FACTORS INFLUENCING YEAR 9 STUDENTS' MATHEMATICS PERFORMANCE RELATED TO LOWER ORDER THINKING (LOT) AND HIGHER ORDER THINKING (HOT) IN ACEH, INDONESIA: A MULTIVARIATE AND MULTILEVEL ANALYSIS

Elizar

B.Ed (Mathematics Education) (Syiah Kuala University)
M.Ed (MT) (The University of Adelaide)

This thesis is submitted in fulfilment of the requirements for
the degree of Doctor of Philosophy

School of Education, Faculty of Arts
University of Adelaide
November 2017

Table of Content

List of Tables

List of Figures

Abstract

This study examines various factors associated with students' mathematics performance, specifically in relation to higher order thinking (HOT) and lower order thinking (LOT). It examines the student-, teacher- and school-level factors, their interrelationships and impact on Year 9 students' mathematics performance in Aceh, Indonesia.

The theoretical basis of this study comes from research on childhood cognitive development and educational theory, educational effectiveness theory, and a review of numerous previous studies related to how variables at student-, teacher- and school-level contribute to students' mathematics performance. The conceptual framework is a multilevel analysis of the factors influencing students' performance related to LOT and HOT designed to examine the possible relationships within and between student-, teacher- and school-level variables. Student-level variables include students' background, attitudes and beliefs, as well as classroom practices as perceived by students. Teacher-level variables include teachers' background, beliefs, and classroom practices as perceived by teachers. School-level variables include school demographics information and resources.

The study employs a quantitative method. Questionnaires and a mathematics test were used to obtain data from students, teachers and schools. Questionnaires were given to students, mathematics teachers and principals/administrators at the schools and a mathematics test administered to the students. The questionnaires were administered to a total of 1135 Year 9 students, 46 Year 9 mathematics teachers and 25 schools from one major city (representing the urban area) and one district (representing the rural area) in the province of Aceh, Indonesia. Scales in the questionnaires were validated using confirmatory factor analysis (CFA) and Rasch analysis. The data was then analysed employing single-level and multilevel analysis techniques. Partial least squares path analysis (PLS-PA) and hierarchical linear modelling (HLM) were employed to examine the relationships between variables tested in this study.

The results from the single-level analysis using PLS-PA show that there are five variables directly influencing students' mathematics performance relating to LOT: (a) students' beliefs concerning mathematics related to LOT; (b) gender; (c) school location; (d) socio-economic status (SES); and (e) students' attitude of liking

mathematics. The multilevel analysis using HLM indicates that there are seven variables (three at student-level, three at teacher-level and one at school-level) that have a direct impact on the students' mathematics performance related to LOT: (a) students' liking of mathematics; (b) students' beliefs concerning mathematics related to LOT; (c) students' beliefs concerning mathematics related to HOT; (d) teachers' professional development; (e) instructional activities; (f) teachers' beliefs concerning mathematics related to HOT; and (g) school resources.

The results from the single-level analysis using PLS-PA indicate that four variables directly influence students' mathematics performance related to HOT, namely: (a) students' mathematics performance related to LOT; (b) students' educational expectations; (c) SES; and (d) school location. The multilevel analysis using HLM indicates seven variables (four at student-level, two at teacher-level and one at school-level) that directly influence student mathematics performance related to HOT, namely: (a) students' mathematics performance related to LOT; (b) students' educational expectations; (c) students' individual judgement of mathematics ability; (d) students' beliefs concerning mathematics related to LOT; (e) teacher certification; (f) teachers' beliefs concerning mathematics teaching related to HOT; and (g) the availability of a 'Mathematics Olympiad' club at the schools.

This study contributes to the literature of how student-, teacher- and school-level variables influence students' mathematics performance related to LOT and HOT, especially in the context of Aceh, Indonesia, a developing nation. This study also provides empirical evidence of Acehnese students' mathematics performance related to LOT and HOT, indicating their poor performance in questions related to both LOT and HOT. While students throughout the world struggle with mathematics problems that require HOT, in Aceh, and Indonesia in general, students are still struggling with LOT. This is clearly a subject of a great concern for the development of mathematics education in Aceh and Indonesia. As the current trends in education have shifted from lower order to higher order thinking, Indonesia as a rapidly developing nation needs to meet the challenge of progressing the nation's education. Thus, the findings of this study have important implications for the improvement of mathematics teaching and learning in Aceh, Indonesia. Mathematics teaching and learning that improve both lower order thinking and higher order thinking skills should be of major concern for Indonesia and the efficient mathematics education of its students.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time

Signed:

Date: 6th November 2017

Acknowledgements

This study would not have been possible without the assistance of many individuals, organisation and schools. I would like to express my deepest gratitude and appreciation to them all.

- The Australian Awards Scholarship (AAS) programme for the full scholarship and support during my study.
- My principal supervisor, Dr I Gusti Ngurah Darmawan, for his guidance and encouragement during my study. His ongoing support does not only make the completion of this thesis possible but also inspires me to be an independent researcher and a lifelong learner.
- My co supervisor, Dr Peggy Lynch, for her time reading my draft despite being away in Ireland.
- Professor John Keeves, for his commitment and enthusiasm in reading my drafts and his constructive feedback on each chapter of this thesis.
- My best friend and 'personal editor', Wendy Baker, for her assistance to read each initial draft of the chapters in this thesis and for her endless support during my highs and lows.
- The 25 schools in the province of Aceh, their teachers and their students for their time participating in my study.
- My late parents who allowed me to go far away for study despite their struggle of missing me each day, without their sincere prayers for my success I would not have come this far. It is very sad that I cannot tell them in person that I have completed my study but I will make sure that they are always in my prayers.
- My siblings, nephews and nieces for making me look forward to coming home.
- My beloved friends: Che Yee, Michelle and Alex for the wonderful 'quantitative group' support system, for the great discussion on our research methods and for the great times we had; Patrick and Giang, for the jokes and laughs we shared each day that make this PhD journey more bearable; Rachel for best companion during my revision days; and Safira for being there for me each time I need.
- All my colleagues and the staff of the school of Education at the University of Adelaide for their support during my study.

Chapter 1 Introduction

1.1. Background of the Study

Throughout history mathematics has been at the centre of scientific progress and is now integral to the technological developments of the digital age. It is a corner stone of modern education and important for everyday life skills (Harackiewicz, Rozek, Hulleman, & Hyde, 2012; Kumar, 2011; Reyna & Brainerd, 2007), as well as career paths (Gowers, 2000; Harackiewicz et al., 2012; Schoenfeld, 2002). Equipping students with mathematics skills prepares them for life in modern society. Some level of understanding of mathematics, mathematical reasoning and mathematical tools are required to comprehend and solve numerous problems and situations in daily life, including in professional contexts (OECD, 2013a). However, as is evidenced by international studies (Mullis, Martin, Foy, & Arora, 2012; OECD, 2013b), many students throughout the world continue to struggle to attain mathematical competence. This phenomenon is not restricted to developing countries, with Australia and the US not ranked highly in international studies such as the Programme for International Student Assessment (PISA) and trends in International Mathematics and Science Study (TIMMS) (Fleischman, Hopstock, Pelczar, & Shelley, 2010; Kelly, Nord, Jenkins, Chan, & Kastberg, 2013; Thomson, De Bortoli, & Buckley, 2013; Thomson, De Bortoli, Nicholas, Hilman, & Buckley, 2013). National education systems are continually renewing their efforts to establish more effective mathematics curricula and teaching programmes in order to improve students' learning outcomes (Li & Lappan, 2014; Mailizar, Alafaleq, & Fan, 2014; Zhang & Stephens, 2013).

Mathematics learning requires various cognitive skills. These are described in Bloom's (revised) taxonomy of educational objectives as six hierarchical cognitive processes: remembering, understanding, applying, analysing, evaluating and creating (Anderson et al., 2001). Remembering, understanding and applying are then categorised as lower order thinking (LOT) and analysing, evaluating and creating are higher order thinking (HOT). This categorisation of thinking and assessment can be applied to all disciplines. However, the sophistication of the modern world places an increasing emphasis on HOT (Mevarech & Kramarski, 2014; J. Osborne, 2013). In real life problems, these thinking skills are interwoven and interdependent.

LOT is a reproduction of previous repetitions (Lewis & Smith, 1993), closely related to memorisation and recalling information (Miri, David, & Uri, 2007), with a simple application of the knowledge to familiar problems or exercises (Zoller, 2002). HOT is more difficult to define. According to Lewis and Smith (1993, p.136), when a person uses both "new information and information stored in memory and interrelates and/or rearranges and extends this information to achieve a purpose or find possible answers in perplexing situations". HOT is linked to problem solving and critical thinking and these are considered as an essential component of skills needed in the 21st century (Edens, 2000; King, Goodson, & Rohani, 1998; Sheffield, 2007; Silva, 2008; Trilling & Fadel, 2009). HOT assists students to face the challenges of a dynamic and innovative world (Forster, 2004; Sendag & Ferhan Odabasi, 2009; Sheffield, 2007). In the field of mathematics, LOT means learning and being able to apply mathematical principals and formulae. While, HOT means understanding the logic behind mathematical principals and being able to apply mathematical reasoning to understand and solve unfamiliar problems in new contexts (OECD, 2013a).

As students encounter more and more unfamiliar challenges in their daily life, where LOT is no longer sufficient, the demand for HOT, including problem solving skills, increases (Zamri, 2016). However, it is reported that “on average in OECD countries, half of the students were unable to solve problems that are more difficult than basic problems” (OECD, 2013a, p.14). The promotion of HOT skills has thus become of great concern to mathematics educators.

Despite the clear articulation of the necessity of equipping students with HOT, it is still seen as a challenge to develop both LOT and HOT skills within the classroom (Marshall & Horton, 2011) and classroom practices related to LOT are still dominating (J. Osborne, 2013). If students’ capacity in mathematics is seen as important to their education what factors contribute to the success of building students’ LOT and HOT skills? This thesis investigates the student-, teacher- and school-level factors influencing Acehnese students’ mathematics performance in relation to both LOT and HOT, as well as the cross-level relationships, providing an overview of the effectiveness of their mathematics education.

1.1.1. Indonesia, a Rapidly Developing Nation

For developing nations, education is critical to progress in the improvement in health, the lowering of poverty and unemployment within a stable government and sustainable socio-economic growth (Suryadarma & Jones, 2013). Indonesia is currently the second fastest developing country in the world, after China. It is the fourth most populous nation, with approximately 261 million in 2016. There are approximately 300 ethnic groups, speaking 700 different languages, across 6000 inhabited islands. In 2011, 32 million people were still living below the national poverty line, often minority groups living in more remote areas, where it is difficult to deliver effective education (Hayden & Martin, 2014).

Throughout the colonial period, few Indonesians received education and with independence in 1945 it was necessary to establish an extensive new education system. In an unstable political system, the government was seeking to unite the nation through the education system (Bjork, 2013). They quickly achieved a situation providing access to school for the whole nation. There are currently more than 50 million students in Indonesian schools, with 3 million teachers in over 300,000 schools. While primary education is now 96%, nearly 80% attend junior high and nearly 60% senior secondary (Hayden & Martin, 2014). The adult literacy rate in 2011 was 95.6%, with 98.7% in younger people between the ages of 15 and 24 years (Hayden & Martin, 2014). While Bahasa Indonesia is the national language of Indonesia and the language used in schools after lower primary, large numbers of its population are not fluent in it (Hayden & Martin, 2014). This, combined with traditional ways of teaching and reliance on multiple choice answers in the national examinations, may go towards explain students' performance in international testing (Bjork, 2013). The focus for Indonesia now is on the quality of its education.

1.1.2. The Indonesian Mathematics Curriculum: an Overview

Mathematics is a compulsory subject from Year 1 to Year 12 in Indonesia. As a core component of the curriculum, it is one of the subjects in the national examinations which are conducted for Year 6, Year 9 and Year 12 students. In 2006, a school-based curriculum, was implemented throughout Indonesia, having been formulated under the supervision of the regional Ministry of Education and Culture and the Ministry of Religious Affairs which are jointly responsible for education. Even though both the problem-solving approach and the introduction of contextual problems was emphasised, there was no clear direction or further elaboration of the teaching and learning processes that would encourage the integration of HOT into mathematics classrooms. An evaluation of this curriculum has indicated that there was limited integration of HOT skills in Indonesian mathematics classrooms (Dewi & Kusumah, 2014; Oktiningrum, Zulkardi, & Hartono, 2015). There was also little provision for educating teachers in this area; the teacher training has been considered for decades as ineffective (Bjork, 2013; Hadi, 2002). In some cases, where teachers received professional development training, they seemed to grasp the knowledge during the training yet they faced many issues in implementing it in their classrooms (Fauzan, 2002; Suhendra, 2015).

In May 2013 the Ministry of Education and Culture introduced another new national curriculum, addressing eight standards of national education: content, process, graduate attributes, educators, educational staff, facilities, management, finance and assessment. The external factors driving this change included the issues of globalisation and the poor performance of Indonesian students in international assessments. Standards for the teaching and learning processes were developed to encourage more innovative teaching, including: (a) moving from a teacher-centred classroom to student-centred learning; (b) moving from a single resource to multi-resources learning; (c) moving from a textbook approach to a scientific approach; (d) moving from content based learning to competence based learning; (e) moving from a single solution learning to multi-solutions learning; and (f) moving from verbal learning to applicable learning. All of these standards can be seen as an attempts to promote HOT skills into the classrooms. It also aimed to encourage lifelong learning, creativity in learning, open space learning (understanding that students may learn in many situations, from many different people), and the use of technology for increasing the effectiveness of learning, and at the same time addressed individual students' needs, background and culture. The standards of assessment in the new curriculum included: (a) assessing students' lower and higher order thinking; (b) emphasising the higher order questions requiring reasoning beyond memorising; (c) assessing processes; and (d) using portfolios of students' learning (Wardhani, Anggraena, & Marfuah, 2015).

The new curriculum stated clearly its direction, employing the scientific approach: observing, questioning, experimenting, associating and communicating in the teaching and learning process. It also aimed to equip students with a better understanding of mathematics concepts and good mathematics reasoning and communication skills for both routine and non-routine problems. The new curriculum aimed specifically for the improvement of mathematics and science teaching and learning in Indonesia. It emphasised HOT skills by recommending the integration of LOT and HOT questions for students throughout the teaching and learning process. It sought to enable students to apply mathematical concepts beyond the classroom. It has been posited forcefully by Bjork (2013), who has studied classroom delivery and teachers' culture in detail in Indonesia and internationally, that while the intentions and policies of the ministry, as well as the curriculum, are aligned with developed standards there is an inability to translate these goals into the schools and classrooms.

1.1.3. Assessment of Indonesian Students' Mathematics Performance

Indonesian Students' Results in International Testing

International studies of student mathematics literacy, along with science and language literacy, are conducted by two major organisations: the Organisation for Economic

Co-operation and Development (OECD) carries out PISA and the International Association for the Evaluation of Educational Achievement (IEA) carries out TIMSS. Although both studies measure students' mathematics they have a different focus, with PISA evaluating students' capability to apply their knowledge to real-life situations which they may meet in their everyday life after leaving school (OECD, 2012) and TIMSS aiming to assess students' mathematics and science achievement and gather information concerning the mathematics curriculum, teachers, schools and the context of mathematics learning across the countries participating (IEA, 2012).

Indonesia has been participating in both these studies since they first started (TIMSS in 1995 and PISA in 2000) and has been consistently ranked as a country with poor outcomes. Based on the results of the TIMSS studies for Year 8 students, in 1999, Indonesia was ranked 34^{th} (out of 38 countries); in 2003, 36^{th} (out of 46 countries); in 2007, 36^{th} (out of 46 countries) (Iryanti, 2010), in 2011, 38^{th} (out of 42 countries) (Mullis et al., 2012). Similar outcomes are also seen in PISA studies: in 2003, 37^{th} (out of 39 countries) (OECD, 2004); in 2006, 50^{th} (out of 57 countries) (OECD, 2007); in 2009, 27^{th} (out of 31 non OECD countries) (Fleischman et al., 2010); in 2012, 64^{th} (out of 65 countries) (OECD, 2013b); and in 2015, 62^{th} (out of 70 countries) (OECD, 2016b). Of concern, PISA studies from the period of 2000 to 2012 show that there had been no significant improvement in Indonesian students' mathematics performance (Firman, 2016). In the latest PISA, 2015, the results reported that Indonesian students' mathematics performance has slightly improved; however the Indonesian ranking remains among the bottom countries (OECD, 2016b).

The results of the PISA 2009, 2012 and 2015 reported that around 70% of the Indonesian students are low performers (below level 1) (OECD, 2016a). PISA 2003 and 2006 results showed that only a very small percentage of Indonesian students were able to successfully complete the more challenging questions which require HOT skills (level 4 and above). Only one per cent of students reached level 4 in 2003 and two per cent in 2006. No students reached level 5 or 6 (Barrera-Osorio, Garcia-Moreno, Patrinos, & Porta, 2011). In PISA 2009, 0.9 per cent of Indonesian students reached level 4 and 0.1 per cent reached level 5, with no students reaching level 6 (OECD, 2010; Stacey, 2014). The results of the latest PISA 2015 also showed similar trends (OECD, 2016a). The description of the mathematical competences of each level used in PISA will be outlined in Chapter 2.

Issues Contributing to the Indonesian Students' Mathematics Performance

The problem of Indonesian students' mathematics performance has not only been recorded in international studies such as PISA and TIMSS but has also been examined in several studies related to mathematics education in Indonesia. Jupri, Drijvers, and van den Heuvel-Panhuizen (2014b) found that Indonesian students were lacking in problem solving-skills, particularly in the development and integration of flexible operational and structural views of equations and algebraic expressions. Furthermore, they were having difficulty in bringing mathematical contexts to mathematical notions (Jupri, Drijvers, & van den Heuvel-Panhuizen, 2014a). Students did mathematical problems without understanding the reasoning involved or identifying the application of the mathematics topics to real life.

Researchers have focused on the poor performance of mathematics students in the PISA and TIMSS studies is a consequence of the lack of HOT skills of the students (Dewi & Kusumah, 2014; Oktiningrum et al., 2015), with students being unfamiliar with the types of questions in PISA and TIMSS testing (Abdul, Muhammad, Syahrullah, & Ikhbariaty, 2014). Furthermore, a secondary analysis of PISA 2012 conducted by Razak and Shafaei (2016) reported that half of Indonesian students used memorisation strategies in studying mathematics. The lack of problem-solving skills being taught in the classroom (Jupri et al., 2014a), the lack of contextual mathematics learning and the lack of training students to be able to present their argumentation and thinking processes (Abdul et al., 2014; Cahyono, 2010) are seen as important factors.

While there are various factors influencing the poor learning outcomes of mathematics teaching in Indonesia, the importance of teaching methods is consistently related to a student's performance (Suhendra, 2015). Even though the Indonesian mathematics curriculum specifies the inclusion of problem solving and HOT, researchers (Fauzan, 2002; Hadi, 2002; Wijaya, 2015), in noting HOT's absence in the classroom, found that teachers are still focused on the LOT skills and emphasising rote learning and the repetition of similar basic problems. Students are not given the opportunity to relate the problem to practical daily life situations or to find solutions using their own methods. Furthermore, continued reliance on a teacher-centred approach has contributed to students' attitudes towards mathematics learning, with them relying entirely on directions from the teachers, without exhibiting critical and creative thinking (Sembiring, Hadi, & Dolk, 2008). The teacher-centred instruction also requires students to reproduce the tasks given by the teacher and there is little encouragement to ask further questions related to the topic (Y. B. Widjaja & Heck, 2003; Zakaria, Solfitri, Daud, & Abidin, 2013). It has also been indicated that Indonesian teachers are generally less confident concerning their ability to integrate and promote HOT skills in the classroom (OECD, 2015).

While Indonesian teachers are criticised for their heavy reliance on rote learning, half of their students do not have to capacity to answer LOT related questions in PISA and TIMSS. At the end of Year 9, all Indonesian students sit for a national examination where over 95% of the students receive a pass. This indicates a discrepancy between the results of the international testing and the national examination. The high pass rate in the national examination may be due to the fact that teachers are teaching students directly for the test (Hendayana, Asep, & Imansyah, 2010). This bring some disadvantages to the learning as teachers tend to focus on curriculum content rather than developing students' beliefs, attitudes and values. They found that covering the breadth of the curriculum often meant that depth of understanding was sacrificed (Fauzan, 2002). The discrepancy in terms of results between the international and national assessments provokes the need for a deeper understanding of what is happening in mathematics education in Indonesia today.

1.1.4. The Case of Aceh

Aceh, at the extreme north-west end of the Indonesian archipelago, was chosen as the focus of this study. International testing, from the researcher's personal knowledge and inquiry, has not been carried out in this province. Aceh is separated by distance and culture from the nation capital and the most populous areas of Indonesia. It has long had a difficult relationship to the central government of Indonesia and for decades, before a peace settlement in 2005, it had been fighting for its independence. Unfortunately this struggle and other issues have kept the province as one of the poorest in the rapidly developing nation. This situation was further exacerbated by the devastating tsunami and earthquake of 2004. As a result of international assistance after the tsunami and the 'special autonomy' granted by the Indonesian government, the province has had huge financial support. As a consequence, a high amount of the Aceh province budget, 30% compared with the national 20% (World Bank, 2006), is being spent on education. The province has had to rebuild a huge number of school and suffered the loss of many teachers in the 2005 tsunami (World Bank, 2008). There is a large discrepancy between the money spent per pupil on education in urban areas and rural areas, and rural areas lack infrastructure that supports effective education. In more remote parts of Aceh with poor infrastructure and resources, teacher qualification and professionalism combined with the students' low socio-economic position are likely contributors to student outcomes.

In Aceh, enrolment levels are higher than the overall figure for Indonesia, across all income levels and types of education. While there are sufficient numbers of teachers, there is also a high level of absenteeism, especially in the rural areas. Improved infrastructure and public services may help alleviate this. The majority of teachers throughout Aceh, however, are not fully qualified (World Bank, 2008). In the national examination, Aceh's performance is not significantly different to the rest of Indonesia, though it would appear that a larger number of Acehnese are failing to complete junior high school (Year 9) (World Bank, 2008).

A small number of studies carried out in Aceh have pointed to the same problems identified throughout Indonesia: teachers are having difficulty incorporating contextual problems (Arsaythamby & Zubainur, 2014); students perceive mathematics as mainly calculation using mathematics operations (Johar & Afrina, 2011); and students are having difficulty engaging with the possibility of creating their own solutions for mathematics problems (Taufik, 2014). Some teachers seem to not fully understand the student-centred learning approach, according to Syah, Fitri, Yani, Qurnati, and Idris (2011) who reported that some mathematics teachers have misconceptions about active learning, not realising it involves students working in small groups.

1.2. Statement of the Research Problem

It would appear that many students throughout the world struggle with mathematics problems that require HOT. But in Indonesia, students are also struggling with LOT (Firman, 2016; OECD, 2013a). Despite the acknowledged importance of equipping students with mathematical skills, the development of both LOT and HOT in the mathematics classroom in Indonesia, including Aceh, continues to be problematic (as described in the previous section). This demands an in-depth analysis and an attempt to discover the various factors that may contribute to the situation.

There has been some research touching on the issues of mathematics teaching in Aceh (Arsaythamby & Zubainur, 2014; Johar & Afrina, 2011; Syah et al., 2011; Taufik, 2014 are studies conducted in Aceh identified in this study). However no study, to the researcher's knowledge, has been empirically based. In order to assess the state of mathematics teaching and learning and student outcomes in Aceh, systematic data collection is necessary. Rather than focusing on one aspect of the education system, and to incorporate an understanding of Aceh's special conditions, the study needs to be multilevel. Are the Acehnese students' mathematics performance in relation to LOT and HOT any different from the overall outcomes presented in previous studies for Indonesia as a whole? What multilevel factors are involved in their performance? How might these analyses help with understanding educational effectiveness? How might students' performance be improved?

There is a wide range of international studies related to HOT skills and the mathematics performance of students (Kolovou, Van den Heuvel-Panhuizen, & Bakker, 2009; Pegg, 2010; Tassell et al., 2012). However, there are limited studies investigating the relationship of mathematics performance related to both LOT and HOT, from both the teachers' and students' perspectives. Most studies focus on either

LOT or HOT, and many are related to teacher effectiveness rather than student outcome (Rooney, 2012; T. Thompson, 2008; Wenglinsky, 2002; Zohar, Degani, & Vaaknin, 2001). Therefore, it is necessary to conduct a multilevel study that caters to three levels of data (student-, teacher- and school-level).

From a methodological perspective, the use of quantitative methods in this study, including a multilevel analysis, will develop a more meaningful model of factors influencing students' mathematics performance related to LOT and HOT. Also, multilevel analyses are at the basis of the growing body of education effectiveness research which seeks to investigate the differing levels (from the institutional, to the local school, to the teachers' classrooms practices to the students) that can explain the variations in students' performance. The hierarchical structure of education that impacts on students requires modelling techniques to overcome the weakness of earlier attempts at analysis that saw just one level being analysed. More complex relationships within and between the various levels and their factors allows for an empirical evaluation of educational effectiveness (Creemers, Kyriakides, & Sammons, 2010).

The limited research in the context of Indonesia (Rakhmani & Siregar, 2016), especially in the area of students' mathematics performance related to LOT and HOT and the fact that there is no reported research in the area of students' mathematics performance for a multilevel analysis of student-, teacher- and school-level factors in the context of Aceh, are the central reasons for conducting this research. In such a large system as exists in Indonesia, it is important to assess the outcome of education within regions. A traditional style of teaching which emphasises memorisation and repetition is not adequate to develop the skills of HOT. It has even been seen to be detrimental to the teaching of LOT (Pogrow, 2005; Soar & Soar, 1976; Zohar & Dori, 2003).

A multilevel analysis, along with the single-level analysis, is expected to provide a more in-depth examination of the effectiveness of mathematics education and factors associated with students' mathematics performance related to LOT and HOT in Aceh. The following section provides the research questions for this study.

1.3. Research Questions

This study aims to examine the student-, teacher- and school-level factors, their interrelationships and impacts on Year 9 students' mathematics performance related to LOT and HOT skills in Aceh, Indonesia, using single and multilevel analyses. The student factors include students' background, students' attitudes and beliefs, and their perception of classroom practices; the teacher factors include teachers' background, beliefs and classroom practices; and school factors include school demographics and resources.

The study addresses the following specific research questions:

1. To what extent are Year 9 Acehnese students able to solve mathematics problems related to LOT and HOT?
2. What are the student-level factors influencing students' mathematics performance related to LOT and HOT?
3. What are the interrelationships among student-level factors influencing students' mathematics performance related to LOT and HOT?
4. What are the teacher-level factors influencing instructional activities in the mathematics classroom?
5. What are the interrelationships among the teacher-level factors influencing instructional activities in the mathematics classroom?

6. Taking into account the hierarchical nature of the data, do the student-, teacher- and school-level factors have direct effects on students' mathematics performance related to LOT and HOT?
7. Are there any cross-level interactions between the student-, teacher- and school-level factors in influencing students' mathematics performance related to LOT and HOT?

1.4. Structure of the Thesis

This thesis is organised into ten chapters. Chapter one provides the background and an overview of the study by discussing the importance of mathematics and the development of cognitive skills, the context of Indonesia and Aceh, including a brief summary of the Indonesian mathematics curriculum and the state of students' performance in Indonesia, followed by the statement of the problem and research questions.

Chapter two provides a review of previous theory and research that provides the basis for this study. It briefly reviews the childhood development theories at the base of the taxonomies developed to assess student learning in the areas of LOT and HOT. The place of LOT and HOT skills in the mathematics classroom is discussed along with issues of assessment. Educational effectiveness theory and research is then reviewed, emphasising the multilevel nature of the education process. Multilevel factors influencing students' mathematics performance are then identified from various studies. The conceptual framework designed for this study is then presented.

Chapter three presents the methods of investigation used in the study, including the ethics approval. The sample techniques and the procedures for data collection are described. The operationalisation, the measurement, and the development of the instruments, including the mathematics test, are then presented. The pilot study is described. Some general methodological considerations are presented and the statistical procedures employed in the study are also discussed. The quantitative techniques used in the study are outlined, including the path analysis for the single-level analyses and the hierarchical linear modelling for the multilevel analysis. The role of confirmatory factor analysis (CFA) and Rasch analysis in the study in assessing the instrument validity is explained.

Chapter four and five present the validation of instruments used in the student questionnaire (Chapter 4) and teacher questionnaire (Chapter 5). The validity and reliability of the instruments are established by employing CFA and Rasch analysis. CFA is carried out to examine the factor structure of the instrument and Rasch analysis is used to confirm the structure that has been previously tested in CFA. The CFA is conducted using IBM SPSS Amos 22 and the Rasch analysis is conducted using ACER Conquest 2.0 software (Wu, Adams, Wilson, & Haldane, 2007). The final structures of the instruments to be used for the subsequent analyses are then made.

Chapter six presents a comprehensive descriptive analysis along with contextual information based on the student, teacher and school questionnaires. The descriptive analyses of students' mathematics performance related to LOT and HOT are also presented. Independent t-tests are conducted for the scales involved in the student and teacher questionnaires and the mathematics performance related to LOT and HOT in order to examine whether there are any gender or school location differences. Significant results are then briefly reported in the chapter to be further discussed in the final chapter. The details of these results are presented in Appendix I. The descriptive results in this chapter were generated using the IBM SPSS 22.

Chapter seven and eight discuss the single-level model of the student-level (Chapter 7) and teacher-level (Chapter 8) models of partial least square path analysis (PLS-PA). The student-level model examines the relationships between the student-level factors in influencing students' mathematics performance related to LOT and HOT. The direct relationships between student-level factors and students' mathematics performance and the interrelationships between student-level factors are illustrated. The teacher-level model examines the interrelationships between teacher-level factors. The path analysis for both student- and teacher-level models are carried out using SmartPLS 3.2.6 (Ringle, Wende, & Will, 2015).

Chapter nine presents the two three-level (student-, teacher- and school-level) HLM models, namely the three-level HLM model of mathematics performance related to LOT and three-level HLM model of mathematics performance related to HOT. The HLM models examine the relationships between variables at each level and the outcome as well as the interaction effects between variables across levels using HLM 6 (Raudenbush, Bryk, & Congdon, 2004). The building and the findings of the HLM models are presented in this chapter.

Chapter ten summarises and discusses the findings from the analyses conducted in this study. Findings are discussed in relation to the research questions and in comparison to the results reported in the previous studies, as well as taking into account the context of this study. A discussion of the limitations and implications of the study as well as some recommendations for future research is also presented in this chapter.

1.5. Summary

The current research is designed to examine the interrelationships between student-, teacher- and school-level factors influencing Year 9 students' mathematics performance related to LOT and HOT in Aceh, Indonesia. This chapter presents the background to the study, the context of the research, its setting in the Indonesian education system. The statement of the problem and the research questions are also provided. The following chapter provides a review of selected research in areas relevant to this thesis.

Chapter 2 Review of Research Studies

2.1 Introduction

This chapter begins with an overview of development of the concepts of LOT and HOT, before looking specifically at their application to mathematics and mathematics teaching methods promoting LOT and HOT, including their assessment. The theoretical basis of this study comes from the research on childhood cognitive development and educational theory. Childhood development theories (particularly Piaget and Bruner) and the association found between physical and cognitive growth resulted in educational theorists attempting to assess the various stages of cognitive learning. The development of taxonomies (particularly Bloom) followed to assist structure the curricula and measure the level of skills attained by students. These skills were roughly divided into the basic 'lower' skills and the more sophisticated 'higher' skills and attempts to understand the best methods of teaching and learning these skills continue to challenge educationalists. From educational effectiveness theory has come the theoretical perspective for this study, focussing on a quantitative assessment of student performance see as being connected to a large set of variables at all levels of the education system. The conceptual model developed from this theory provides the basis of a multilevel analysis of the factors related to student performance related to LOT and HOT.

2.2 Lower Order Thinking (LOT) and Higher Order Thinking (HOT)

2.2.1. Childhood Development and Learning

Foundational Theories and Earlier Development of Cognitive Skills

The work of Piaget and Bruner, two major developmental theorists, have influenced childhood psychology and education throughout the latter half of the 20th century, with educational thought and practice relying heavily on the pioneering work of Piaget (Biggs & Moore, 1993; Seddon, 1978). In his theory of cognitive development Piaget identified four stages: the sensorimotor stage (0 – 18 months old); the symbolic stage (18 months – 7 or 8 years old); the concrete-operational stage (7 or 8 – 12 years old); and the stage of formal operations (12 – 15 years old). Piaget claimed that each successive stage was built upon the earlier stage (Bigge & Shermis, 1992). Piaget explained that during the sensorimotor period children only understood direct action because they had a limited ability to recognise symbols. A child starts to recognise symbols and differentiate things at the symbolic stage, when the child begins to think and classify things as well as combine and classify concepts. The symbolic stage and the concrete-operational stage, were seen as prerequisites for the period of formal operations, when adolescents developed the capacity to evaluate their thoughts and create ideas about the future use of their reasoning skills. Piaget's emphasis was on the development of more complex thinking resting upon the acquisition of the earlier stage.

Bruner (1977), similarly identified cognitive growth according to the ability to use differing modes of representation. The enactive mode was the initial mode, which involved the 'knowing' skill of a child with the absence of images. The iconic middle stage mode developed the knowledge with the presence of images enabling the understanding of concepts. The symbolic mode was the last mode of cognitive growth and was indicated by the ability of adolescents to understand abstract concepts with the presence of symbols and using language as the medium of thought (Bigge & Shermis, 1992). Bruner's modes, unlike Piaget's stages, were not simply linked to age as environment also played a significant role in either slowing down or accelerating the processes of cognitive growth (Bigge & Shermis, 1992). Neither Piaget nor Bruner explicitly discussed LOT and HOT skills in their cognitive developmental theories.

Vygotsky's constructivist theory, which became influential in the late 20th century, was based on the notion that the environment was a crucial factor in children's development, expanding upon Piaget's and Bruner's theories, which were focused on cognitive development as an inherently individual process. Vygotsky emphasised that teachers and peers had a great influence on students' learning (Jaworski, 1994; Leonardo & Manning, 2017). In his theory of learning Vygotsky introduced the idea of 'the zone of proximal development' which pointed to the learning a student could do when assisted by the teacher or classroom peers. The activity of assisting students' development was also known as 'scaffolding' (Fernández, Wegerif, Mercer, & Rojas-Drummond, 2002). Vygotsky's theory of development justified the use of collaborative learning (Bonk & Cunningham, 1998; Nyikos & Hashimoto, 1997). While Vygotsky's constructivism was also not specifically focussed on the development of LOT and HOT skills, it highlighted the potential of peer-group learning.

Educational Taxonomies

In 1956 Bloom's taxonomy was published and, with various modifications, it is still influential today. This taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) is an explicit classification of the goals and objectives of education and is applicable to all subject areas. It provides a basis for assessment of students' performance in relation to the stated educational goals and is thus used to compare student performance across states, nationally and internationally. It became the basis for the division of learning into lower order or basic thinking and higher order thinking.

Bloom's taxonomy consisted of six hierarchical levels within the cognitive domain of students' learning: 'knowledge' (the ability to recall facts); 'comprehension' (the ability to understand, manipulate and interpret facts); 'application' (the ability to use understanding to solve problems); 'analysis' (the ability to recognise patterns within the facts); 'synthesis' (the ability to go beyond the facts to produce a new idea); and 'evaluation' (the ability to judge the quality of a solution or theory) (Seddon, 1978). This structure indicated an ascension of processing, from simplicity to complexity, and conceptualising, from concrete to abstract, with the lower categories being important for the development of the upper categories (Krathwohl, 2002). Bloom's taxonomy was subsequently revised by Anderson et al. (2001), members of the original team who developed the taxonomy. In this revision, emphasising the active processes of thinking, the levels are converted to verbs and the end goals are interchanged with the ultimate goal being 'creating' (rather than 'synthesis') (Krathwohl, 2002).

A further taxonomy, Structure of Observed Learning Outcomes, or SOLO, was created by Biggs and Collis (1982) and was based on Piaget's stages of cognitive development, and the hierarchical structuring of the levels on 'increasing cleverness' (Biggs & Moore, 1993). In this, learning is seen as a cycle with five levels, known as 'prestructural', 'unistructural', 'multistructural', 'relational' and 'extended abstract', and three modes, known as 'previous', 'target' and 'next'. SOLO taxonomy classifies the depth of thinking demonstrated by learners; the first three categories correspond to surface thinking and the last two being are associated with deep thinking (Hunt, Walton, Martin, Haigh, & Irving, 2015; Murphy, 2017). The main aim of the SOLO taxonomy was to classify students' outcomes or responses (Hunt et al., 2015; Murphy, 2017), differentiating it from Bloom's taxonomy which was mainly used in order to classify test items.

Current Developments

Since the foundational work of Bloom and others, the concepts of LOT and HOT have been continuously explored. Finding adequate definitions and understanding the relationship between the two orders has challenged educational theorists. However, the emphasis on learning has increasingly moved from remembering and repeating information to thinking, reasoning and problem-solving, with critical thinking are at the base of even the elementary levels of knowledge. Education practice has shifted to encouraging all learners at all levels to thinking and understanding.

With the current trend in education worldwide, the shift is from LOT to HOT and the definition of HOT has developed into a broader context. LOT is defined as reproductive thinking involving the application of routine and algorithmic procedures (Lewis & Smith, 1993; Miri et al., 2007; Newmann, 1990). It should however be recognised that learning and teaching LOT (see later) do not have to be routine and unimaginative; learning involves a thinking and questioning that can also be extended to learning LOT.

Newmann (1988) defined HOT as a thinking process involving challenges which required interpretation and analysis being applied to a problem which could not be answered using a simple routine procedure. This definition was later expanded by contrasting it with LOT which mainly involved routine problems with mechanistic or repetitive procedures (Newmann, 1990). It has been noted that students' abilities to demonstrate such skills were influenced by the type of tasks given and the student's history in performance (Lewis & Smith, 1993). Later, Miri et al. (2007) conceptualised HOT as a non-algorithmic sophisticated mode of thinking associated with the capability to produce multiple solutions. Recently, Alexander et al. (2011) proposed that it is:

> the mental engagement with ideas, objects, and situations in an analogical, elaborative, inductive, deductive, and otherwise transformational manner that is indicative of an orientation toward knowing as a complex, effortful, generative, evidence-seeking, and reflective enterprise (p.53).

Brookhart (2010) classified HOT in terms of its association with the ability to transfer knowledge, with the ability to think critically and to solve problems. The ability to transfer knowledge involves conveying (or translating) learned skills in one area to new situations; the critical thinking component involves the ability to reason, reflect and make decisions; the problem solving component involves the ability to encounter and resolve a new problem which then resulted in the creation of a new solution.

It can be seen from the pioneering work of Piaget and Bruner to the later taxonomies that there has been a shift of developing towards a clearer differentiation between LOT and HOT. Current international trends in education are focussing on HOT skills. However, there is still a great need, especially in less developed countries, for LOT skills to be developed. LOT and HOT skills in this study will be described generally using Bloom’s taxonomy: LOT skills involve remembering, understanding and applying and HOT skills involve analysing, evaluating and creating skills (Anderson & Sosniak, 1994; Pegg, 2010).

The terms ‘lower order’ and ‘higher order’ can be interpreted as one being ‘superior’ to the other. It has been pointed out by Zohar and Dori (2003) that the pyramid model of Bloom’s presents pictorially an emphasis on LOT and suggests a hierarchy of educational goals. However, they also emphasise that the specifying of cognitive levels that are distinguished from one another is still useful. Both LOT and HOT skills are likely to be interwoven (Lewis & Smith, 1993) and both need addressing in the classroom (Sangwin, 2017), including the mathematics classroom. Earlier studies (Cardelle-Elawar, 1992; Zohar & Dori, 2003) questioned whether HOT could be taught to low-achieving students but more recent analysis points to their interdependence (Pogrow, 2005). It is now recognised that thinking, rather than memorising, is a part of all students learning at all levels of education and that teachers, through their practices in the classroom (rather than their subject content) can foster both these skills.

The attempt to encourage more HOT in the curriculum, as is evidence in the last decades, can result in the neglect of LOT skills. Yet, LOT is essential in mathematics education. The foundational skills in mathematics enable HOT because once the basic skills are mastered students are free to become involved in problem solving (Tikhonova & Kudinova, 2015). The basic skills cannot be neglected. Mathematics students who do not have the basic skills are confused by the simplest problems; students who perform well in mathematics will have all the basic skills.

