evaluation and Accountability*, *31*(2), 177–200. https://doi.org/10.1007/s11092-019-09297-5

Joncas, M., & Foy, P. (2010). *Sample design in TIMSS and PIRLS*. Boston, MA: TIMSS and PIRLS International Study Center.

Khine, M. S. (Ed.). (2013). Application of structural equation modeling in educational research and practice. Sense Publishers.

Klassen, R., & Tze, V. M. (2014). Teacher's self-efficacy, personality and teacher's effectiveness: Ameta-analysis. *Educational Research Review, 12,* 59–76.

Klassen, R., Tze, V., Betts, S., & Gordon, K. (2011). Teacher efficacy research 1998–2009: signs of progress or unfulfilled promise? *Educational Psychology Review, 23*(1), 21–43.

Kline, R. B. (2011). Principles and Practice of Structural Equation Modeling (3rd ed.). New York: Guilford Press.

Kline, R. B. (2015). Principles and Practice of Structural Equation Modeling, Fourth Edition. Guilford Publications.

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, *9*(3), 231–251. https://doi.org/10.1007/s10984-006-9015-7

Kyriakides, L., Campbell, R. J., & Gagatsis, A. (2000). The Significance of the Classroom Effect in Primary Schools: An Application of Creemers' Comprehensive Model of Educational Effectiveness. *School Effectiveness and School Improvement*, *11*(4), 501–529.

Kyriakides, L., Charalambous, E., Creemers, H. P. M. (Bert), & Dimosthenous, A. (2019). Improving quality and equity in schools in socially disadvantaged areas. *Educational Research*, *61*(3), 274–301.

Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, *36*, 143–152.

LaRoche, S., & Foy, P. (2016). Sample implementation in TIMSS 2015. In Martin, M.O., Mullis, I. V. S., & Hooper, M. (Eds.), *Methods and procedures in TIMSS 2015* (pp. 5.1–5.175). Retrieved from http://timss.bc.edu/publications/timss/2015-methods/chapter-5.html

Laukaityte, I., & Wiberg, M. (2017). Using plausible values in secondary analysis in large-scale assessments. *Communications in Statistics - Theory and Methods, 46*(22), 11341–11357. https://doi.org/10.1080/03610926.2016.1267764

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34*(2), 120–131.

Luque, M., Marcenaro-Gutierrez, O. D., & Ruiz, A. B. (2020). Evaluating the global efficiency of teachers through a multi-criteria approach. *Socio-Economic Planning Sciences*, *70*, 100676. https://doi.org/10.1016/j.seps.2018.12.003

Luschei, T., & Chudgar, A. (2011). Teachers, student achievement and national income: A cross-national examination of relationships and interactions. *Prospects (00331538)*, *41*(4), 507–533. https://doi.org/10.1007/s11125-011-9213-7

Martin, M. O., Mullis, I. V. S., Hooper, M., Yin, L., Foy, P., & Palazzo, L. (2016). Creating and Interpreting the TIMSS 2015 Context Questionnaire Scales. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 15.1-15.312). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timss.bc.edu/publications/timss/2015-methods/chapter-15.html

Mislevy, R. J. (1991). Randomization–based inference about latent variables from complex samples. *Psychometrika 56*:177–96.

Mueller, C. W., & Parcel, T. L. (1981). Measures of Socioeconomic Status: Alternatives and Recommendations. *Child Development*, *52*(1), 13–30.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics***.** Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/international-results/

Mullis, I.V.S., Martin, M.O., & Loveless, T. (2016). *20 Years of TIMSS: International Trends in Mathematics and Science Achievement, Curriculum, and Instruction*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA).

Mullis, I., Martin, M., Ruddock, G., O'Sullivan, C., & Preuschof, C. (2009). *TIMSS 2011 Assessment Frameworks. Chapter 3*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22*(3), 376–398.

Muthén, L.K. and Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

Nachbauer, M., & Kyriakides, L. (2019). A review and evaluation of approaches to measure equity in educational outcomes. *School Effectiveness and School Improvement,* 1–26.

Nachtigall, C., Kroehne, U., Funke, F., & Steyer, R. (2003). (Why) should we use SEM? Pros and cons of structural equation modeling. *Psychological Research Online, 8*(2), 1-22.

Nilsen, T., & Gustafsson, J. E. (2016). *Teacher quality, instructional quality and student outcomes*. (Vol. 2, IEA Research for Education). Cham: Springer International Publishing.