Staples and Truxaw (2010) argue that HOT skills can be seen as the ability of students to think mathematically in their reasoning when solving problems. A review of HOT skills in three countries by Fullan and Watson (2011), found that these skills in mathematics emerged in problem solving skills and the ability to communicate the solution mathematically. Being able to reason mathematically and solve problems (utilising, analysing, evaluating skills) allow HOT mathematical problems to be understood. Thus, while a student's performance can still be assessed in terms of LOT and HOT, a teacher's practice needs to understand and apply both in the classroom (Lewis & Smith, 1993; Zohar et al., 2001; Zohar & Dori, 2003).

Research on LOT and HOT is summarised below.

The characteristics of LOT involve the use of:

a) Routine problems (King et al., 1998; Newmann, 1988; T. Thompson, 2011)
b) Memorisation or rote learning (Miri et al., 2007; Newmann, 1990)
c) Reproduction of previous repetitions (Brookhart, 2010; Lewis & Smith, 1993)

The characteristics of HOT involve the use of:

(a) Non-routine problems (Kolovou et al., 2009; Mullis et al., 2003; Newmann, 1988, 1990; Schoenfeld, 1992);
(b) Problem solving (Brookhart, 2010; Fullan & Watson, 2011; Sheffield, 2007; Silva, 2008; Trilling & Fadel, 2009);
(c) Reasoning mathematically (Bigge & Shermis, 1992; Fullan & Watson, 2011; Staples & Truxaw, 2010); and
(d) Open-ended questions (Bobis, Anderson, Martin, & Way, 2011; Puchner & Taylor, 2006; Staples & Truxaw, 2010; Watson, Collis, Callingham, & Moritz, 1995).

2.2.2. Teaching Methodology Promoting Lower Order Thinking (LOT) and Higher Order Thinking (HOT) in Mathematics

The introduction of universal education in the West saw the emphasis being placed on the basic skills of reading, writing and arithmetic. Teachers often had little education themselves; classes were large with few resources. These conditions are often still prevalent in developing countries. Questions concerning pedagogy have dominated educational research and practice over the last 60 years in an attempt to equip students with the appropriate skills. How a teacher teaches mathematics is integral to a student's understanding of mathematical concepts. Even basic skills may be taught in such a way that the student's thinking and not just routine memorising are engaged. Some simple examples can indicate the difference: a) Using values between 100 and 2000, what would be 2 of the hardest subtraction sums you can think of? Say why you think they are hard; b) A room is $48m^2$, draw 4 different shapes it could be. These examples are developing LOT skills in a HOT way.

Incorporating HOT in the classroom and integrating it with subject content and assessing the skills are challenging tasks for teachers (Forster, 2004; Schulz & FitzPatrick, 2016). In a video study of mathematics lessons of Year 8 students in the US and Australia, researchers found that approximately 79% of all questions in the US classes were related to LOT, while in the Australia it was 73% (Shahrill & Mundia, 2014). However, there are some teaching methods which are believed to be conducive to encouraging these skills. It has been suggested that student-centred instruction plays an important role in these (King et al., 1998; Savery, 2006). A student-centred approach provided students with more time to develop HOT skills, as well as educating them to be independent thinkers. King et al. (1998) highlighted the importance of instructional communications, scaffolding, learning and thinking strategies, direct instruction, questioning strategies, feedback, team activities and computer mediation for facilitating these skills in students.

Problem-based learning, where teachers acts as facilitators who both encourage students to develop their HOT and give information regarding the problems, as well as inquiry-based learning, where students are expected to develop their thinking without any information provided for the problems, are other aspects of student-centred teaching (T. Thompson, 2011; Weiss, 2003). A collaborative learning approach, derived from Vygotsky's research (Vygotsky, 1978), also promotes HOT skills through peer collaboration that elicits discussion and learning engagement

(Cicconi, 2014). Furthermore, classroom practices involving a learning strategy that asked students to connect their learning experiences, including asking students to relate what they have learnt to other subjects they studied, are seen to have a positive effect on students' problem solving skills, in particular in Indonesia and Malaysia (Razak & Shafaei, 2016).

Wilks (2005) found that teachers could cultivate students' HOT by acknowledging that it is possible for every student to be able to examine abstract ideas and develop their thinking, by providing suitable resources to trigger inquiry and investigation, as well as establishing a classroom climate which enabled students to explore the content of the subject. Teachers need to take students' responses into account in order to extend their skills. The classroom atmosphere established by the teacher also influences students' participation, engagement and self-confidence in dealing with problems related to HOT.

The types of tasks given in the classroom is also significant. The use of open-ended questions for mathematical tasks in the classroom is one of the approaches which could be utilised for promoting these skills (Bobis et al., 2011; Puchner & Taylor, 2006; Staples & Truxaw, 2010; Watson et al., 1995). Open-ended questions are defined as "those that require a student to think more deeply and to give a response that involves more than recalling a fact or reproducing a skill" (Sullivan & Lilburn, 2002, p.1) and students need to be given enough time to think about and create appropriate answers (Hoskins, 2005). Teachers also need to be able to stimulate students to investigate problems. One approach suggested for developing students' HOT was the adoption of philosophical inquiry which was defined as 'thinking about thinking' (metacognition). This approach allows students to explore and present their ideas in an atmosphere that encouraged the appreciation of others' ideas and opinions (Abbott & Wilks, 2005).

Promoting HOT in the mathematics classroom does not displace the need for LOT skills, such as teaching algorithms and formulae. Once the significance of HOT is acknowledged, the goal is to create classroom practices which allow students to experience and develop both LOT and HOT skills (Hoskins, 2005).

2.2.3. Assessing Lower Order Thinking (LOT) and Higher Order Thinking (HOT) Skills in Mathematics

T. Thompson (2008) suggested that mathematics teaching has changed little in the last two decades in the US, still relying largely on routine tasks through mechanical techniques. He found that while teachers were able to classify the tasks involved in Bloom’s taxonomy, many were unable to write test questions related specifically to HOT. While integrating HOT in the classroom is challenging, assessing students’ HOT is even more challenging where still mainly focused on LOT tasks (Hoskins, 2005).

Caygill and Eley (2001) created categories to specifically examine mathematical and scientific processes and skills on the basis of Bloom’s taxonomy: (a) recalling knowledge (general knowledge, facts and basic information); (b) calculating or following formulae (ability to use certain algorithms or formulae to answer); (c) experimenting or investigating (using a range of mathematical procedures in solving the problems); (d) comparing or contrasting (using various methods and ideas and then comparing results before deciding on the best solution); and (e) concluding or explaining or justifying (synthesising ideas, presenting rigorous reasoning or conclusions in order to solve the problems). In assessing HOT, the tasks developed need to emphasise the last two categories. This is consistent with the levels in the SOLO taxonomy which could also be used to determine the tasks required to assess HOT in the mathematics classroom. The last two categories of the SOLO taxonomy, namely relational level and extended abstract level, might also be used as the base to create assessment tasks for HOT. The relational level requires students to integrate various related information to produce a better understanding and to solve the problems, while the extended abstract level requires students to make a generalisation based on the given information to create a more abstract dimension (Biggs & Moore, 1993).

In Japan, a high ranking achiever in PISA and TIMMS, a system called ‘lesson study’ has been devised, an approach to improve classroom teaching in Japan. The tasks are designed following particular principles: (a) they are appropriate and mathematically valuable in terms of the aims of the lesson; (b) they are engaging for students; (c) they are at the appropriate level of difficulty; (d) they have several possible methods of solution; (e) they have implications for other mathematical problems or real life problems; and (f) they can elicit valuable basic wisdom (Fujii, 2015). Students’ answers and responses to the tasks are part of the evidence of what students learn and understand about the concepts which in turn reshapes teaching strategies.

Caygill and Eley (2001) suggested a teacher-student interview as that an effective and appropriate way to assess HOT. This method not only allowed students to provide sophisticated explanations and present various responses but it also allows teachers to examine both the processes and the products of students’ thinking. In a similar vein, Hoskins (2005) suggested the use of a student journal.

In PISA 2012, the detailed analysis of which in that particular year was focused on mathematics, a definition of mathematical literacy was used in the measurement of students’ performance across six levels, the first three being based on LOT and the next three on HOT:

> mathematical literacy is an individual’s capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning

> mathematically and using mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens (OECD, 2013a, p. 25).

PISA classifies students' mathematics competence into six level of mathematics proficiency, each level can be summarised as: (a) level 1, students are able to solve routine problems involving familiar context using routine procedures and the direct information provided; (b) level 2, students are able to solve problem requiring direct inference by employing basic algorithms, formulae, procedures, or conventions, as well as being able to make a basic interpretation of the solution; (c) level 3, students are able to apply simple problem solving strategies using their own interpretation and representations of the given information, as well as being able to show their interpretation and reasoning skills; (d) level 4, students are able to solve complex problems requiring simple decision-making by integrating different representations and making a connection to real-life situations, as well as being able to explain their argument using their own interpretations, reasons and actions; (e) level 5, students are able to solve more complex problems by utilising their well-develop thinking and reasoning skills to select, compare, and evaluate the appropriate strategies to be used, as well as being able to reflect on their own work, formulate and explain their own interpretation and reasoning; (f) level 6, students are able to solve more complex non-routine problems by using their own investigation and capacity of advance mathematical thinking and reasoning, applying their understanding to develop new approaches and strategies for non-familiar situations, as well as being able to reflect and communicate their findings, interpretations and reasons in relation to the appropriateness of these to the original situation (OECD, 2013c).

The description of the six levels mathematics proficiency established in PISA shows the hierarchy of competences: from routine to non-routine; from common to unfamiliar context; from direct instruction to own investigation; from relying on given information to creating new approaches and solutions; from the absence of reasoning to the demonstration of advance reasoning skills in communicating mathematical ideas and findings. Even though PISA does not literally classify the level of the mathematical proficiency into two streams of LOT and HOT, the first three levels generally reflect LOT and the next three, HOT. When referring to international comparison of students' performance, this research uses the assessment levels devised by PISA.

2.3 Educational Effectiveness Theory and the Dynamic Model

The field of educational effectiveness theory (EET), research (EER) and models, now identified with the work of Creemers and Kyriakides, attempts to build a quantitatively based model of school effectiveness with the goal of improving student outcomes. It has been extensively researched and reviewed over the last decades, undergoing shifts of methodology and theory, sometimes as a result of software development.

EER was built upon earlier studies by Coleman (1968) using an equality of educational opportunities survey that was initially created to examine the magnitude and the factors contributing to the inequality of educational opportunities between races in the US. The variables used in this study found that students' background are the main factors influencing students' performance with the minimum effects of the school, suggesting that the school impact was minimal on students' outcome. Later, Jencks et al. (1972) conducted a study, based on Coleman's study, reassessing the effect of family and schooling on inequality in the US. Hanushek (1986) also argued that school factors, seen here as teacher-student ratios, teacher education or teacher experience, had little or no positive influence on student performance.

These findings initiated extensive debate. Walberg (1984) argued that factors related to students (ability, development and motivation), to teaching (quality and quantity) and a student's environment (home background, peers, classroom environment and media) all had an impact on each other and the students' performance. Other researchers followed (Bidwell & Kasarda, 1980; Bosker & Dekkers, 1994; McCormack-Larkin, 1985; McCormack-Larkin & Kritek, 1982; Taylor, 1990), disputing the findings of Coleman (1968) and Jencks et al. (1972).

As the research concerning educational effectiveness evolved, several models have been developed. This brought together factors related to students' performance, establishing the importance of the school's impact on students' academic development and seeing the need for all schools to review their practices (Reynolds, Teddlie, Creemers, Scheerens, & Townsend, 2000). Researchers, using large scale data, began to focus on causal effects (Creemers & Reezigt, 1999). Furthermore, the emphasis on the multilevel nature of the educational process consolidated to involve students, classroom (teacher), school, district and national levels (Teddlie, Reynolds, & Pol, 2000).

Creemers and Reezigt (1996) proposed the significance of the school level where the student learning outcomes are influenced by the school factors (including the quality of education, time and opportunity); classroom factors (including quality of instruction and opportunity to learn); and student factors (including time on task and opportunities used, motivation as well as aptitude and background). This proposition of the importance of the school level was incorporated into the basic model of educational effectiveness designed by Creemers (1994). In this model, classroom instruction was the emphasis.

The theory, research and model building of EER has further developed, student outcome being the central concern:

> "the main research question of EER is which 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, Socio Economic Status and prior attainment" (Creemers & Kyriakides, 2008, p. 12).

Creemers and Kyriakides (2008) have now developed a more complex model called a dynamic model of educational effectiveness, explaining the interrelations of the factors within and between levels (above school, school, teacher and student) and the impact they have on students' outcome. The dynamic model, with the multiple factors operating at different levels in contributing to students learning outcome, is seen as an integrated approach to educational effectiveness modelling. Its theoretical basis understands that: a) the influences on student performance is multilevel; b) the factors in the different levels are interrelated; c) the school and context level factors are defined and measured in a different way to the classroom level factors; d) the relationships between some of the factors in the model may not be linear; e) the necessity of carefully investigating the relationships between factors at the same level; and f) the model employs five dimensions in measuring each factor including: frequency, focus, stage, quality, and differentiation (Creemers & Kyriakides, 2010).

This theory and model of educational effectiveness is not without its problems (Sandoval-Hernandez, 2008). While it may be useful in large scale comparative studies, its framework may not be able to provide assistance to individual schools to facilitate students' outcomes. Schools are organisations that are involved in complex links with their education systems and broader socio-cultural context. The model's effectiveness depends on the variables and their operationalisation selected for study. These concepts (such as professionalism, social capital, self-efficacy and motivation) require a theoretical understanding from various disciplines and often need qualitative data to legitimise their status. Sandoval-Hernandez (2008, p. 38) clarifies that the data collected for these studies is not objective as "the decisions on what and how information (is) collected necessarily imply personal choices and the output of the analyses always requires some interpretation". However, at this point in time, the dynamic educational effectiveness model is being applied in large-scale testing concerning students' outcomes. It should be noted that this model registers cognitive development only.

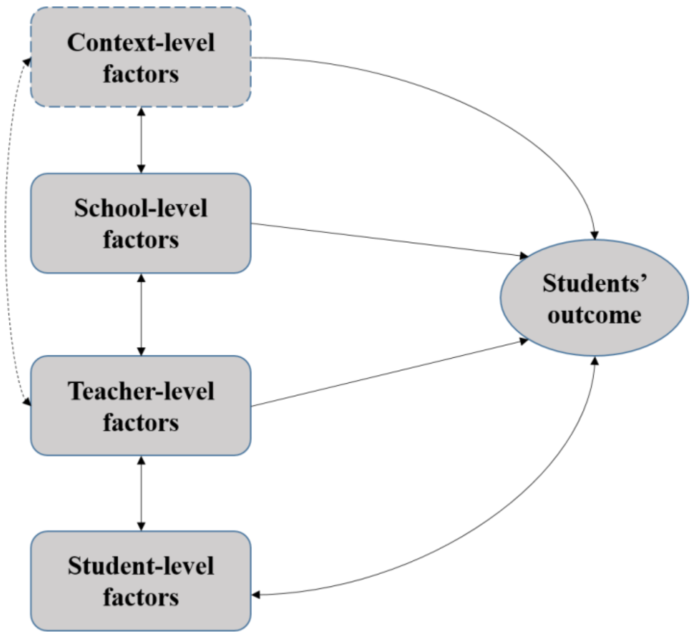


Figure 2.1 Simplified diagram of educational effectiveness

The particular value of quantitative methods, especially when analysing large samples, is that their results may be used to further develop theory which may generate further qualitative research. The dynamic effective model of educational effectiveness by

Creemers and Kyriakides (2008) is simplified and adapted to this study. The simplified diagram is presented in Figure 2.1. The context level is not studied empirically in this study. It is noted that new directional arrows are placed here as teachers, schools and their institutional context are able to interact and influence each other. The review of factors included at the school, teacher and student levels are presented in the next section.

2.4 Factors Affecting Students' Mathematics Performance

Students' achievement is an important dimension of both teaching and learning. Studies have tried to discover ways to raise students' performance through identifying the factors influencing their performance. As students are nested within classrooms and classrooms are nested within schools, the factors influencing students' mathematics performance may be categorised as school-level factors, teacher-level factors and student-level factors. However, findings concerning the relationships between factors at the three levels and students' mathematics performance have varied, depending on the context and the set of variables involved in the study (Chen, 2016). The following sections present the factors that are related to students' mathematics performance in each category.

2.4.1. School-level Factors

Many previous studies have been conducted concerning school-related factors associated with students' mathematics performance and researchers have varied in what they classified as these factors. Fuller (1987) suggested that the association between school-related factors and students' mathematics performance was more significant in developing countries than it was in developed countries, with the school-related factors explaining a higher percentage of the variance in developing countries than it did in developed countries. The school-related factors included the number of textbooks, class size, school size, instructional materials and media, hours devoted to the subject in the timetable and the school library facilities (Fuller, 1987). Entwisle and Alexander (1992) listed student socio-economic background, whether the school was segregated, and racial composition as school-related factors. Sutton and Soderstrom (1999) reported that: (a) school-related factors had varying impacts on the graduation rate; (b) drop-out rates had a strong impact on students' achievement; (c) attendance, mobility and student-teacher ratio had a medium impact on achievement; and (d) class size and expenditure per students had a small impact on achievement. Schreiber (2002) reported that school resources and school size were associated with students' achievement in advanced mathematics and Mji and Makgato (2006) reported that teaching strategies, content knowledge, laboratory use and non-completion of the syllabus in a year were some school-related factors directly influencing mathematics performance. Furthermore, school type, school size and student-teacher ratio significantly influenced students' mathematics performance in Singapore (Chen, 2016).

The location of the school (whether it was in privileged or underprivileged area), and the type of mathematics programme adopted by the school were also reported to be associated with students' mathematics performance (S. E. Graham & Provost, 2012; Mohammadpour & Ghafar, 2014). School location has also been one of the factors influencing students' mathematics performance in Malaysia (Abdullah, Zain, Nair, Abdullah, & Ismail, 2016; Zamri, 2016) and Vietnam (Ha, 2016), with students from the urban schools having higher mathematics performance than students from the rural schools. Other studies have found that student demographic information and the school's socio-economic standing (Demir & Kilic, 2010; Lamb & Fullarton, 2002;

McCoach et al., 2010; McConney & Perry, 2010), as well as the tone of the school, also contributed to mathematics performance (Choi & Chang, 2011; McCoach et al., 2010; Shin, Lee, & Kim, 2009).

Based on these previous studies, the school-related factors selected for analysis of students' mathematics performance in Aceh are: school location, school and class size, school resources (including instructional materials, media and the number of textbooks), types of mathematics programmes and demographic information.

2.4.2. Teacher-level Factors

Teachers' Background

At the teacher-level, teacher background has been reported to have an impact on students' mathematics performance. In relation to teachers' background, factors commonly linked with students' mathematics achievement are the qualifications of teachers (Clotfelter, Ladd, & Vigdor, 2010; Croninger, Rice, Rathbun, & Nishio, 2007; Darling-Hammond, 2000; Demir & Kilic, 2010; Dodeen, Abdelfattah, Shumrani, & Hilal, 2012; Palardy & Rumberger, 2008) and teachers' professional development (Blank & de las Alas, 2009; Laura, McMeeking, Orsi, & Cobb, 2012; Wallace, 2009). Teacher certification is also found to be strongly related to students' mathematics performance (Darling-Hammond, 2000; Hill, Rowan, & Ball, 2005). Experience, usually indicated by years of teaching, is associated with teachers' certification also contribute to students' achievement (Hill et al., 2005; Kini & Podolsky, 2016; Ladd & Sorensen, 2017). There is a lack of evidence that the teacher's gender influences students' mathematics achievement (Chudgar & Sankar, 2008). However, the link between a teacher's gender and their students' mathematics achievement might exist in a particular country, such as was found in research in Pakistan (Warwick & Jatoi, 1994).

Teachers' Beliefs

While research related to teachers' beliefs has been conducted over a long period, no precise definition has been established (Goldin, Rösken, & Törner, 2009; A.G Thompson, 1992). The definitions and classification of beliefs are highly dependent on the discipline under research (Goldin et al., 2009; Leder, Pehkonen, & Törner, 2003). Philipp (2007, p. 259) created a working definition of beliefs as "physiologically held understanding premises or propositions about the world that are thought to be true". Kagan (1992, p. 73) defined teachers' beliefs as knowledge which would "reflect the actual nature of the instruction the teacher provides for the students". Furinghetti and Morselli (2009, p. 62) outlined three main components of mathematics teachers' beliefs as "the nature of mathematics, the nature of mathematics teaching, and the process of learning mathematics". In this study 'teachers' beliefs' focuses on teachers' understanding and teaching of LOT and HOT.

Research into teachers' beliefs in mathematics education is important as this provides a link between the curriculum and classroom practice (McLeod, 1992; Pajares, 1992). A. G. Thompson (1984) established that teachers' beliefs influence the direction of teaching practices. Teachers play important roles in the implementation of curriculum and decision-making related to classroom practices. Nathan and Koedinger (2000) looked at the accuracy of teachers' beliefs concerning difficulties in algebra and the student's performance, discovering that middle school teachers were most accurately able to predict students' performance. Stipek, Givvin, Salmon, and MacGyvers (2001) worked on linking teachers' beliefs and practices in mathematics instruction and discovered that teachers' beliefs were usually consistent with their practices. Staub and

Stern (2002) carried out a study concerning teachers' beliefs in terms of the teachers' knowledge of pedagogical content and its association with student learning outcomes, finding that when teachers believed strongly in a cognitive constructivist orientation, the students' performance was high. Muijs and Reynolds (2002) established that teachers' beliefs were linked indirectly to students' learning outcome through the teachers' practices in the classroom. Tschannen-Moran and Barr (2004) confirmed the previous study, finding that teachers' beliefs had a positive link with students' achievement. These studies indicate that teachers' beliefs make an important contribution to students' learning outcomes either directly or indirectly.

Classroom Practices

Turner, Christensen, and Meyer (2009) noted that teachers' beliefs strongly influenced teachers' decisions about their classroom instructions. Research into classroom practices in mathematics classrooms indicates their substantial influence on students' achievement (Kane, Taylor, Tyler, & Wooten, 2010; Rodriguez, 2004; Saxe, Gearhart, & Seltzer, 1999; Wenglinsky, 2001, 2002). Classroom practices include the choice of teaching methods and approaches, the types of learning activities, the types of tasks and the types of assessments. Seeing assessment as part of the classroom practices, Martínez, Stecher, and Borko (2009) accepted that teachers' conceptions concerning students' performance was influenced by the assessment practices. However, this relationship may also be inverted to indicate that teachers' conceptions of students' capabilities influence the assessment practices.

Brown, Kennedy, Fok, Chan, and Yu (2009) reported that teachers paid more attention to training students for a test when they believed that the students had a low level of abilities, modifying their usual classroom practice in order to improve students' performance. In another context, Tan and Saw Lan (2011) reported that when English was new as the medium of instruction, teachers' beliefs influenced the classroom practices through influencing the method of teaching: teachers used translation and the repetition of keywords in teaching mathematics and science. These studies highlight the influence of teachers' classroom practices on students' achievement but also indicate that the impacts may vary according to the situation.

Classroom practices can be defined in various ways. Stuart and Thurlow (2000) considered teaching methods as classroom practices. Stillman et al. (2009) described the types of mathematical tasks as classroom practices, while Cross (2009) defined classroom practices as how classroom activities were organised, teacher-student interaction and assessment of students' learning outcomes. The differing definitions of classroom practices reflect the objectives of the studies. In this current research study, classroom practices are described as teaching approaches, including teaching strategies, used in the classroom. It seeks to find: (a) whether teacher-centred or student-centred instructional activities predominate; (b) whether LOT or HOT activities are predominate; and (c) whether the types of assessment tasks relate to routine mathematics problems or non-routine problems.

As already outlined, there are various frameworks for understanding classroom practices. The factors can be related to how classroom practices are defined in a particular study. Carpenter, Fennema, Peterson, Chiang, and Loef (1989) argued that providing teachers with professional development workshops, which were related to research-based knowledge, changed both teachers' beliefs and classroom practices. Supovitz and Turner (2000) found that the number of professional development courses teachers were involved in had a strong influence on their classroom practices. This finding was followed up by several studies which indicated that teachers' beliefs concerning the nature of mathematics greatly influenced classroom practices (Cross, 2009; Stuart & Thurlow, 2000). The years of experience, teachers' qualification and gender might also influence the classroom practices but the influence of gender was not significance (Nisbet & Warren, 2000). Even though teachers' beliefs were strongly linked to classroom practices, there were other factors found to have a strong influence, namely: (a) the current school situation and previous teaching experience (Barkatsas & Malone, 2005); and (b) textbooks, standardised testing, and district requirements (Grouws, Good & Dougherty, 1990 in Da Ponte & Chapman, 2006). Furthermore, teachers' perceptions of students' abilities in learning (Stuart & Thurlow, 2000) and the integration of technology into the classroom (Kozma, 2003) might also result in a change in classroom practices.

Based on these previous studies, the teacher-level factors selected for analysis of students' mathematics performance in Aceh are: teacher background (experience, gender, age, education, professional development, and certification), teachers' beliefs (related to LOT and HOT) and classroom practices (instructional activities, teacher engaging students, types of mathematics questions used and teaching resources).

2.4.3. Student-level Factors

Students' Background

It is generally accepted that factors related to students have more predictors. These background factors associated with students' mathematics achievement include gender, socio-economic status (SES), parents' education, parents' occupation, wealth, technological devices available at home and number of books available at home. Gender has consistently been investigated as one of the so called 'predictors' of mathematics achievement, being reported as a significant factor in some studies (Lloyd, Walsh, & Yailagh, 2005; Mohammadpour, 2012), yet not significant in others

(Chen, 2016; Lindberg, Hyde, Petersen, & Linn, 2010; Petty, Wang, & Harbaugh, 2013; Steinmayr & Spinath, 2008; Tsui, 2007). Female students had lower achievement than males in some studies (Guiso, Monte, Sapienza, & Zingales, 2008) but also outperformed male students in some others (Hyde, Fennema, & Lamon, 1990). Gender has been found in one study to be a significant factor for higher achieving students, with male students having performed better (Tsui, 2007). Some studies reported that, male students performed better in problem solving (Hyde et al., 1990) and novel problems (Halpern, Wai, & Saw, 2005). However, the PISA 2012 study also reported the inconsistency of gender effect to students' mathematics performance in the South East Asian context. Female students were recorded to outperformed male students in Thailand (Dechsri, 2016) and Malaysia (Darmawan, 2016). Conversely, male students were reported to have higher mathematics performance in Vietnam (Ha, 2016) and Indonesia (Darmawan, 2016).

Socio-economic status (SES) (Chen, 2016; McConney & Perry, 2010; Mohammadpour & Ghafar, 2014; Oh, 2013; Thien & Darmawan, 2016), and parents' education (Howie, 2002; Kanyongo, Schreiber, & Brown, 2007; Mullis, Martin, Gonzalez, & Chrostowski, 2004; Schreiber, 2002; Wang, Wildman, & Calhoun, 1996) positively influenced students' mathematics achievement. The impact of SES tends to be stronger on students with high and average level SES than on those with lower level SES (Farooq, Chaudhry, Shafiq, & Berhanu, 2011). Parents' occupation has also widely been used as an indicator of SES, with students from higher SES having higher mathematics achievement, corresponding to the availability of technological devices at home. The availability of technological resources has been seen to influence students' achievement (Papanastasiou, 2000). Furthermore, the number of books at home indicated how important literacy was to a family and was also associated with mathematics achievement, with students who had a larger number of books at home tending to have higher achievement (Papanastasiou, 2000). Parents' education has been found to be more important than parents' occupation in explaining students' mathematics achievement (Farooq et al., 2011). In addition, the family composition, indicated by whether students were raised by single parents or two parents, was also associated with mathematics performance in certain countries. PISA results indicated that Indian students from single parent families had higher mathematics scores due to the culture of single parents living with the students' grandparents which resulted in the children receiving more help (Areepattamannil, 2014). A further factor in the students' backgrounds relates specifically to the individual student's educational aspirations, with how far they want to go in their schooling correlating with their achievement (Papanastasiou, 2000).

Students' Affective Factors

Affective factors, which include students' attitudes towards mathematics (valuing and liking), students' self-efficacy and beliefs concerning mathematics are prominent student-related factors associated with mathematics achievement. Affective factors have been seen to include "students' feeling about mathematics, aspects of the classroom or about (how they saw) themselves as learners of mathematics" (Reyes, 1984, p.558). These factors were not only important predictors of student mathematics achievement but also were important for the students' process of decision making when solving mathematical problems and their choices of courses and career paths (Reyes, 1984). Affective factors have also been found to be significant when looking at gender differences in mathematics achievement (Fennema, 1977; Fennema & Sherman, 1977; Hyde et al., 1990).

Students' positive attitudes toward mathematics (Areepattamannil, 2014; Mohamed & Waheed, 2011; Papanastasiou, 2000), students' high valuing of mathematics (Lay, Ng, & Chong, 2015) and students' positive self-efficacy (Ferry, Fouad, & Smith, 2000; Mohammadpour, 2012; Papanastasiou, 2000) were more likely to be positively associated with students' mathematics performance. In some studies, self-efficacy had the strongest association with mathematics achievement (Liu, 2009; Pajares & Graham, 1999; Papanastasiou, 2000). However, this might not necessarily be the case when different cultural and background characteristics are examined. Liu (2009), in her study comparing standardised mathematics performances, found that while Hong Kong students had a very low mathematics self-concept, their mathematics performance was exceptionally high, whereas students in the US had very high mathematics confidence, yet their performance was weak. Shin et al. (2009), in a study of Korean and Japanese students found that the attitudes toward mathematics were stronger predictors for mathematics performance while, for the US students, valuing mathematics was a stronger predictor.

Attitude has been defined as the student 'evaluation', involving 'liking or disliking', or providing a 'positive or negative ' response to a particular subject (Aiken, 1970, p. 551; Shrigley, Koballa, & Simpson, 1988, p. 675). When students have a positive attitude toward mathematics then they have high performance or they perform well in mathematics; positive performance then further reinforced positive attitude. One of the attitudinal variables is motivation which include both intrinsic and extrinsic motivation. Intrinsic motivation involves interest in learning and liking mathematics while extrinsic motivation concerns the utility values of whether mathematics is perceived as important and useful (Guay et al., 2010; Pintrich, 1999; Ryan & Deci, 2000).

Self-efficacy has the power to change individual behaviour both directly and indirectly (Bandura, 2006) and is defined as "people's judgement about their abilities" (Bandura, 1986, p. 94). Performance accomplishment is a main trigger, with high self-efficacy being correlated to highly successful performance while low self-efficacy is related to failure (Bandura, 1977). Despite the main core of self-efficacy being an individual judgement of one's performance, one's confidence to execute a particular task is also involved (Pintrich, 1999). Thus, in an educational setting, self-efficacy could be defined as what students thought of their abilities and how confident of success they are in a particular subject. Self-efficacy is a subject-specific matter, with no single self-efficacy scale being applicable across all subjects. Bandura (2006) advised that a generic form of self-efficacy scale would result in a poor fit of the measurement for a particular subject. Thus for this research, a specific scale for mathematics self-efficacy needed to be designed.

Students' beliefs concerning mathematics is a further affective factor influencing mathematics performance. Schoenfeld (1992, p.68) defined beliefs related to mathematics as "an individual's understandings and feelings that shape the ways that the individual conceptualizes and engages in mathematical behaviour". Furthermore, Op't Eynde, De Corte, and Verschaffel (2003, p.16) provided a working definition of students' mathematics-related beliefs as "the implicitly or explicitly held subjective conceptions students hold to be true that influenced their mathematical learning and problem solving". In this study, students' beliefs related to mathematics are defined as students' conceptions of the nature of mathematics and students' conceptions on how they learn mathematics.

Despite students' beliefs related to mathematics being found to have an influence on students' mathematics performance, it is rarely reported as having a direct effect.

Earlier, Schoenfeld (1989) found that students believed that mathematics helped the development of logic and allowed for creativity but at the same time, they believed memorisation was the best strategy to master mathematics. Garofalo (1989) found that students who held the belief that mathematics textbook exercises could be solved only by the methods presented in the textbook would not be able to see themselves creating their own solution for mathematics problems. Students' beliefs concerning mathematics influenced their decision-making in their approaches to and methods of dealing with mathematics tasks (Garofalo, 1989b). Recently, Schommer-Aikins, Duell, and Hutter (2005) reported that students' beliefs concerning mathematics influenced their mathematical problem solving performance; the less students believing in quick or fixed learning, the more likely they were to have higher achievement in mathematics problem solving tasks. Thus, research indicates that even though students' beliefs concerning mathematics has no direct association with mathematics performance, it does influence their practice in learning mathematics, which in turn has an impact on their performance.

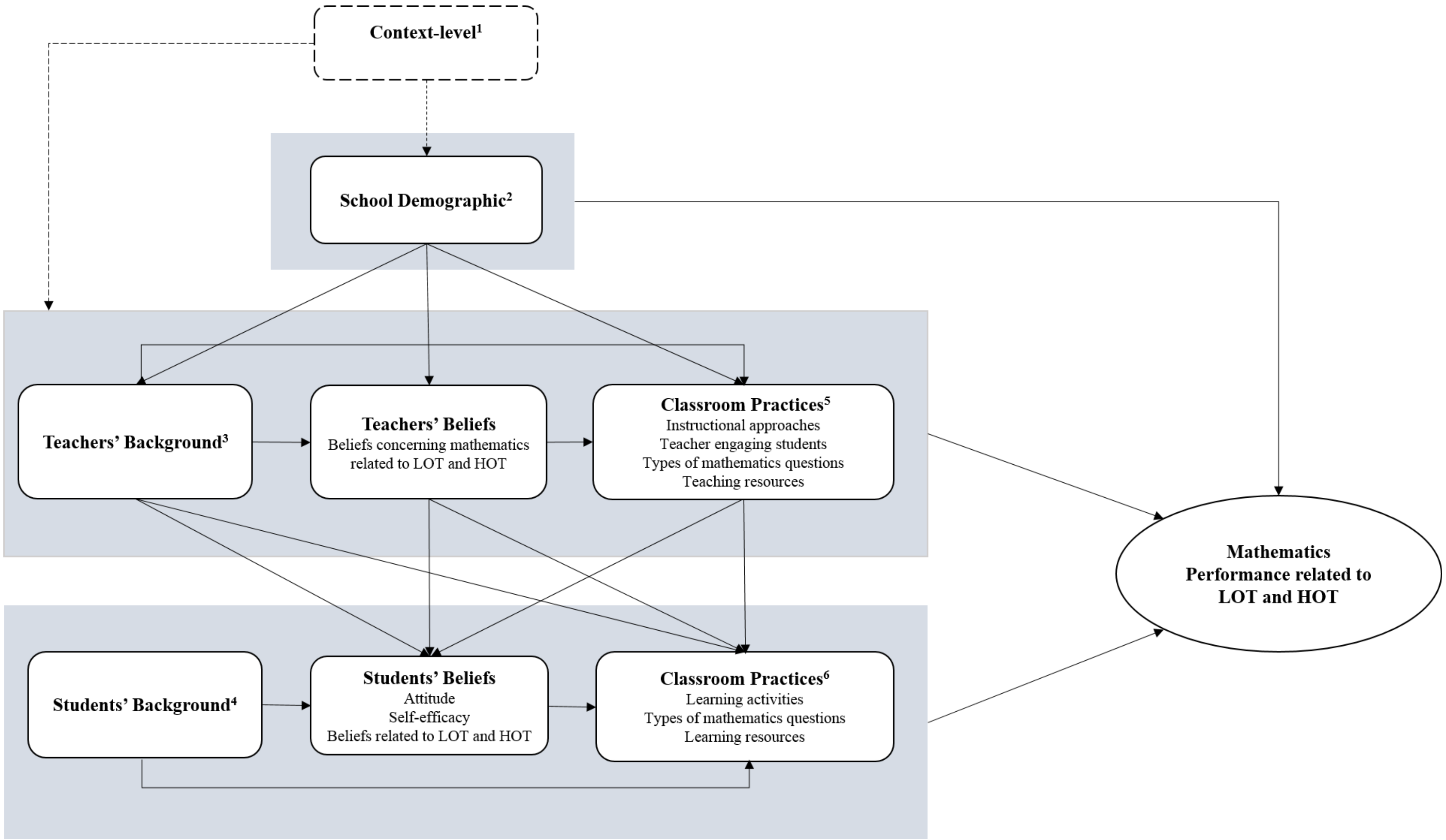
Based on these previous studies, the student-level factors selected for analysis of students' mathematics performance in Aceh are: students' background (gender, parents' occupation and educational level, home possession, and educational expectation), students' attitude, self-efficacy, beliefs related to LOT and HOT, students learning activities, types of mathematics questions and learning resources.

2.5 Conceptual Framework

From the review of previous studies relating to LOT and HOT, from the research into educational effectiveness theory and its emphasis on a multilevel analysis, and factors influencing students' mathematics performance, the conceptual framework for the present study is presented. The framework is created by incorporating constructs which represent student-, teacher-and school-level factors examining students' mathematics performance related to LOT and HOT. While there have been many studies investigating factors predicting students' mathematics performance by examining school, teacher, and school characteristics, they have often been a single level analysis rather than a multilevel analysis, despite the hierarchical nature of the data obtained.

This present study examines the causal relationships between factors at the student-, teacher- and school-level and how they influence students' mathematics performance related to LOT and HOT. The dynamic model of educational effectiveness developed by Creemers and Kyriakides (2006) is adapted and modified for this study. Only three out of four levels of the original model are used in this study, namely, student-, teacher- and school-level. While the context-level is not examined empirically in this study (as presented by dash line in Figure 2.2), it will be discussed in relation to the findings of this study in the final chapter. The variables or scales in each level have been selected for this particular study.

The conceptual framework in Figure 2.2 illustrates the relationships between student, teacher- and school-level factors as well as the interrelationships between factors in each level (student and teacher level) in influencing students' mathematics performance related to LOT and HOT. The conceptual framework proposes that school-level factors (defined by school demographic), teacher-level factors (defined by teacher background, teacher beliefs and classroom practice as perceived by teacher), and student-level factor (defined by students' background, student affective factors and classroom practices as perceived by students) have a direct impact on students mathematics performance related to LOT and HOT. The conceptual framework also implies the interrelationships of factors at single-level (teacher and students), as well as the relationships among factors across levels.



Note: [1]*National and regional socio-cultural characteristics and educational systems.* [2]*Types of schools, total enrolment, admission criteria, class size and classification, number of teachers, school resources and additional mathematics programme.* [3]*Experience, gender, age, education, professional development, and certification.* [4]*Gender, parents' occupation and educational level, home possession, and educational expectation.* [5]*Classroom practices as perceived by teacher.* [6]*Classroom practices as perceived by student.*

Figure 2.2 Conceptual framework of factors influencing students' mathematics performance related to LOT and HOT

2.6 Summary

Students' mathematics performance is an important focus of studies in the area of mathematics education. The findings may be useful for educational effectiveness involving the improvement of classroom practices and mathematics learning which in turn can benefit in advancing students' mathematics performance. Furthermore, advancing mathematics performance related to HOT is also being given attention in all countries, as students increasingly need these skills to be able to compete and fit in a global society.

The factors contributing to students' mathematics performance may be categorised into student-, teacher-, school-level factors. In this study, the school-level factors include school demographics information; teacher-level factors include background, beliefs and classroom practices as perceived by teachers; and the student-level factors include background, beliefs, and classroom practices as perceived by students. Wherever appropriate, a distinction is made between LOT and HOT so as to ascertain a more accurate assessment of these factors' influences on performance. The nested nature of these levels and the possibility of impacts across the levels requires both single-level and multilevel analysis. This allows the examination of relationships among factors in each level as well as across levels. Therefore, the conceptual framework is created to incorporate the explanatory factors associated with students, teachers and schools. The research methods employed to examine this conceptual framework follows this chapter.

Chapter 3 Methods of Investigation

3.1 Introduction

This chapter outlines the research methods used in this study, including the process of obtaining ethics approvals from both the University of Adelaide and the Education Department in Indonesia. It also provides the details of the pilot study in Aceh and the process of sampling and data collection. There are several scales used within the study and the operationalisation and instrumentation of these scales is presented. The statistical procedures associated with analyses employed are also discussed, together with the validity and reliability issues associated with the instruments. Confirmatory factor analysis (CFA), structural equation modelling (SEM), partial least squares (PLS) path analysis and hierarchical linear modelling (HLM) are examined. Moreover, the process of validation of the scales of measurement, using CFA and Rasch analysis, are discussed.

3.2 Methods Employed in this Study

The methods chosen for this study derive from the framework and research questions. The design of this study is fundamentally cross-sectional, with data collection being conducted at one point in time but with an underlying longitudinal purpose. A cross-sectional study design provides the best range of descriptive results within a short time at a relatively low cost and also allows the researcher to obtain data concerning causal processes and to examine a hypothesised causal model (De Vaus & de Vaus, 2001). A quantitative research method is necessarily employed in this study. Bryman (2012, p.160) described a quantitative method as "entailing the collection of numerical data as exhibiting a view of relationship between theory and research as deductive and a prediction for a natural science approach". The numerical data is obtained through compiling questionnaires to schools, teachers and students, as well as a mathematics test for students to assess performance. The questionnaires were used to enable the researcher to obtain a description of trends in a population by investigating a sample of the population (Creswell, 2013).

3.3 Ethics Approval

Ethics approval was required before the data collection could take place. Approval was sought from the University of Adelaide, Indonesia requiring no further ethics approval. However, it was necessary to seek permission from the Department of Education and Culture as well as from the Department of Religious Affairs of Republic Indonesia in the province of Aceh to conduct the study. The application for approval of the ethics was submitted to the Research and Ethics Committee (UAHREC) at Adelaide University and approved on 21 May 2013 (project number HP-2013-034, refer to Appendix A). The approvals from the Aceh Department of Education and Culture and from the Department of Religious Affairs of Republic Indonesia were obtained a month later (refer to Appendix B).

3.4 Sampling and Data Collection

3.4.1. Population

The study is conducted in the province of Aceh, located in the far west of Sumatra, Indonesia. The province consists of 18 districts and five cities, with 567 public junior high schools, 113 Islamic public junior high schools, 93 private junior high schools, 189 private Islamic junior high school and 40 special needs junior high schools. The special needs schools, which do not use the standard curriculum, are excluded from the target population of this study.

3.4.2. Sample and Sampling Method

The sample of this study is Year 9 students enrolled in 25 schools and their mathematics teachers, from both rural and urban areas. The urban area in the study is the capital of Aceh, Banda Aceh, where 13 schools were co-opted. The rural area chosen is Aceh Besar, the district area surrounding Banda Aceh, where 12 schools were co-opted. The hinterland of Banda Aceh was chosen as the rural area to enable the researcher to visit the schools involved throughout the data collection period, further distances requiring time and transport. The number of Year 9 students participating in this study is 1135 (581 from urban schools and 554 from rural schools) with 46 mathematics teachers. The number of female and male students is approximately equal (630 female and 505 male) but the number of female and male teachers is not (39 female and seven male), as female teachers in this sample predominate at this level. The summary of the sample is presented in Table 3.1 (refer to Chapter 6 for a more detailed description).

Table 3.1

Distribution of the Sample according to Location

	Frequency		Total
	Urban	Rural	
School	13	12	25
Teacher	23	23	46
Student	581	554	1135

A stratified sampling method was used to obtain the appropriate number of schools and students in the desired sample. The advantages of stratified sampling is that it raised the "accuracy in sample estimate" without raising the cost (Ross, 1978, p. 10). Stratified sampling also enables the researcher to have a sample containing a representation of the general population (Bryman, 2012). The stratification variables are the type of schools, which has four categories: (a) public school, (b) public Islamic school, (c) private school, and (d) private Islamic school; and school location, which has two categories: city of Banda Aceh (urban) and district of Aceh Besar (rural).

In this study, students are nested within teachers, and the teachers are nested within the schools. The sampling process is started by randomly selecting the schools in each of the strata. Then the mathematics teachers who teach Year 9 in the respective school are approached, with their Year 9 students making up the student data set. Usually only two mathematics teachers are teaching Year 9 in each school. When there were more than two mathematics teachers teaching Year 9 students in one school, one female and one male teacher were chosen in order to balance the gender distribution. Where the teacher had more than one Year 9 classroom the teacher chose which class would participate in the study. All students in the chosen class are included in the study.

3.4.3. Data Collection

The data collection period was from July 2013 until mid-October 2013. The initial phase of the data collection involved administering the school questionnaire and the teacher questionnaire. The second and third phases involved administering the student questionnaire and mathematics test to the students. Sometimes, according to school flexibility, the test was administered before the questionnaire.

3.5 The Questionnaires

This study administered three sets of questionnaires: to the school; to the teachers; and to their students. These are included in the appendices (Appendix C to E).

3.5.1. School Questionnaire

The school questionnaire, completed by a school administrator, consists of 14 questions concerning demographic information: the total enrolment of the schools, the number of teachers, school admission criteria, classroom classification, available learning resources and also any available extra mathematics programmes at the school. No validation of the items in the questionnaire is considered to be needed.

3.5.2. Teacher Questionnaire

The teacher questionnaire is self-reported with 17 closed-ended questions. The first section contains information relating to gender, qualifications, certification, professional development, years of experience, and the use of technology. The second section concerns teachers' beliefs related to the nature of mathematics, mathematics learning, and mathematics teaching distinguishing LOT and HOT. LOT characteristics are characterised by the use of routine problems, repetitive or mechanistic work, closed questions, the use of a teacher-centred approach, and the development of the skills of remembering, understanding and applying skills. HOT is characterised by the use of non-routine problems, open-ended questions, the use of a student-centred learning environment, as well as the development of analysing, evaluating, creating, transferring, and reasoning skills involving problem solving and critical thinking. The third section of the teacher questionnaire consists of questions related to their classroom practices, as well as learning resources in the classroom.

3.5.3. Student Questionnaire

The student questionnaire consists of 17 closed-ended questions to provide information on student's gender, socio-economic status (SES) and affective dimensions of self-efficacy towards mathematics, attitudes toward mathematics and beliefs concerning mathematics. SES is indicated by the parents' jobs and education and the possession of particular items in the home (e.g. computer, mobile phone, car, internet connection, a room of your own, etc.). Judgements of self-efficacy toward mathematics are indicated by statements of their self-perceived abilities in doing mathematics, whether they like mathematics or not and whether they enjoy learning mathematics or not. Attitudes toward mathematics are indicated by statements regarding students' opinions concerning mathematics relating to their daily life, as well as views on the necessity of learning mathematics in order to understand other subjects, or to enter university, or to obtain a particular job. Beliefs toward mathematics involve both LOT and HOT aspects and are indicated by questions regarding students' beliefs towards mathematics such as: 'mathematics has only one correct answer', 'it is best to solve mathematics problems individually', and 'only advanced students can learn mathematics'.

3.6 Operationalisation and Measurement

This section presents information on the scales used in school, teacher and student questionnaires.

3.6.1. Variables and Scales used in the School Questionnaire

The school questionnaire consists of variables related to demographic information, namely: types of schools, total enrolment of the schools, school admission criteria, class classification, class size, the number of teachers in the schools, the number of teachers with certification, school resources, and any additional mathematics programmes. The types of school refers to the category of the school such as public or private school (there are four different types of schools included in the study). The total enrolment of the schools refers to the total number of the students in the school.

School admission criteria refers to the means of selecting students to be admitted in the school (such as being a residence in particular area).

Class size refers to the average numbers of students in the classroom and class classification refers to the method of classifying students into the classroom (whether students are grouped based on their ability or not). The number of teachers refers to the total number of teachers in the school, the total number of mathematics teachers and the total of mathematics teacher teaching Year 9. Teacher certification refers to the total number of teachers with teaching certification and the total number of mathematics teachers with teaching certification.

Table 3.2

Development of Items of School Questionnaire and Expressions Used in School Measurement Scales

Variables	Items	Expression used in the category and coding	Sources
Types of schools	1	Yes (1), No (0)	Researcher own constructed item
Total enrolment of the schools	2	Number	Foy, Arora, and Stanco (2013)
School admission criteria	3-5	Yes (1), No (0)	Foy et al. (2013)
Class classification	6-8	Yes (1), No (0)	Foy et al. (2013)
Class size	9	Number	Researcher own constructed item
Number of teachers in the schools	10, 12, and 14	Number	As above
Number of teachers with certification	11 and 13	Number	As above
School resources	15-25	Number	Foy et al. (2013)
Additional mathematics programmes	26-29	Yes (1), No (0) ≤ 2 hours (1) to ≥4 hours (4)	Researcher own constructed item

School resources refer to the number of resources (such as mathematics textbook and computers for mathematics instruction) that are available at schools. Additional mathematics programmes refers to mathematics class after school hours and club as ‘Mathematics Olympiad Club’ provided in the school and the frequency of the programmes conducted. The development of the items of the school questionnaire, including the expression used in each variable, is summarised in Table 3.2.

3.6.2. Variables and Scales used in the Teacher Questionnaire

Teachers' Background

Teachers' background variables include experience, gender, age, education, teacher professional development and teacher certification. Experience refers to the number of teaching years. Gender refers to the sex of the teachers, i.e. male or female. Age refers to the age of teachers and education refers to teacher's highest level of education and subject major. Teachers' professional development refers to the regularity of attending professional development programmes for mathematics teachers and the topic of professional development training or seminars the teachers attended during the last two years. Teacher certification refers to whether or not teachers have passed the certification examination.

Beliefs concerning Mathematics

Teachers' beliefs concerning mathematics refers to teachers' personal knowledge and understanding concerning the nature of mathematics, mathematics learning and mathematics teaching associated to both LOT and HOT (Furinghetti & Morselli, 2009; Goldin et al., 2009; Philipp, 2007). The questions related to beliefs concerning mathematics are used to assess teachers' beliefs concerning both LOT and HOT and have three subscales: beliefs concerning the nature of mathematics; beliefs concerning mathematics learning; and beliefs concerning mathematics teaching. The questions, developed by Perry, Tracey, and Howard (1999) for mathematics generally, were adapted. These questions are registering the teachers' self-reported beliefs. Some of the items from Perry et al. (1999) already met the characteristics of LOT or HOT (as previously outlined in Chapter 2). The adapted scale consisted of 34 items with a four-point Likert response scale: nine items of beliefs concerning the nature of mathematics; 12 items of beliefs concerning mathematics learning; and 13 items of beliefs concerning mathematics teaching. Each statement has four possible responses: 'disagree a lot', 'disagree a little', 'agree a little' and 'agree a lot'. The original scale used by Perry et al. (1999) employed only three Likert-scale responses: 'agree', 'undecided' and 'disagree'. The modified instrument also allowed for the characterisation of statements related to LOT and HOT.

The scale of beliefs concerning the nature of mathematics was composed with four items from the original scale developed by Perry et al. (1999); one item taken from the scale of conceptions of mathematics by Crawford, Gordon, Nicholas, and Prosser (1998); and four items created for the current study. Five statements relate to HOT and four statements relate to LOT. The scale of beliefs concerning mathematics learning included five items from Perry et al. (1999); two items taken from Hart (2002); one item taken from Peterson, Fennema, Carpenter, and Loef (1989); and three items created for the current study. Eight statements relate to HOT and four items relate to LOT. The scale of beliefs concerning mathematics teaching included four items from Perry et al. (1999); two items taken from Peterson et al. (1989); two items taken from Hart (2002); one item taken from (Zakaria & Musiran, 2010); and four items created for the current study. Eight statements relate to HOT and five statements relate to LOT.

Classroom Practices

Classroom practices consist of the types of mathematics questions used in classrooms and examinations (Martínez et al., 2009); instructional activities for students (Cross, 2009); teacher engaging students (Franke, Kazemi, & Battey, 2007; Stuart & Thurlow, 2000); and teaching resources (Kozma, 2003; Remillard, 2005). The scales concerned with classroom practices are adapted from the TIMSS 2011 teacher questionnaire (Foy et al., 2013). The scale was modified and employed in this study using the responses of a five-point Likert scale: 'never', 'rarely', 'sometimes', 'usually', and 'always'.

Table 3.3

Development of Items of Teacher Questionnaire and Expressions Used in Teacher Measurement Scales

Scales	Subscales	Items	Expression used in the category and coding	Sources
Teachers' background	Experience	1	Year	
	Gender	2	Female (1), Male (0)	
	Age	3	Year	Foy et al. (2013)
	Education	4-13	Yes (1), No (0)	
	Teachers' professional development	14-26	Yes (1), No (0)	
	Certification exam	27	Yes (1), No (0)	Items created for this study
Teachers' beliefs concerning mathematics (TBM)	Teachers' beliefs concerning nature of mathematics (TBNM)	28-35		Crawford et al. (1998); Perry et al. (1999); four items created for this study
	Teachers' beliefs concerning mathematics learning (TBML)	36-45	Disagree a lot (1) Disagree a little (2) Agree a little (3) Disagree a lot (4)	Crawford et al. (1998);Hart (2002); Perry et al. (1999); three items created for this study
	Teachers' beliefs concerning mathematics teaching (TBMT)	46-56		Hart (2002); Perry et al. (1999); Peterson et al. (1989); Zakaria and Musiran (2010); four items created for this study
Classroom practices	Types of questions used in the classroom	57-60		
	Types of questions used in the exam	61-65	Never (1) Rarely (2) Sometimes (3) Usually (4) Always (5)	Foy et al. (2013)
	Instructional activities for student (IAS)	66-75		
	Teacher engaging students (TES)	76-81		
	Teaching resources	82-85	Not used (0) Supplement (1) Basis of instruction (2)	

The types of mathematics questions used in the classroom refers to the frequency of teachers using certain types of questions (such as word problems and non-routine problems) in mathematics classrooms. The types of mathematics questions used in the examination refers to the frequency with which teachers use certain types of questions. Teaching resources refers to the frequency of teachers using certain types of teaching resources (such as mathematics textbooks and concrete materials) in the mathematics classroom. Instructional activities in the classroom refers to the frequency with which teachers ask students to do various activities and teacher engaging students refers to the frequency with which teachers attempt to increase students' interest and engagement in their learning. The development of the items of the teacher questionnaire, including the expression, used in each measurement scale, is summarised in Table 3.3.

3.6.3. Variables and Scales used in the Student Questionnaire

Students' Background

Students' background consists of gender, family structure, parents' occupation, parents' educational level, home possessions and student's educational expectation. Gender refers to the sex of students, i.e. male or female. Family structure refers to the family students' living with. Parents' occupations refer to the occupation of mother and father or guardian. Parents' educational levels refer to the highest level of education held by mother and father. Home possessions refers to two categories: (a) several items of educational resources and general possessions in the students' home; and (b) the quantity of (so called) luxury items that are available in the student's home. Students' educational expectation refers to the highest level of education that students aspire to achieve.

Students' attitude Toward Mathematics

Students' attitude toward mathematics refers to students' evaluation of mathematics as a subject, which involves their liking or disliking it (Aiken, 1970; Shrigley et al.,

1988) and whether it is important or useful (Guay et al., 2010; Pintrich, 1999; Ryan & Deci, 2000). In order to assess students' attitudes toward mathematics, this study uses an adaptation of the scale of attitude toward mathematics employed by TIMSS 2011 (Foy et al., 2013). The original items explicitly aim to measure students' attitudes toward mathematics; therefore it is assumed that the scale meets the purpose of the scale required for this study. The adapted instrument includes nine statements relating to two subscales of attitude toward mathematics: students' liking of mathematics (five statements) and students' valuing mathematics (four statements). The original instruments use four-point Likert scale responses: 'agree a lot', 'agree a little', 'disagree a little', and 'disagree a lot'. The adapted instrument retains four-point Likert scale but in the reverse order: 'disagree a lot', 'disagree a little', 'agree a little' and 'agree a lot'. The reversing of the responses is made to assist in coding the responses.

Students' Self-efficacy toward Mathematics

Student self-efficacy toward mathematics refers to what students thought of their abilities and how confident of success they are in handling mathematics as a subject (Bandura, 1986, p. 94; Pintrich, 1999). The students' self-efficacy toward mathematics scale is used to assess their confidence in learning mathematics and their assessment of their mathematical ability. The scale is made up of two subscales: confidence (four statements) and individual judgement (five statements). The instrument used is adapted from the TIMMS 2011 student questionnaire (Foy et al., 2013). The original instrument intended the nine items to assess mathematics confidence only and classified mathematics confidence as part of student attitudes toward mathematics. In this study the scale fits the description of self-efficacy, which also involves a judgement of one's own ability. The original instruments uses a four-point Likert scale responses: 'agree a lot', 'agree a little', 'disagree a little', and 'disagree a lot'. The adapted instrument also uses four-points Likert scale responses but in the reverse order: ‘disagree a lot’, ‘disagree a little’, ‘agree a little’ and ‘agree a lot’. The reversing of the responses is made to assist in coding the responses.

Students’ Beliefs concerning Mathematics

Students’ beliefs concerning mathematics refers to students’ understanding of the nature of mathematics (Schoenfeld, 1992, p.68) and students’ conceptions on how they can learn mathematics (Op’t Eynde et al., 2003) in relation to LOT and HOT. This scale aims to assess the tendency in which students’ beliefs concerning mathematics are related primarily LOT or HOT. As no previous scale was available fitting the focus on LOT and HOT as in this research, these items include statements associated with the nature of mathematics and mathematics learning and was formed by selecting items from beliefs related to the nature of mathematics and mathematics learning employed by previous researchers (Kloosterman, 2002; Op't Eynde & De Corte, 2003; Presmeg, 2002; Schoenfeld, 1992; A.G Thompson, 1992). The scale consists of 14 statements related to the nature of mathematics and mathematics learning which have two subscales: beliefs concerning mathematics related to LOT and beliefs concerning mathematics related to HOT. The scale uses a four-point Likert scale response: ‘disagree a lot’, ‘disagree a little’, ‘agree a little’ and ‘agree a lot’.

Students’ Learning Activities

The scale of students’ learning activities refers to students’ engagement in learning activities in the mathematics classroom (Cross, 2009; Franke et al., 2007; Stuart & Thurlow, 2000). It provides a link with the teacher scale of instructional approaches. It also provides a comparison between how students perceive their engagement in learning and how teachers perceive their engagement in the classroom. The instrument is adapted from TIMSS 2011 teacher questionnaire (Foy et al., 2013), using its statements of teacher instructional approaches. It consists of eight statements which are then classified into two subscales of student learning activities related to LOT and student learning activities related to HOT. However, the student learning activity scale responses are simplified to a three-point Likert scale: ‘never or almost never’, ‘sometimes’ and ‘always or almost always’. The development of the items of the student questionnaire, including the expression used in each measurement scale, is summarised in Table 3.4.

Table 3.4

Development of Items of Student Questionnaire and Expressions Used in the Student Measurement Scales

Scales	Subscales	Items	Expression used in the category and coding	Sources
Student's background	Gender Family structure Mother's job Mother's education Father's job Father's education Home possessions Educational expectation	1 2-7 8 9-16 17 18-25 26-43 44-49	Yes (1), No (0)	Foy et al. (2013)
Students’ attitude toward mathematics (SAM)	Liking mathematics Valuing mathematics	48-54 55-59	Disagree a lot (1) Disagree a little (2) Agree a little (3) Disagree a lot (4)	Foy et al. (2013)
Students’ self-efficacy toward mathematics (SSM)	Mathematics confidence Individual judgement of mathematics ability	60, 63, 65-67 61-62, 64, 68	As above	Foy et al. (2013)
Student beliefs concerning mathematics (SBM)	Students’ beliefs concerning mathematics related to LOT and Students’ beliefs concerning mathematics related to LOT	69-82	As above	Kloosterman (2002); Op’t Eynde et al. (2003); Presmeg (2002); Schoenfeld (1992); A.G Thompson (1992)
Types of questions used in the classroom		83-86	Never or almost never (0) Sometimes (1) Always or almost always (3)	Foy et al. (2013)
Learning resources		87-90	As above	Foy et al. (2013)
Learning activity		91-98	As above	Foy et al. (2013)

3.6.4. The Mathematics Test

The mathematics test is employed to examine the students' performance related to LOT and HOT, consisting of eight questions. The test questions combine items for Grade 8, adapted from both *TIMSS* and *PISA.* Two items are adapted from TIMSS 1999 (IEA, 2001), one item from TIMSS 2003 (IEA, 2005), one item from TIMSS 2007 (IEA, 2009), two items from TIMSS 2011 (IEA, 2013) and two items from PISA released items (OECD, 2006). As Indonesia has been participating in TIMSS and PISA testing, it was decided that collating items from these tests would provide a valid assessment. The choice of questions taken from each original instrument is based on the task level (lower or higher order) and on the content areas covered in the curriculum by Year 9 students in Aceh, Indonesia at the time of data collection. The classification of the test items as LOT or HOT was according to Bloom's revised taxonomy (Krathwohl, 2002) in which remembering, understanding, and applying are categorised as LOT and analysing, evaluating and creating are categorised as HOT. The lists of the items, the content domain and the cognitive domain are given in Appendix F and G.

3.7 The Pilot Study

In order to ensure that the instruments are functional a pilot study was conducted. The pilot study provided the content validity of the instrument and also exposed any problems with the questions, format and scales (Creswell, 2013). Such a study is especially important when self-administered questionnaires are involved, as there is no possibility of clearing up any areas of confusion as the forms are filled in by the participants (Bryman, 2012). Back-to-back translation from English to Bahasa Indonesia was carried out for all the tests and questionnaires used in the study prior to the pilot study. The following section provides detailed information relating to the pilot study.

3.7.1. The School Questionnaire

The school questionnaire was concerned with demographic information related to the schools, with both closed and open-ended questions. Three teachers and one university scholar in Aceh provided feedback on the format and content of the school questionnaire. Their feedback was then used to improve the questionnaire.

3.7.2. The Teacher Questionnaire

The teacher questionnaire consisted of three parts, seeking information about the teachers' backgrounds, their beliefs and classroom practices. Four different schools participated in the pilot study and ten mathematics teachers were involved. Two schools in the rural area and two schools in the urban area of the Aceh province, Indonesia were chosen. These schools were chosen for convenience, as the researcher had contacts within these schools. However, in these schools there were few male mathematics teachers. The participants were asked to respond to the questionnaires and provide feedback on statements or words they found confusing or unclear. They were also asked to provide further suggestions for the questionnaires. The feedback was then collected and amendments were made to improve the questionnaire.

3.7.3. The Student Questionnaire

The initial student questionnaire had 17 questions and was also categorised in three parts including students' backgrounds, beliefs about mathematics and classroom practices. After the demographic information, all questions were in a closed-ended format using Likert scale responses. There were five scales tested in the questionnaire:

‘students liking mathematics’, ‘valuing mathematics’, ‘self-efficacy’, ‘beliefs concerning nature of mathematics’, and ‘beliefs concerning mathematics learning’.

One school was chosen for the pilot study, a public school in the urban area of the Aceh province, Indonesia. This school was chosen because of its convenience and suitability and was then excluded from the sampling plan for the main study. A class of 30 Year 9 students were chosen from this school. Although class was not chosen randomly, it did consist of equal numbers of female and male students. It was important to discover how long it took the students to fill in the questionnaire and if the students had any confusion with questions. On average, students took around 15-20 minutes to complete the questionnaire.

The data obtained from the student questionnaire were then keyed in and processed into the IBM SPSS 22 which was employed to analyse the data in order to examine the consistency and reliability of the items for each scale in the instrument.

3.7.4. The Mathematics Test

The student mathematics test consisted of eight problems related to LOT and HOT. The student sample for piloting the mathematics test was the same sample as that used for piloting the student questionnaire. The pilot study was carried out to ensure that the test format and wording were appropriate for Year 9 students. It was also necessary to check the time needed, as well as the level of difficulty of the test (whether the test was too easy or too difficult for students). It took approximately 35 to 45 minutes for students to answer the questions. The results from the pilot study showed that the test was adequate for Year 9 students. However, most students were having problems with the HOT questions. Before administering the pilot test, the mathematics questions had already passed through checks on the content and face validity. One mathematics education university scholar from Indonesia and one mathematics education university scholar in Australia participated in the examination of content and face validity. Amendments were then made based on the feedback received.

3.8 General Methodological Considerations

Several statistical procedures were prepared for use in this study with a range of statistical software which was created purposely for each procedure to investigate causal relationships between student-, teacher- and school-level factors and student mathematics performance related to LOT and HOT. The use of appropriate statistical procedures was vital to ensure valid and reliable results. Prior to choosing the appropriate analyses for the purpose of this study, there are some general methodological issues that have to be considered: (a) missing values, (b) the notion of causality and (c) the level of analysis.

3.8.1. Missing Values

Despite the great effort to obtain complete datasets, missing values are unavoidable in research based on surveys (Kline, 2011). This is for varied reasons, such as: (a) the respondents' failing to provide answers for certain questions (Darmawan, 2003); (b) the respondents' lack of knowledge concerning certain questions; and (c) the respondents' failing to complete the questions due to the length of the questionnaires (Vriens & Melton, 2002). Kline (2011) noted that a small number of missing values (such as less than 5%) for a single variable of a large sample was not a great concern, especially when data loss occurred accidentally or not systematically. There are four methods for dealing with the missing values (Vriens & Melton, 2002):

(1) The available case method: This analyses only the data available involving a list-wise deletion (deleting cases that have missing values on any variable) or pair-wise deletion (deleting cases that have missing values on the variables involved in a certain analysis).
(2) The single imputation method: This analyses data after replacing the missing values with the mean of each variable based on the cases available.
(3) The model-based multiple imputation method: This analyses data after replacing the missing values with multiple values "obtained by a draw from the predictive distribution of the variable" (Vriens & Melton, 2002, p.14).

The main advantage of the available case method list-wise deletion is its simplicity and directness (Hair, Black, Babin, & Anderson, 2014). However, discarding any case with any missing values might result in a large decrease in the sample size, causing substantial loss of information (Vriens & Melton, 2002) and the pair-wise deletion would result in a non-uniform sample size for each variable (Kline, 2011). The single imputation method that employed the replacement of missing values by the mean is considered superior to the available case method (Vriens & Melton, 2002). This method is easily implemented and widely used, the rationale behind it being that the mean value is the best single replacement value (Hair et al., 2014). However, it could result in the distortion of the underlying distribution of the data, with the variance reduced and the distribution being centred more on the mean (Vriens & Melton, 2002). Within the literature, the model-based multiple imputation is the most recommended and advanced method of managing missing values (J. W. Graham & Coffman, 2012; Vriens & Melton, 2002), imputing "the missing value several times by drawing from a normal distribution with the estimated mean and variance" (Vriens & Melton, 2002, p.14). The major advantages of this method are that it is able to handle both non-random and random missing data and provide new complete cases that are more representative of the original data and have least bias (Hair et al., 2014). However, this method is complex and special software is required except for the expected maximisation method (EM) in the IBM SPSS 22 (Hair et al., 2014).

The single-imputation method of managing missing data is used in this study, as the percentage of missing values in the datasets is less than five per cent and is not systematic (Hair et al., 2014; Kline, 2011). It is also less complex than the other methods. The missing values are replaced by the mean of each variable in this study.

3.8.2. Notion of Causality

Causal analysis is widely conducted in the social sciences. Such analysis has "increasingly developed consensus on applying the potential outcome model" (Wolf & Best, 2015, p. 76). The application of causal analysis leads to a particular inference drawn from the survey data called 'a causal inference'. The notion of causality arises when the existence of one event could be a basis to expect the production of another (Heise, 1975). Heise (1975, p. 12) provided the guidelines for the application of the causality principle in theory construction and the design of research:

> An event C, causes another event, E, if and only if:
>
> (a) An operator existed which generated E which responds to C, and which was organised so the connection between C and E could be analysed into a sequence of compatible components with overlapping event fields;
>
> (b) Occurrences of event C were coordinated with the presence of such an operator, such an operator exists within the fields of C;
>
> (c) When condition (a) and (b) were met, when the operator was isolated from the fields of events and other than C, and neither C nor E was present to begin with, then occurrences of C invariably started before the beginning of an occurrence of E;
>
> (d) When condition (a) and (b) were met, C implies E; that was, during some time interval occurrences of C were always accompanied by occurrences of E, though E might be present without C or both events might be absent.

Further, Heise (1975) elaborated that condition (a) and (b) were the basis of causality where a highly structured situation ought to have existed for the possibility of a particular causal relation to be present, and the incidents ought to be synchronised with such conditions for the effect to exist.

Recently, Pearl (2010, p.108) proposed a five-step process of methodological principles of causal inference: (a) Define: indicating the target quantity Q as a function Q(M) that could be computed from any model M; (b) Assume: expressing the causal assumptions in ordinary scientific language and creating a graphical representation of the structural part; (c) Identify: finding out whether the target quantity is identifiable; (d) Test: determining any testable implication of M and test the necessary implications for the identifiability of Q; (e) Estimate: estimating the identifiable target quantity or approximating the non-identifiable ones.

Based on the earlier guidelines provided by Heise (1975) and the five-step process of Pearl (2010), it is necessary to set up the outcome variables and any variables to be included in a model, hypothesising the possible causal relationships between variables, determining how to measure each variable, undertaking the analysis for any possible relationships and estimating the outcomes. Furthermore, with the assumption that the causal inference requires the precedent event establishing an expectation for the second event to happen, in some particular circumstances deleting the impossible or non-significant relationship is acceptable. The principles mentioned before are reflected in the model building of any causal analysis employed in this study and in the model trimming where the elimination of the impossible or non-significant relationships is necessary.

3.8.3. Level of Analysis

The data collected in this study includes variables gathered at the student-, teacher- and school-level. This data is of a hierarchical nature, with the individual students, the professional teachers and the institution of the school. The model of students' mathematics performance of both LOT and HOT skills as the outcome is influenced by certain student-, teacher- and school-level variables. A causal relationship may be found using a test of statistical significance. However, bringing all three levels of data into one dataset may raise some problems and create some bias. Therefore, it is necessary to pay attention to the level of analysis of the data.

Commonly, data from two levels is brought together by aggregating (taking lower level data to a higher level) or disaggregating (taking higher level data to a lower level) (Snijders & Bosker, 1999). Hox and Roberts (2011, p.4) have stated that both aggregating and disaggregating "were flawed because the analysis either ignored the different level or treated them inadequately". Some of the potential problems of the aggregation method have been listed by Snijders and Bosker (1999): (a) the shift of meaning, where the variables aggregated to the upper meaning could not be directly referred to the lower level, instead referring only to the upper-level; (b) the ecological fallacy, which was related to the previous problem, where any correlations between the upper-level variables could not be used to make any claim about the lower-level; (c) the neglect of the original data structure; and (d) the absence of cross-level interaction examination. The problems of the disaggregation involved a high increase of the number of units, as the number of units at the lower-level were added to the number of units brought from upper-level (Snijders & Bosker, 1999).

The problems resulting from aggregation and disaggregation methods are avoided in this study. In order to obtain meaningful estimates and conclusions, the analyses are conducted at each individual level, as well as at the multilevel. This multilevel analysis incorporates the three levels of hierarchical data (student, teacher and school) in a hierarchical linear modelling (HLM) analysis, using HLM software developed by Raudenbush et al. (2004).

3.9 Statistical Procedure Employed in this Study

This section discusses the methods of analysis employed in this study. Confirmatory factor analysis (CFA) and Rasch measurement are used for the validation of the scales. Structural equation modelling (SEM), in particular, the SEM based on partial least square (PLS), is used for examining relationships at each level (student and teacher level). Hierarchical linear modelling (HLM) analysis is employed to examine variables from the three-levels of data. The details of each analysis employed and the statistical software used are presented.

3.9.1. Confirmatory Factor Analysis (CFA)

Factor analysis is used to examine the relationship between a set of manifest variates and a latent variable by calculating the covariance within the set of manifest variates to obtain information on the implicit trait (Byrne, 2013; Field, 2013). The main goal of factor analysis is to examine the correlations between the observed and underlying latent variables (Bollen, 1989) and to reduce the number of initial variables into a smaller set of factors without losing much information (Hair et al., 2014).

Exploratory factor analysis (EFA) is conducted with the assumption that the relationship between a set of observed variates and the latent variable is not yet known or established and performed to obtain statistical information on how many factors are required for the model to best represent the data, where the factors are only named after the analysis has been carried out (Hair et al., 2014). In contrast, the confirmatory factor analysis (CFA) is conducted with the assumption that the relationship between the set of observed variates and the latent variable has already been established and therefore the analysis is carried out to test the hypothesised structure of the relationship (Byrne, 2013; Schumacker & Lomax, 2012). In addition, CFA is performed to examine how well the constructs were represented by the observed variates (Hair et al., 2014).

CFA allows relationships to be established between observed variates and latent variables (Bollen, 1989). It is used to investigate whether the constructs fitted and operated well in the population of the study as well as to provide comparisons of the use of constructs across different studies and populations (Harrington, 2008). CFA is therefore employed in this study as most sets of items used are adapted from the finding of existing research studies and the researcher has already specified the model and the number of constructs that can result from a set of observed variates. In this study, CFA is performed using the IBM SPSS Amos (Analysis of Moment Structure) 22.

CFA is used in this study, at the validation stage, to examine the construct validity and reliability of each scale. In particular, the model comparison approach is employed, where "a researcher specifies a number of alternative a priori models and fits each model to the same dataset" (MacCallum, 1995, p. 34). The main purpose of the model comparison approach is to examine whether one or several alternative models are "more consistent with the data" (MacCallum, 1995, p. 186). Therefore, the scales can be structured into several different models, namely: (1) a single-factor model, when all the observed variables load onto a single-factor; (2) an N-orthogonal factor model, when the observed variables load onto two or more uncorrelated factors; (3) an N-correlated factor model, when the observed variables load onto two or more correlated factors; (4) an hierarchical factor model, when two or more single-factor models reflect a higher order factor; and, (5) a nested factor model, when the observed variables reflect both a set of an N-orthogonal factor model as well as a single-factor model. The nested factor model is not used in this study as it is not possible to incorporate the nested structure in the subsequent analyses of path analysis and hierarchical linear modelling analysis. Once the alternative models are tested, the results are then evaluated and compared with respect to the goodness-of-fit (details of goodness of fit is presented in Appendix H).

3.9.2. Rasch Measurement

The Rasch model is used in this study to analyse and scale both the mathematics test and attitudinal questionnaire items. It is used in the validation process to examine the factor structures of a set of items which have been previously been examined by CFA. Rasch analysis complements the CFA in the validation process as it enables further data analysis, such as item fit (Masters & Keeves, 1999).

The Rasch model involves item response theory (IRT), and is a method used widely to analyse and scale both test and attitudinal data (Alagumalai & Curtis, 2005). IRT is seen as an improvement over classical test theory (CTT) which is based on the true score theory where the raw score was the combination of the true score and the measurement error (Alagumalai & Curtis, 2005). The major drawbacks of CTT include the limited possibility to compare test items and examinees when: (a) different tests that investigated the same construct were given to a different cohort of students; (b) the ambiguity of the raw score with the limitation of comparing student ability; and (c) the dependency of the observed and true score for the test that resulted in the fluctuation of a score based on the modification of test difficulty (Alagumalai & Curtis, 2005). Therefore, CTT is not used in this study.

Rasch scaling is seen as a technique to ensure a more rigorous instrument for measuring a certain construct (Boone, Staver, & Yale, 2014). The purpose of the Rasch model is to: “measure the ability of a person in a domain by testing it against a set of items of calibrated difficulty, so that the person’s ability can be placed appropriately on the scale” (Stacey & Steinle, 2006, p.77). Furthermore, Bond and Fox (2015) noted that the key feature of the Rasch model enabled the examination of the possibility of a person with certain ability to solve an item with identified difficulty correctly. Wu and Adams (2007) described the Rasch model as a particular IRT model which applies the transformation process of a raw score in order to maintain that interval between two people was not dependant on the specific items tested. Rasch scaling could then be used in the analysis of data using the Rasch model. One of the requirements of a Rasch analysis is unidimensionality (Bond & Fox, 2015; Stacey & Steinle, 2006).

The Rasch model was originally formulated for dichotomous items, while now it also includes the rating scale, partial credit, facets models, the Saltus model and the unfolding model (Alagumalai & Curtis, 2005). In Rasch modelling, a mathematical function is applied to illustrate the probability of a person’s response to an item, as a function of that person’s ability level. This probability function is referred to as an item characteristic curve (Wu & Adams, 2007). The original Rasch model, involving an item characteristics curve for a dichotomous item, was created by Rasch (1960). The model is given below by the exponential function:

$$p = P(X = 1) = \frac{\exp(\theta - \delta)}{1 + \exp(\theta - \delta)} \qquad [3.1]$$

Where,

P is the probability that the person provides a correct response to an item which depends on the person’s ability and the item difficulty;
X is the success or failure of the item demonstrated by a random variable; success is recorded by X=1 and failure is recorded by X=0;
θ represents the ability of the person on the latent variable scale; and

δ represents the item difficulty on the same latent variable scale.

The Rasch model for polytomous items was later created for the partial credit scale (Masters, 1982) and the rating scale (Andrich, 1978) derived from the original equation (3.1). Andrich (1978) argued that each threshold of the responses discriminated equally for all items in the rating scale model. In this study, the rating scale model is employed for examining the questionnaires which use Likert-scale responses such as ‘strongly disagree’, ‘disagree’, ‘agree’, and ‘strongly agree’. The partial credit model is also employed in this study for the mathematics test. The partial credit model is an alternative to Andrich’s rating scale model which was developed and was based on the difficulty level of the steps in every single item (Masters, 1982). Allowing the steps to discriminate in numbers and structure in the partial model result in alternative estimates of a person’s ability (Masters, 1982). The mathematics test administered in this study does not use a fixed response. Instead, each item has different independent steps. Therefore, the partial credit model is used. Both the rating scale and the partial credit model employed in this study are carried out using ACER Conquest 2.0 software (Wu et al., 2007).

3.9.3. Structural Equation Modelling (SEM)

SEM is a form of multivariate analysis which, using the application of statistical methods, concurrently examines causal relationships between variables. Grace (2006, p.12) defined SEM as “modelling hypotheses with structural equations”, while Byrne (2013, p. 3) provided a clearer definition of SEM as “a statistical methodology that takes a confirmatory (i.e., hypothesis-testing) approach to the analysis of a structural theory bearing on some phenomenon”. SEM combines regression, factor analysis, statistical modelling, and model evaluation (and requires the selection of software) (Grace, 2006). There are two emphases in SEM: (a) examining causal relationships between variables which are indicated by a series of structural equations (Byrne, 2013; Grace, 2006); and (b) the graphic presentation through path analyses of the structural relationships which allows a better understanding and conceptualisation of the framework tested (Byrne, 2013). The main goals of SEM analysis are to examine a theoretical model hypothesised by incorporating both observed variates and unobserved variables and to provide a thorough understanding of the relationships between the constructs under study (Schumacker & Lomax, 2012).

While SEM analysis is considered confirmatory, path is exploratory, which means that the interrelationships between variables have previously been determined in SEM. SEM allows the model to include both latent variables and observed variates (Byrne, 2013). Latent (or unobserved) variables cannot be measured directly so they are measured using a set of observed variates. SEM consists of two sub-models: a measurement model and a structural model. A measurement model hypothesises relationships between latent variables and observed variates by connecting the scores of the observed variates and the latent constructs that they are designed to measure. A structural model examines the relationships between the latent variables by estimating the extent of one latent variable's direct or indirect influence on other latent variables (Byrne, 2013).

SEM analysis is conducted and based on the five sequence steps as suggested by Schumacker and Lomax (2012): (a) model specification, (b) model identification, (c) model estimation, (d) model testing and (e) model modification. Model specification is the initial step where the researcher builds the theoretical model based on relevant theory and information from previous research. Specifying the relationship between variables in the model is also part of model specification. Model identification is the expansion of the previous model when the researcher identifies the model on the basis of whether all parameters could be estimated by the sampled data. There are three possibilities in model identification: (a) a model could be under-identified when there is not sufficient information in the matrix S to estimate one or more parameters; (b) a model could be just-identified when there is just sufficient information available in the matrix S to estimate all parameters; and (c) a model could be over-identified when there is a surplus of information producing more than one way of estimation for a parameter or parameters. Model estimation is the middle step where estimating the parameters in the model is carried out. This step involves the effort to reduce the difference between $\sum$ and S including unweighted or ordinary least squares (ULS or OLS), partial least squares (PLS), generalised least squares (GLS) and maximum likelihood (ML). Model testing is a way of examining whether the data fits the model and how the sampled data fits the theoretical model specified. The data fit might be examined using the general types of fit indices for the whole model as well as the individual parameters within the model. Model modification is then the final step, where the researcher might need to modify the original model if the model has a poor fit. The modification of the model should not only be based on the better fitting model but also on the more meaningful model (Schumacker & Lomax, 2012).