Nilsen, T., Gustafsson, J.-E., & Blömeke, S. (2016). Conceptual Framework and Methodology of This Report. In T. Nilsen & J.-E. Gustafsson (Eds.), *Teacher Quality, Instructional Quality and Student Outcomes: Relationships Across Countries, Cohorts and Time* (pp. 1–19). Springer International Publishing. https://doi.org/10.1007/978-3-319-41252-8_1

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257.

Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics, 130,* 105–119.
Pedhazur, E.L. & Schmelkin, L.P (1991). *Measurement, design and analysis. factor analysis.* Psychometrika, 38, 1–10.
Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.
OECD. (2013). *PISA 2012: Assessment and analytical framework. Mathematics, reading, science, problem solving and financial literacy*. Paris, France: OECD Publishing.
Raudenbush, S.W., and A.S. Bryk. 2002. *Hierarchical linear models*. 2nd ed. Thousand Oaks, CA: Sage.
Reddy, V., Visser, M., Winnaar, L., Arends, F., Juan, A and Prinsloo, C.H., and Isdale, K. (2016). *TIMSS 2015: Highlights of Mathematics and Science Achievement of Grade 9 South African Learners*. Human Sciences Research Council.
Reddy, V., Juan, A., Isdale, K., & Fongwa, S. (2019). Mathematics Achievement and the Inequality Gap: TIMSS 1995 to 2015. In N. Spaull & J. D. Jansen (Eds.), *South African Schooling: The Enigma of Inequality: A Study of the Present Situation and Future Possibilities* (pp. 169–187). Springer International Publishing. https://doi.org/10.1007/978-3-030-18811-5_9
Reimer, D., Sortkear, B., Oskarsson, M., Nilsen, T., Rasmusson, M., & Nissinen, K. (2018). *Northernlights on TIMSS and PISA 2018*. Copenhagen, Denmark: Nordisk Ministerrad.
Reyes, L. H. (1984). Affective variables and mathematics education. *The Elementary School Journal*, *84*(5), 558–581.
Reynolds, D., Sammons, P., Fraine, B. D., Damme, J. V., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement*, *25*(2), 197–230.
Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.
Rutkowski, L., Gonzalez, E., Joncas, M., &, von Davier, M. (2010). Methodological challenges with international large scale assessment data. *Educational Researcher, 39,* 142 151.
Sayed, Y., & Ahmed, R. (2011). Education quality in post-apartheid South African policy: Balancing equity, diversity, rights and participation. *Comparative Education*, *47*(1), 103–118.
Scheerens, J., & Blömeke, S. (2016). Integrating teacher education effectiveness research into educational effectiveness models. *Educational Research Review*, *18*, 70–87.
Scherer, R., & Nilsen, T. (2016). The relations among school climate, instructional quality, and achievement motivation in mathematics. *Teacher Quality, Instructional Quality and Student Outcomes*, *2*, 51–80.
Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453.
Slater, H., Davies, N. M., & Burgess, S. (2012). Do teachers matter? Measuring the variation in teacher effectiveness in England. *Oxford Bulletin of Economics & Statistics*, *74*(5), 629–645. https://doi.org/10.1111/j.1468-0084.2011.00666
Spaull, N. (2019). Equity: A Price Too High to Pay? In Nic Spaull & J. D. Jansen (Eds.), *South African Schooling: The Enigma of Inequality: A Study of the Present Situation*