3.9.4. Path Analysis

Path analysis is a procedure for structural equation modelling (SEM) that examines and tests the structural model involving the hypothesised causal modelling of manifest variates (Kline, 2011). It was initially developed in 1918 by the American geneticist Sewall Wright and the techniques have since been further developed far beyond its initial usages (Hoyle, 2012; Kline, 2011). Path analysis involves graphic presentation (which is referred to as a path diagram) and a path model (which is referred as to the development and estimation of the path coefficients in the model) (Loehlin, 1998). A path model usually consists of circles or ovals for the representation of latent variables and rectangles for representing manifest variates (Hair, Hult, Ringle, & Sarstedt, 2013). Path analysis is performed to examine the strength of the path displayed in the path model, the strength of the path being the path coefficient (Hair et al., 2014). The model development of path analysis involves three major steps: (a) creating a path diagram, (b) identifying the structural model and (c) specifying the measurement model.

Partial Least Square Path Analysis (PLS-PA)

Partial least squares path analysis (PLS-PA) is a SEM procedure, involving partial least squares (PLS) theory to examine causal relationships between variables and observed or manifest variates (Hair et al., 2013). PLS was originally created to allow theoretical conclusions to be drawn from rich-datasets (Wold, 2006). The term ‘PLS-PA analysis’ is also referred to as PLSPATH, PLS path modelling and PLS-SEM. However, in this study, PLS-PA is used throughout. PLS-PA employs variance-based SEM, for estimating the path coefficients by maximising the variance explained of the dependent variables (Hair et al., 2013), as opposed to a covariance-based path analysis which operates by maximising the likelihood of a covariance matrix. PLS-PA is “primarily used to develop theories in exploratory research” (Hair et al., 2013, p.4). Wold (2006) described PLS as a flexible procedure, broadly applicable for the data analysis with a diverse set of acceptable data. The data could be nominal, ordinal or categorical and have the capacity to work with both large and small sample sizes. PLS-PA involves predictive logic, meaning that it is able to aim towards maximising prediction rather than the maximum likelihood of explanatory power and the “statistical accuracy of estimates” (Vinzi, Trinchera, & Amato, 2010, p.48). However, PLS-PA does not replace the covariance-based path analysis but rather complements it, with each method compensating for the other's drawbacks (Hair, Ringle, & Sarstedt, 2012). The advantages of PLS-PA are not restricted to the normal distribution of the data and its capacity to incorporate both formative and reflective variates in the measurement models (Chin, 1998b; Hair et al., 2013). PLS-PA also allows the inclusion of both predictive and causal models in the analysis, this being one of the primary reasons for its use in this study.

As one of the features of SEM, the path analysis modelling procedure, including PLS-PA, has two sub-models: the measurement model and the structural model. The relationships between each unobserved or latent variable and the corresponding observed or manifest variates are accounted for in the measurement model and the relationships between unobserved or latent variables are employed in the structural model (Vinzi et al., 2010). The structural model is also referred to as the inner-model, where a path model is formed by the linear relationships between latent variables and the measurement model, also referred to as the outer-model (Lohmöller, 2013).

The analysis begins with drawing a path diagram consisting of circles or ellipses representing the unobserved or latent variables and rectangles representing the observed or manifest variates. The relationships between the latent variables in the inner model and the latent variables and manifest variates in the outer model are represented by one-directional arrows. The outer measurement model may involve either: (a) 'formative measurement', where the one-directional arrow representing influence is coming from the manifest variates to the construct or latent variable; or (b) 'reflective measurement', where the one-directional arrow representing influence is coming from the construct or latent variable to the manifest variates. There are some main differences between reflective and formative measurement models. The manifest variables are explained by the construct in the reflective model and the construct of the formative model is formed by the manifest variables (Chin, 1998a). Furthermore, the manifest variables of the reflective measurement model are interchangeable and the elimination of one variable does not change the meaning of the construct. However, formative measurement models consist of variables that are not interchangeable and the elimination of one variable might change the meaning of the construct (Jarvis, MacKenzie, & Podsakoff, 2003).

The inner structural model is made up of the exogenous and endogenous variables. An exogenous variable is a variable that acts as an independent variable with the directional arrow coming from the variable. An endogenous variable is a variable that acts as either a dependent variable or both as a dependent and independent variable, indicating that the variables may have the directional arrows either coming to or to and from the variable. The causal influence in a path diagram is usually illustrated from the left to the right, starting from the exogenous variables. The requirement of PLS-PA algorithms is recursive, meaning that a reciprocal or loop relationship between the latent variables in the structural model is not acceptable. This influence in the relationship operates only in one direction. In the path analyses carried out in this study, SmartPLS 3.2.6 (Ringle et al., 2015), is employed.

Evaluating the Results of PLS-PA

One of the acknowledged limitations of PLS-PA is the lack of established fit indices. According to Henseler and Sarstedt (2013, p.56), “the lack of a global scalar function and the consequent lack of global goodness-of-fit measures has long been considered a drawback of PLS path modelling”. This absence of a well identified global optimisation is the reason why PLS-PA has no global fit indices to examine how well the model worked or operated (Vinzi et al., 2010). In fact, PLS-PA operates by maximising the variance in order to allow prediction. Thus, the validation of the model is more about the model's predictive capacity (Vinzi et al., 2010). Vinzi et al. (2010) outlined how the validation of PLS-PA is examined in the measurement model, the structural model, and the overall model. Consequently, PLS-PA requires three fit indices including the communality, the redundancy and the goodness-of-fit (GoF) index. Tenenhaus, Amato, and Vinzi (2004) suggested that a global GoF index for PLS-PA, which catered for the redundancy and communality, is the equivalent of the chi-square test. The GoF index, the geometric mean of the average communality index and the average of the R^2 value are created to evaluate the operation of both the measurement and structural models, as well as the overall power model (Vinzi et al., 2010). Furthermore, a modified version of GoF, called relative GoF or $GoF_{rel,}$ was also developed. The GoF_{rel} is acquired by comparing "the communalities obtained from PLS with the communalities obtained from a principal component analysis, and the R^2 values obtained from PLS with the R^2 values obtained from a canonical correlation analysis" (Henseler & Sarstedt, 2013, p. 571). The values of both GoF and GoF_{rel} range from 0 to 1, with a value closer to 1 indicating a good performance of the model. However, it is not possible to examine the statistical significance of both values (Vinzi et al., 2010).

Both the GoF and GoF_{rel} have limitations. The GoF is used for the reflective measurement model (Henseler & Sarstedt, 2013). While Vinzi et al. (2010) claimed that it is also possible to use the GoF for formative measurement by interpreting it differently from reflective measurement. Henseler and Sarstedt (2013) suggested not to use GoF for formative measurement, explaining that using the formative model is not the main purpose of PLS-PA. Another limitation of the GoF is that it might only operate for path models with endogenous latent variables (Henseler & Sarstedt, 2013). This limitation is overcome by the development of the GoF_{rel} (Vinzi et al., 2010).

However, the simulation of data conducted by Henseler and Sarstedt (2013) showed a non-successful result for both GoF and GoF_{rel} in distinguishing between the valid and non-valid model. Therefore, neither GoF or GoF_{rel} are appropriate indices for indicating the strength of a model. However, they might be used for judging how well the path model explains the different set of data (Henseler & Sarstedt, 2013).

In conclusion, while the covariance-based path analysis has global fit indices, it is acknowledged that there is no single goodness-of-fit criterion for evaluating the PLS-PA. As Hair et al. (2013) made clear, the meanings of the term 'fit' in the PLS-PA and in the covariance-based path analysis are different:

> fit statistics for CB-SEM are derived from the discrepancy between the empirical and the model implied (theoretical) covariance matrix, whereas PLS-SEM focuses on the discrepancy between the observed (in the case of manifest variables) or approximated (in the case of latent variables) values of the dependent variables and the values predicted by the model in question (Hair et al., 2013, p. 96).

The evaluation as to whether a model is good in PLS-PA is mainly based on measures related to the model's strength to predict or explain. In particular, the evaluations of the measurement and structural models are based on a set of nonparametric evaluation using techniques such as bootstrapping and blindfolding (Efron & Tibshirani, 1994; Hair et al., 2013).

When conducting a PLS-PA, especially when using SmartPLS 3.2.6 (Ringle et al., 2015), there are several requirements for evaluating both the measurement and structural models. In this research, it is proposed to employ a systematic evaluation of the results of PLS-PA in two stages: the evaluation of the measurement model (stage 1) and the evaluation of the structural model (stage 2). The evaluation of the measurement model is divided into the evaluations of reflective and formative measurement models. The evaluation of the reflective measurement model involves internal consistency (composite reliability), indicator reliability, convergent validity (average variance extracted) and discriminant validity. Convergent validity, colinearity among indicators as well as significance and relevance of outer weights are all required for the evaluation of the formative measurement model. The assessment of structural models involves reporting the coefficients of determination (R^2), predictive relevance (Q^2), size and significance of path coefficients, f^2 effect sizes and q^2 effect sizes. The detailed guidelines for evaluating the PLS-PA measurement and structural models, adopted from (Hair, Sarstedt, Ringle, & Mena, 2012), are presented in Table 3.5.

Table 3.5

Guidelines of Evaluating PLS Path Models (Hair, Sarstedt, et al., 2012)

Criterion	Recommendation and reference
Outer model evaluation: reflective	
Indicator reliability: The square of a standardised indicator's outer loading that represents how much of the variation in an item is explained by the construct	Standardised indicator loadings ≥ 0.70; or 0.40 for exploratory studies are acceptable (Hulland, 1999)
Internal consistency reliability**:** One of the reliability indicators that is used to indicate the consistency of results across items on the same test	Composite reliability ≥ 0.70; 0.60 is also acceptable for exploratory research (Bagozzi & Yi, 1988)
Convergent validity: The extent to which a measure correlates positively with alternative measures of the same construct	AVE ≥ 0.50 (Bagozzi & Yi, 1988)
Discriminant validity (Fornell-Larcker criterion): The extent to which a construct is truly distinct from other constructs, in terms of how much it correlates with other constructs	The AVE's for each construct should be higher than its squared correlation with any other construct (Fornell & Larcker, 1981)
Cross loadings	The loading of each indicator should be higher in the intended construct compared to any other construct (Chin, 1998b; Grégoire & Fisher, 2006)
Outer model evaluation: formative	
Indicators' relative contribution to the construct	Report indicator weights
Significance of weights	Report t-values, p-values or standard errors
Multicollinearity: The condition where two or more indicators are highly correlated	VIF < 5; Tolerance > 0.20; Condition index < 30 (Hair et al., 2014)
Inner model evaluation	
R^2	The acceptable level is based on research context (Hair et al., 2014)
Effect size f^2	0.02 is weak, 0.15 is medium and 0.35 is strong effects (Cohen, 1988)
Path coefficient estimates	Provide confidence interval through bootstrapping (Chin, 1998b; Henseler, Ringle, & Sinkovics, 2009);
Predictive relevance Q^2 ***and*** q^2: A measure of predictive relevance based on the blindfolding technique to accurately predict the data points of indicators in reflective measurement models of endogenous constructs	Q^2 0.02 is weak, 0.15 is moderate and 0.35 is a strong degree of predictive relevance; the predictive relevance is obtained by blindfolding (Chin, 1998b; Henseler et al., 2009)

3.9.5. Hierarchical Linear Modelling (HLM)

Hierarchical Linear Modelling (HLM) refers to a sophisticated statistical technique that allows an analysis of variance for cross-level variables. HLM is defined by two characteristics: (a) the structure of the data used for the model is hierarchical, meaning that the first level variables are nested within the second level variables and the second level variables are nested within the third level variables (in the case of three-level hierarchical model); and (b) the framework for the models is considered as a hierarchical linear structure (Raudenbush, 1993). HLM particularly fits the structure of most educational data which is hierarchical, where students are nested within classrooms, and classrooms are nested within schools (Cheung & Keeves, 1990).

The major advantages of HLM is in addressing the problems caused by the disaggregation and aggregation processes. When data is hierarchically structured, the most common method applied prior to HLM data analysis was bringing the data into a one level structure by disaggregating or aggregating the data (see section 3.8.3). However, such a method creates bias and results in serious errors in estimating the precision and the meaning of the unit of analysis (Bryk & Raudenbush, 2002). Furthermore, such a technique also resulted in some potentially meaningful effects at certain levels to be dismissed (Hofmann, 1997). A further advantage of HLM is that it provides the partitioning of the variance of the lower level and upper level, resulting in more accurate and meaningful estimates (Cheung & Keeves, 1990). HLM is aimed at the analysis of data within each hierarchical level, as well as at examining the relationships and effects between variables across various levels (Bryk & Raudenbush, 2002). It enables the examination of how the higher level variables influence the lower level outcomes while ensuring the appropriate level of analysis (Hofmann, 1997).

HLM analysis involves both the fully unconditional or null model, and the conditional model of a two-level, and three-level model (depending on the number of levels of the data). A fully unconditional model is the simplest model without any predictors at each level, to estimate the available variance in an outcome measured across all levels (Bryk & Raudenbush, 2002). The conditional model is the following model, with predictors at each level, from the lowest to the highest level, to examine the predictors at each level that explains the variability obtained across different levels of the unconditional model (Bryk & Raudenbush, 2002).

In a two-level model, where level-1 is the student level and level-2 is the school level, the level-1 model and level-2 model can be shown as the following regression model equations:

Level-1 model

$$Y_{ij} = \beta_{0j} + \sum \beta_{qj} X_{qij} + r_{ij} \qquad [3.3]$$

Where, Y_{ij} is the outcome for students i in school j , X_{qij} is the students' characteristics and r_{ij} is the random error. β_{oj} is the intercept and β_{qj} is the regression coefficient indicating the magnitude of effects between each X_{qij} and the outcome in school j.

Level-2 model

$$\beta_{qj} = \theta_{q0} + \sum \theta_{qs} W_{sj} + u_{qj} \qquad [3.4]$$

Where, β_{qj} is the outcome variable, W_{sj} is the school characteristics and u_{qj} is the random error. θ_{q0} is the intercept and θ_{qs} is the regression slope indicating the magnitude of effects between each W_{sj} and the outcome β_{qj}.

HLM fits the structure of data in this present study and is, therefore, employed. The analysis is carried out using HLM 6.08 software (Raudenbush et al., 2004).

3.10 Validity and Reliability of Instruments

In order to ensure robust research, the standard of validity and reliability needs to be high. Bryman (2012) pointed out that validity is concerned with whether the range of indicators created to measure an underlying construct operates with strength to examine the intended construct. Validity involves an investigation into whether the scores of the instrument could be used to derive meaningful and strong inferences. Cronbach (1946, p.475) had argued that a valid test was when "what the test measures is determined by the content of the items". Later, Messick (1987, p.1) defined validity as "an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness and actions based on test scores". In this study, validity involves the strength of a range of items in measuring the proposed constructs as indicated by the validity score to assist the researcher draw a significant and practical conclusion.

Validity can be assessed in several ways, including content or face validity, construct validity, concurrent or criterion validity, predictive validity and convergent validity. Face validity involves using someone's expertise to check whether the items of the instrument measure the proposed construct (Bryman, 2012; Heale & Twycross, 2015). The construct validity is assessed by investigating the relationships between the constructs under study based on the relevant concepts (Bryman, 2012; Roberts, Priest, & Traynor, 2006). Construct validity of the instrument is concerned with the meaningful inferences that could be drawn from the validity test score (Heale & Twycross, 2015). Criterion validity is assessed by examining the correlations of the constructs in the intended instrument with constructs in other instruments that examined the same variable (Bryman, 2012; Heale & Twycross, 2015). Criterion validity could be assessed in three different ways: (a) convergent validity, indicating that the construct in the intended instrument is highly correlated to the same construct in another instrument developed in a different way; (b) divergent validity, indicating that the construct in the intended instrument is weakly correlated to the different construct in another instrument; and (c) predictive validity, indicating that the construct in the intended instrument is highly correlated to a future criterion measure (Bryman, 2012; Heale & Twycross, 2015). However, finding the appropriate instrument with a matching criterion is a major problem in employing the criterion-related validity (Darmawan, 2003). Therefore, this study employs two types of validity assessment: face and construct validity. The construct validity is conducted using CFA and Rasch analysis.

When referring to reliability, Cronbach (1947, p. 2) argued that it was “a concept which could not be directly observed” . Reliability was defined as “the coefficient of stability and equivalence, and hypothetical self-correlation” (Cronbach, 1947, p. 5) Reliability also relates to the examination of measurement error, namely the random error (Guion, 2002). A reliability estimate could be obtained in several ways, including internal consistency, stability and equivalence. The internal consistency could be examined by item-to-total correlation, split-half reliability, Kuder-Richardson coefficient and Cronbach alpha. The split-half reliability was developed by Spearman, and was obtained by using two different scores from a single testing by splitting the odd and even number items and then correlating the two scores (Cronbach & Shavelson, 2004). The Kuder-Richardson coefficient reliability was developed and based on Spearman split-half reliability, but was only applicable for an item with a zero and one scoring (Cronbach & Shavelson, 2004). Among the four internal consistency reliability techniques, Cronbach alpha is the most frequently used and reported. The Cronbach alpha is a reformulation of a Kuder-Richardson coefficient that provides an average of all split-half coefficients obtained from any possible split of a test (Cronbach, 1951). Despite the wide application of Cronbach alpha, Cronbach and Shavelson (2004) stated that Cronbach alpha is not necessarily the best way of examining reliability. They emphasised that their current view (in 2004) of the Cronbach alpha was "to cover only a small perspective of the range of measurement uses for which reliability information was needed and was now seen to fit within a much larger system of reliability analysis" (Cronbach & Shavelson, 2004, p. 416). Therefore, the Cronbach alpha is not used in this study. This study employs Rasch analysis to evaluate the reliability of the instrument. Rasch analysis provides a number of reliability indices including person and item reliability, as well as person separation and item separation reliability with the value closer to one, making it a more internally consistent measure (Boone et al., 2014).

3.11 Summary

The focus of this study is to investigate the relationships between student-, teacher-, and school-level factors which have an impact on students' mathematics performance related to LOT and HOT. This chapter describes and justifies the research methods employed in this study. Three sets of questionnaires (student, teacher and school) and a mathematics test for students are developed for investigation. The student and teacher questionnaires have been developed to seek information related to students' or teachers' background, as well as information related to attitudes, beliefs and classroom practices. The items in both student and teacher questionnaires are the combination of items adapted from previously used scales and items developed specifically for this study. The school questionnaire is made up of information related to school demographics. All items in the mathematics test are from PISA and TIMSS released items that are chosen and are based on the appropriateness of the items for the purposes of the study concerning mathematics performance related to LOT and HOT.

The pilot study was carried out prior to the actual data collection. The results of the pilot study were used to make some amendments to the original questionnaires and mathematics test.

The sample in this study is Year 9 students and mathematics teachers in the province of Aceh, Indonesia. Stratified purposive sampling is employed to select 25 schools. The schools chosen are from one city representing an urban area and one district representing a rural area. The total samples involved in the study are 1135 Year 9 students, 46 Year 9 mathematics teachers and 25 schools. All sets of the questionnaires and mathematics test were translated into Bahasa Indonesia after their final development. Back-to-back translation and face validity were carried out for all the instruments before the pilot testing. The ethics approval was obtained from the University of Adelaide. The approval for conducting the study at the randomly chosen schools was also sought from the Department of Education and Culture as well as the Department of Religious Affairs in the province of Aceh, Indonesia.

A number of statistical techniques are employed in this study. The descriptive analysis of the data is carried out using the IBM SPSS 22. CFA, using the IBM SPSS Amos 22, and Rasch analysis, using Conquest 2.0 software (Wu et al., 2007), are employed to validate the scales. PLS-PA is undertaken to examine relationships between variables at the student and the teacher levels using SmartPLS 3.2.6 (Ringle et al., 2015). The analysis of the relationship across three levels of data (student, teacher and school) is also carried out by employing HLM. The following chapter presents the validation procedures conducted preceding the analysis and examination of the causal models that are hypothesised for investigation.

Chapter 4 Research Instruments: Students

4.1 Introduction

This study investigates the student-, teacher- and school-level factors, their interrelationships and their impacts on students' mathematics performance related to LOT and HOT. In order to provide accurate estimates for the subsequent analyses, the reliability and validity of the instrument measuring each construct needs to be established. In this chapter, the validation of the instruments involved in measurement are presented. The validation processes are performed using confirmatory factor analysis (CFA) and the Rasch analysis.

CFA is carried out to examine the construct validity and reliability of each scale. It is conducted to confirm the factor structure as well as the relationship between the observed variable and its respective underlying traits. CFA also enables the examination of the unidimensionality of the scales that is required for the Rasch analysis (Uher et al., 2008). Rasch analysis is conducted to examine whether the structure, which has been previously validated through CFA, fits the Rasch model. The internal consistency of the scales is measured by the separation reliability and the expected a posteriori/plausible value (EAP/PV) separation reliability of the Rasch analysis. The factor structures to be used in the subsequent analyses are then made.

The student questionnaire is in three parts. Part 1 consists of personal background questions seeking information on parents' occupations and education, home possessions and educational expectations. Part 2 consists of questions investigating students' attitudes, values and beliefs concerning mathematics. Part 3 covers students' views and perceptions related to their mathematics classroom practices, including learning resources and the types of mathematical problems used. There are 1135 students in the data set.

4.2 Confirmatory Factor Analysis (CFA)

CFA is conducted in order to confirm the structure of the relationship between the observed variable and their underlying traits. It is carried out using the IBM SPSS Amos 22. One of the advantages of using this software is the highly developed graphical interface for drawing the model. The software is able to perform extensive bootstrapping and administered analyses for datasets with missing data through employing a special maximum likelihood procedure (Kline, 1998).

In this study, a model comparison approach is employed to identify the best structure of the scales. Several alternative models are tested and the results are compared. Four different models are examined through CFA: a one-factor model; an N-orthogonal factor model; an N-correlated factors model; and a hierarchical model. The fit for each factor model is reported and the factor model that has the best fit is then included in further analyses of the measurement model.

It needs to be noted that the terms 'latent variable', 'construct' and 'trait' are interchangeable, as are the terms 'observed variate', 'manifest variate' and 'indicator variate'.

4.2.1. Construct validity, reliability and model fit indices

CFA enables the assessment of the construct validity and reliability of the scales. The construct validity includes convergent and discriminant validity. The convergent validity is established through various components: (a) the loadings of each observed variable onto its construct; (b) the average variance extracted (AVE) for each construct; and (c) the construct reliability. The cut-off value of the factor loadings used in this study is 0.32 (Tabachnick & Fidell, 2013). An AVE of 0.50 or greater indicates adequate convergent validity (Hair et al., 2014). The construct reliability is indicated by the composite reliability, with a value between 0.60 and 0.70 as acceptable and a value greater than 0.70 regarded as good reliability (Hair et al., 2014).

CFA also enables the examination of discriminant validity in a model-comparisons approach (Brackett & Mayer, 2003; Darmawan, 2017) involving competing CFA models. The measurement model is evaluated on its model fit. Model fit indices may be used as criteria for judging the statistical significance and the meaningfulness of the parameter estimates of the CFA results. There are several fit indices that are widely used: chi square (χ^2), root-mean-square error of approximation (RMSEA), comparative fit index (CFI), goodness-of-fit-index (GFI), adjusted GFI (AGFI), root-mean-square residual (RMR), standardised RMR (SRMR), Tucker-Lewis index (TLI), normed fit index (NFI), parsimony fit index (PNFI) and Akaike information criterion (AIC). The model fit indices are used as guidelines to evaluate whether the model fits the observed data. While there are numerous model fit indices, each with its own benefits and drawbacks (a detailed description of each model fit index is presented in Appendix H), this study will report the chi-square, RMSEA, CFI and PGFI as the criteria for model fit (see Hooper, Coughlan, & Mullen, 2008). However, TLI is also included in this study because of its power and robustness, as suggested by Iacobucci (2010). A summary of model fit indices, the acceptable ranges and the interpretation, as devised by Schumacker and Lomax (2012), is presented in Table 4.1.

Table 4.1

Summary of Model Fit Indices, Acceptable Range and Interpretation (Schumacker & Lomax, 2012, p.76)

Model fit indices	Acceptable range	Interpretation
χ^2	Based on the χ^2 table value	Compares the obtain χ^2 value with tabled value for given *df*
GFI	0 (no fit) to 1 (perfect fit)	Values close to 0.90 or 0.95 indicate a good model fit
AGFI	0 (no fit) to 1 (perfect fit)	Values adjusted for *df* Value close to 0.90 or 0.95 indicate a good model fit
RMR	Depending on the researcher	Indicates the closeness of ∑ to S matrices
SRMR	< 0.05	Values less than 0.05 indicate a good model fit
RMSEA	0.05 to 0.08	Values close to 0.5 to 0.08 indicate close fit
TLI	0 (no fit) to 1 (perfect fit)	Values close to 0.90 or 0.95 indicate a good model fit
NFI	0 (no fit) to 1 (perfect fit)	Values close to 0.90 or 0.95 indicate a good model fit
PNFI	0 (no fit) to 1 (perfect fit)	Compares values in alternative models
AIC	0 (perfect fit) to positive value (poor fit)	Compares values in alternative models

4.3 Rasch Analysis

Rasch analysis is conducted in order to examine whether the structure, which has been previously validated in CFA, fits the Rasch model. The Rasch analysis conducted in this study is based on the data from 1135 student respondents. ACER Conquest 2.0 software (Wu et al., 2007) is employed to run the analysis using the rating scale model. Rasch analysis has criteria to evaluate the item fit statistics. Infit mean square (INFIT MNSQ) and outfit mean square (OUTFIT MNSQ) are the two main criteria used for examining the fit of items to the Rasch scale. The INFIT MNSQ assesses whether the fit of the respondents is uniform with respect to the item characteristic curve for each item (Afrassa, 2005) and with consideration of the item difficulty and person ability (Boone et al., 2014). The OUTFIT MNSQ assesses whether the fit of the respondents is uniform with the item characteristic curve for each item (Afrassa, 2005) after putting more emphasis on extreme cases or outliers (Boone et al., 2014).

The acceptable ranges of INFIT and OUTFIT MNSQ have been established by several researchers. Adams and Khoo (1993) advised a strict cut-off limit for the fit, within the range from 0.70 to 1.30. Values which fell outside the acceptable ranges are a sign of misfitting (above 1.3) and redundancy (below 0.7) (Afrassa, 2005). Other researchers suggest that the acceptable range is between 0.60 to 1.40 for a Likert scale (survey) (Bond & Fox, 2015; Wright & Linacre, 1994). This study adopts the range suggested by Bond and Fox (2015), with the acceptable range within 0.60 to 1.40.

The t-value is also used for examining the item fit, as suggested by Wright and Stone (1979) that a misfitting item being recognised when the t-value was greater than 5, while Boone et al. (2014) explained that the t-value could be ignored when the INFIT or OUTFIT MNSQ fell within the acceptable ranges. In this study, the INFIT and OUTFIT MNSQ as well as the t-value will be used to assess the item fit statistics.

4.4 Student Attitudes towards Mathematics (SAM) Instrument

The items assessing Student Attitude toward Mathematics (SAM) used in this study are adapted from the TIMSS 2011 student questionnaire. SAM consists of nine items, each using a four-point Likert-scale response of ‘disagree a lot’, ‘disagree a little’, ‘agree a little’ and ‘agree a lot’, coding being 1, 2, 3, and 4. Those items that are negatively worded are recoded in reverse for the purposes of the analysis.

Table 4.2

Student Attitudes towards Mathematics (SAM) Instrument Subscales

Subscales	Item label	Nature of statement	Reverse scoring	Item text
Liking of Mathematics (LIKE_MATH)	ATM_A	Positive	None	I enjoy learning mathematics
	ATM_B	Negative	ATM_BR	I wish I did not have to study mathematics
	ATM_C	Negative	ATM_CR	Mathematics is boring
	ATM_D	Positive	None	I learn many interesting things in mathematics
	ATM_E	Positive	None	I like mathematics
Valuing of Mathematics (VALUE_MATH)	SVM_A	Positive	None	I think learning mathematics will help me in my daily life
	SVM_B	Positive	None	I need mathematics to learn other school subjects
	SVM_C	Positive	None	I need to do well in mathematics to get into the university or college of my choice
	SVM_D	Positive	None	I need to do well in mathematics to get the job I want

The nine items assessing SAM is made up of two subscales and are presented in Table 4.2. Five items involve Students' Liking of Mathematics (LIKE_MATH) and four items involve Students' Valuing of Mathematics (VALUE_MATH). Table 4.1 presents the nine items of the SAM including the subscales, item labels, the nature of the items, and item labels after recoding. The wording of the statement is also displayed.

4.4.1. Confirmatory Factor Analysis (CFA) of the SAM

The CFA for the nine items of the Student Attitude toward Mathematics (SAM) scale is undertaken by testing a series of models. The models tested are: a single-factor model; a two-orthogonal factors model; a two-correlated factors model; and a two-hierarchical factors model.

The first model tested for SAM is a single-factor model involving all nine items included in one construct named 'Student Attitude toward Mathematics' (SAM) (see Figure 4.1). The second model is a two-orthogonal factors model with items classified into two uncorrelated constructs (see Figure 4.1). Figure 4.2 shows the models, with the two-factor models that are correlated and then error terms that are also correlated. The errors of ATM_CR (e3) and ATM_BR (e4) are correlated, as well as the errors of SVM_B (e8) and SVM (e9). The correlated errors in this factor analysis is data-driven as suggested by modification indices.

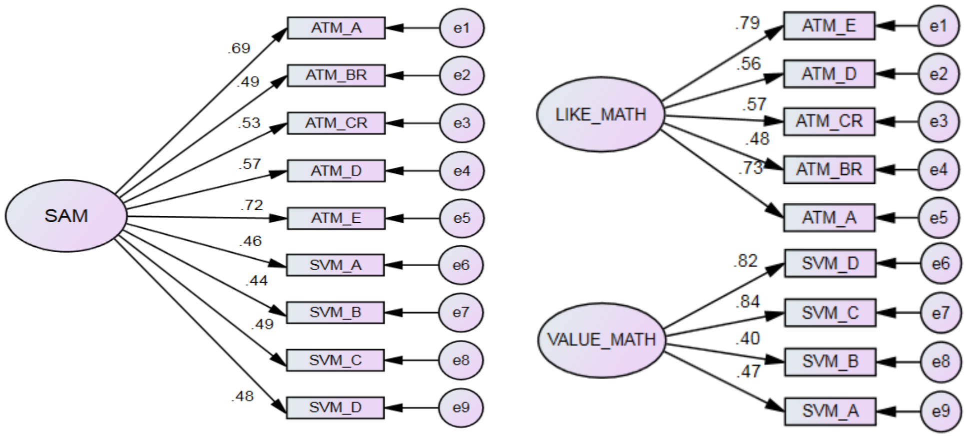


Figure 4.1 Single-factor model and two-orthogonal factors model of SAM

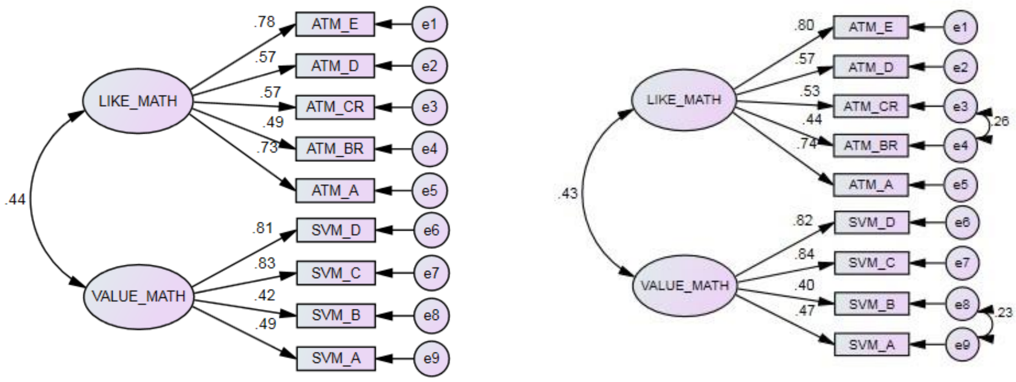


Figure 4.2 Two-correlated factors model and two-correlated factors model of SAM with two-correlated errors (residual terms)

The models shown in Figure 4.3 are two-hierarchical factors models. A hierarchical model structure involves two or more different factors that are presented to reflect a higher order factor (Curtis, 2005). Figure 4.3 shows the two-hierarchical factor model and the two-hierarchical factors model with correlated errors of SAM, reflecting the scale of SAM as the first or higher order and the subscales of Liking Mathematics and Valuing Mathematics as the second order.

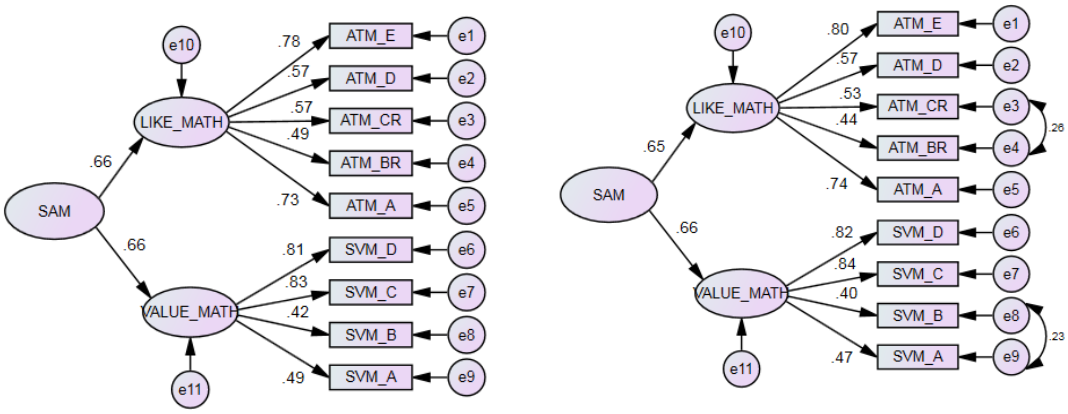


Figure 4.3 Two-hierarchical factors model and two-hierarchical factors model of SAM with two-correlated errors (residual terms)

The initial stage of evaluating a model involves examining the factor loading for each item, defined as the correlation between the original variates and latent variables. In this evaluation, the cut-off value for the factor loading is 0.32 (Tabachnick & Fidell, 2013). The estimated factor loadings of both the single-factor model and the two-factor models, including the hierarchical model, are presented in Table 4.3. Based on the criteria for evaluating the factor loadings, all items in the single-factor model and the two-factor models (including the two-orthogonal factors model, the two-correlated factors model, two-correlated factors model with errors correlated, two-hierarchical factors model as well as the two-hierarchical factors model with errors correlated) satisfy the requirement that all factor loadings exceed or are equal to 0.32. The factor loadings of the two-correlated factors model are the same as the loadings of the two-hierarchical factors model. Furthermore, the factors loadings for the two-correlated factors model with errors correlated (e3 and e4; e8 and e9) are the same as the loadings of the two-hierarchical factors model with the errors correlated (Figure 4.2 and Figure 4.3).

The factor loadings for the two-factor models are better than the loadings for the single-factor model. The correlations between the subscales (LIKE_MATH and

VALUE_MATH) are moderate in the two-factor model (correlated) and the two-factor model (correlated with errors) (0.44 and 0.43 respectively). In addition, the factor loading for the subscales (LIKE_MATH and VALUE_MATH) in the two-hierarchical factors model and the two-hierarchical factors model (with errors correlated) also fall within the acceptable range.

The next stage of evaluating the model is to examine the model fit based on the fit indices. The model fit indices for all models tested in Table 4.2 are presented in Table 4.3. The model fit for the two-orthogonal factors model is moderate, where the TLI value is 0.79 and is close to an acceptable value (a value close to or greater than 0.90 is acceptable). The CFI value is close to 0.90 which can be considered as acceptable. However, the RMSEA value is high (a RMSEA value less than or equal to 0.05 is acceptable).

The single-factor model has the poorest fit (TLI is less than 0.50, CFI is less than 0.70 and RMSEA is 0.17). The composite reliabilities are acceptable in all models. The single-factor model has the best composite reliability despite the fact that it has the poorest AVE. The model fit values of the two-correlated factors model and the two-hierarchical factors model are similar. Both have an acceptable fit value (TLI is close to 0.90, CFI is greater than 0.90 and RMSEA is 0.08). The best fitting models are the two-correlated and the two-hierarchical factors models with errors correlated (TLI and CFI values are greater than 0.9, and RMSEA value is 0.06). The two-correlated factors model is now to be examined in the Rasch analysis.

Table 4.3

Factor Loadings of Items, Average Variance Extracted (AVE) and Composite Reliability (CR) of Single-factor Model and Two-factor Models of SAM

Items	Subscales	Single-factor model			Two-factor models														
					Two-orthogonal			Two-correlated			Two-correlated (errors correlated)			Two-hierarchical			Two-hierarchical (errors correlated)		
		Loading	AVE	CR	Loading	AVE	CR	Loading	AVE	CR	Loading	AVE	CR	Loading	AVE	CR	Loading	AVE	CR
ATM_A	LIKE_MATH	0.49	0.30	0.79	0.73	0.41	0.77	0.78	0.41	0.71	0.74	0.40	0.76	0.78	0.41	0.71	0.74	0.40	0.76
ATM_BR		0.69			0.48			0.57			0.44			0.57			0.44		
ATM_CR		0.53			0.57			0.57			0.53			0.57			0.53		
ATM_D		0.57			0.56			0.49			0.57			0.49			0.57		
ATM_E		0.72			0.79			0.73			0.80			0.73			0.80		
SVM_A	VALUE_MATH	0.46			0.47	0.44	0.74	0.49	0.44	0.74	0.47	0.44	0.74	0.49	0.44	0.74	0.47	0.44	0.74
SVM_B		0.44			0.40			0.42			0.40			0.42			0.40		
SVM_C		0.49			0.84			0.83			0.84			0.83			0.84		
SVM_D		0.48			0.82			0.81			0.82			0.81			0.82		

Table 4.4

Summary of Model Fit Indices for Single-factor Model and Two-factor Models of SAM

Indices	Single-factor model	Two-factor Model				
	Single-factor model	Two-orthogonal Factor Model	Two-correlated	Two-correlated (correlated errors)	Two-hierarchical model	Two-hierarchical model (correlated errors)
Chi-square (χ^2)	874.46	325.26	224.21	106.36	224.21	106.36
Degree of freedom (*df*)	27	27	26	24	26	24
χ^2/df	32.39	13.90	8.62	4.43	8.62	4.43
Goodness of Fix Index (GFI)	0.84	0.93	0.96	0.98	0.96	0.98
Adjusted Goodness of Fix Index (AGFI)	0.74	0.89	0.92	0.96	0.92	0.96
Parsimony Goodness of Fix Index (PGFI)	0.51	0.57	0.55	0.52	0.55	0.52
Tucker-Lewis Index (TLI)	0.48	0.79	0.87	0.94	0.87	0.94
Comparative Fit Index (CFI)	0.69	0.88	0.93	0.97	0.93	0.97
Root Mean Square Error of Approximation (RMSEA)	0.17	0.11	0.08	0.06	0.08	0.06

4.4.2. Rasch Analysis of SAM

The Rasch analysis is carried out for the multidimensional SAM scale which corresponds to the two-correlated factor model, in order to examine to what extent the factor structure of the scale as validated in CFA fit the Rasch model. The summary of the response model parameter estimates of the scale, including the INFIT and OUTFIT MNSQ, confidence interval (CI) and *t* value, is presented in Table 4.5.

Table 4.5

Response Model Parameter Estimates of Two-correlated Factor Model of SAM

Items			Unweighted Fit			Weighted Fit		
N=1135	Estimates	Error	OUTFIT MNSQ	CI	*t*	INFIT MNSQ	CI	*t*
ATM_A	-0.08	0.04	0.79	(0.92, 1.08)	-5.2	0.78	(0.91, 1.09)	-5.1
ATM_BR	-0.57	0.04	1.19	(0.92, 1.08)	4.3	1.22	(0.91, 1.09)	4.6
ATM_CR	0.33	0.04	1.15	(0.92, 1.08)	3.5	1.11	(0.91, 1.09)	2.2
ATM_D	0.16	0.04	0.97	(0.92, 1.08)	-0.6	0.98	(0.91, 1.09)	-0.5
ATM_E	0.15	0.08	0.91	(0.92, 1.08)	-2.1	0.96	(0.91, 1.09)	-0.9
SVM_A	-0.20	0.04	0.98	(0.92, 1.08)	-0.6	1.01	(0.92, 1.08)	0.3
SVM_B	1.22	0.04	0.94	(0.92, 1.08)	-1.5	0.92	(0.91, 1.09)	-1.7
SVM_C	-0.64	0.05	0.94	(0.92, 1.08)	-1.5	1.06	(0.92, 1.08)	1.4
SVM_D	-0.38	0.08	1.02	(0.92, 1.08)	0.4	1.11	(0.92, 1.08)	2.4

Separation reliability =1,
Chi-square test of parameter equality = 1351.96
df = 7
Significance level = 0.00

The results of the Rasch analysis of the two-correlated factors model of SAM in Table 4.5 shows that the INFIT and OUTFIT MNSQ of all items fall within the acceptable range between 0.60 and 1.40. While the *t*-value of ATM_A is slightly greater than 5, the weighted fit estimates of the INFIT and OUTFIT MNSQ is within the acceptable range.

The Rasch analysis not only provides the INFIT and OUTFIT MNSQ but also the reliability values. There are two reliability values available from Conquest 2.0 software (Wu et al., 2007) namely, separation reliability and EAP/PV separation reliability. The separation reliability reported in this study refers to item separation reliability and the EAP/PV reliability refers to person reliability. Person separation indicates how efficiently a set of items is able to separate those persons measured while item separation indicates how well a sample of people is able to separate those items used in the test (Wright & Stone, 1999). The value of the reliabilities ranges from 0 to 1, where a value closer to 1 indicates a better separation and a more accurate measurement (Wright & Stone, 1999).

The separation reliabilities for the two-correlated factor models of SAM scale is high, the value being 1. While the EAP/PV reliability values for Liking Math subscale and Valuing Math subscale are 0.80 and 0.74 respectively, the values are all within the satisfactory range of the required benchmark of 0.7 or above (Jähnig, 2013).

4.4.3. Summary of SAM instrument

Based on the examination of factor loadings, AVE, composite reliability and model fit indices in the CFA, the INFIT and OUTFIT MNSQ and the reliability in the Rasch analysis, it is evident that all items in the SAM scale can be retained. The AVE of the SAM scale is below 0.50 and is to be reassessed in the subsequent path analysis by taking into account the relationships with other scales involved. The two-correlated factors model of the SAM scale is used for the subsequent analyses.

4.5 The Student Self-efficacy towards Mathematics (SSM) Instrument

The Student Self-efficacy towards Mathematics (SSM) instrument is made up of nine items, with four of them being negatively worded. The instrument uses the same four-point Likert scale categories as previously used in this chapter. The items of the SSM instrument used in this study are also adapted from the 2011 TIMMS student questionnaire. The SSM instrument consists of two subscales: Confidence and Individual Judgement. Table 4.6 presents the nine items of SSM including the subscales, item labels, the nature of the items and item labels after recoding. Item texts are also displayed. The validation processes of the SSM instrument used in this section follow the same argument as for the previous instrument.

Table 4.6

Student Self-efficacy towards Mathematics (SSM) Instrument Subscales

Subscales	Item label	Nature of statement	Item label to indicate reverse	Item text
Mathematics Confidence (MCONF)	SCM_A	Positive	None	I usually do well in mathematics
	SCM_D	Positive	None	I learn things quickly in mathematics
	SCM_F	Positive	None	I am good at working out difficult mathematics problem
	SCM_G	Positive	None	My teacher thinks I can do well in mathematics classes with difficult materials
	SCM_H	Positive	None	My teacher tells me I am good at mathematics
Individual Judgement (INDI_JUD)	SCM_B	Negative	SCM_BR	Mathematics is not one of my strengths
	SCM_C	Negative	SCM_CR	Mathematics is more difficult for me than for many of my classmates
	SCM_E	Negative	SCM_ER	Mathematics makes me confused and nervous
	SCM_I	Negative	SCM_IR	Mathematics is harder for me than any other subject

4.5.1. Confirmatory Factor Analysis (CFA) of SSM

The model test in the CFA for SSM includes the single-factor model, the two-orthogonal factors model, the two-correlated factors model and the two-hierarchical factors model. All models tested are presented in Figures 4.4, 4.5 and 4.6. All factor loadings meeting the requirement of the acceptable ranges (all loadings are greater than 0.32). The loadings of SSM in the single-factor model are the poorest, yet still meet the requirement of the acceptable loadings (above 0.32). The loadings of the two-correlated factors model are similar to the loadings of the two-hierarchical factors model and the loading of the two-correlated factors models with errors correlated (Figure 4.5 and Figure 4.6) are similar to the loadings of the two-hierarchical factors model with errors correlated. Loadings of the two-correlated factors model are the most satisfactory.

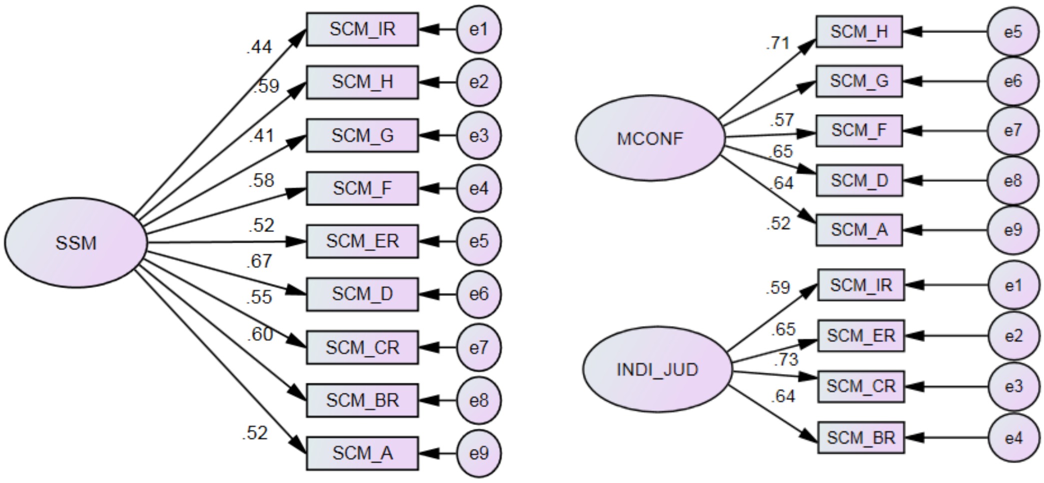


Figure 4.4 Single-factor model and two-orthogonal factors model of SSM

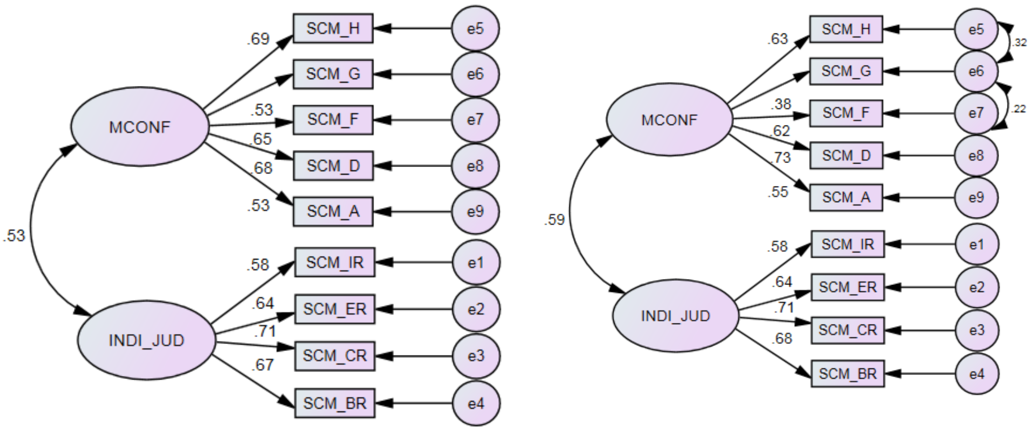


Figure 4.5 Two-correlated factors model and two-correlated factors model with errors correlated of SSM

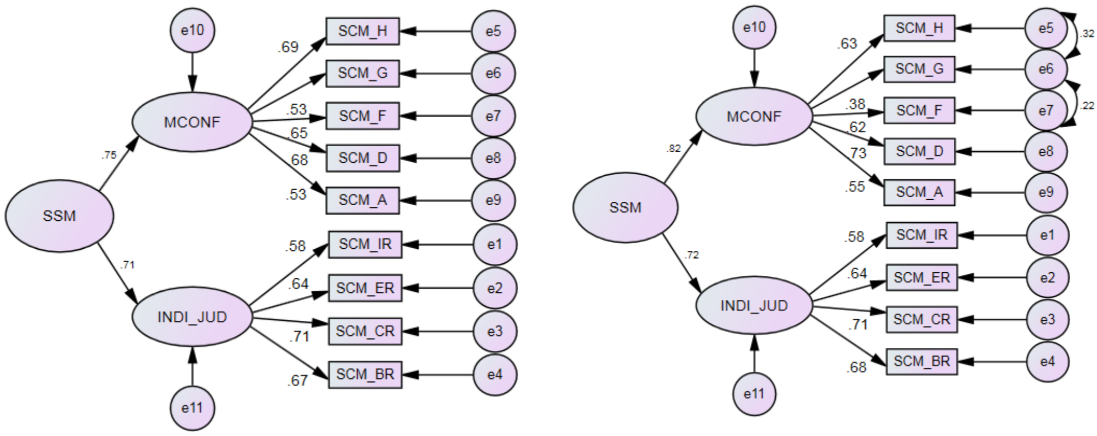


Figure 4.6 Two-hierarchical factors model and two-hierarchical factors model with errors correlated of SSM

The average variance extracted (AVE) and composite reliability (CR) values for all models tested are also recorded in Table 4.7. All models have acceptable composite reliability (above 0.7). None of the models reach the AVE value of 0.5 or above, with the single-factor model having the lowest AVE. The summary of the loadings, AVE and CR values for all models tested is presented in Table 4.7.

Table 4.8 presents the model fit indices of all models tested for SSM. The two-orthogonal factors model is the least acceptable (TLI < 0.8, CFI < 0.9 and RMSEA >1). The model fit of the two-correlated and two-hierarchical models are the same, both have a better fit than the single-factor model (TLI > 0.8, CFI > 0.9 and RMSEA < 1). Furthermore, the model fit of the two-correlated and two-hierarchical models with errors correlated are similar and both models have the most satisfactory fit (TLI > 0.9, CFI > 0.9 and RMSEA < 1). The two-correlated factors models is to examined in the Rasch analysis.

Table 4.7

Factor Loadings of Items, Average Variance Extracted (AVE) and Composite Reliability (CR) Values of Single-factor Model and Two-factor Models of SSM

Items	Subscales	Single-factor model			Two-factor model														
					Two-orthogonal			Two-correlated			Two-correlated (correlated errors)			Two-hierarchical			Two-hierarchical (correlated errors)		
		Loading	AVE	CR	Loading	AVE	CR	Loading	AVE	CR	Loading	AVE	CR	Loading	AVE	CR	Loading	AVE	CR
SCM_A	Confidence	0.52	0.28	0.78	0.52	0.39	0.76	0.53	0.38	0.76	0.55	0.35	0.72	0.53	0.39	0.76	0.55	0.35	0.72
SCM_D		0.55			0.64			0.68			0.73			0.68			0.73		
SCM_F		0.58			0.65			0.65			0.62			0.65			0.62		
SCM_G		0.41			0.57			0.53			0.38			0.53			0.38		
SCM_H		0.59			0.71			0.69			0.63			0.69			0.63		
SCM_BR	Individual judgement	0.60			0.64	0.43	0.75	0.67	0.42	0.75	0.68	0.43	0.75	0.67	0.43	0.75	0.68	0.43	0.75
SCM_CR		0.55			0.73			0.71			0.71			0.71			0.71		
SCM_ER		0.52			0.65			0.64			0.64			0.64			0.64		
SCM_IR		0.44			0.59			0.58			0.58			0.58			0.58		

Table 4.8

The Summary of Model Fit Indices for Single-factor Model and Two-factor Models SSM

Indices N=1135	Single-factor model	Two-factor Model				
		Two-orthogonal	Two-correlated	Two-correlated (correlated errors)	Two-hierarchical model	Two-hierarchical (correlated errors)
Chi-square (χ^2)	72.63	381.59	194.20	91.97	194.20	91.97
Degree of freedom (*df*)	5	27	26	24	26	24
χ^2/df	14.53	14.13	7.47	3.82	7.47	3.82
Goodness of Fix Index (GFI)	0.86	0.95	0.97	0.98	0.97	0.98
Adjusted Goodness of Fix Index (AGFI)	0.77	0.92	0.95	0.97	0.95	0.97
Parsimony Goodness of Fix Index (PGFI)	0.52	0.52	0.56	0.53	0.56	0.53
Tucker-Lewis Index (TLI)	0.83	0.75	0.88	0.95	0.88	0.95
Comparative Fit Index (CFI)	0.94	0.85	0.93	0.95	0.93	0.95
Root Mean Square Error of Approximation (RMSEA)	0.11	0.11	0.08	0.05	0.08	0.05

4.5.2. Rasch Analysis of SSM

The Rasch analysis is carried out for the nine items of the Student Self-efficacy toward Mathematics (SSM) scale using the rating scale model. The model examined is the multidimensional model of the two-correlated factor model of SSM. The summary of the item response parameter estimates of the scale and subscales, including the INFIT and OUTFIT MNSQ, is presented in Table 4.9. The INFIT and OUTFIT MNSQ of all items are within the acceptable range. One item (SCM_IR) has a *t*-value above 5. However, the INFIT and OUTFIT MNSQ is still within the acceptable range.

Table 4.9

Item Response Parameter Estimates of the Two-correlated Factor Models of Student Self-efficacy towards Mathematics (SSM)

Items N=1135	Estimates	Error	Unweighted Fit			Weighted Fit		
			OUTFIT MNSQ	CI	*t*	INFIT MNSQ	CI	*t*
SCM_A	-1.06	0.04	0.87	(0.92, 1.08)	-3.2	0.87	(0.92, 1.08)	-3.4
SCM_BR	0.14	0.04	0.92	(0.92, 1.08)	-2	0.92	(0.92, 1.08)	-2
SCM_CR	0.77	0.04	0.95	(0.92, 1.08)	-1.3	0.95	(0.92, 1.08)	-1.1
SCM_D	0.15	0.04	0.98	(0.92, 1.08)	-0.4	0.98	(0.92, 1.08)	-0.4
SCM_ER	0.00	0.08	0.83	(0.92, 1.08)	-4.2	0.83	(0.92, 1.08)	-4.3
SCM_F	-0.02	0.04	1	(0.92, 1.08)	-0.1	0.98	(0.92, 1.08)	-0.4
SCM_G	-0.35	0.04	0.98	(0.92, 1.08)	-0.5	0.99	(0.92, 1.08)	-0.2
SCM_H	-0.18	0.04	0.97	(0.92, 1.08)	-0.6	0.98	(0.92, 1.08)	-0.5
SCM_IR	0.55	0.07	1.35	(0.91, 1.09)	7.4	1.35	(0.92, 1.08)	7.7

Separation Reliability = 0.99
Chi-square test of parameter equality = 1333.71
df = 7
Significance Level = 0.00

The separation reliability for the two-correlated factors model of SSM is high (0.99). The EAP/PV reliabilities of the confidence in mathematics subscale and individual judgement subscale are 0.75, and 0.79 respectively. Thus, all the EAP/PV reliability values of the scale and subscales are still above the required benchmark of 0.7 (Jähnig, 2013).

4.5.3. Summary of SSM Instrument

Based on the examination of factor loadings and model fit indices in the CFA, the INFIT and OUTFIT MNSQ and the reliability in the Rasch analysis, it is decided that all items in the SSM scale are to be retained. The AVE of the scale is reassessed in the subsequent path analysis by taking into account its relationship with other scales. The model of the SSM scale to be used for further analyses are the two-correlated factors model.

4.6 The Student Beliefs towards Mathematics (SBM) Instrument

Student Beliefs towards Mathematics (SBM) instrument is made up of items assessing Beliefs in the nature of mathematics and beliefs in mathematics learning (Kloosterman, 2002; Op’t Eynde et al., 2003; Presmeg, 2002; Schoenfeld, 1992; A.G Thompson, 1992). SBM is then further classified into beliefs related to LOT and HOT. Table 4.1 presents the 14 items assessing SBM including corresponding subscales, item labels, the nature of the items, and item labels after recoding. Statements of the items are also given. The validation process of the instrument using CFA is similar to the procedure previously employed.

Table 4.10

Student Beliefs towards Mathematics (SBM) Instrument Subscales

Subscales	Item label	Nature of statement	Item label to indicate reverse scoring	Item text
SBM_LOT	SBNM_A	LOT	None	Mathematics is just about addition, subtraction, multiplication and division
	SBNM_B	LOT	None	Mathematics problems should be quickly solvable in a few steps
	SBNM_D	LOT	None	In mathematics, a correct answer is more important than the way to get it
	SBNM_E	LOT	None	All mathematics problems can be solved in one way only
	SBNM_F	LOT	None	Mathematics is just a collection of rules and formulas
	SBML_A	LOT	None	The way teacher solves mathematics problem is the only correct way to solve the problem
	SBML_B	LOT	None	Memorising is the most important thing in learning mathematics
	SBML_C	LOT	None	I should always follow the procedures the teacher taught in solving the mathematics problem
	SBML_D	LOT	None	The main goal of doing mathematics problems is to obtain a correct answer
	SBML_F	LOT	None	I wish my teacher only provided me with those problems which I am familiar with
	SBML_G	LOT	None	I think my teacher should always show how to solve a problem before she/he asks me to do the task
SBM_HOT	SBNM_C	HOT	None	There are several ways to solve a mathematics problem
	SBML_E	HOT	None	I usually try to create my own solution for mathematics problem
	SBML_H	HOT	None	I like working on a mathematics problem where the solution is not obvious

4.6.1. Confirmatory Factor Analysis (CFA) of SBM

There are two models tested for the SBM instrument: a single-factor model and a two-orthogonal factors model. It is important to note that there is no n-correlated factors model and n-hierarchical factors model presented for the SBM as the correlation between the subscales is small (0.21). The models tested are displayed in Figures 4.7.

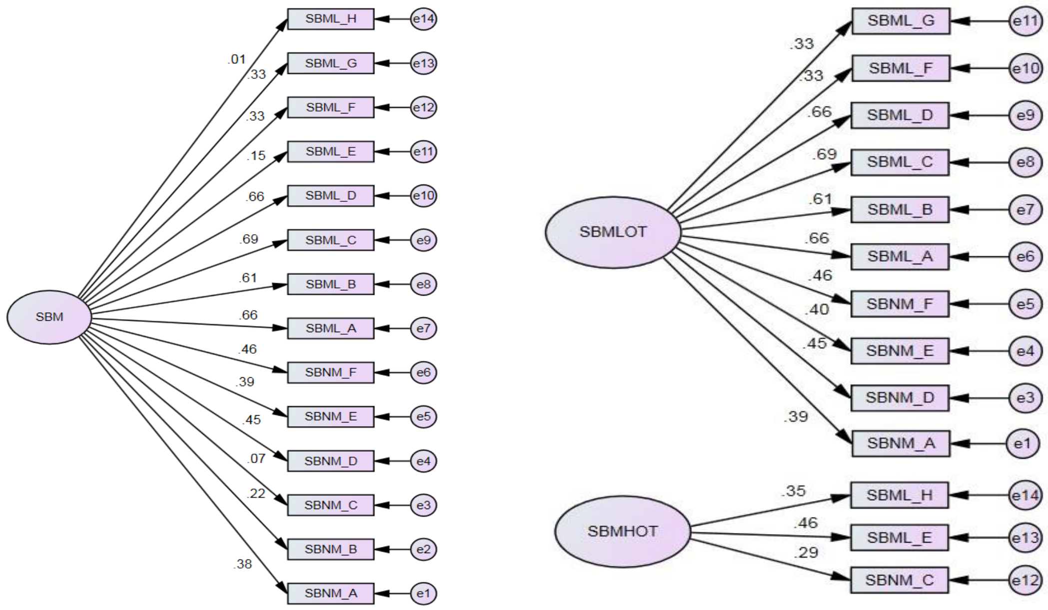


Figure 4.7 Single-factor and two-orthogonal factors models of SBM

Some items of the single-factor model have poor factor loadings (SBNM_C, SBML_E, and SBML_H with a loading less than 0.20). The loading of SBNM_B in the two-orthogonal factors model is also consistently low (below 0.32). Therefore, SBNM_B is excluded from the two-orthogonal factors model. SBNM_C also has low loadings (0.29), in the two-orthogonal factors model. However it is decided to retain SBNM_C as the subscale has only three items. The summary of the factor loading, AVE and CR, for all models is presented in Table 4.11. The composite reliability of the single-factor model is acceptable but the AVE is low. The AVE for one subscale in the two-orthogonal model is higher (0.30) but it is very low for the other subscale (0.14). This low AVE may be due to the small number of items included in that particular subscale. The model fits statistics for the models tested are presented in Table 4.12. The model fit indices also shows that the two-orthogonal factors model has the best fit. The two-orthogonal model of SBM is to be examined using the Rasch analysis.

Table 4.11

Factor Loadings of Items, Average Variance Extracted (AVE) and Composite Reliability (CR) of Single-factor Model, Two-factor Model of SBM

Items	Subscales	Single-factor model			Two-orthogonal factors model		
		Loading	AVE	CR	Loading	AVE	CR
SBNM_A	SBM related to LOT	0.38			0.39		
SBNM_B		0.22			Excluded		
SBNM_D		0.45			0.45		
SBNM_E		0.39			0.40		
SBNM_F		0.46			0.46		
SBML_A		0.66			0.66	0.30	0.77
SBML_B		0.61			0.61		
SBML_C		0.69	0.20	0.72	0.69		
SBML_D		0.66			0.66		
SBML_F		0.33			0.33		
SBML_B		0.33			0.33		
SBNM_C	SBM related to HOT	0.07			0.29		
SBNM_E		0.15			0.46	0.14	0.32
SBNM_H		0.01			0.35		

Table 4.12

Summary of Model Fit Indices for Single-factor Model, Two-factor and Three Factor Models of SBM

Indices	Single-factor model	Two-orthogonal factors model
Chi-square (χ^2)	610.51	497.57
Degree of freedom (*df*)	77	65
χ^2/df	7.93	7.66
Goodness of Fix Index (GFI)	0.92	0.93
Adjusted Goodness of Fix Index (AGFI)	0.89	0.90
Parsimony Goodness of Fix Index (PGFI)	0.68	0.67
Tucker-Lewis Index (TLI)	0.69	0.73
Comparative Fit Index (CFI)	0.77	0.81
Root Mean Square Error of Approximation (RMSEA)	0.08	0.08

4.6.2. Rasch Analysis of SBM Instrument

Rasch analysis is conducted on the 13 item SBM instrument (SBNM_B has been excluded). The two separate models of the SBM scale, corresponding to the two-orthogonal factors model, is examined. The summary of INFIT and OUTFIT MNSQ is presented in Table 4.13 for SBM LOT and Table 4.14 for SBM HOT.

Table 4.13

Response Model Parameter Estimates of SBM_LOT

Items N=1135	Estimates	Error	Unweighted Fit			Weighted Fit		
			OUTFIT MNSQ	CI	*t*	INFIT MNSQ	CI	*t*
SBNM_A	0.37	0.03	1.22	0.92, 1.08	5	1.20	0.92, 1.08	5.4
SBNM_D	0.87	0.03	1.09	0.92, 1.08	2.1	1.08	0.92, 1.08	2.1
SBNM_E	1.49	0.03	0.92	0.92, 1.08	-1.8	0.97	0.92, 1.08	-1.7
SBNM_F	0.25	0.03	0.94	0.92, 1.08	-1.3	0.96	0.92, 1.08	-1.4
SBML_A	-0.07	0.03	0.96	0.92, 1.08	-1.0	0.98	0.92, 1.08	-1.0
SBML_B	0.04	0.03	0.96	0.92, 1.08	-0.9	0.97	0.92, 1.08	-1.3
SBML_C	-0.60	0.03	0.74	0.92, 1.08	-6.7	0.76	0.92, 1.08	-7.0
SBML_D	-0.64	0.03	0.88	0.92, 1.08	-2.9	0.89	0.92, 1.08	-3.0
SBML_F	-0.30	0.03	1.10	0.92, 1.08	2.2	1.10	0.92, 1.08	2.2
SBML_G	-1.33	0.10	1.15	0.92, 1.08	3.5	1.13	0.92, 1.08	3.1

Separation Reliability = 1
Chi-square test of parameter equality = 3735.03
df = 10
Significance Level = 0.00

Table 4.14

Response Model Parameter Estimates of SBM_HOT

Items N=1135	Estimates	Error	Unweighted Fit			Weighted Fit		
			OUTFIT MNSQ	CI	*t*	INFIT MNSQ	CI	*t*
SBNM_C	-0.86	0.03	0.84	0.92, 1.08	-3.9	0.85	0.92, 1.08	-3.6
SBML_E	0.41	0.03	0.92	0.92, 1.08	-1.9	0.92	0.92, 1.08	-2.0
SBML_H	0.46	0.05	1.16	0.92, 1.08	3.5	1.15	0.92, 1.08	3.4

Separation Reliability = 1
Chi-square test of parameter equality =868.79
df = 2
Significance Level = 0.000

Tables 4.13 and 4.14 show that the INFIT and OUTFIT MNSQ of all items of SBM LOT and SBM HOT subscales are within the acceptable range. The reliabilities of the subscales involve separation reliability and EAP/PV reliability. The Rasch analysis records that the separation reliability for the SBM LOT and SBM HOT are very good (all have a reliability of 1). The EAP/PV reliabilities of the subscales are lower compared to the separation reliabilities, with the EAP/PV reliability values being 0.74 (SBM LOT) and 0.30 (SBM HOT). The EAP/PV reliabilities of the SBM LOT is within the satisfactory range. However, the EAP/PV reliabilities for the SBM_HOT subscale are below the acceptable value of 0.70. This is likely to be as a consequence of the small number of items involved in this subscale.

4.6.3. Summary of SBM

Based on the examination of factor loadings and model fit indices in the CFA, the INFIT and OUTFIT MNSQ and the reliability in the Rasch analysis, the conclusion is made to exclude item SBNM_B and that the best model for in the subsequent analyses is the two-orthogonal factors model of SBM.

4.7 Student Learning Activities (SLA) Instrument

The SLA scale is related to how often students perceive themselves as engaged in particular activities during mathematics lessons. The items assessing Student Learning Activities used in the study are adapted from the TIMMS 2011 student questionnaire. The SLA instruments consist of eight items which are divided into two subscales: SLA related to LOT and HOT. The instrument uses a three point Likert-type scale to express the frequency: ‘never’, ‘sometimes’, and ‘always or almost never’.

Table 4.15

Student Learning Activities (SLA) Instrument Subscales

Subscales	Item label	Nature of statement	Item label to indicate reverse scoring	Item text
SLA related to LOT	SLA_A	LOT	None	Listening to the teacher explaining how to solve problems
	SLA_B	LOT	None	Memorising formulas
	SLA_C	LOT	None	Working on problems with your classmates
	SLA_D	LOT	None	Using your knowledge from previous topics to solve problems
SLA related to HOT	SLA_E	HOT	None	Explaining your answer
	SLA_F	HOT	None	Relating what you learn in mathematics to your daily life
	SLA_G	HOT	None	Using your own way to solve difficult problem
	SLA_H	HOT	None	Working on problems which you cannot find the solution to straight away

The validation processes of the SLA instrument is undertaken in a similar way to the previous instruments in the student questionnaire. The detailed description of the instrument, including the item labels, nature of the items and texts of the items, is shown in Table 4.15.

4.7.1. Confirmatory Factor Analysis (CFA) of SLA Instrument

The models tested for CFA include a single-factor model, a two-orthogonal factors model, a two-correlated factors model, and a two-hierarchical factors model. Figures 4.8 and 4.9 show the models tested and the summary of the loadings is displayed in Table 4.16.

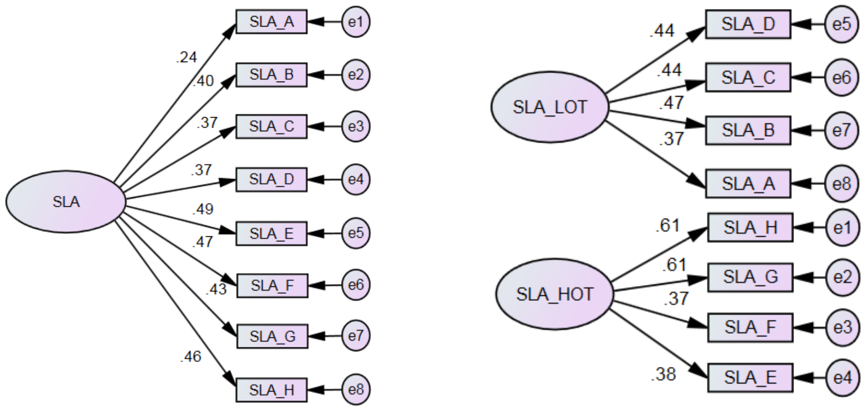


Figure 4.8 Single-factor model and two-orthogonal factors models of SLA

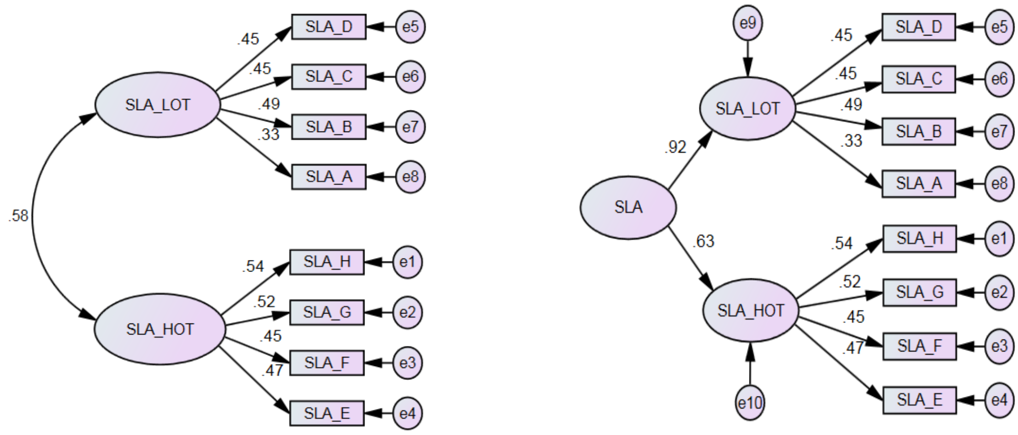


Figure 4.9 Two-correlated factors and two-hierarchical factors models of SLA

Table 4.16 shows that the single-factor model has the poorest factor loading, but only one item loading falls outside the acceptable range (SLA_A with a loading of 0.24). The loadings of the remaining models are acceptable. The two-correlated factors and two-hierarchical factors models have similar loadings and also the most satisfactory loadings. The AVE and CR values are also presented in Table 4.16. Both AVE and CR for all models are below the acceptable values.

Table 4.16

Factor Loading of Items, Average Variance Extracted (AVE) and Composite Reliability (CR) of the Single-factor Model and Two-factor Models of SLA

Items	Subscales	Single-factor model			Two-factor model								
					Two-orthogonal			Two-correlated			Two-hierarchical		
		Loading	AVE	CR	Loading	AVE	CR	Loading	AVE	CR	Loading	AVE	CR
SLA_A	SBM related to LOT	0.24	0.34	0.61	0.44	0.19	0.48	0.45	0.19	0.48	0.45	0.19	0.48
SLA_B		0.40			0.44			0.45			0.45		
SLA_C		0.37			0.47			0.49			0.49		
SLA_D		0.37			0.37			0.33			0.33		
SLA_E	SBM related to HOT	0.49			0.61	0.26	0.57	0.54	0.25	0.57	0.54	0.25	0.57
SLA_F		0.47			0.61			0.52			0.52		
SLA_G		0.43			0.37			0.45			0.45		
SLA_H		0.46			0.38			0.47			0.47		

Table 4.17

Summary of Model Fit Indices for the Single-factor Model and Two-factor Models of SLA

Indices	Single-factor model	Two-factor model		
		Two-orthogonal	Two-correlated	Two-hierarchical
Chi-square (χ^2)	180.12	212.25	118.56	118.56
Degree of freedom (*df*)	20	19	19	19
χ^2/df	9.01	6.24	6.24	6.24
Goodness of Fix Index (GFI)	0.96	0.95	0.97	0.97
Adjusted Goodness of Fix Index (AGFI)	0.93	0.92	0.95	0.95
Parsimony Goodness of Fix Index (PGFI)	0.54	0.53	0.51	0.51
Tucker-Lewis Index (TLI)	0.59	0.51	0.73	0.73
Comparative Fit Index (CFI)	0.77	0.73	0.86	0.84
Root Mean Square Error of Approximation (RMSEA)	0.08	0.09	0.07	0.07

The model fit indices for each model are presented in Table 4.17. As with the factor loading assessment, the single-factor model has the least satisfactory fit values (TLI <

0.60, CFI < 0.70 and RMSEA > 0.08). The best fitting models are the two-correlated factors and two-hierarchical factors models (TLI > 0.70, CFI is close to 0.90 and RMSEA is 0.07). While the TLI values of the two models are outside the acceptable range, the CFI and RMSEA are within the acceptable range. The two-correlated factors model is to be examined in the Rasch analysis.

4.7.2. Rasch Analysis of SLA Instrument

A Rasch analysis is conducted for the multidimensional model of the two-correlated factors model of SLA. Tables 4.18 shows the summary of the response model parameter estimates of the scale, including the INFIT and OUTFIT MNSQ, as well as the *t* values for the two-correlated factors model of SLA scale. Table 4.18 records the response model parameter estimates of SLA as a two-correlated factor model. All items have INFIT and OUTFIT MNSQ value which fall within the acceptable range.

Table 4.18

Response Model Parameter Estimates of Rating Scale of the Two-correlated Factors Model of SLA

Items N=1135	Estimates	Error	Unweighted Fit			Weighted Fit		
			OUTFIT MNSQ	CI	*t*	INFIT MNSQ	CI	*t*
SLA_A	-1.88	0.05	1.06	(0.92, 1.08)	1.5	1.09	(0.92, 1.10)	1.7
SLA_B	0.73	0.04	0.96	(0.92, 1.08)	-0.9	0.97	(0.92, 1.08)	-0.8
SLA_C	0.87	0.04	0.81	(0.92, 1.08)	-4.9	0.81	(0.92, 1.08)	-4.9
SLA_D	0.27	0.07	1.01	(0.92, 1.08)	0.4	1.02	(0.92, 1.07)	0.5
SLA_E	-0.64	0.04	0.96	(0.92, 1.08)	-1	0.96	(0.92, 1.08)	-1.1
SLA_F	-0.43	0.04	0.98	(0.92, 1.08)	-0.4	0.99	(0.92, 1.08)	-0.3
SLA_G	0.38	0.04	1.03	(0.92, 1.08)	0.8	1.04	(0.92, 1.08)	1
SLA_H	0.69	0.07	1.06	(0.92, 1.08)	1.5	1.06	(0.92, 1.08)	1.5

Separation Reliability = 1
Chi-square test of parameter equality = 2844
df = 6
Significance Level = 0.00

The Rasch analysis records that the separation reliability for the two-correlated factors model of SLA is very good, the value is 1. The EAP/PV reliabilities of the subscales are lower than the separation reliabilities (0.51 and 0.61 for SLA LOT and SLA HOT respectively, below the acceptable range of 0.70). The low EAP/PV reliabilities of the subscales is likely to be a consequence of the small number of the items (only four items in each scale).

4.7.3. Summary of SLA Instrument

Based on the examination of factor loadings and model fit indices in the CFA, the INFIT and OUTFIT MNSQ and the reliability in the Rasch analysis, it is concluded that the two-correlated factors model is the best model and consequently, this model is used in the subsequent analysis. The AVE and composite reliability are reassessed in the subsequent path analysis by taking into account the relationships of the scales with the other scales involved.

4.8 Summary

The validation processes of the four instruments included in the student questionnaire are discussed in this chapter: Student Attitudes toward Mathematics (SAM); Student Self-efficacy toward Mathematics (SSM); Student Beliefs concerning Mathematics (SBM); and Student Learning Activities (SLA). The validation processes are similar across all instruments, involving both CFA and Rasch analysis. The results of the competing CFA models, together with the INFIT and OUTFIT MNSQ of the Rasch analysis, are used to judge how well the models fit the data. The best models are chosen for the subsequent analyses. Rasch scale scores are used for the subsequent analysis when appropriate. The chosen model from each instrument is summarised as follows:

1. Student Attitudes toward Mathematics (SAM)

   SAM measures student attitudes toward mathematics. After examining the CFA and Rasch analysis results of all alternative models, the two-correlated factors model is used for the subsequent analysis.

2. Student Self-efficacy toward Mathematics (SSM)

   SSM measures student self-efficacy toward mathematics. The two-correlated factors model is used for the subsequent analysis.

3. Student Beliefs concerning Mathematics (SBM)

   SBM measures student beliefs concerning mathematics. The two-orthogonal factors model is used for the subsequent analysis

4. Student Learning Activities (SLA)

   SLA measures student engagement in the mathematics classroom. The two-correlated factors model is used for the subsequent analysis

Chapter 5 Research Instruments: Teacher

5.1 Introduction

The validation of the instrument used in the teacher questionnaire is presented in this chapter, carried out using the same methods as that done in Chapter 4 for students. The teacher questionnaires contain three sections: personal and educational information; information related to the teachers' beliefs: and information about classroom practices. The validation is carried out on the basis of the 46 teachers in the data set for the subscales in the questionnaire. The first section is not included in the validation process as it consists of background information, while the second and third sections are concerned with the measurement models related to teachers' beliefs and classroom practices.

5.2 Teacher Beliefs concerning Mathematics (TBM) Instrument

These scales involving teacher beliefs are self-perceived and self-reported by the participants. There are six subscales (as presented in Table 5.1) namely: (a) Teacher Beliefs concerning Nature of Mathematics (TBNM) related to HOT; (b) Teacher Beliefs concerning Nature of Mathematics (TBNM) related to LOT; (c) Teacher Beliefs concerning Mathematics Learning (TBML) related to HOT; (d) Teacher Beliefs concerning Mathematics Learning (TBML) related to LOT; (e) Teacher Beliefs concerning Mathematics Teaching (TBMT) related to HOT; and (f) Teacher Beliefs Concerning Mathematics Teaching (TBMT) related to LOT. Most items related to

Teacher Belief toward Mathematics (TBM) are adapted from the questionnaires developed by previous researchers.

TBM is made up of 29 items, including ten items from Perry et al. (1999), three items from Hart (2002), two items from Peterson et al. (1989), one item from Zakaria and Musiran (2010) and one item from Battista (1994). The remaining 13 items are specifically developed for this study. As this study is distinguishing LOT and HOT, which no known researcher has previously investigated, a new instrument is created by using items derived from several previous instruments and matching each item to the criteria of LOT or HOT. The responses in the TBM instrument are made on a four-point Likert-scale: ‘disagree a lot’, ‘disagree a little’, ‘agree a little’ and ‘agree a lot’, coding being 1, 2, 3, and 4. All items are positive with respect to the intended subscales; therefore the agreement or disagreement of each item indicates agreement or disagreement with the corresponding subscales.

Table 5.1 presents the 29 items associated with the TBM instrument and includes the subscales, item label, source of items, original items if the item is reworded and the item text.