and Future Possibilities (pp. 1–24). Springer International Publishing. https://doi.org/10.1007/978-3-030-18811-5_1
Spaull, N. (2013a). South Africa's education crisis: The quality of education in South Africa 1994-2011. *Johannesburg: Centre for Development and Enterprise,* 1-65.
Spaull, N. (2013b). Poverty & privilege: Primary school inequality in South Africa. *International Journal of Educational Development*, *33*(5), 436–447. https://doi.org/10.1016/j.ijedudev.2012.09.009
Takyi, S. A., Amponsah, O., Asibey, M. O., & Ayambire, R. A. (2019). An overview of Ghana's educational system and its implication for educational equity. *International Journal of Leadership in Education*, 1–26.
Toropova, A., Johansson, S., & Myrberg, E. (2019). The role of teacher characteristics for student achievement in mathematics and student perceptions of instructional quality. *Education Inquiry*, 1–25. https://doi.org/10.1080/20004508.2019.1591844
Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher Efficacy: Its Meaning and Measure. *Review of Educational Research, 68*(2), 202–248.
Van Damme, J., & Bellens, K. (2017). Countries strive towards more quality and equity in education: Do they show success or failure? Evidence from TIMSS 2003 and 2011, for Grade 4. In M. Rosén, K. Yang Hansen, & U. Wolff (Eds.), *Cognitive Abilities and Educational Outcomes: A Festschrift in Honour of Jan-Eric Gustafsson* (pp. 127–148).
Van der Berg, S. (2008). How effective are poor schools? Poverty and educational outcomes in South Africa. *Studies in Educational Evaluation, 34*(3), 145–154. https://doi.org/10.1016/j.stueduc.2008.07.005
Visser, M., Juan, A., & Feza, N. (2015). Home and school resources as predictors of mathematics performance in South Africa. *South African Journal of Education, 35*(1). https://doi.org/10.15700/201503062354
Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI monograph series, 2,* 9-36.
Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology*, *108*(5), 705–721.
Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A Review. *Review of Educational Research*, *73*(1), 89–122.
White, K. R. (1982). The relation between socio-economic status and academic achievement. *Psychological Bulletin*, *91*(3), 461–481. https://doi.org/10.1037/0033-2909.91.3.461
Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics, 100,* 61–78.
Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation, 31*(2–3), 114–128.
Wyatt, M. (2014). Towards a re-conceptualization of teachers' self-efficacy beliefs: Tackling enduring problems with the quantitative research and moving on. *International Journal of Research and Method in Education, 37*(2), 166–189.
Yang-Hansen, K. (2003). Dimensions of socio-economic status and their relationship to mathematics and science achievement at individual and collective levels. *Scandinavian Journal of Educational Research, 47*(1), 21–41. https://doi.org/10.1080/00313830308609

Yang-Hansen, K., & Gustafsson, J. E. (2004). Measuring socioeconomic status at individual and collective levels. *Educational Research and Evaluation, 10*(3), 259–288. https://doi.org/10.1076/edre.10.3.259.30268

Yang Hansen, K., & Gustafsson, J. E. (2016). Causes of educational segregation in Sweden–School choice or residential segregation. *Educational Research and Evaluation, 22*(1-2), 23-44.

Yang-Hansen, K., Gustafsson, J. E., Rosen, M., Sulkunen, S., Nissinen, K., Kupari, P., Olafsson, R. F., Bjornsson, J. K. et al. (2014). *Northern lights on TIMSS and PIRLS 2011: Differences and similarities in the Nordic countries*. Copenhagen, Denmark: Nordic Council of Ministers.

Zee, M., & Koomen, H. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research, 86,* 981–1015.

14 Appendices

14.1 Appendix 1. Selection of relevant articles for literature review

First, and potentially most importantly, the study is drawn from an interdisciplinary body of scholarly literature pertaining to educational quality and equity, with a focus on the South Africa experience and the rest of the world. The search strategies for the most popular databases such as Google Scholar, Scopus, and Eric (Education Resources Information Center) were used. The University of Gothenburg library Super Search was adhered to as far as possible. Therefore, the use of at least three databases gives safe mechanisms that are more likely to overcome the limitations of anyone database.

The search for an interdisciplinary body of scholarly literature was guided by keywords such as educational quality, equity, and TIMSS through the University of Gothenburg library (Super Search). This revealed over five thousand and nine scholarly literature results to be analyzed. This number of publication was too large at this point; for that reason, the same search strategy was confined to scholarly literature in Scopus and Eric database with the keywords: educational quality, equity, and TIMSS providing eight research publications of which six is an article, one is a book and the other is book chapter. As such, only two publications were chosen since they are relevant to the purpose of this study. However, these two publications were not enough to form the sample or corpus of this review, hence additional searches were made in Scopus, Eric and Google Scholar pertaining to educational quality and equity on several occasions with the search words South Africa, TIMSS, quality in education South Africa, equity in education South Africa, and mathematics. Unexpectedly, Google Scholar pulled out twenty-three thousand and hundred literature results whiles publications are unlikely to be found in Scopus and Eric.