Table 5.1

Teachers' Beliefs concerning Mathematics (TBM) Instrument Subscales

Subscales	Item label	Source	Item text	The original text
TBNM related to LOT	TBNM_A	Battista (1994)	Mathematics is not a flexible process where accuracy, speed, and memory is important	Mathematics as a rigid system of externally dictated rules governed by standards of accuracy, speed, and memory
	TBNM_B	Researcher	Mathematics is just addition, subtraction, multiplication and division	
	TBNM_C	Perry et al. (1999)	Mathematics problems given to students should be quickly solvable in a few steps	
	TBNM_F	Perry et al. (1999)	Right answers are more important in mathematics than the ways in which you get them	
	TBNM_G	Researcher	Mathematics is about remembering the rules	
TBNM related to HOT	TBNM_D	Perry et al. (1999)	Mathematics is the dynamic searching for order and pattern in the learner's environment	
	TBNM_E	Perry et al. (1999)	Mathematics is a creative human endeavour that is both a way of knowing and a way of thinking	Mathematics is a beautiful, creative, useful human endeavour that is both a way of knowing and a way of thinking
	TBNM_H	Researcher	Analysis is important in solving mathematics problems	
	TBNM_I	Researcher	Mathematics problems could be solved in various ways	
TBML related to LOT	TBML_A	Researcher	Mathematics learning is enhanced by activities which build upon students' thinking skills	
	TBML_B	Perry et al. (1999)	Mathematics learning is enhanced by challenges within a supportive environment	
	TBML_C	Researcher	Students' critical thinking skills are enhanced when they work on an open-ended mathematics problem	
	TBML_D	Researcher	Mathematics learning is being able to transfer the skills to a new unfamiliar problem	
	TBML_F	(Hart, 2002)	A demonstration of good reasoning should be seen as more important than a student's ability to find the correct answer	
	TBML_G	(Hart, 2002)	Mathematics learning is enhanced when students are given enough opportunities to discover their own solutions	Students can be creative in solving mathematics problems when teachers give them enough time to discover things by themselves

Table 5.1 *(continued)*

Subscales	Item label	Source	Item text	The original text
TBML related to HOT	TBML_E	Researcher	Mathematics learning is enhanced by exposing students to repetitive routine problems only	
	TBML_H	Perry et al. (1999)	Being able to memorise facts is very critical in mathematics learning	
	TBML_I	Perry et al. (1999)	Mathematics learning is being able to get the right answer	
TBMT related to LOT	TBMT_E	Peterson et al. (1989)	Teachers should always demonstrate how to solve simple problems before students are allowed to solve problems	
	TBMT_G	Hart (2002)	Teachers should teach students the exact procedures for solving problem	Teachers should show students the exact way to answer the questions which they will be tested on
	TBMT_H	Researcher	Mathematics should be taught as a collection of rules, procedures and algorithms	
	TBMT_J	Researcher	Higher order thinking tasks should not be given to low achieving students	
	TBMT_K	Perry et al. (1999)	The role of the mathematics teacher is to transmit mathematical knowledge and to verify that learners have received this knowledge	
TBMT related to HOT	TBMT_A	Perry et al. (1999)	Teachers should provide instructional activities which result in challenging situations for learners	Teachers should provide instructional activities which result in problematic situations for learners
	TBMT_B	Perry et al. (1999)	Teachers should allow students to work in a cooperative learning environment with their peers	Teachers should negotiate social norms with the students in order to develop a cooperative learning environment in which students can construct their
	TBMT_C	Peterson et al. (1989)	Teachers should allow students the opportunity to analyse problems before they try to find solutions	
	TBMT_D	Perry et al. (1999)	Teachers should not tell students if their answers are correct or incorrect. Rather, they should challenge them to explain their strategies	
	TBMT_F	(Zakaria & Musiran, 2010)	Students should be encouraged to justify their solutions and reasoning	Students should be encouraged to justify their solutions, thinking and conjectures
	TBMT_I	Researcher	Mathematics should be taught using a combination of routine and non-routine problems to develop students' thinking skills	

5.2.1. Confirmatory Factor Analysis (CFA) of TBM

There are seven alternative models tested: a single-factor model, a six-orthogonal factors model, a four-orthogonal factors model of the TBM and four separate models. The separate models include a two-orthogonal factors model of TBM LOT, a two-correlated factors model of TBM LOT, a two-orthogonal factors model of TBM HOT and a two-correlated factors model of TBM HOT.

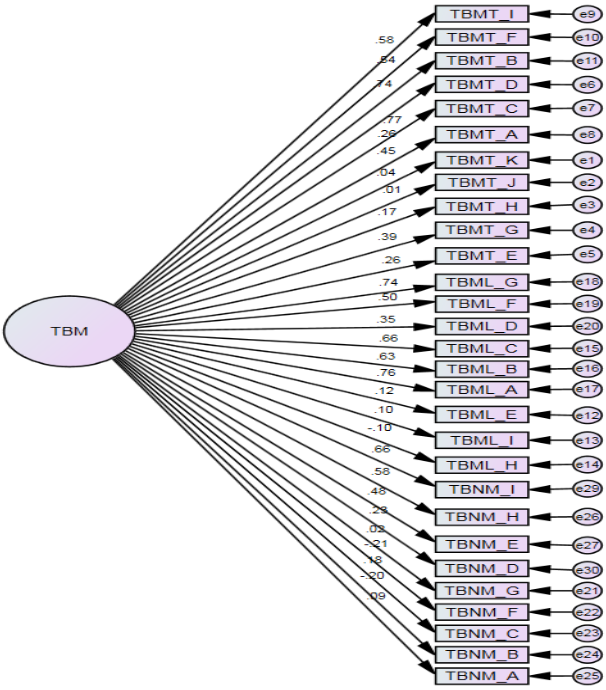


Figure 5.1 Single-factor model of TBM

A single-factor model, shown in Figure 5.1, is the first model tested with all 29 items involved as one construct. The second model examined in Figure 5.2, is the six-orthogonal factors model breaking the TBM into six subscales: (1) Teacher Beliefs of Nature of Mathematics (TBNM) related to HOT, (2) Teacher Beliefs of Nature of Mathematics (TBNM) related to LOT, (3) Teacher Beliefs of Mathematics Learning (TBML) related to HOT, (4) Teacher Beliefs of Mathematics Learning (TBML) related to LOT, (5) Teacher Beliefs of Mathematics Teaching (TBMT) related to HOT, and (6) Teacher beliefs of Mathematics Teaching (TBMT) related to LOT. The third model examined, also shown in Figure 5.2, is the four-orthogonal factors model, merging the TBNM HOT together with TBML HOT and TBNM LOT with TBML LOT. The fourth set of models tested are the separate models, after splitting TBM related to LOT and HOT into two separate models with two subscales each.

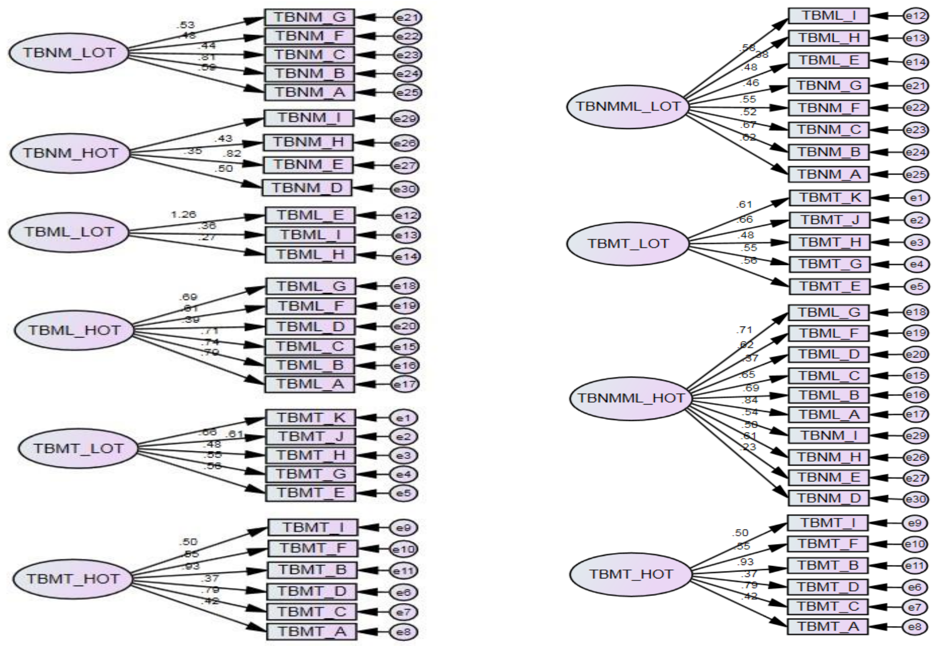


Figure 5.2 Six-orthogonal factors model and four-orthogonal factors model of TBM

Table 5.2 presents the factor loadings, the average variance extracted (AVE) and the composite reliability (CR) values for all items for each model tested and Table 5.3 presents the model fit indices of all models examined. The evidence presented in Table 5.2 indicates that the single-factor model has the poorest loadings. Some items are outside the acceptance range (below 0.32). The loadings for item TBNM_D is consistently low in the first three models examined (less than 0.32), except for the six-orthogonal factors model (0.5). TBNM_D has an acceptable loading in the model but item TBML_E has a loading that exceeds 1, which is not acceptable. Therefore, TBNM_D is excluded from the separate models of TBM (see Figure 5.4).

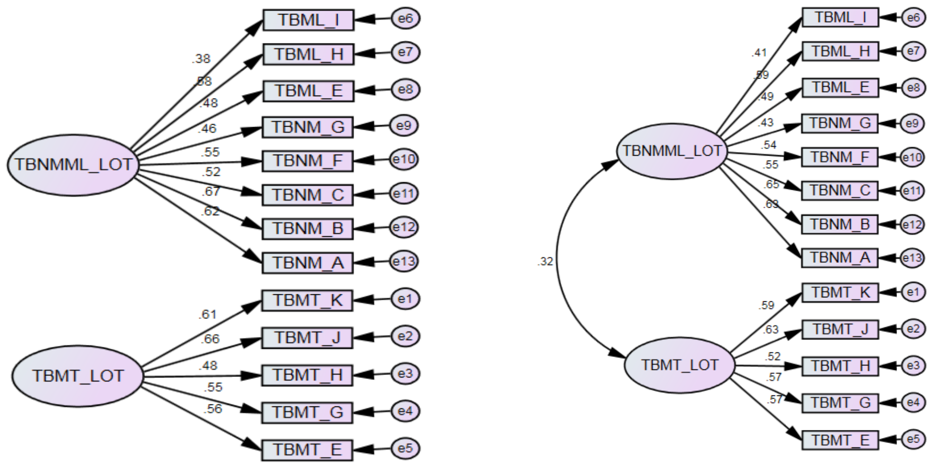


Figure 5.3 Separate model: two-orthogonal factors model and two-correlated models of TBM related to LOT

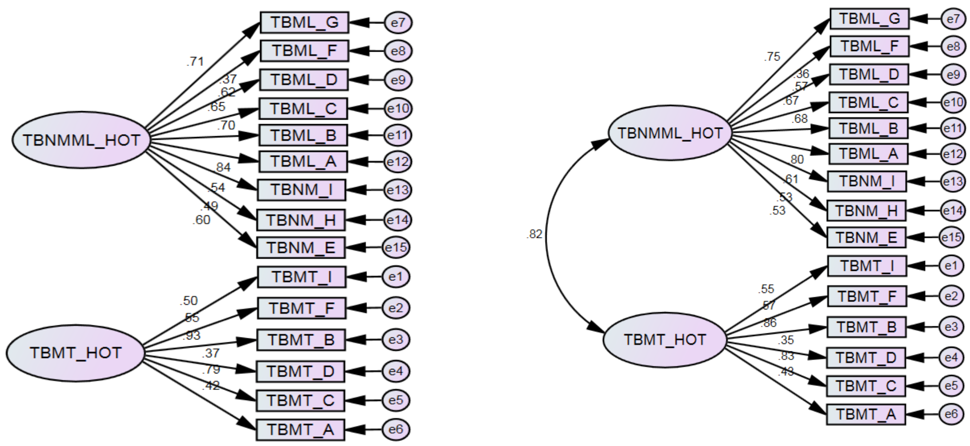


Figure 5.4 Separate model: two-orthogonal and two-correlated factors models of TBM related to HOT (TBNM_D is excluded)

The separate models of TBM LOT and TBM HOT are examined, in Figures 5.3 and 5.4, because of the poor fit indices of the first three models tested. The breaking of TBM into two separate scales of TBM LOT and TBM HOT improve the fit indices. The loadings of each item in the separate models of TBM LOT and TBM HOT are acceptable (above 0.32). The model fit indices recorded in Table 5.3 indicate that the separate two-correlated factors models for TBM LOT and HOT (with the exclusion of TBNM_D), with two subscales in each factor, are the best models. These models are then examined in the Rasch analysis.

Table 5.2

Factor Loadings, Average of Variance Extracted (AVE) and Composite Reliability (CR) of Items of Single-factor Model, Separate Models, Four-factor Models, and Six-factor Models of TBM

Scale	Variables	Subscales	Single-factor model			Six-orthogonal factor models			Four-orthogonal factor models			Two-orthogonal factor models			Two-correlated factor models		
			Loading	AVE	CR	Loading	AVE	CR	Loading	AVE	CR	Loading	AVE	CR	Loading	AVE	CR
TBM	TBNM_A	TBNM_LOT	0.09	0.20	0.84	0.59	0.34	0.71	0.62	0.29	0.76	0.62	0.29	0.76	0.63	0.29	0.76
	TBNM_B		0.20			0.81			0.67			0.67			0.65		
	TBNM_C		0.18			0.44			0.52			0.52			0.55		
	TBNM_F		0.21			0.48			0.55			0.55			0.54		
	TBNM_G		0.02			0.53			0.46			0.46			0.43		
	TBML_E	TBML_LOT	0.12			1.26	0.60	0.75	0.48			0.48			0.49		
	TBML_H		0.10			0.27			0.58			0.58			0.59		
	TBML_I		0.10			0.36			0.38			0.38			0.41		
	TBMT_E	TBMT_LOT	0.26			0.56	0.33	0.71	0.56	0.33	0.70	0.56	0.33	0.70	0.57	0.33	0.70
	TBMT_G		0.39			0.55			0.55			0.55			0.57		
	TBMT_H		0.17			0.48			0.48			0.48			0.52		
	TBMT_J		0.01			0.66			0.66			0.66			0.63		
	TBMT_K		0.04			0.61			0.61			0.61			0.59		
	TBNM_D	TBNM_HOT	0.23			0.50	0.31	0.61	0.23	0.36	0.84	Excluded	0.34	0.84	Excluded	0.39	0.84
	TBNM_E		0.48			0.35			0.61			0.60			0.53		
	TBNM_H		0.58			0.82			0.50			0.49			0.53		
	TBNM_I		0.66			0.43			0.54			0.54			0.61		
	TBML_A	TBML_HOT	0.76			0.79	0.42	0.87	0.84			0.84			0.80		
	TBML_B		0.63			0.74			0.69			0.70			0.68		
	TBML_C		0.66			0.71			0.65			0.65			0.67		
	TBML_D		0.35			0.79			0.37			0.62			0.57		
	TBML_F		0.50			0.61			0.62			0.37			0.36		
	TBML_G		0.74			0.69			0.71			0.71			0.75		
	TBMT_A	TBMT_HOT	0.45			0.42	0.39	0.78	0.42	0.39	0.78	0.42	0.39	0.78	0.43	0.39	0.78
	TBMT_C		0.77			0.79			0.79			0.79			0.83		
	TBMT_D		0.46			0.37			0.37			0.37			0.35		
	TBMT_B		0.74			0.93			0.93			0.93			0.86		
	TBMT_F		0.54			0.55			0.55			0.55			0.57		
	TBMT_I		0.68			0.50			0.50			0.50			0.55		

Table 5.3

Summary of Model Fit Indices for Single-factor Model, Separate Models, Four-factor Models, and Six-factor Models of TBM

Indices	Two-orthogonal factors of TBM_LOT	Two-correlated factors of TBM_LOT	Two-orthogonal factors of TBM_HOT	Two-correlated factors of TBM_HOT	Single-factor model	Four-orthogonal factors model	Six-orthogonal factor models
Chi-square (χ^2)	98.90	96.82	141.26	112.64	661.63	603.96	633.86
Degree of freedom (*df*)	65	64	90	89	377	377	377
χ^2/df	1.52	1.51	1.57	1.26	1.76	1.60	1.68
Goodness of Fix Index (GFI)	0.73	0.73	0.76	0.77	0.53	0.59	0.58
Adjusted Goodness of Fix Index (AGFI)	0.62	0.62	0.68	0.70	0.45	0.53	0.52
Parsimony Goodness of Fix Index (PGFI)	0.52	0.51	0.57	0.58	0.46	0.51	0.50
Tucker-Lewis Index (TLI)	0.55	0.53	0.68	0.85	0.28	0.43	0.35
Comparative Fit Index (CFI)	0.66	0.67	0.76	0.89	0.39	0.50	0.44
Root Mean Square Error of Approximation (RMSEA)	0.11	0.11	0.11	0.08	0.13	0.12	0.12

5.2.2. Rasch Analysis of Teacher Beliefs concerning Mathematics (TBM) Instrument

A Rasch analysis of the TBM is conducted for the separate models of two-correlated factors model of TBM LOT (TBNMML LOT and TBMT LOT) and two-correlated factors model of TBM HOT (TBNMML HOT and TBMT HOT). Table 5.4 records the response model parameter estimates of the TBM LOT scale, including the INFIT and OUTFIT MNSQ. Table 5.5 records the summary of model parameter estimates of the scales for the two-correlated factors model of TBM HOT (TBNMML HOT and TBMT HOT).

Table 5.4

Response Model Parameter Estimates of Two-correlated Factors Model of TBM LOT

Variables	Estimates	Error	Unweighted Fit			Weighted Fit		
			OUTFIT MNSQ	CI	*t*	INFIT MNSQ	CI	*t*
TBNM A	-0.43	0.17	1.23	(0.58, 1.42)	1.1	1.25	(0.59, 1.41)	1.2
TBNM B	1.26	0.18	0.75	(0.59, 1.41)	-1.3	0.75	(0.59, 1.41)	-1.3
TBNM C	-1.20	0.17	0.61	(0.59, 1.41)	-2.1	0.61	(0.62, 1.38)	-2.3
TBNM F	1.01	0.17	1.3	(0.59, 1.41)	1.4	1.3	(0.59, 1.41)	1.4
TBNM G	-0.04	0.17	1.2	(0.59, 1.41)	1	1.21	(0.59, 1.41)	1
TBML E	0.05	0.17	0.65	(0.59, 1.41)	-1.8	0.66	(0.59, 1.41)	-1.8
TBML H	0.88	0.17	0.7	(0.59, 1.41)	-1.5	0.7	(0.58, 1.42)	-1.5
TBML I	-1.54	0.46	1.07	(0.59, 1.41)	0.4	1.08	(0.62, 1.38)	0.5
TBMT E	0.68	0.18	1.03	(0.58, 1.42)	0.2	1.05	(0.59, 1.41)	0.3
TBMT G	-0.05	0.17	0.76	(0.59, 1.41)	-1.2	0.74	(0.61, 1.39)	-1.3
TBMT H	-0.03	0.18	1.14	(0.57, 1.43)	0.7	1.09	(0.60, 1.40)	0.5
TBMT J	-0.17	0.17	1.53	(0.59, 1.41)	2.3	1.52	(0.61, 1.39)	2.3
TBMT K	-0.43	0.35	0.62	(0.58, 1.42)	-2	0.59	(0.61, 1.39)	-2.4

Separation Reliability = 0.94
Chi-square test of parameter equality = 183.81
df = 11
Significance level = 0.00

Table 5.4 shows that items apart from TBMT_J in the two-correlated factors models of TBM LOT are within the acceptable range of INFIT and OUTFIT MNSQ. The separation reliability of the two-correlated factors models of TBM LOT is high (0.96). However, the EAP/PV separation reliability of the subscales of TBNMML LOT and

TBMT LOT are slightly below the acceptable value of 0.70 (0.68 and 0.66 respectively).

Table 5.5

Response Model Parameter Estimates of Two-correlated Factors Model of TBM HOT

Variables	Estimates	Error	Unweighted Fit			Weighted Fit		
			OUTFIT MNSQ	CI	*t*	INFIT MNSQ	CI	*t*
TBNM E	-0.18	0.25	0.99	(0.59, 1.41)	0	0.94	(0.59, 1.41)	-0.2
TBNM H	-0.06	0.25	1.07	(0.59, 1.41)	0.4	1.21	(0.58, 1.42)	1
TBNM I	-0.81	0.25	0.89	(0.59, 1.41)	-0.5	0.9	(0.61, 1.39)	-0.4
TBML A	-1.33	0.25	0.54	(0.59, 1.41)	-2.6	0.62	(0.61, 1.39)	-2.1
TBML B	-0.44	0.25	0.71	(0.59, 1.41)	-1.4	0.71	(0.60, 1.40)	-1.5
TBML C	-0.27	0.25	0.97	(0.59, 1.41)	-0.1	0.98	(0.58, 1.42)	0
TBML D	1.68	0.24	1.26	(0.59, 1.41)	1.2	1.13	(0.51, 1.49)	0.6
TBML F	1.46	0.24	1.65	(0.59, 1.41)	2.6	1.50	(0.51, 1.49)	1.8
TBML G	-0.06	0.70	1.02	(0.59, 1.41)	0.2	1.02	(0.58, 1.42)	0.2
TBMT A	0.52	0.25	1.54	(0.59, 1.41)	2.3	1.54	(0.50, 1.50)	1.9
TBMT B	-0.85	0.25	0.68	(0.59, 1.41)	-1.7	0.8	(0.65, 1.35)	-1.1
TBMT C	-1.36	0.25	0.67	(0.59, 1.41)	-1.7	0.76	(0.68, 1.32)	-1.5
TBMT D	1.14	0.25	1.68	(0.58, 1.42)	2.7	1.54	(0.45, 1.55)	1.7
TBMT F	0.40	0.25	0.67	(0.59, 1.41)	-1.7	0.69	(0.51, 1.49)	-1.3
TBMT I	0.15	0.56	0.67	(0.59, 1.41)	-1.7	0.69	(0.54, 1.46)	-1.4

Separation Reliability = 0.94
Chi-square test of parameter equality = 198.15
df = 13
Significance level = 0.00

Table 5.5 shows that the majority of the items in the two-correlated factors model of TBM HOT fall within the acceptable range of INFIT and OUTFIT MNSQ, with three items having INFIT and OUTFIT MNSQ outside the range (TBMT_D, TBMT_A, TBMT_F). The separation reliability of two-correlated factors model of TBM HOT is high (0.94) and the EAP/PV of the subscales of TBNMML HOT and TBMT HOT are satisfactory (0.89 and 0.70 respectively).

Values presented in Tables 5.4 and 5.5 indicate that the Rasch analysis of the two-correlated factors model of TBM LOT and TBM HOT are not the best models. As the separate model of the two-orthogonal factors models of TBM LOT and TBM HOT are the second best models according to the CFA results, Rasch analyses are also to be conducted for these two models.

Table 5.6

Response Model Parameter Estimates of Teacher Beliefs concerning Nature of Mathematics and Mathematics Learning related to LOT (TBNMML LOT)

Variables	Estimates	Error	Unweighted Fit			Weighted Fit		
			OUTFIT MNSQ	CI	*t*	INFIT MNSQ	CI	*t*
TBNM_A	-0.41	0.18	1.37	(0.57, 1.43)	1.6	1.40	(0.58, 1.42)	1.8
TBNM_B	1.36	0.18	0.81	(0.58, 1.42)	-0.9	0.81	(0.59, 1.41)	-0.9
TBNM_C	-1.23	0.17	0.75	(0.59, 1.41)	-1.2	0.76	(0.61, 1.39)	-1.3
TBNM_F	1.05	0.18	1.33	(0.59, 1.41)	1.5	1.35	(0.59, 1.41)	1.6
TBNM_G	-0.13	0.18	1.30	(0.59, 1.41)	1.4	1.29	(0.60, 1.40)	1.4
TBML_E	0.03	0.18	0.71	(0.59, 1.41)	-1.5	0.72	(0.59, 1.41)	-1.4
TBML_H	0.91	0.18	0.78	(0.59, 1.41)	-1.1	0.78	(0.59, 1.41)	-1.0
TBML_I	-1.56	0.47	1.10	(0.59, 1.41)	0.5	1.11	(0.62, 1.38)	0.6

Separation Reliability = 0.96
Chi-square test of parameter equality = 172.65
df = 7
Significance level = 0.00

Table 5.7

Response Model Parameter Estimates of Teacher Beliefs concerning Mathematics Teaching related to LOT (TBMT LOT)

Variables	Estimates	Error	Unweighted Fit			Weighted Fit		
			OUTFIT MNSQ	CI	*t*	INFIT MNSQ	CI	*t*
TBMT_E	0.74	0.18	1.19	(0.58, 1.42)	0.9	1.22	(0.60, 1.40)	1.1
TBMT_G	-0.07	0.18	0.94	(0.59, 1.41)	-0.2	0.93	(0.60, 1.40)	-0.3
TBMT_H	-0.04	0.18	0.97	(0.57, 1.43)	-0.1	0.97	(0.59, 1.41)	-0.1
TBMT_J	-0.19	0.18	1.25	(0.59, 1.41)	1.2	1.26	(0.60, 1.40)	1.3
TBMT_K	-0.44	0.36	0.90	(0.58, 1.42)	-0.4	0.81	(0.60, 1.40)	-1.0

Separation Reliability = 0.83
Chi-square test of parameter equality = 18.23
df = 4
Significance level = 0.00

The results of the two-orthogonal factors models of TBM LOT are presented in Tables 5.6 (for the TBNMML subscales) and 5.7 (TBMT LOT subscale). The results of the two-orthogonal factors models of TBM HOT are presented in Table 5.8 (TBNMML HOT subscale) and Table 5.9 (TBMT HOT subscale). The results show that the separate models of the two-orthogonal factors of TBM LOT and the two-orthogonal factors of TBM HOT have a better fit than the two-correlated factors model. All items are within the acceptable range between 0.60 and 1.40. The separation reliabilities for the TBNMML_LOT, TBMT_LOT, TBNMML_HOT, TBMT_HOT subscale are high (close to 1 except for TBMT LOT which is 0.83). The EAP/PV reliability of the subscales are 0.80, 0.73, 0.81 and 0.75 for TBNMML_LOT, TBMT_LOT, TBNMML_HOT, and TBMT_HOT respectively. The values are all within a satisfactory value of 0.70, as suggested by Jähnig (2013).

Table 5.8

Response Model Parameter Estimates of Teacher Beliefs concerning Nature of Mathematics and Mathematics Learning related to HOT (TBNMML HOT)

Variables	Estimates	Error	Unweighted Fit			Weighted Fit		
			OUTFIT MNSQ	CI	*t*	INFIT MNSQ	CI	*t*
TBNM_E	-0.00	0.23	1.01	(0.59, 1.41)	0.1	0.92	(0.58, 1.42)	-0.3
TBNM_H	-0.11	0.23	1.27	(0.59, 1.41)	1.2	1.11	(0.58, 1.42)	0.6
TBNM_I	-0.54	0.23	1.02	(0.59, 1.41)	0.2	1.06	(0.60, 1.40)	0.3
TBML_A	-0.99	0.24	0.66	(0.59, 1.41)	-1.8	0.80	(0.62, 1.38)	-1.0
TBML_B	-0.55	0.23	0.75	(0.59, 1.41)	-1.2	0.78	(0.60, 1.40)	-1.1
TBML_C	-0.29	0.23	0.91	(0.59, 1.41)	-0.4	0.95	(0.58, 1.42)	-0.2
TBML_D	1.44	0.23	1.15	(0.59, 1.41)	0.8	1.05	(0.55, 1.45)	0.3
TBML_F	1.14	0.23	1.39	(0.59, 1.41)	1.7	1.33	(0.55, 1.45)	1.4
TBML_G	-0.11	0.66	0.86	(0.59, 1.41)	-0.6	0.90	(0.58, 1.42)	-0.4

Separation Reliability = 0.93
Chi-square test of parameter equality = 92.29
df = 8
Significance level = 0.00

Table 5.9

Response Model Parameter Estimates of Teacher Beliefs concerning Mathematics Teaching related to HOT (TBMT HOT)

Variables	Estimates	Error	Unweighted Fit			Weighted Fit		
			OUTFIT MNSQ	CI	*t*	INFIT MNSQ	CI	*t*
TBMT_A	0.38	0.23	1.40	(0.59, 1.41)	1.8	1.28	(0.47, 1.53)	1.0
TBMT_C	-0.47	0.23	1.01	(0.59, 1.41)	0.1	1.09	(0.55, 1.45)	0.5
TBMT_D	-1.14	0.23	0.64	(0.59, 1.41)	-1.9	0.69	(0.61, 1.39)	-1.7
TBMT_B	0.88	0.22	1.33	(0.58, 1.42)	1.5	1.17	(0.47, 1.53)	0.7
TBMT_F	0.17	0.23	0.65	(0.59, 1.41)	-1.8	0.63	(0.48, 1.52)	-1.5
TBMT_I	0.17	0.51	0.78	(0.59, 1.41)	-1.0	0.75	(0.48, 1.52)	-0.9

Separation Reliability = 0.92
Chi-square test of parameter equality = 46.82
df = 5
Significance level = 0.00

5.2.3. Summary of Teacher Beliefs concerning Mathematics (TBM) Instrument

After examining the results of the CFA and Rasch analysis, one item (TBNM_D) is excluded as the loading is below 0.32. It is concluded that the best fitting models are the separate models, the two-orthogonal factors model of TBM LOT and the two-orthogonal factors model of TBM HOT. These are the models used in the subsequent analyses.

5.3 Instructional Activities for Students (IAS) Instrument

There are 10 items in the Instructional Activities (IAS) instrument, are self-reported by the participants. The instrument endeavours to provide a picture of the sorts of task the teachers expect their students to participate in while in their classroom. The IAS items are adapted from the 2011 TIMSS teacher questionnaire. The responses are measured using a five-point Likert scale with the choices being: 'never' (1), 'rarely' (2), 'sometimes' (3), 'usually' (4) and 'always' (5). All items are positive with respect to the intended scale. Table 5.10 presents the IAS items, the item labels and the item texts.

Table 5.10

Instructional Activities Approach for Students (IAS) Instrument

Item label	Item text
IAS_A	Listen to me explaining how to solve problems
IAS_B	Memorise rules, procedures and facts
IAS_C	Work problems with peers with my guidance
IAS_D	Work problems together in whole class with direct guidance from me
IAS_E	Apply fact, concepts and procedures to solve routine problems
IAS_F	Explain their answer
IAS_G	Relate what they are learning in mathematics to their daily lives
IAS_H	Decide on their own procedures for solving complex problems
IAS_I	Work on problems for which there is no immediate obvious method of solution
IAS_J	Take a written test or quiz

5.3.1. Confirmatory Factor Analysis (CFA) of IAS Instrument

The single-factor model and two-factor models are examined. However, only the single-factor model is examined in detail as the fit indices for the two-factors model are too low for further consideration in the analysis. Figure 5.5 presents the models examined. The factor loadings for all items and the model fit indices are recorded in Tables 5.11 and 5.12.

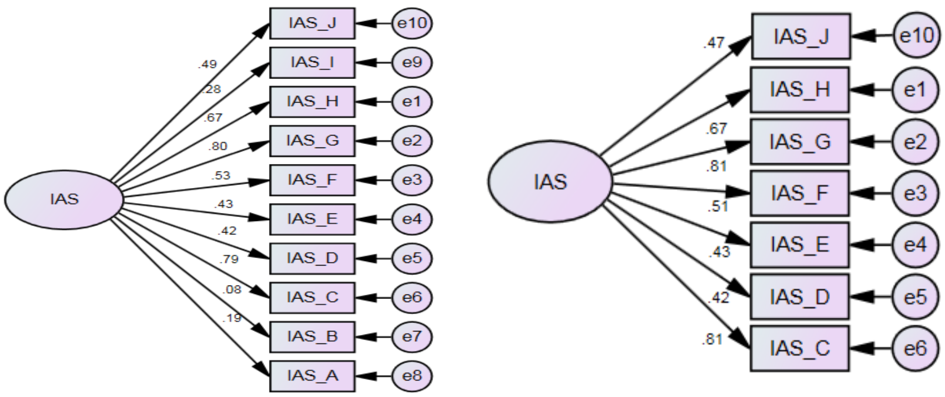


Figure 5.5 Single-factor model with 10 items and 7 items of the IAS

Table 5.11

Factor Loading of Items, Average Variance Extracted (AVE) and Composite Reliability (CR) of Single-factor Model of IAS

Variable	Single-factor model 10 items			Single-factor model 8 items		
	Loading	AVE	CR	Loading	AVE	CR
IAS_A	0.19			deleted		
IAS_B	0.08			deleted		
IAS_C	0.79			0.81		
IAS_D	0.42			0.42		
IAS_E	0.43			0.43		
IAS_F	0.53	0.27	0.75	0.51	0.37	0.79
IAS_G	0.80			0.81		
IAS_H	0.67			0.67		
IAS_I	0.28			deleted		
IAS_J	0.49			0.47		

Table 5.11 records the factor loadings for all items in the IAS instrument. Three items have low factor loadings (below the cut-off point of 0.32), namely items IAS_A,

IAS_B and IAS_I. The remaining items have acceptable loadings. The single-factor model is then re-examined with the exclusion of items A, B and I. In the single-factor model with seven items, all factor loadings fall within the acceptable range (above 0.30). The composite reliability value is acceptable (above 0.70), with the AVE of the scale higher but the value is still below 0.50. The summary of the model fit indices for the two models is recorded in Table 5.12. The second model shows better fit than that of the first model (CFI > 0.9, TLI and GFI is close to 0.9 and RMSEA < 1). The second model of IAS with seven items is to be examined in a Rasch analysis.

Table 5.12

Summary of Fit Indices for Single-factor Model with 10 Items and 7 Items of IAS

Indices	Single-factor model 10 items	Single-factor model 7 items
Chi-square (χ^2)	75.79	19.73
Degree of freedom (*df*)	35	14
χ^2/df	2.17	1.41
Goodness of Fix Index (GFI)	0.75	0.89
Adjusted Goodness of Fix Index (AGFI)	0.61	0.78
Parsimony Goodness of Fix Index (PGFI)	0.48	0.44
Tucker-Lewis Index (TLI)	0.39	0.84
Comparative Fit Index (CFI)	0.61	0.92
Root Mean Square Error of Approximation (RMSEA)	0.16	0.09

5.3.2. Rasch Analysis for IAS Instrument

A Rasch analysis is conducted for the second model of IAS with seven items. Table 5.13 records the summary of the measurement model parameter estimates of the IAS instrument, including INFIT and OUTFIT MNSQ for all items. Table 5.13 presents the item analysis of IAS with seven items. All items are just within the acceptable range of 0.60 to 1.40 and are accepted. IAS_J is the only item with the INFIT MNSQ slightly above 1.40 but as the OUTFIT MNSQ is still within the acceptable range the item is retained. The separation reliability for the IAS scale of seven items is high

(0.91). While the EAP/PV reliability is lower than the separation reliability (0.75), it is acceptable (above 0.70).

Table 5.13

Response Model Parameter Estimates of IAS Instrument for 7 Items of IAS

Variables	Estimates	Error	Unweighted Fit			Weighted Fit		
			OUTFIT MNSQ	CI	*t*	INFIT MNSQ	CI	*t*
IAS_C	0.11	0.15	0.7	(0.59, 1.41)	-1.6	0.72	(0.58, 1.42)	-1.3
IAS_D	-0.18	0.15	1.40	(0.59, 1.41)	1.9	1.24	(0.58, 1.42)	1.1
IAS_E	0.07	0.15	1.28	(0.59, 1.41)	1.3	1.18	(0.58, 1.42)	0.8
IAS_F	-0.21	0.15	0.94	(0.59, 1.41)	-0.2	1.02	(0.57, 1.43)	0.2
IAS_G	-0.41	0.15	0.92	(0.59, 1.41)	-0.3	0.99	(0.57, 1.43)	0
IAS_H	1.02	0.14	1.13	(0.59, 1.41)	0.7	1.14	(0.59, 1.41)	0.7
IAS_J	-0.40	0.36	1.40	(0.59, 1.41)	1.9	1.52	(0.57, 1.43)	2.1

Separation Reliability = 0.91
Chi-square test of parameter equality = 66.02
df = 6
Significance level = 0.00

5.3.3. Summary of IAS Instrument

Based on the result of the CFA and Rasch analysis, it is concluded that the single-factor model with seven items is the best model for IAS and therefore three items are excluded: IAS_A, IAS_B, and IAS_I. The single-factor model with seven items is used for the next stage of the analyses.

5.4 Teacher Engaging Student (TES) Instrument

Teacher Engaging Student (TES) is the third scale included in the classroom practice section of the teacher questionnaire. The instrument endeavours to provide a picture of the methods the teachers use to involve and encourage their students into classroom activities. The instrument consists of six items, adapted from the 2011 TIMSS Teacher Questionnaire. The responses involve a five-point Likert scale of ‘never’, ‘rarely’, ‘sometimes’, ‘usually’ and ‘always’. Table 5.14 records the items of the TES instrument including the item label and item text. None of the items are negative, therefore none are recoded.

Table 5.14

Teacher Engaging Student (TES) Instrument

Item label	Item text
TES_A	Summarise what students should have learned from the lesson
TES_B	Relate the lesson to student’s daily life
TES_C	Use questioning to elicit reasons and explanation
TES_D	Encourage all students to improve their performance
TES_E	Praise students for good effort
TES_F	Bring interesting materials to class

5.4.1. Confirmatory Factor Analysis (CFA) of TES Instrument

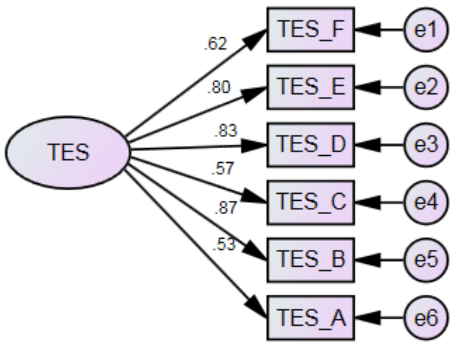


Figure 5.6 Single-factor model of TES

CFA is carried out for the items of TES. As TES consists of only one scale, the only model examined is a single-factor model. Figure 5.6 shows the model and Table 5.15 shows the factor loadings, average variance extracted (AVE) and composite reliability (CR). All six items of the TES instrument have acceptable factor loadings, all are above 0.32. The AVE is acceptable (above 0.50) and the reliability is satisfactory (above 0.70). The model fit indices recorded in Table 5.16 indicate that the single-factor model has a moderately good fit, with both the CFI and TLI being close to 0.90. However, the RMSEA falls outside the acceptable range (above point 0.08).

Table 5.15

Loadings of items, Average Variance Extracted (AVE), and Composite Reliability (CR) of Single-factor Model of TES

Variable	Loading	AVE	CR
TES_A	0.53	0.51	0.86
TES_B	0.87		
TES_C	0.57		
TES_D	0.83		
TES_E	0.80		
TES_F	0.62		

Table 5.16

Summary of Fit Indices for Single-factor Model of TES

Indices	Single-factor model
Chi-square (χ^2)	22.16
Degree of freedom (*df*)	9
χ^2/df	2.47
Goodness of Fix Index (GFI)	0.87
Adjusted Goodness of Fix Index (AGFI)	0.69
Parsimony Goodness of Fix Index (PGFI)	0.37
Tucker-Lewis Index (TLI)	0.82
Comparative Fit Index (CFI)	0.89
Root Mean Square Error of Approximation (RMSEA)	0.18

5.4.2. Rasch Analysis for TES Instrument

The Rasch analysis for the TES instrument is recorded in Table 5.17 with the INFIT and OUTFIT MNSQ presented for all items.

Table 5.17

Response Model Parameter Estimates of TES

Variables	Estimates	Error	Unweighted Fit			Weighted Fit		
			OUTFIT MNSQ	CI	*t*	INFIT MNSQ	CI	*t*
TES_A	0.20	0.16	1.40	(0.59, 1.41)	1.9	1.83	(0.54, 1.46)	2.90
TES_B	-0.11	0.16	0.94	(0.59, 1.41)	-0.2	1.09	(0.51, 1.49)	0.40
TES_C	0.53	0.15	1.01	(0.59, 1.41)	0.1	1.06	(0.56, 1.44)	0.30
TES_D	-1.40	0.18	0.63	(0.59, 1.41)	-2	1.2	(0.33, 1.67)	0.70
TES_E	-0.62	0.17	0.66	(0.59, 1.41)	-1.8	0.84	(0.46, 1.54)	-0.50
TES_F	1.41	0.36	0.95	(0.59, 1.41)	-0.2	1.00	(0.59, 1.41)	0.10

Separation Reliability = 0.95
Chi-square test of parameter equality = 90.76
df = 5
Significance level = 0.00

The INFIT and OUTFIT MNSQ of most items fall within the acceptable range (0.60 to 1.40), except for TES_A (the INFIT MNSQ is 1.83, but the OUTFIT MNSQ is acceptable). The separation reliabilities for the TES scale is high (0.95), and the EAP/PV reliability is also high (0.84), with both values within the satisfactory range and above the of the required cut-off value of 0.70 (Jähnig, 2013).

5.4.3. Summary of TES Instrument

Based on the result of CFA and Rasch analysis, none of the items is excluded. Even though one item (TES_A) has an INFIT MNSQ that falls outside the acceptable range, the item is retained as the loading in the CFA is above 0.32 and the Rasch analysis shows good reliability. The decision is made to use the single factor model of TES for the subsequent analyses.

5.5 Summary

The validation of the three instruments included in the teacher questionnaire is discussed in this chapter. The three instruments involve (a) Teacher Beliefs concerning Mathematics (TBM), (b) Instructional Activities Approach used for Students (IAS), and (c) Teacher Engaging Student (TES). The same processes are employed for each instrument. The validation process involves CFA and Rasch analysis. The final models to be used for the subsequent analyses are based on the results of the CFA and the Rasch analysis. The following models are used in subsequent analyses.

1. Teachers Beliefs concerning Mathematics (TBM)

   TBM measures teacher beliefs concerning the nature of mathematics and mathematics learning as well as mathematics teaching related toLOT and HOT. The best model for further analysis involves the two separate models of beliefs related to LOT and HOT namely: two-orthogonal factors model of TBM LOT and two-orthogonal factors model of TBM HOT.

2. Instructional Activities for student (IAS)

   IAS measures the instructional activities used by teachers in teaching mathematics. The best model is the single-factor model of seven items and is used for the next stage of analysis.

3. Teacher Engaging Student (TES)

   TES measures ways the teachers endeavour to engage students in the mathematics classroom. The best model for further analysis is the single-factor model.

Chapter 6 Descriptive Analysis and Contextual Information

6.1 Introduction

Based on the background data obtained from school, teacher and student questionnaires as well as a mathematics test, this chapter presents and discusses the descriptive analysis of these data and contextual information. The analyses reported in this chapter are conducted using the IBM SPSS 22. This preliminary information provides the basis for the next level of analysis which includes path analysis and hierarchical linear modelling.

6.2 School Demographic Information

The school data consist of the total enrolment, admission criteria, classroom classification, learning resources available and also any additional mathematics programmes available in the schools.

6.2.1. Types of Schools

A summary of the schools involved in the study is presented in Table 6.1. Four types of schools are managed by the Ministry of Education and Culture: general public schools, general private schools, some Islamic private schools and other religious private schools (such as Catholic private schools). The Ministry of Religious Affairs of the Republic of Indonesia usually manages both public and private Islamic schools. The general public schools and the Islamic public schools are the most numerous.

Table 6.1

Types of Schools Involved in the Study (N=25)

Type of School	Frequency		Per cent
	Urban	Rural	
General public school	4	4	32.0
General private school	2	0	8.0
Islamic private school[a]	1	2	12.0
Islamic public school[b]	4	4	32.0
Islamic private school	2	2	16.0
Total	13	12	100.0

Note. [a]the Islamic private school under the Ministry of Education and Culture
[b]the Islamic private school under the Ministry of Religious Affairs

6.2.2. Total Enrolment

The total number of students enrolled in the schools when the research was conducted is presented in Table 6.2. The average school enrolment is 400 students. Three schools (*n*=3, 12%) have below 100 students and four schools (*n*=4, 16%) have enrolments of more than 500 students. The school location, urban or rural is also indicated.

Table 6.2

Total Enrolment in Schools (N=25)

Total enrolment	Frequency		Per cent
	Urban	Rural	
Less than 100	3	0	12
100 to 299	2	5	28
300 to 499	3	5	32
500 to 699	1	2	12
700 to 899	3	0	12
More than 900	1	0	4
Total	13	12	100

6.2.3. School Admission Criteria

Table 6.3 presents the school admission criteria. Each school has its own criteria for the admission of the students and may include the location of student's residence, academic record and achievement on a placement test. However, these criteria are not necessarily applied in all schools and some schools apply more than one criteria. The placement test is the most common criterion applied by the school (*n*=19, 76%), followed by the students' academic record (*n*=13, 52%), location of residence is applied in five schools (*n*=5, 20%).

Table 6.3

School Admission Criteria (N=25)

Criteria		Frequency		Per cent
		Urban	Rural	
Residence in a particular area	No	9	11	80
	Yes	4	1	20
Students' academic record	No	7	5	48
	Yes	6	7	52
Placement test	No	6	0	24
	Yes	7	12	76

6.2.4. Class Size and Class Classification

The data on class size is presented in Figure 6.1. The most common class size in both rural and urban schools is between 26 and 30 students, although 31-35 students is also very common in urban schools.

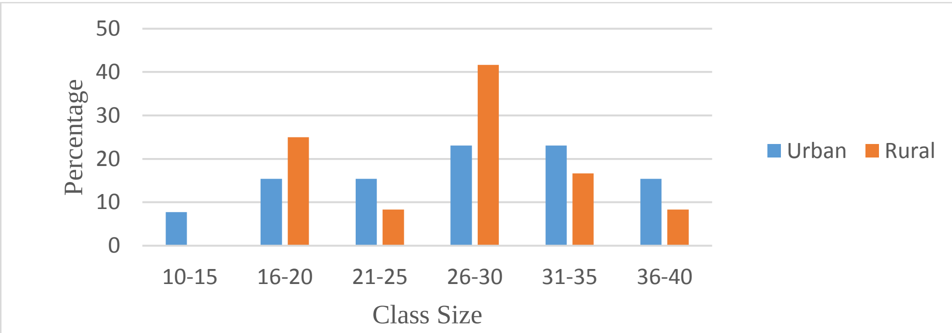


Figure 6.1 Schools class size distribution (*N*=25)

Figure 6.2 records the classification of classes within schools. The most common method of classification is a mixed grouping of students by their ability within their classes (*n*=13, 52%), where the schools spread the high achieving students and low achieving students in each class. Eight schools (32%) group students in classes based on their ability, indicating that high achieving students are grouped together in one class and low achieving students are grouped in another class. Four schools (*n*=4, 16%) do not employ any special method for classification, with the students being assigned to classes randomly.

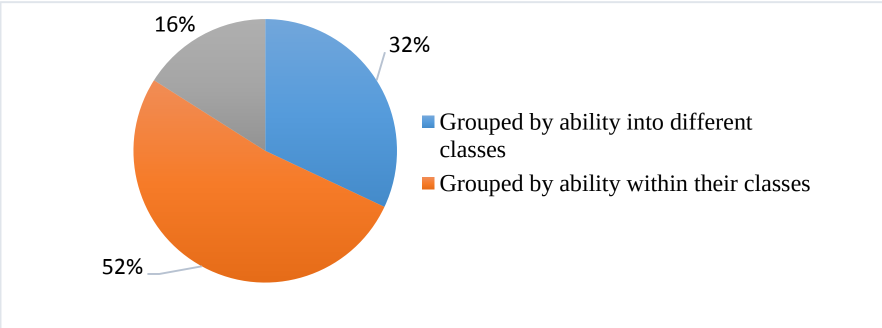


Figure 6.2 Class classification methods within schools (*N*=25)

6.2.5. Number of Teachers

Table 6.4 records information on the total number of teachers in each school, the total number of mathematics teachers for Year 9, and the number of teachers with certification. The number of teachers in schools is closely related to the total enrolment of the schools in both urban and rural schools.

Table 6.4

Characteristics of the Teachers in Schools (N=25)

	Minimum	Maximum	Mean	Std. Error	Std. Deviation
Total number of teachers	12	67	35.72	3.12	15.61
Total number of teacher with certification	0	65	24.96	4.77	23.87
Total number of mathematics teachers	1	8	3.80	0.40	1.98
Total number of mathematics teachers with certification	0	7	2.88	0.43	2.17
Total number of mathematics teachers teaching Year 9	1	3	1.68	0.14	0.75

6.2.6. School Resources

Table 6.5 presents the teaching and learning resources for mathematics available in the schools. Twenty four of the school (*n*=24, 96%) have mathematics textbooks and 21

(84%) have concrete materials to help students understand mathematics better. Eleven schools (*n*=11, 44%) have library materials related to mathematics (other than the mathematics textbooks). However, only four schools (*n*=4, 16%) have computers available for use in mathematics teaching and only three (*n*=3, 12%) have an internet connection. No school in the study has computer software for mathematics instruction.

Table 6.5

Schools Having Specific Teaching and Learning Resources (N=25)

	Frequency		Per cent
	Urban	Rural	
Computer for mathematics teaching (SR_A)	3	1	16
Computer with internet connection for mathematics teaching (SR_B)	3	0	12
Concrete material for mathematics teaching (SR_C)	11	10	84
Computer software for mathematics instruction (SR_D)	0	0	0
Library material relevant to mathematics instruction (SR_E)	5	6	44
Mathematics text books (SR_F)	12	12	96

Of 13 urban schools, three (*n*=3, 23%) have computers for mathematics instruction but only one out of 12 rural schools (*n*=1, 8.33%) has. Only three urban schools (*n*=3, 23%) have an internet connection while none of the rural schools have an internet connection. Twelve urban schools (*n*=12, 92.30%) and 12 rural schools (*n*=12, 100%) have textbooks available, with 11 urban schools (*n*=11, 91.7%) and 12 rural schools (92.30%) having concrete materials. However, only five urban (*n*=5, 38.50%) and six rural schools (*n*=5, 50%) have library material for mathematics learning, excluding the textbooks in use.

6.2.7. Additional Mathematics Programme

Additional mathematics programmes include: (a) extra lessons given at schools for the students in after school hours, and (b) an “Mathematics Olympiad Club”, a special programme organised by the school for training students who have a greater interest in advanced mathematics or are willing to participate in a mathematics competition. Table 6.6 presents information on an additional mathematics programme, while the number of urban and rural schools that have additional mathematics program after school hours are similar, more urban schools have a 'Mathematics Olympiad Club'.

Table 6.6

Additional Mathematics Programme at Schools (N=25)

		Frequency		Per cent
		Urban	Rural	
Additional mathematics programme after school	No	3	3	24
	Yes	10	9	76
Mathematics Olympiad Club	No	8	10	72
	Yes	5	2	28

6.3 Teachers' Information

Teachers' background, teachers' beliefs and their classroom practices are presented. Teachers' background considers gender, age, education, years of teaching, and professional development. Teachers' beliefs include beliefs concerning mathematics related to LOT and HOT. The teachers' classroom practices include their instructional activities, how the teachers engage their students, the types of problems used in the classroom and in the assessment, as well as the resources used. Independent *t*-tests related to school location are conducted for teachers' beliefs, classroom practices and teaching resources with the significant results being recorded in this chapter (a detailed report of the independent *t*-test is provided in Appendix I).

6.3.1. Teachers' Background

Teachers' Gender and Teaching Experience

The description of teacher gender and teaching experience are presented in Table 6.7. The total number of teacher respondents in this study is 46 consisting of 39 female (85%) and 7 male (15%). Teachers' teaching experience is assessed by the length of time they have been teaching the subject of mathematics. Most teachers (*n*=25, 54%) have been teaching for between 10 and 20 years. More than 30% of teachers have less than ten years teaching experience, with the majority of these teachers being in the rural schools (*n*=11, 73%).

Table 6.7

Teachers' Gender and Years of Teaching (N=46)

	Frequency		Per cent
	Urban	Rural	
Gender			
Female	18	21	84.80
Male	5	2	15.20
Total	23	23	100
Years of Teaching			
Less than 10 years	4	11	33
10 to 20 years	16	9	54
21 to 30 years	1	3	9
31 to 40 years	2	0	4
Total	23	23	100

Educational Background

The descriptive data on teachers' educational background and major are presented in Table 6.8. Almost all mathematics teachers included in this study hold a Bachelor degree (*n*=44, 96%), with just two (*n*=2, 4%) having completed a Master degree (one urban and one rural teacher). In terms of their major, 44 (*n*=44, 96%) of the teachers majored in mathematics education, one (*n*=1, 2%) majored in pure mathematics and one (*n*=1, 2%) majored in another subject.

Table 6.8

Teachers' Education and Major (N=46)

	Frequency	Per cent
Teachers' education		
Completed bachelor degree	44	96
Completed master degree	2	4
Total	46	100
Teachers' major		
Mathematics education	44	95.70
Pure Mathematics	1	2.15
Others	1	2.15
Total	46	100

Teachers' Professional Development and Certification

Teachers' professional development is assessed by their participation in training and seminars during the last two years. Regular professional development programmes are known as 'mathematics teacher forums' and are usually organised within regions. Table 6.9 summarises the regularity of the attendance at the professional development programmes.

Table 6.9

Regularity of Attending Professional Development Programme for Mathematics Teachers (N=46)

	Frequency		Per cent
	Urban	Rural	
Every 2 weeks	15	2	37.0
Every month	1	1	4.3
Every 2 months	0	1	2.2
Every 3 months	1	1	4.3
Every 6 months	1	4	10.9
Less regularly	5	14	41.3
Total	23	23	100.0

More than one-third (n=17, 37%) of teachers attended forums fortnightly, with approximately ten per cent (n=5) saying that in their area they take place between one and six monthly intervals. However, more than 41% (n=19) report that the forum is organised less regularly in their area (more than 70% who report this are from a rural area).

Table 6.10

Teachers' Professional Development (N=46)

Professional development	Frequency		Per cent
	Urban	Rural	
Mathematics content	10	13	50
Mathematics pedagogy/instruction	10	13	50
Mathematics curriculum	9	13	47.8
Integrating information technology to mathematics	6	8	30.4
Improving students critical thinking and problem	6	6	26.1
Mathematics assessment	9	7	34.8
Addressing individual students needs	4	1	10.9

Table 6.10 shows that half of the respondents (*n*=23, 50%) reported that they had attended professional development programme related to mathematics content, pedagogy and mathematics curriculum during the last two years. A lesser number of teachers stated that they had been in a professional development programme related to: (a) integrating information technology to mathematics (*n*=14, 30%); (b) improving students' critical thinking and problem solving (*n*=12, 26%); and (c) mathematics assessment (*n*=16, 35%). Only five teachers (*n*=5, 11%) reported that they had attended professional development programmes related to addressing individual students' needs.

The majority of the teachers (*n*=43, 74%) participating in the study had passed the certification test. The number of mathematics teachers with certification in urban areas is 25% higher than in rural areas. The information is presented on Table 6.11.

Table 6.11

Teachers' Certification (N=46)

	Frequency		Per cent
	Urban	Rural	
Not certified	3	9	26.1
Certified	20	15	73.9
Total	23	23	100.0

6.3.2. Teachers' Beliefs Concerning Mathematics

Teachers were asked about their beliefs concerning mathematics related to LOT and HOT.

Beliefs concerning Mathematics related to LOT

The summary of the index of teacher beliefs concerning mathematics related to LOT is presented on Table 6.12. An index of beliefs concerning mathematics related to LOT is created from 13 items (refer to Chapter 3). The average is computed across the 13 items based on a four-point Likert scale: 'disagree a lot' =1, 'disagree a little' =2, 'agree a little' = 3, and 'agree a lot' = 4. A highly positive level is indicated by an average score of more than or equal to 3, corresponding to teachers agreeing with the statements 'a little' or 'a lot'. A negative level is indicated by an average score of equal to or less than 2, corresponding to teachers disagreeing with the statements 'a little' or 'a lot'. A somewhat positive level is indicated by an average score of greater than 2 but less than 3.

Table 6.12

Index of Teacher Beliefs concerning Mathematics related to LOT (N=46)

	Frequency		Per cent
	Urban	Rural	
Negative	9	5	30
Somewhat positive	13	18	68
Highly Positive	1	0	2
Total	23	23	100

Most teachers (*n*=31, 68%) indicate that they hold somewhat positive beliefs concerning mathematics related to LOT. Only two per cent of teachers hold highly positive beliefs (*n*=1, 2%) and nearly a third (*n*=14, 30%) hold negative beliefs concerning mathematics related to LOT.

Beliefs concerning Mathematics related to HOT

An index of beliefs concerning mathematics related to HOT is created from 16 items (refer to Chapter 3). The average is computed across the 16 items based on a four-category Likert-scale : ‘disagree a lot’ =1, ‘disagree a little’ =2, ‘agree a little’ = 3, and ‘agree a lot’ = 4. A highly positive level indicates an average score of more than or equal to 3, corresponding to teachers’ agreeing with the statements ‘a little’ or ‘a lot’. The procedure of creating the index follows the procedure used in the previous section. The summary of the index of teacher beliefs concerning mathematics related to HOT is presented in Table 6.13.

Table 6.13

Index of Teacher Beliefs concerning Mathematics related to HOT (N=46)

	Frequency		Per cent
	Urban	Rural	
Negative			
Somewhat positive	3	5	17
Highly Positive	20	18	83
Total	23	23	100

Most teachers (*n*=38, 83%) have highly positive, and a small number of teachers (*n*=8, 17%) have somewhat positive beliefs concerning mathematics related to HOT. None of the teachers has negative beliefs concerning mathematics related to HOT.

6.3.3. Classroom Practices

The data concerning classroom practices focus on activities in the mathematics classroom including the types of problems used during the lesson and in testing, instructional activities, how students are engaged, and the resources used.

Types of Mathematics Problems Used in Classroom

Data was collected to assess the regularity with which teachers presented particular types of mathematical problems to the students during lessons and testing. The problems are itemised as:

1. Problems similar to those demonstrated by the teacher (TQ_A).
2. Problems involving the application of mathematical procedures (TQ_B).
3. Problems which can be solved in many ways (TQ_C).
4. Unfamiliar problems (TQ_D).
5. Open-ended questions (TQ_E).
6. Word problems (TQ_F).

The responses of the questions are assessed by a six-category Likert-scale ranging from never (0) to always (5). The information on the regularity of the types of problems used in the lesson is presented in Table 6.14. The most frequent types of problems used are: the problem that can be solved in many ways (TQ_C) (*M*= 4.07, *SE*=0.13); word problems (TQ_F) (*M*=3.93, *SE*=0.12); problems involving the application of mathematical procedures (TQ_B) (*M*=3.91, *SE*=0.16); and problems similar to those demonstrated by the teacher (TQ_A) (*M*=3.73, *SE*=0.19). The least frequently used problem is that involving unfamiliar problems (TQ_D) (*M*=2.64, *SE*=0.16); followed by open-ended questions (TQ_E) (*M*= 3.49, *SE*=0.15). The regularity of the type of questions in mathematics classroom is similar across rural and urban school.

Table 6.15 presents information on the frequency of types of problems used in mathematics tests. Similar to the types of problems used during the lesson, questions involving unfamiliar problems (TQ_D) (*M*=2.74, *SE*=0.17) and open-ended questions (TQ_E) (*M*=3.16, *SE*=0.20) are least frequently used in testing. Word problems

(TQ_F) (*M*=3.93, *SE*=0.13) and questions involving the application of mathematical procedures (TQ_B) (*M*=3.80, *SE*=0.15) are the most frequent problems used in testing.

Table 6.14

Regularity of Types of Questions used in Mathematics Classroom

	N	Mean	Std. Error	Std. Deviation
TQ_A	44	3.73	0.19	1.23
TQ_B	45	3.91	0.16	1.06
TQ_C	45	4.07	0.13	0.86
TQ_D	44	2.64	0.16	1.08
TQ_E	45	3.49	0.15	1.04
TQ_F	45	3.93	0.12	0.81

Questions similar to those solved by the teacher in the classroom (TQ_A) are used more frequently during the lesson (*M*=3.73, *SE*=0.19) than in tests (*M*= 3.44, *SE*=0.21). While the trend of regularity of using the types of problems in the test is similar across the urban and rural school, teachers in rural schools (*M*=3.83, *SE*=0.30) use the questions similar to those the teachers have solved in the classroom (TQA) more frequently than teachers in the urban schools (*M*=2.91, *SE*=0.29). This difference is significant $t(44) = -2.2$, $p < 0.05$.

Table 6.15

Regularity of Types Questions used in Mathematics Examination

	N	Mean	Std. Error	Std. Deviation
TQ_A	45	3.44	0.21	1.39
TQ_B	45	3.80	0.15	1.04
TQ_C	45	3.78	0.15	1.04
TQ_D	43	2.74	0.17	1.12
TQ_E	45	3.16	0.20	1.36
TQ_F	45	3.93	0.13	0.86

Instructional Activities

Data related to instructional activities involve examining how often the teachers use the instructions for the students in their classrooms. The responses are assessed by a six-category Likert-scale ranging from never (0) to always (5). The instructional activities used by teachers in the classroom are presented by the following items (refer to Chapter 5).

1. Work problems with peers with my guidance (IAS_C).
2. Work problems together in whole class with direct guidance from me (IAS_D).
3. Apply fact, concepts and procedures to solve routine problem (IAS_E).
4. Explain their answer (IAS_F).
5. Relate what they are learning in mathematics to their daily lives (IAS_G).
6. Decide on their own procedures for solving complex problems (IAS_H).
7. Work on problems for which there is no immediate obvious method of solution (IAS_I).
8. Take written tests or quizzes (IAS_J)

Table 6.16

Instructional Approaches used for Students in the Mathematics Classroom

	N	Mean	Std. Error	Std. Deviation
IAS_C	46	3.96	0.13	0.87
IAS_D	46	4.09	0.14	0.92
IAS_E	46	4.02	0.13	0.86
IAS_F	45	4.13	0.13	0.84
IAS_G	46	4.22	0.12	0.84
IAS_H	46	3.37	0.17	1.12
IAS_I	41	2.24	0.20	1.28
IAS_J	46	4.20	0.14	0.96

The information on the frequency of use of instructional activities is presented in Table 6.16. The most frequently used approaches are relating mathematics to students' daily lives (IAS_G) (*M*=4.22, *SE*=0.12); giving tests or quizzes (IAS_J) (*M*=4.20, *SE*=0.14); and asking students to explain their answers (IAS_F) (*M*= 4.13, *SE*=0.13). Working on problems for which there is no immediately obvious method of solution (IAS_I) (*M*=2.24, *SE*=0.20) and asking students to decide on their own procedures for solving complex problems (IAS_H) (*M*=3.37, *SE*=0.17) are the least frequent approaches given to the students. The trends are similar across the locations.

Teacher Engaging Students

'Engaging students' reflects teachers' efforts in reinforcing concepts by summarising the lesson, relating lessons to daily life, and asking for reasoning from the students. It also involves feedback to students and bringing interesting materials to the class. The approaches used by teachers to engage students in the classroom are presented by the following items:

1. Summarise what students should have learned from the lesson (TES_A)
2. Relate the lesson to student's daily life (TES_B)
3. Use questioning to elicit reasons and explanation (TES_C)
4. Encourage all students to improve their performance (TES_D)
5. Praise students for good effort (TES_E)
6. Bring interesting materials to class (TES_F)

Table 6.17

Engaging Students (N=46)

	Mean	Std. Error	Std. Deviation
TES_A	4.37	0.14	0.95
TES_B	4.50	0.11	0.72
TES_C	4.24	0.13	0.90
TES_D	4.76	0.08	0.57
TES_E	4.57	0.10	0.65
TES_F	3.76	0.13	0.90

The teachers' responses to the frequency with which they used these methods are measured by a six-category Likert-scale ranging from never (0) to always (5). Table

6.17 presents the information on teachers engaging students. Generally, all teachers used all these methods frequently. The mean for 5 out of the 6 items are more than 4, corresponding to 'nearly always'. The least frequently used approach is bringing interesting materials to the classroom (M=3.76, SE=0.13).

Teaching Resources

The question relating to teaching resources examines whether teachers use such things as textbooks, worksheets or workbooks, concrete objects and computer software as a supplement or basis of instruction. Figure 6.3 provides information indicating that textbooks are the basic means of instruction for three-quarters of the teachers (n=34, 74%). Concrete objects and worksheets are used mainly as a supplement by almost half of the teachers while the other half used them as the basis of instruction. The least commonly used resource is computer software. More than 50% of teachers (n=26) never use computer software for mathematics instruction, while 41% of teachers (n=19) use it as a supplement. Only one teacher (n=1, 2%) said that they use it as the basis of instruction. In terms of school location, the trends of the usage of textbooks, worksheets and concrete materials are similar across the areas. The gap of usage appears for the usage of computer software, teachers from an urban area use the computer software more frequently (M=0.65, SE=0.12) compared to those from a rural area (M=0.26, SE=0.09). An independent t-test confirms that the difference in the usage of computer software is significant across the locations, $t(44)=2.58$, $p<0.05$. It should be noted that only four of the school have computers available for mathematics instruction.

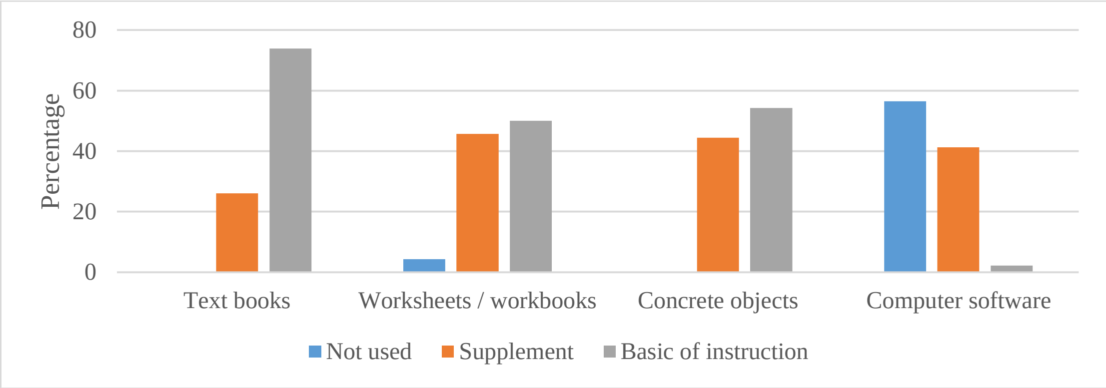


Figure 6.3 Percentage of teaching resources used in mathematics classroom (N=46)

6.4 Students' Information

Students' background, students' attitude, students' beliefs concerning mathematics and classroom practices as perceived by students are presented. Independent *t*-test in relation to school location and gender are also carried out. However, only significant results are explained in this chapter (the details report of the independent t-test are available in the Appendix I).

6.4.1. Students' Background

Gender and School Location

The total number of student respondents in this study is 1135: 505 (44.5%) are male and 630 (55.5%) female; 581(51.2%) are urban students and 554 (48.8%) are rural students. Table 6.18 provides the details concerning student gender and school location.

Table 6.18

Gender and School Location (N=1135)

		Gender of student		Total
		Female	Male	
School location	Urban	314	267	581
	Rural	316	238	554
Total		630	505	1135

Parents' Education

The highest levels of education achieved by the students' mothers and fathers are presented in Table 6.19. Approximately 73% of mothers (*n*=831) have completed Year 12 of schooling, with 40.9% (*n*=464) having that as their highest level, and 19.7% (*n*=224) holding bachelor degrees.

Table 6.19

Mothers' and Fathers' Highest Level of Education

	Mother		Father	
	Frequency	Per cent	Frequency	Per cent
Did not complete Year 6	21	1.90	19	1.70
Completed Year 6	82	7.20	105	9.30
Completed Year 9	182	16	156	13.70
Completed Year 12	464	40.90	441	38.90
Vocational or technical certificate after high school	14	1.20	11	1
Diploma	77	6.80	39	3.40
Bachelor's degree	224	19.70	222	19.60
Postgraduate	52	4.60	84	7.40
Valid N	1116	98.30	1077	94.90
Missing	19	1.70	58	5.10
N	1135	100	1135	100

The same trend occurs for fathers' education, with 70% (*n*=797) having completed Year 12, 38.9% (*n*=441) having that as their highest level, and 19.6% (*n*=222) holding bachelor degrees. More fathers (*n*=84, 7.4%) than mothers (*n*=52, 4.6%) have completed post graduate degrees.

The level of parental education varies with school location with urban parents being more educated than rural parents. Approximately three times the number of rural mothers completed Year 6 only (*n*=62, 11%); twice the number completed Year 9 only (*n*=113, 20%); and nearly one-fifth more completed Year 12 only (*n*=250, 45%). The number of urban mothers who have completed a bachelor degree (*n*=145, 25%) is almost double the number of rural mothers (*n*=79, 14%). Nearly five times more urban (*n*=43, 7%) mothers completed a postgraduate degree. Table 6.20 presents the mother's educational level.

Table 6.20

Mother's Educational Level and School Location (N=1116)

Mother's Education	School location			
	Urban		Rural	
	Frequency	Per cent	Frequency	Per cent
Did not complete Year 6	8	1.38	13	2.3
Completed Year 6	20	3.44	62	11.2
Completed Year 9	69	11.88	113	20.4
Completed Year 12	214	36.83	250	45.1
Vocational or technical certificate after high school	13	2.24	1	0.2
Diploma	60	10.33	17	3.1
Bachelor's degree	145	24.96	79	14.3
Postgraduate	43	7.40	9	1.6
Valid N	572	98.45	544	98.2
Missing	9	1.55	10	1.8
N	581	100	554	100

A similar trend occurs in fathers' education. The educational level of the urban fathers is higher than that of the rural fathers. The number of urban fathers who completed Year 6 only (*n*=19, 3%) is approximately a quarter the number of those in rural areas (*n*=86, 15.5%); those who only completed Year 9 (*n*=48.8%) is more than half of the number of those in rural areas (*n*=108, 19.50%).

Table 6.21

Fathers' Educational Level and School Location (N=1077)

Father's education	School location			
	Urban		Rural	
	Frequency	Per cent	Frequency	Per cent
Did not complete Year 6	5	0.90	14	2.50
Completed Year 6	19	3.30	86	15.50
Completed Year 9	48	8.30	108	19.50
Completed Year 12	214	36.80	227	41
Vocational or technical certificate after high school	11	1.90	0	0
Diploma	28	4.80	11	2
Bachelor's degree	158	27.20	64	11.60
Postgraduate	70	12	14	2.50
Valid N	553	95.20	524	94.60
Missing	28	4.80	30	5.40
N	581	100	554	100

The trend continues within the higher education sector. The number of urban fathers (*n*=158, 27%) who have a bachelor degree is more than double the number of rural fathers (*n*=64, 12%), while the number of urban fathers (*n*=70, 12%) with a postgraduate degree is five times the number of rural fathers (*n*=14, 2.5%). The detailed numbers of urban and rural fathers' education is presented in Table 6.21.

Home Possessions

Home possessions are divided into two categories. The first category involves several items of educational resources and general possessions in the students' home. The second category is the quantity of luxury items that are available in the student's home.

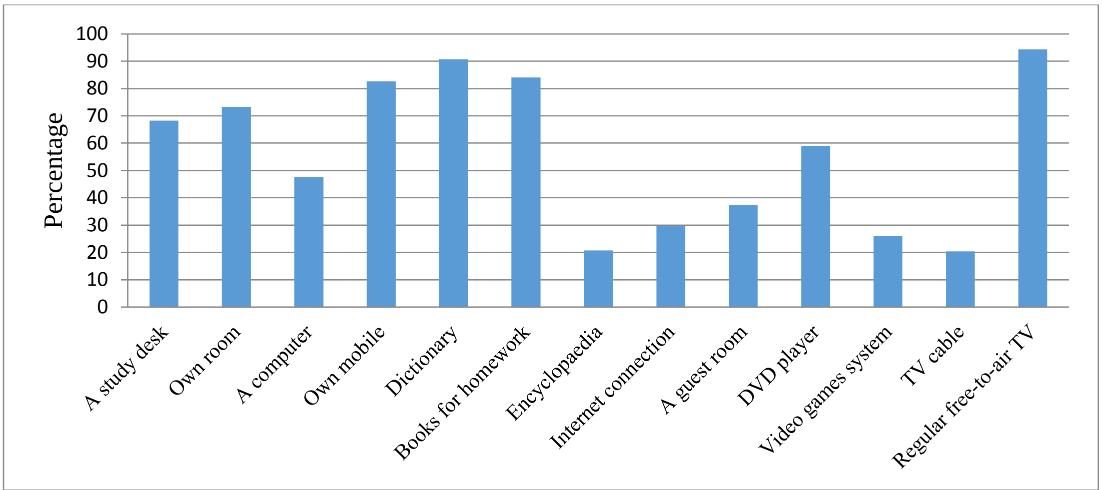


Figure 6.4 *Number of students who live in a home and the first category of home possessions (N=1135)*

Figure 6.4 present the data related to the category of home possessions. The majority of students have their own study desk (*n*=775, 68%), their own room (*n*=832, 73%), a mobile phone (*n*=939, 83%), a dictionary (*n*=1029, 90.7%) and books for doing homework (*n*=953, 84%). However, less than half of the students have access to a computer for doing homework (*n*=540, 48%) and less than a third have an internet connection at home (*n*=337, 30%). Even though most students say that they have books which they can use for their homework at home, only one-fifth (*n*=235, 21%) say they own an encyclopaedia. The data also show that most students have a regular free-to-air TV at home (*n*=1072, 94%).

Figure 6.5 provides further data for information of the number of each item in the luxury of home possessions. Television and mobile phone are the items that are commonly available in the student's home. Less than five per cent (*n*=127, 4.5%) of the total students report that there is no mobile phone at home and approximately 70% (*n*=801) report that there are at least three or more mobile phones in the household. Computers and cars are the least frequently owned items; nearly half of the homes (*n*=510, 45%) have no computer and 68 % (*n*=774) have no car.

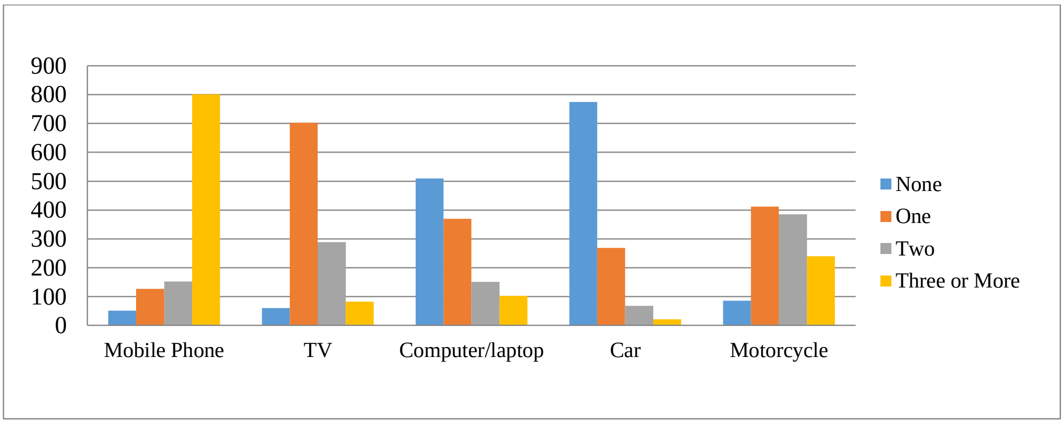


Figure 6.5 *Number students who have the second category of home possessions (N=1135)*

Figure 6.6 compares the first category of home possessions and student location in urban or rural areas. As with the educational levels of parents, in which urban parents have higher educational achievements than rural parents, the possessions in the home present a similar pattern.

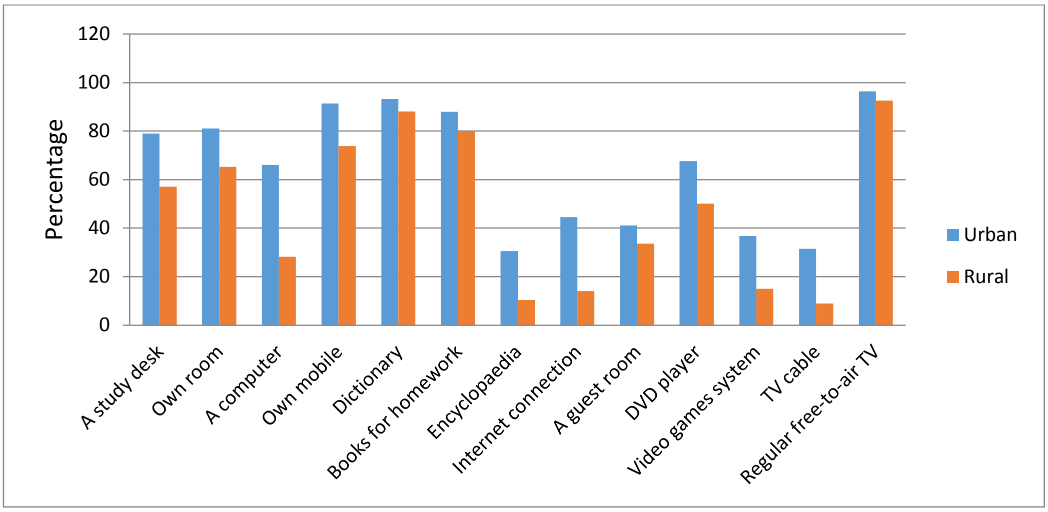


Figure 6.6 *First category of home possessions across school location (N=1135)*

The highest disparities in terms of ownership are with computers, internet connection and cable, encyclopaedias and video games. The numbers of the urban students owning a computer (*n*=384, 66%) is almost three times that of the rural students (*n*=156, 28%), while those with an internet connection are three times greater (*n*=259, 45% of urban and *n*=78, 14% of rural). The number of the rural students owning an encyclopaedia (*n*=57, 10%) is only a third that of the urban students (*n*=259, 45%).

Table 6.22

Second Category of Students' Home Possessions and School Location (N=1135)

Home possessions	None		One		Two		Three or More	
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
Mobile Phone	11	40	37	90	54	98	479	322
TV	18	42	292	410	198	91	73	9
Computer/laptop	164	346	217	153	112	39	88	14
Car	316	458	187	82	57	11	21	1
Motorcycle	36	50	183	229	220	175	142	98

Table 6.22 shows the second category home possessions according to locality. The quantity of luxury items owned by students' families from the urban area outnumber those owned by the rural students'. The number of urban students whose families own a car is almost triple (*n*=265, 46%) the number of rural students (*n*=95, 17%). While

this disparity is large, the gap in motorcycle ownership is not. This is likely to reflect the low purchase and maintenance costs of motorcycles.

Students' Educational Expectations

Students have a very high expectation of how far they want to go in schooling. The detailed numbers of students' educational expectations are presented in Table 6.23. Nearly 60% of the students (*n*=658) expect that that will be able to do a doctoral degree and less than five per cent of the students (*n*=56) see themselves finishing their education at the end of high school.

Table 6.23

Students' Educational Expectations (N=1114)

Students' educational expectation	Frequency	Per cent
High school	56	4.9
Vocational or technical education after high school	92	8.1
Diploma	11	1
Bachelor's degree	192	16.9
Master's degree	105	9.30
Doctoral degree	658	58.0
Valid N	1114	98.1
Missing	21	1.9
N	1135	100

Table 6.24 presents the data of students' expectation according to location. Generally, expectations are consistent across rural and urban areas. Most students from both areas expect to continue their education after high school. While more rural students are content with high school or some vocational course, overwhelmingly all students aim to graduate and go on to post graduate studies.

Table 6.24

Students' Educational Expectation across School Location (N=1135)

Students' educational expectation	Urban		Rural	
	Frequency	Per cent	Frequency	Per cent
High school	16	2.8	40	7.3
Vocational or technical education after high school	34	6.0	58	10.6
Diploma	7	1.2	4	0.7
Bachelor's degree	82	14.4	110	20.2
Master's degree	71	12.5	34	6.2
Doctoral degree	359	63.1	299	54.9

6.4.2. Students' Attitude

This study includes several questions to investigate student attitudes and beliefs towards mathematics. The attitudes include their liking of mathematics and valuing of mathematics, while the beliefs include their mathematical confidence and judgement, and their beliefs concerning the nature of mathematics and mathematical learning. Note that the procedure of creating the index in the liking mathematics is also used to obtain the index of all scales included in the student questionnaire.

Liking Mathematics

There are five items in the liking mathematics scale. Two of them are negatively worded and therefore recoded for analysis purposes. An average is calculated across the five items belonging to this scale (refer to Chapter 4, page 2), based on a four-category scale, namely: 'disagree a lot' =1, 'disagree a little' =2, 'agree a little' = 3, and 'agree a lot' = 4. A highly positive level indicates an average score of more than or equal to 3, and corresponds to students' agreeing with the statements 'a little' or 'a lot'. A negative level indicates an average score of equal to or less than 2, corresponds to students' disagreeing with the statements 'a little' or 'a lot'. A somewhat positive level indicates an average score of greater than 2 but less than 3.

Table 6.25 presents the summary of student liking mathematics. Most students (n=885, 78%) admit they really like mathematics, while fewer students (*n*=219, 19%) say they somewhat like mathematics and even fewer students reported to not liking mathematics (*n*=30, 2.60%). An independent *t*-test also indicates that there is a significant difference in the number of students liking mathematics between urban and rural areas, $t(1121) = -2.94$, $p<0.01$; students in rural areas (*M*=2.71, *SE*=0.02) have a higher positive attitude regarding liking mathematics. It is found that the difference between male and female of liking mathematics is also significant, $t(962) = 2.88$, $p<0.01$. Female students (*M*=2.79, *SE*=0.02) have a higher positive attitude regarding the liking of mathematics than male students (*M*=2.71, *SE*=0.02).

Table 6.25

Students' Liking of Mathematics (N=1134)

	Frequency		Per cent
	Urban	Rural	
Negative	19	11	2.60
Somewhat positive	129	90	19.30
Highly positive	433	452	78
Missing		1	0.10
Total	581	553	100

Valuing Mathematics

The valuing of mathematics scale consists of four items. The summary of the index of student valuing mathematics is presented in Table 6.26. The data indicates that, in general, most students (*n*=991, 87%) have a highly positive attitude in terms of valuing mathematics. They agree that they need mathematics for their daily life, for learning other subjects, entering university and getting a job. A small number of students (*n*=127, 11%) somewhat value mathematics, with less than two per cent of the students (*n*=17) indicating that they do not value mathematics.