Also, as teacher and student background characteristics are proxies for educational quality and equity in this study. Several searches were employed to gather the best possible pool of studies to represent a large number of existing studies on aspects of teacher quality and student's family SES background. The search was conducted using Scopus, and Eric. For teacher background factors, the search terms *teacher quality*, *mathematics achievement*, and *TIMSS* were used. The other search words were *teacher instructional quality* and *mathematics achievement*. For students' SES background the terms *socioeconomic status*, mathematics achievement, and TIMSS were used. The search yielded thirty-eight, sixty-one, and twenty-seven publications respectively. Furthermore, additional scholarly literature was identified from the reference lists of research publications found in the database search including studies that employed *two-level structural equation modelling* as the main analytical method. All the identified publications were uploaded into Zotero, and duplicates were removed. Therefore, in order to obtain at least twenty scholarly articles for my corpus, the available publications at my possession were screened based on full text; and those which were unrelated to my topic were removed by skimming and critically appraised the titles and abstracts, resulting in about at least twenty scholarly articles, including academic book chapters for my review. The selection criteria would be further discussed below.

14.2 Selection Criteria

The scope of this review was determined by a number of search parameters informing my decision to consider selection criteria for including and excluding research publication. Thus, the framework for inclusion and exclusion were outlined below:

14.2.1 Inclusion criteria

- Studies representing the teacher and teaching concept, and student's family background characteristics.
- Studies that constitute outcome measure; student achievement on a standardized test.
- Studies representing the basic descriptive information. These are geographical scope such as studies published in English only. For instance, more than 90 percent of sciences publications are written in English (Hamel, 2007). Studies focused on South Africa and across the region of the world, also reflect a geographical scope. Thus, research across the rest of the world was necessary in order to be able to draw lessons from research in other countries and may identify gaps in the South Africa-based research literature.
- Time scale such as studies published from either 2000-2010 or 2011-2020, deemed necessary to investigate the most recent research findings.

14.2.2 Exclusion criteria

- Studies published prior to 2000.
- Incomplete publication details.
- A master thesis was not included due to time constraints.
- Studies dedicated to book reviews and non-English language.

However, it is important I mention here that after doing a lot of reading on my topic, a decision was made to include a few older publications that although outside my critical timescale (2000-2010/2011-2020), but I refer to them because they provide interesting findings and they also provide needed context for subsequent studies. In addition, relevant publications could appear in a number of different journals indexed in different databases. However, my selection of databases was informed by advice from the University of Gothenburg librarians and with some IMER Doctors and Professors, preliminary searches, and careful review of individual titles.

14.3 Appendix 2. Inputs of measurement model of teacher instructional quality

```
TITLE: Confirmatory Factor Analysis of Teacher's Instructional Quality;
  DATA:  File is TEACHERQ14_Q17.dat;
  VARIABLE:
      NAMES ARE IDTEACH BTBG14A_R_mean BTBG14B_R_mean BTBG14C_R_mean
    BTBG14D_R_mean BTBG14E_R_mean BTBG14F_R_mean
    BTBG14G_R_mean BTBM17A_R_mean BTBM17B_R_mean
    BTBM17C_R_mean BTBM17D_R_mean BTBM17E_R_mean
```

```
BTBM17F_R_mean BTBM17G_R_mean BTBM17H_R_mean BTBM17I_R_mean;

USEVARIABLES ARE BTBG14A_R_mean BTBG14B_R_mean BTBG14C_R_mean
BTBG14D_R_mean BTBG14E_R_mean BTBG14F_R_mean
BTBG14G_R_mean;

MISSING ARE all (-99);

ANALYSIS:
ESTIMATOR=MLR;

MODEL:
InQua BY BTBG14A_R_mean BTBG14B_R_mean BTBG14C_R_mean
BTBG14D_R_mean BTBG14E_R_mean BTBG14F_R_mean
BTBG14G_R_mean;

OUTPUT: Standardized Modindices;
```

14.4. Appendix 3. Inputs of the measurement model of teacher's confident (TCM) in mathematics

```
TITLE: Confirmatory Factor Analysis of Teacher's Confident in Mathematics;
DATA:  File is TEACHERQ14A17.DAT;
VARIABLE:
  NAMES ARE IDTEACH BTBG14A_R_mean BTBG14B_R_mean BTBG14C_R_mean
  BTBG14D_R_mean BTBG14E_R_mean BTBG14F_R_mean
  BTBG14G_R_mean BTBM17A_R_mean BTBM17B_R_mean
  BTBM17C_R_mean BTBM17D_R_mean BTBM17E_R_mean
  BTBM17F_R_mean BTBM17G_R_mean BTBM17H_R_mean BTBM17I_R_mean;

USEVARIABLES ARE BTBM17A_R_mean BTBM17B_R_mean
  BTBM17C_R_mean BTBM17D_R_mean BTBM17E_R_mean
  BTBM17F_R_mean BTBM17G_R_mean BTBM17H_R_mean BTBM17I_R_mean;

  MISSING ARE all (-99);

  ANALYSIS:
  ESTIMATOR=MLR;

  MODEL:
  TeachC BY BTBM17A_R_mean BTBM17B_R_mean
  BTBM17C_R_mean BTBM17D_R_mean BTBM17E_R_mean
  BTBM17F_R_mean BTBM17G_R_mean BTBM17H_R_mean BTBM17I_R_mean;

  OUTPUT: Standardized Modindices;
```

14.5. Appendix 4. Inputs of the structural model.

```
TITLE: Two-Level Confirmatory Factor Analysis Model for South Africa;
DATA: FILE IS SA_TIMSS_5PV.TXT;
 TYPE = IMPUTATION;
      VARIABLE:
       NAMES ARE IDSCHOOL IDCLASS IDSTUD IDTEACH IDLINK IDGRADER IDGRADE
       NMTEACH NTEACH MATWGT BSMMAT BTBG01 BTBG02 BTBG03 BTBG04 BTBG14A
       BG14A_R BTBG14B BG14B_R BTBG14C BG14C_R BTBG14D BG14D_R BTBG14E BG14E_R
       BTBG14F BG14F_R BTBG14G BG14G_R BTBM16 BTBM17A BM17A_R BTBM17B BM17B_R
       BTBM17C BM17C_R BTBM17D BM17D_R BTBM17E BM17E_R BTBM17F BM17F_R
       BTBM17G BM17G_R BTBM17H BM17H_R BTBM17I BM17I_R BTBGEAS BTBGSOS BTBGSCR
       BTBGTJS BTBGCFT BTBGLSN BTDM05 BSBG01 BSBG03 BSBG03_R BOOKS BSBG05
       BSBG06A BSBG06B BSBG06C BSBG06D BSBG06E BSBG06F BSBG06G BSBG06H BSBG06I
       BSBG06J BSBG06K BSBG08 BSDAGE WGTADJ1 WGTADJ2 WGTADJ3 WGTFAC1 WGTFAC2
       WGTFAC3 BSBGHER BSBGSSB BSBGSB BSBGSLM BSBGEML BSBGSCM BSBGSVM BSBGSCP
       HSS BSDGEDUP PEDU BSDMWKHW BSDMWKHW_R;

 USEVARIABLES ARE PEDU BOOKS HSS BG14A_R BG14B_R BG14C_R
       BG14D_R BG14E_R BG14F_R BG14G_R BM17I_R BM17A_R BM17B_R
       BM17C_R BM17D_R BM17E_R BM17F_R BM17G_R BM17H_R
       BSMMAT BTBG01 BTBG04 BTDM05 HOUWGT;

       BETWEEN = BG14A_R BG14B_R BG14C_R BG14D_R BG14E_R BG14F_R
       BG14G_R BM17I_R BM17A_R BM17B_R BM17C_R BM17D_R BM17E_R
       BM17F_R BM17G_R BM17H_R BTBG01 BTBG04 BTDM05;
       USEOBSERVATIONS ARE (IDTEACH NOT EQ 16301);

         MISSING ARE all (-99);
         CLUSTER=IDTEACH;
         WEIGHT = HOUWGT;
         BWEIGHT =MATWGT;

   DEFINE:
          BSMMAT=BSMMAT/100;
          HOUWGT= WGTADJ3*WGTFAC3;

ANALYSIS:
         TYPE=TWOLEVEL;
         ESTIMATOR=MLR;

 MODEL:
            %WITHIN%
             SESW BY PEDU BOOKS HSS;
             BSMMAT ON SESW;
            %BETWEEN%
             SESB BY PEDU BOOKS HSS;
            INQUA BY BG14A_R BG14B_R BG14C_R BG14D_R
             BG14E_R BG14F_R BG14G_R;
             TCM BY BM17A_R BM17B_R BM17C_R BM17D_R BM17E_R
             BM17F_R BM17G_R BM17H_R BM17I_R;
             BSMMAT ON SESB InQua TCM BTBG01 BTBG04 BTDM05;
             InQua TCM BTBG01 BTBG04 BTDM05 ON SESB;
             InQua ON TCM BTBG01 BTBG04 BTDM05;
             TCM WITH BTBG01 BTBG04 BTDM05;
```

```
BTBG01 BTBG04 BTDM05 WITH BTBG01 BTBG04 BTDM05;
```

OUTPUT: SAMPSTAT STANDARDIZED MODINDICES;

MODEL INDIRECT:
BSMMAT IND SESB;

14.6. Appendix 5. Summary of analysis

Number of groups	1
Average number of observations	12444
Number of replications	
Requested	5
Completed	5
Number of dependent variables	23
Number of independent variables	0
Number of continuous latent variables	4

Observed dependent variables

Continuous

BG14A_R	BG14B_R	BG14C_R	BG14D_R	BG14E_R	BG14F_R
BG14G_R	BM17I_R	BM17A_R	BM17B_R	BM17C_R	BM17D_R
BM17E_R	BM17F_R	BM17G_R	BM17H_R	BTBG01	BTBG04
BTDM05	PEDU	BOOKS	HSS	BSMMAT	

Continuous latent variables

SESW	SESB	INQUA	TCM

Variables with special functions

Cluster variable IDTEACH
Weight variable (cluster-size scaling)
HOUWGT
Between weight variable (sample-size scaling)
MATWGT

Between variables

BG14A_R	BG14B_R	BG14C_R	BG14D_R	BG14E_R	BG14F_R
BG14G_R	BM17I_R	BM17A_R	BM17B_R	BM17C_R	BM17D_R
BM17E_R	BM17F_R	BM17G_R	BM17H_R	BTBG01	BTBG04
BTDM05					

Estimator	MLR
Information matrix	OBSERVED
Maximum number of iterations	100
Convergence criterion	0.100D-05
Maximum number of EM iterations	500
Convergence criteria for the EM algorithm	

Loglikelihood change	0.100D-02
Relative loglikelihood change	0.100D-05
Derivative	0.100D-03
Minimum variance	0.100D-03
Maximum number of steepest descent iterations	20
Maximum number of iterations for H1	2000
Convergence criterion for H1	0.100D-03
Optimization algorithm	EMA

Input data file(s)
Multiple data files from
SA_TIMSS_5PV.txt
Input data format FREE

SUMMARY OF DATA FOR THE FIRST DATA SET

Number of missing data patterns	61
Number of clusters	326
Average cluster size	38.172

Estimated Intraclass Correlations for the Y Variables

Variable	Intraclass Correlation	Variable	Intraclass Correlation	Variable	Intraclass Correlation
PEDU	0.094	BOOKS	0.158	HSS	0.162
BSMMAT	0.623				

14.7 Appendix 6. STDYX Standardization

Standardized Total, Total Indirect, Specific Indirect, and Direct Effects

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value	Rate of Missing
WITHIN					
Effects from SESB to BSMMAT					
Total	0.000	0.000	999.000	0.000	NaN
Total indirect	0.000	0.000	999.000	0.000	NaN
BETWEEN					
Effects from SESB to BSMMAT					
Total	0.879	0.030	29.669	0.000	0.009
Total indirect	0.006	0.010	0.587	0.557	0.002
Specific indirect 1					
BSMMAT					
BTBG01					

SESB	0.000	0.003	0.081	0.935	0.000
Specific indirect 2					
BSMMAT					
BTBG04					
SESB	0.007	0.006	1.303	0.192	0.004
Specific indirect 3					
BSMMAT					
BTDM05					
SESB	0.000	0.001	-0.103	0.918	0.005
Specific indirect 4					
BSMMAT					
INQUA					
SESB	-0.001	0.007	-0.164	0.870	0.000
Specific indirect 5					
BSMMAT					
TCM					
SESB	0.000	0.002	0.177	0.859	0.001
Specific indirect 6					
BSMMAT					
INQUA					
BTBG01					
SESB	0.000	0.001	0.081	0.936	0.000
Specific indirect 7					
BSMMAT					
INQUA					
BTBG04					
SESB	0.000	0.001	0.344	0.731	0.001
Specific indirect 8					
BSMMAT					
INQUA					
BTDM05					
SESB	0.000	0.000	0.188	0.851	0.000
Specific indirect 9					
BSMMAT					
INQUA					
TCM					
SESB	-0.001	0.005	-0.272	0.786	0.002
Direct					
BSMMAT					
SESB	0.874	0.032	27.218	0.000	0.009