Table 6.26

Students' Valuing Mathematics (N=1135)

	Frequency		Per cent
	Urban	Rural	
Negative	7	10	1.50
Somewhat positive	65	62	11.20
Highly positive	509	482	87.30
Total	581	554	100

In terms of school location and gender, both urban and rural students value mathematics positively and both female and male students also value mathematics positively. There is significant difference in terms of gender, $t(989) = 3.23$, $p<0.01$, with female students placing a higher positive value (M=2.89, SE=0.01) on mathematics than male students (M=2.82, SE=0.02).

6.4.3. Students' Self-efficacy

Mathematics Confidence

The mathematics confidence scale consists of five items and is administered to examine students' confidence in mathematics. Table 6.27 presents the summary of the index of students' mathematics confidence. Generally students (n=670, 59%) are confident about their mathematics ability. However, more than 17% of the students (n=88) indicate that they are not confident in their mathematics ability and more than 20% of the students (n=264) said they are very confident of their mathematics ability. Confidence in mathematics is similar in both urban and rural areas, with most students being confident and smaller numbers being very confident and not confident. The trend is also similar across gender.

Table 6.27

Students' Confidence in Mathematics (N=1133)

	Frequency		Per cent
	Urban	Rural	
Negative	111	88	17.60
Somewhat positive	346	324	59.10
Highly Positive	124	140	23.30
Missing		2	0.20
Total	581	552	100

Individual Judgement of Mathematics Abilities

Table 6.28 presents the information of the index of students' individual judgement of mathematics ability. Even though the scale of judgement of ability is similar to the mathematics confidence scale, the results are different. Nearly half consider themselves in the middle when it comes to ability (*n*=528; 47%). Thirty-two per cent of the students (*n*=365) judge themselves as having a high ability in mathematics while 21% (*n*=239) judge themselves as having a low ability in mathematics.

Table 6.28

Students' Individual Judgement of Mathematics Ability (N=1132)

	Frequency		Per cent
	Urban	Rural	
Negative	115	124	21.10
Somewhat positive	270	258	46.60
Highly Positive	194	171	32.20
Missing	2	1	0.30
Total	579	553	100

6.4.4. Students' Beliefs concerning Mathematics

Students were asked concerning their beliefs in relation to LOT and HOT in mathematics.

Belief concerning Mathematics related to LOT

The beliefs concerning mathematics related to HOT scale consists of 11 items and the index is summarised in Table 6.29. Nearly three-quarters of the students have somewhat positive beliefs (*n*=830, 73%). Some students have high positive beliefs (*n*=213, 18%) and only eight per cent of students have negative beliefs.

Table 6.29

Students' Beliefs concerning Mathematics related to LOT (N=1132)

	Frequency		Per cent
	Urban	Rural	
Negative	66	23	7.90
Somewhat positive	433	397	73.30
Highly positive	81	132	18.80
Missing	1	2	
Total	580	552	100

There is no significant difference indicated by an independent *t*-test for gender. However, there is a significant difference between school locations, $t(1129.15)=-5.80$, $p<0.05$, students from rural schools (M=2.20, SE=0.02) having more positive beliefs concerning mathematics related to LOT than the urban (M=2.03, SE=0.02).

Belief concerning Mathematics related to HOT

The beliefs concerning mathematics related to HOT scale consists of three items and the index is summarised in Table 6.30. Half students (*n*=592, 52%) have highly positive beliefs and a further large number (*n*=486, 43%) have somewhat positive beliefs, and a small number of students (*n*=56, 5%) have negative beliefs.

Table 6.30

Students' Beliefs concerning Mathematics related to HOT (N=1134)

	Frequency		Per cent
	Urban	Rural	
Negative	26	30	4.9
Somewhat positive	235	486	42.8
Highly positive	319	592	52.2
Missing	1		0.10
Total	580	554	100

6.4.5. Classroom Practices

The student questionnaire involved questions regarding the types of mathematical problems they usually work on, the learning resources they usually use and the activities they mostly do in mathematics classrooms.

Types of Mathematics Problems

The students are asked how often they work on particular types of mathematical problems in the classroom: questions similar to what the teacher solves in the classroom; questions applying their knowledge from a previous topic; questions which can be solved in many ways; and word problems. The responses are categorised in a three-category Likert-scale, 'never or almost never', 'sometimes' and 'always' and the summary of the mean is presented in Table 6.31. Based on the mean, it can be concluded that most students said that they sometimes work on these types of questions.

The frequency of types of questions students work on in the urban and rural area is slightly different. The urban reported that three types of questions (*M*=1.35, *SE*=0.02; *M*=1.25, *SE*=0.02; and *M*=1.49, *SE*=0.02 for SFWQ_A, SFWQ_B and SFWQ_C respectively) are used more frequently than in the rural (*M*=1.27, *SE*=0.02; *M*=1.16, *SE*=0.02; and *M*=1.41, *SE*=0.02 for SFWQ_A, SFWQ_B and SFWQ_C respectively).

The differences are significant for three items, $t(1131)=2.59$, $p<0.05$ for SFWQ_A; $t(1130.90)=2.64$, $p<0.05$ for SFWQ_B; and $t(1133)=2.44$, $p<0.05$ for SFWQ_C.

Table 6.31

Types of Questions in Mathematics Classroom

	N	Mean	Std. Error mean	Std. Deviation
Questions similar to what the teacher solves in the classroom (SFWQ_A)	1111	1.31	0.016	0.53
Questions which apply my knowledge from previous topics (SFWQ_B)	1093	1.21	0.017	0.57
Questions which can be solved in many ways (SFWQ_C)	1096	1.45	0.017	0.58
Word problems (SFWQ_D)	1100	1.17	0.016	0.54

Learning Resources

Students are asked of how often they used the following learning resources in their mathematics classroom. The responses consist of a three-category Likert-scale namely ‘never or almost never’, ‘sometimes’ and ‘always’.

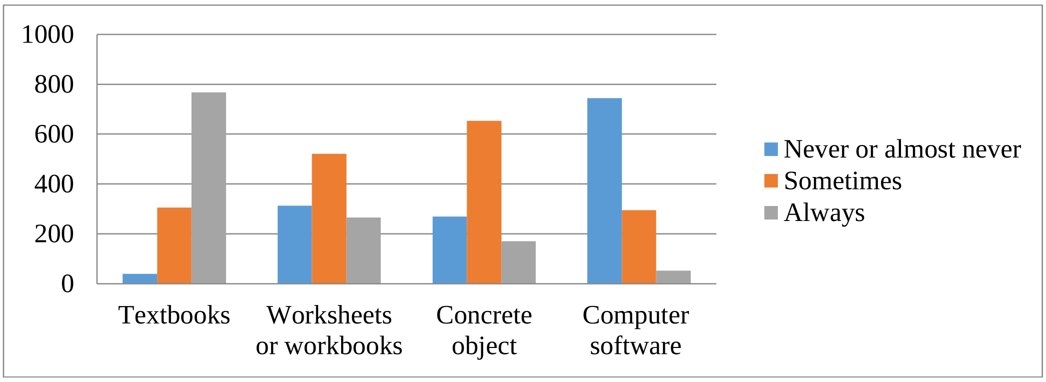


Figure 6.7 Learning resources used in mathematics classroom (*N*=1135)

Figure 6.7 shows that textbooks are the most frequently used learning resources followed by worksheets and concrete objects. Almost 70% of students (*n*=768) say they always use textbooks in mathematics classroom, while half of the students said that they sometimes use worksheets or workbooks (*n*=556, 49%) and concrete object (*n*=653, 57.5%) in the classroom. Computer software is the least frequent used resources in mathematics classroom with 66% of the students (*n*=744) saying they never or almost never use it.

Comparing rural and urban results, an independent *t*-test shows there is a significant difference for both the usage of concrete objects, $t(1133)=-3.99$, $p<0.01$; and computer software, $t(1040.73)=-9.53$, $p<0.01$, in mathematics classroom between the areas. Rural students (*M*=0.98, *SE*=0.03) use concrete materials compared to the urban students (*M*=0.84, *SE*=0.02), while urban students (*M*=0.51, *SE*=0.03) use computer software more frequently compared to the rural students (*M*=0.21, *SE*=0.02).

Student Learning Activities (SLA)

Students learning activities asks how often students engage in learning activities in the mathematics classroom to make an assessment of their involvement. LOT and HOT were considered separately.

SLA LOT

The SLA LOT subscale consists of four items. An average is computed across the five items belong to this scale, based on a three-category scale: ‘never’=0, ‘sometimes’=1, ‘always’=2. Nearly half of the students (*n*=529, 47.2%) see themselves as fully engaged in LOT activities, while a half (*n*=563, 50%) said they are somewhat engaged. Less than three per cent of students (*n*=28) say that they are not engaged in LOT-related activities. Table 6.32 presents the summary of students’ engagement related to LOT. Comparing the students’ LOT engagement based on school location, while overall the trend is even, the independent t-test finds there is a significant difference in relation to engagement, $t(1129.15)=-5.80$, $p<0.01$, with students in urban areas (*M*=2.03, *SE*=0.02) being less engaged with LOT related activities compared to the rural (*M*=2.20, *SE*=0.02).

Table 6.32

Students' Engagement related to LOT (N=1120)

	Frequency	Per cent
Not engaged	28	2.5
Somewhat engaged	563	50.3
Engaged	529	47.2
Valid N	1120	99.50
Missing	15	0.50
N	1135	100.0

SLA HOT

The SLA HOT subscale consists of four items. An average is computed across the four items belonging to this scale, based on a three-category scale: 'never'=0, 'sometimes'=1, always=2. Students' engagement related to HOT is presented in Table 6.33. The majority of the students (*n*=664, 66%) indicate that they are somewhat engaged; nearly 20% of the students (*n*=218) indicate that they are not engaged with activities associated with HOT; and 14% of the students indicate they are engage with activities related to HOT (*n*=158).

Table 6.33

Students' Engagement with the Learning Activities related to HOT (N=1110)

	Frequency	Per cent
Not engaged	218	19.6
Somewhat engaged	734	66.0
Engaged	158	14.2
Valid N	1110	99.8
Missing	25	.2
N	1135	100.0

6.4.6. Mathematics Performance

The mathematics test consists of eight questions including four questions related to HOT and four questions related to LOT. The partial credit Rasch model is used to analyse the items. The raw scores of the mathematics test is transformed to measures using the weighted likelihood estimation (WLE) obtained from the Conquest 2.0 software (Wu et al., 2007). Measures derived from the WLE estimation are further transform into *W* score using the formula that used a base $_{(9)}$log transformation of the performance scale using the following equation: *W*= 9.1024 logits + 500 (Woodcock, 1999).

The transformation of WLE score to *W* score is to: (a) eliminate the negatives values by setting the centring constant at 500 and (b) group ability is made easier as the distances along the W score is clearer than the distances along the logits scale. The measures are then classified into three categories of 'low', 'average' and 'high achieving' students: (a) the 'high achieving' corresponds to the students who obtain at least 75% of the total score, (b) the 'low achieving' corresponds to the students whose scores are less than 50% of the total score, (c) the 'average achieving' corresponds to the students whose scores are in between the two previous groups. The summaries of students' mathematics performance related to LOT and HOT are presented in Table 6.34 and Table 6.35.

Table 6.34

Student Mathematics Performance related to LOT (N=1135)

	Frequency		Per cent
	Urban	Rural	
Low achieving	364	466	73.1
Average achieving	130	80	18.5
High achieving	87	8	8.4
Total	581	554	100.0

In general students have a low mathematics performance. More than 70% of the students fall into the category of 'low achieving' (*n*=830, 73%) for mathematics performance related to LOT and for the mathematics performance related to HOT (*n*=876, 77%). The number of average achieving students for mathematics performance related to LOT and HOT are approximately 18% (*n*=210) and 20% (230) respectively. There is less than ten per cent of high achieving students (*n*=85, 8.4%) of mathematics performance related to LOT and there is less than three per cent high achieving students (*n*=29, 2.5%) of mathematics performance related to HOT.

Table 6.35

Student Mathematics Performance related to HOT (N=1135)

	Frequency		Per cent
	Urban	Rural	
Low achieving	389	487	77.18
Average achieving	163	67	20.26
High achieving	29	0	2.56
Total	581	554	100.0

To compare the students' mathematics performance related to LOT and HOT, independent *t*-tests are conducted. The independent *t*-test result shows that there is a significant different of mathematics performance related to LOT, $t(1133)=11.26$, $p<0.01$, with students in urban areas (*M*=495.50, *SE*=0.55) having higher performance related to LOT than the rural (*M*=487.73, *SE*=0.41). There is also a significant different of mathematics performance related to HOT, $t(1133)=11.85$, $p<0.01$, with students in urban areas (*M*=485.05, *SE*=0.77) having higher performance related to HOT than the rural (*M*=472.88, *SE*=0.67).

6.5 Summary

This chapter provides demographic information of the schools, teachers and students. There are 25 schools involved in the study representing five types of schools from both urban and rural areas, including both Islamic public schools and general public schools. More than half of the schools classify the classes so that there are high achieving and low achieving students in each class. In terms of resources, nearly all schools have mathematics textbook available for students. However, most schools have no access to computer with internet connection that can be used for mathematics teaching. Moreover, there is no school having special computer software for mathematics instruction. Most schools provide additional mathematics lesson after school hours but less than a third of the schools have a mathematics Olympiad club.

The 46 teachers involved in the study include 39 female and 7 male teachers. Most of the teachers have a bachelor of mathematics education as their highest level of education. About half of the teachers involved have 10 to 20 years of teaching experience. Teacher professional development is more prevalent in the urban schools. More than 40% of teachers say they attend these professional development programmes less than once in six months. The majority of those who attend less regularly are from the rural area. More than 70% of teachers involved in the study have passed the certification test of teaching. Teachers generally have positive beliefs concerning mathematics related to HOT and generally have somewhat positive beliefs concerning mathematics related to LOT. Among the types of problem given to students in the classroom, teacher indicate that problem which can be solved in many ways are the most frequently used (further discussion of this in chapter 11). Among the instructional activities given to students, asking students to decide on their own procedures for solving complex problems is the least frequently used. In line with the availability of mathematics textbook in most school, most teachers use them as the basic of instruction.

There are 1135 students involved in the study, comprising of 630 female and 505 male. Students who come from the urban area have parents of higher educational level compared to those from the rural area. A similar trend occurs for the home possessions.

In general, students from both urban and rural have a high expectation of the educational level they want to achieve, with more than 70% having a highly positive attitude towards mathematics and more than 80% placing a very high value of mathematics. Students judge their mathematics ability highly, with less students who admit they have a high confidence in mathematics. In general, students are having positive beliefs concerning mathematics related to LOT and HOT, with fewer students having highly positive beliefs concerning mathematics related to LOT compared to students having a highly positive beliefs concerning mathematics related to HOT. In line with teachers answer on the most frequent types of questions used in the classroom, students also accept that the problem that can solved in many ways is the most frequently used. A similar trend also occurs for textbook, students also admit that the most frequently used resource is mathematics textbook.

The mathematics test assessed students' performance related to both LOT and HOT. Generally, students have a low performance of mathematics related LOT and an even lower of mathematics performance related to HOT. More than 70% of the students are categorised as low achieving students, with less than ten per cent of the students achieving an average performance related to LOT. Less than three per cent of the students are achieving highly in mathematics related to HOT.

The information discussed in this chapter provides the necessary preliminary findings for the following chapters of path analysis and hierarchical linear modelling (HLM) analysis.

Chapter 7 Single-Level Path Analysis: Student-Level

7.1 Introduction

The theoretical framework presented in Chapter 2 describes the variables included in the study and their possible relationships. Variables at the student-level and teacher-level are presented, showing their influence both within each level, as well as the variables influencing across levels. At the student-level, student related variables are hypothesised to have an influence on students' mathematics performance related to both LOT and HOT. A hypothesised path model is then developed to illustrate the relationships among variables at the student-level and to examine how these variables influence students' mathematics performance related to LOT and HOT. The path analysis procedure used in this study is the partial least squares path analysis (PLS-PA), undertaken using SmartPLS 3.2.6 (Ringle et al., 2015). The definition, background, and evaluation of PLS-PA is described in Chapter 3. The results of the analyses for both the measurement and structural models are then presented and discussed, followed by a summary. All the analyses are based on the student data set (1135 students).

7.2 Model Building in the PLS-PA

The process of model building with PLS-PA starts with the drawing of the schematic representation of the model, called a 'path diagram' or 'path model'. A path diagram is a description of the assumed relationships between variables included in the study and the particular causal processes (Byrne, 2013; Hair et al., 2013). Before developing the path diagram, it is necessary to identify the latent variables and observed variables included in the analysis. The causal relationships between the latent variables also need to be specified based on both theory and logic.

The model tested in this chapter is of the student-level factors influencing students' mathematics performance related to LOT and HOT. The hypothesised model indicates that there are 15 latent variables included. The variables are drawn from students' background, their perception of classroom practices, attitudes and beliefs concerning mathematics and their mathematics performance related to LOT and HOT. These latent variables are listed in Table 7.1, using their acronym, along with their mode (unity, formative, or reflective), the manifest variates that are associated with each latent variable and their description, and finally the coding.

The latent variables in the path model are classified into exogenous and endogenous variables. Exogenous variables have no directional arrow coming to them and endogenous variables have one or more directional arrows coming to them. The hypothesised path model of factors influencing students' mathematics performance related to LOT and HOT is presented in Figure 7.1. The final model is then presented in Figure 7.2.

Table 7.1

Variables at Student-level Model

Latent Variable		Mode	Manifest Variate	Description	Coding
GENSTU	Respondent's gender	Unity	Sex	0=Female 1=Male	0=F 1=M
IDAREA	The school location of the respondent is enrolled in	Unity	Location	Urban Area Rural Area	1=U 2=R
EXP_EDU	Respondent's expected education	Unity	Educational level	High School, Vocational or technical, education after high school, Diploma, Bachelor's degree, Master's degree and Doctoral degree	0 to 5 0= High school 5= Doctoral degree
SES	Respondent's Socio-Economic Status	Inward	HomePos H EDU H JOB	Home possessions Parents' highest education Parents' highest job	Raw score
LR	Learning resources used in respondent's mathematics classroom	Inward	SFLR A SFLR B SFLR C SFLR D	Frequency of textbooks, worksheets or workbooks, concrete objects and computer use	
Q_TYPE	The type of mathematics question used in respondent's mathematics classroom	Inward	SFWQ A SFWQ B SFWQ C SFWQ D	Frequency of four types of questions use in mathematics classroom	
SLA HOT	Learning activities related to HOT in respondent's mathematics classroom	Outward	SLA A SLA B SLA C SLA D	Frequency of Learning activities related to HOT in mathematics classroom	
SLA HOT	learning activities related to LOT in respondent's mathematics classroom	Outward	SLA E SLA F SLA G SLA H	Frequency of learning activities related to LOT in mathematics classroom	
SBM HOT	Respondent's beliefs concerning mathematics related to HOT	Outward	SBML A SBML B SBML C SBML D SBML F SBML G SBML H SBNM A SBNM D SBNM E SBNM F	Beliefs concerning mathematics related to HOT	
SBM HOT	Respondent's beliefs concerning mathematics related to LOT	Outward	SBML E SBML B SBML C	Beliefs concerning mathematics related to LOT	
INDI_JUD	Respondent's individual judgement of mathematics ability	Outward	SCM BR SCM CR SCM ER SCM IR	The self-judgement of mathematics ability	
MCONF	Respondent's confidence in mathematics	Outward	SCM A SCM D SCM F SCM G SCM H	Confidence in mathematics	
LIKE_MATH	Respondent's attitude toward mathematics	Outward	ATM A ATM BR ATM CR ATM D ATM E	Liking or disliking mathematics	
VALUE_MATH	Respondent's valuing toward mathematics	Outward	SVM A SVM B SVM C SVM D	Valuing or devaluing mathematics	
MATH_LOT	Respondent's mathematics performance related to LOT	Unity	T LOT	Total score of questions related to LOT	Rasch score
MATH_HOT	Respondent's mathematics performance related to HOT	Unity	T HOT	Total score of questions related to HOT	Rasch score

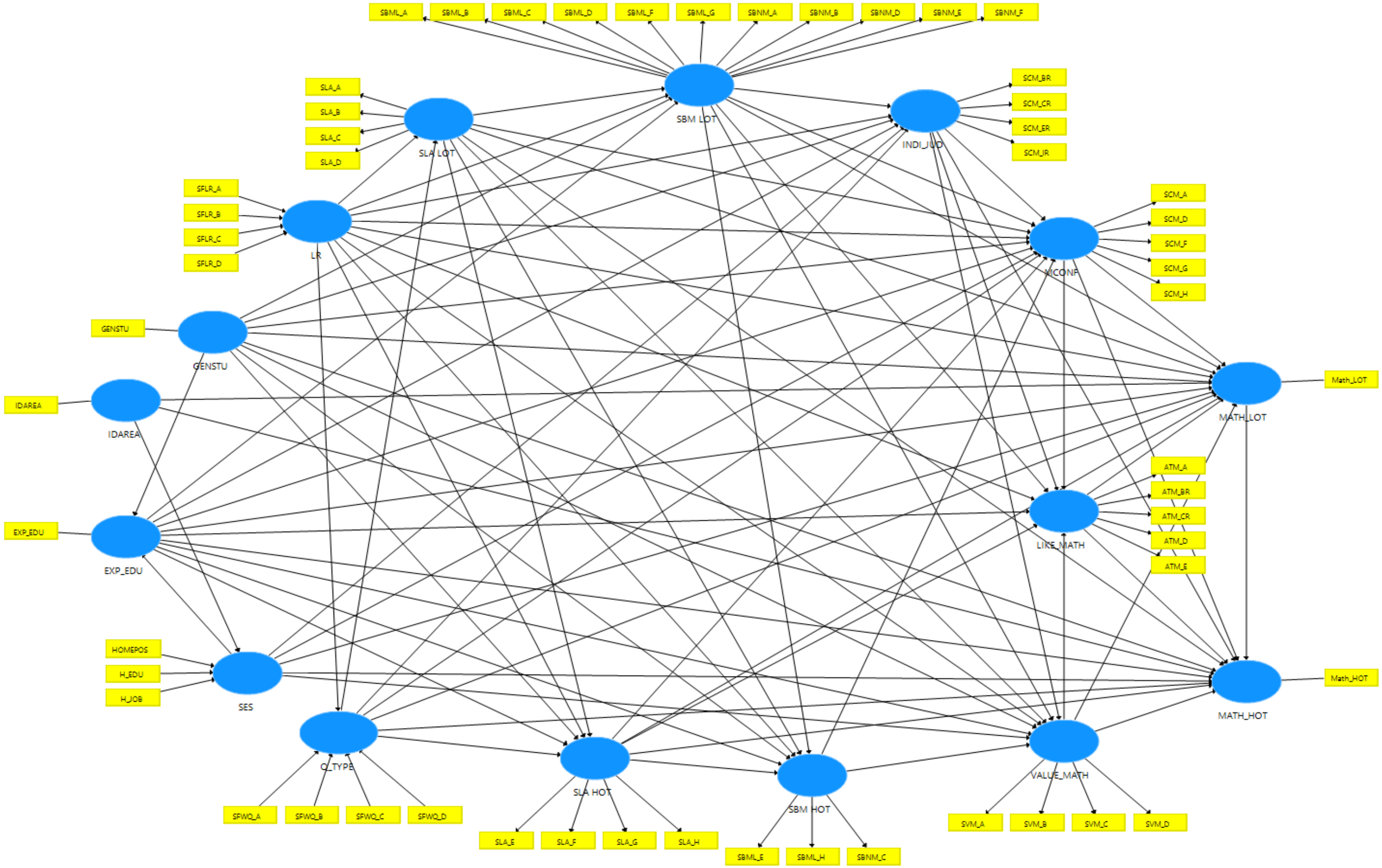


Figure 7.1 Hypothesised model for student-level model (N=1135)

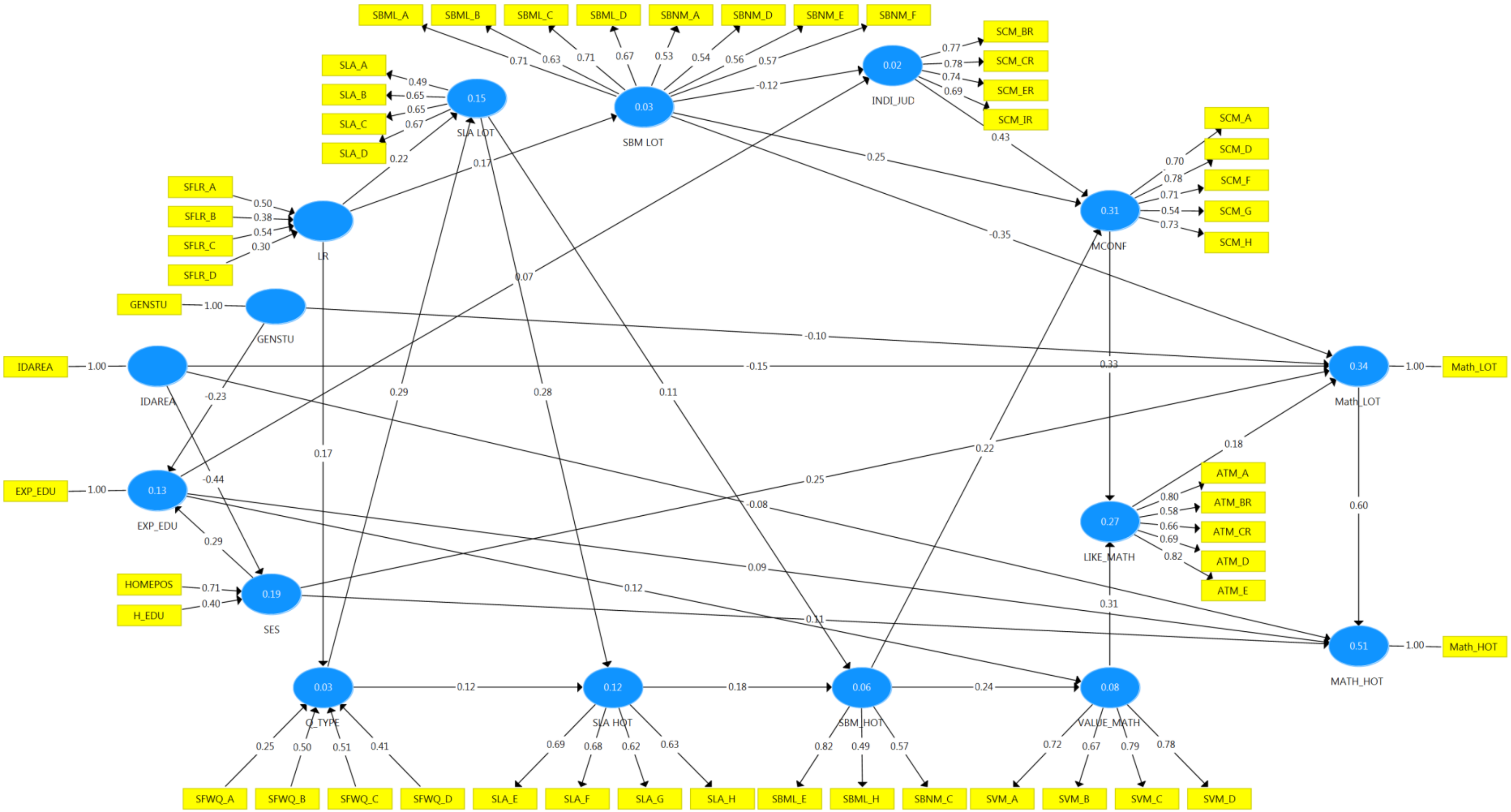


Figure 7.2 Final model for student-level model (*N*=1135)

7.3 Measurement Model (Outer Model)

The measurement or outer model describes the relationship of the latent variables to the manifest variates. The initial path model building is undertaken to create the hypothesised model. The next stage involves estimating the measurement model, followed by the comprehensive evaluation of the results and refinement of the measurement model.

In this study, the PLS-PA is carried out employing SmartPLS 3.2.6, where the algorithm settings used is the path weighting scheme (initial weight is 1.0), with the maximum iteration being 300 and the abort criterion 1.0E-5. This indicates that the PLS_PA algorithm stops when the maximum number of 300 iterations has been reached, or the cut-off criterion of 0.00001 has been reached.

The assessment of the measurement or outer model is divided into the assessment of reflective and formative constructs.

7.3.1. Reflective Constructs of Measurement Model

The examination of the reflective measurement model begins with reporting the estimated loadings. Hulland (1999) suggested that loadings should be ≥ 0.70 to be acceptable, with ≥ 0.40 being acceptable for exploratory research. Most of the indicators' loadings of the four reflective constructs (being INDI_JUD, MCONF, LIKE_MATH and VALUE_MATH) meet Hulland's criteria, being slightly below 0.70, with two being slightly below 0.40 (SBML_F and SBML_G), which were not removed in the validation stage (chapter 4) as the cut-off value used was 0.32. However, it is evident in this analysis that the exclusion of the two items (SBML_F and SBML_G) slightly improve the model. Therefore, the two items (SBML_F and

SBML_G) with indicator loadings below 0.40 are removed in the final model in order to achieve a better model.

The next step involves assessing the 'convergent validity' which indicates whether or not a scale is correlated with other known measures of the concepts. The average variance extracted (AVE) can be one of the criteria of convergent validity with the recommendation that AVE is to be ≥ 0.50 (Bagozzi & Yi, 1988). The AVE of the reflective constructs vary in this study, with three constructs satisfying the criteria of AVE ≥ 0.50 (these being INDI_JUD, LIKE_MATH and VALUE_MATH), three constructs have the AVE slightly below 0.50 (these being MCONF, SBM HOT, and SLA HOT) and two constructs have the AVE less than 0.40 (these being SBM LOT and SLA LOT).

The discriminant validity is then assessed, which examines how distinct or unique a construct is compared to other constructs. The discriminant validity is examined based on the cross loadings of the indicators, where the loading of the indicators of the constructs is higher for the intended construct compared with its loading on other constructs, and the Fornell-Larcker criterion, where the square root of AVE for each construct is higher than its correlation with other constructs (Fornell & Larcker, 1981). The discriminant validity of all reflective constructs in the measurement model is established as indicated by the cross loadings and the Fornell-Larcker criterion. The details of the cross loadings and the calculation of Fornell-Larcker criterion are recorded in Table 7.2.

Table 7.2

Relationship between Correlations among Constructs and Square Root of AVEs (Fornell-Larcker Criterion) for Discriminant Validity

	EXP_EDU	GENSTU	IDAREA	INDI_JUD	LIKE_MATH	LR	MATH_HOT	MCONF	Math_LOT	Q_TYPE	SBM_LOT	SBM_HOT	SES	SLA HOT	SLA LOT	VALUE_MATH
EXP_EDU	**SC**															
GENSTU	-0.21	**SC**														
IDAREA	-0.15	-0.03	**SC**													
INDI_JUD	0.09	-0.04	-0.03	**0.75**												
LIKE_MATH	0.09	-0.10	0.09	0.44	**0.72**											
LR	0.03	-0.03	-0.05	-0.02	0.08	**FMS**										
MATH_HOT	0.27	-0.07	-0.33	0.14	0.08	-0.06	**SC**									
MCONF	0.04	-0.01	0.06	0.43	0.42	0.12	-0.04	**0.70**								
Math_LOT	0.22	-0.14	-0.32	0.16	0.12	-0.11	0.69	0.03	**SC**							
Q_TYPE	0.14	-0.16	-0.11	0.08	0.09	0.17	0.11	0.11	0.08	**FMS**						
SBM LOT	-0.20	0.11	0.23	-0.13	0.10	0.17	-0.43	0.21	-0.45	-0.02	**0.62**					
SBM_HOT	0.10	0.00	-0.08	0.16	0.23	0.07	0.15	0.31	0.17	0.16	0.07	**0.64**				
SES	0.28	0.05	-0.44	0.01	-0.08	-0.01	0.40	-0.02	0.40	0.10	-0.29	0.10	**FMS**			
SLA HOT	0.04	0.01	-0.07	0.13	0.12	0.15	0.08	0.23	0.07	0.22	0.05	0.21	0.11	**0.65**		
SLA LOT	0.15	-0.15	-0.07	0.08	0.13	0.27	0.14	0.13	0.13	0.33	-0.02	0.17	0.14	0.32	**0.62**	
VALUE_MATH	0.15	-0.11	0.00	0.17	0.41	0.08	0.05	0.29	0.02	0.11	0.14	0.26	0.02	0.13	0.13	**0.75**

Note. The square root of AVE values is shown on the diagonal and printed in bold (reflective construct only); nondiagonal elements are the latents variable correlations.
SC= Single-item construct
FMS=Formative measurement model

Table 7.3

Result Summary for Reflective Measurement (Outer) Model

Latent variable	Indicator	Loading	Composite reliability	AVE
SLA LOT	SLA A	0.49	0.71	0.38
	SLA B	0.65		
	SLA C	0.65		
	SLA D	0.67		
SLA HOT	SLA E	0.69	0.75	0.43
	SLA F	0.68		
	SLA G	0.62		
	SLA H	0.63		
SBM LOT	SBML A	0.71	0.83	0.38
	SBML B	0.63		
	SBML C	0.71		
	SBML D	0.67		
	SBNM A	0.53		
	SBNM D	0.54		
	SBNM E	0.56		
	SBNM F	0.57		
SBM HOT	SBML E	0.82	0.67	0.41
	SBML B	0.49		
	SBML C	0.58		
INDI_JUD	SCM BR	0.77	0.84	0.56
	SCM CR	0.78		
	SCM ER	0.74		
	SCM IR	0.69		
MCONF	SCM A	0.71	0.83	0.49
	SCM D	0.78		
	SCM F	0.71		
	SCM G	0.54		
	SCM H	0.73		
LIKE_MATH	ATM A	0.80	0.84	0.52
	ATM BR	0.58		
	ATM CR	0.66		
	ATM D	0.69		
	ATM E	0.82		
VALUE_MATH	SVM A	0.72	0.83	0.56
	SVM B	0.67		
	SVM C	0.79		
	SVM D	0.78		

Even though some constructs do not satisfy the criterion of convergent validity (as indicated by the AVE), all constructs meet the requirement of discriminant validity and, therefore, all constructs are considered to be adequately acceptable. Moreover, all constructs meet the criteria of internal consistency reliability as indicated by the composite reliability being ≥ 0.7 (Bagozzi & Yi, 1988). A summary of the reflective measurement model evaluation is presented in Table 7.3.

7.3.2. Formative Constructs of Measurement Model

The evaluation of the formative measurement model does not follow the same procedures as the evaluation of the reflective measurement model. While assessment of the reflective model is mainly based on the correlations, this is not applicable for the formative model assessment as the indicators of the formative constructs are not required to correlate with each other. Consequently, the evaluation concerns collinearity and the significance test. In addition, the decision to remove or retain an indicator of formative model is made based on the outer weight and loading, as well as its significance. When the weight is low and both outer weight and loading are not significant, the indicator should be removed from the model (Hair et al., 2013).

Table 7.4 records the collinearity test results, it indicates that there is no collinearity issues as the variance inflation factor (VIF) value for each indicators is below 5, satisfying the criteria suggested by Hair, Ringle, and Sarstedt (2011) that the VIF value of 5 or higher indicates a greater chance of collinearity problems.

Table 7.4 also presents the outer weights and outer loading as well as their significance level for each formative construct, obtained through the bootstrapping procedure using SmartPLS 3.2.6 (Ringle et al., 2015). The student's SES is initially made up of three indicators: (a) possessions within the home possession (HomePos); (b) parent's highest educational level (H_EDU); and (c) parent's highest job level (H_JOB). H_JOB was removed from the final models as the outer weight was too low (below

0.1) and neither its outer weight nor outer loading was significant. As for the rest of the formative constructs, the weights and/or loadings are significant and therefore these indicators are retained.

Table 7.4

Outer Weights, Outer Loading Significance and Collinearity Testing Results for Formative Measurement Model

Formative Construct	Formative Indicators	Outer Weights (Outer Loadings)	Significance Level	VIF
SES	HomePos	0.71 (0.95)	***(***)	1.53
	H EDU	0.40 (0.82)	***(***)	1.53
LR	SFLR A	0.50 (0.60)	***(***)	1.04
	SFLR B	0.38 (0.55)	***(***)	1.04
	SFLR C	0.53 (0.71)	***(***)	1.06
	SFLR D	0.30 (0.38)	*** (***)	1.04
Q_TYPE	SFWQ A	0.25 (0.40)	***(***)	1.03
	SFWQ B	0.50 (0.66)	***(***)	1.05
	SFWQ C	0.51 (0.68)	***(***)	1.05
	SFWQ D	0.41 (0.55)	**(***)	1.03

Note: NS= Not significant
** $p < 0.05$, *** $p < 0.01$

In addition to the constructs included in the reflective and formative outer models, the model also has four constructs in the unity mode. These are: (a) highest expected education (EXP_EDU), (b) gender (GENSTU), (c) students' mathematics performance related to LOT (MATH_LOT) and (d) students' mathematics performance related to HOT (MATH_HOT). Each of these constructs is reflected by a single indicator indicating that the construct is reflected by a single indicator. The loading, the composite reliability and the communality of the unity mode are designated as 1.

7.4 Structural (Inner) Model

The structural or inner model is the path model, representing the magnitude of relationships between one variable and another in the hypothesised model. The examination of the structural (inner) model is conducted after finalising the outer model. It is presented in two subsections, the first section presents the overall examination of the structural (inner) model and the second section discusses the details of the relationships between variables in the model.

7.4.1. The Evaluation of the Structural (Inner) Model of the Student-level Model

The assessment of the inner model may lead to the elimination of some path arrows and of some variables from the model. Latent variables which have weak path coefficients and low significance need to be removed from the model. The steps for evaluating the structural (inner) model are: (a) assessing the collinearity between latent variables in the model; (b) assessing the significance and relevance of the structural model relationships; (c) assessing the level of R square (R^2); (d) assessing effect size (f^2) and the predictive relevance (Q^2).

The collinearity assessment is based on the value of tolerance being ≥ 0.2 and the value of VIF ≤ 5 in order to have an adequate collinearity; if values are outside these ranges this is an indication of a collinearity problem (Menard, 2002). Such a problem requires consideration of removing one or more constructs, merging predictors into a single construct and creating a higher order construct. The assessment of significance and relevance of the structural model relationships involves the evaluation of standardised path coefficients. The path coefficient represents the magnitude of relationships between the latent variables in the structural model. The value of any standardised path coefficient is between -1 and +1, with a value closer to 1 (either positive or negative) indicating a stronger relationship and a value closer to 0 indicating a weaker relationship. Only significant path coefficients can be retained in the structural model. The significance of the path coefficients in PLS-PA using SmartPLS 3.2.6 (Ringle et al., 2015) can be obtained using the bootstrapping procedure which, in this study, is based on the population of 1135 students. A path coefficient of a value of 0.05 or greater for a large sample and 0.10 or greater for a small sample are acceptable as being meaningful (Sellin, 1990).

Table 7.5 presents the summary of the results of the structural (inner) model including the VIF value, path coefficients (β), R^2 and Q^2. The VIF values for all latent variables are below the recommended threshold of 5, indicating that there is no problem with collinearity. The path coefficients of the final model are also recorded in Table 7.5 (the non-significant coefficients and the path coefficients below 0.05 are removed from the model). It should be noted that only standardised path coefficients (β) are provided by SmartPLS 3.2.6 (Ringle et al., 2015), and therefore, only these are presented. The weakest of the path coefficients is the path between students' expected education (EXP_EDU) and students' individual judgement of mathematics ability (INDI_JUD) (β = 0.07). The strongest path coefficient is the path between students' mathematics performance related to LOT (MATH_LOT) and students' mathematics performance related to HOT (MATH_HOT) (β = 0.60). Most of the path coefficients are significant with one per cent or less probability of error ($p \leq 0.01$), and only one path coefficient (EXP_EDU to INDI_JUD) being significant with a five per cent probability of error ($p \leq 0.05$).

Table 7.5

Summary of Results of Structural (Inner) Model

Criterion N=1135	Predictor	VIF	*β*	*SE*	*p*	R^2	f^2	Q^2
EXP_EDU	SES	1	0.29	0.02	***	0.13	0.10	0.12
	GENSTU	1	-0.23	0.03	***		0.06	
SES	IDAREA	1	-0.44	0.02	***	0.19	0.24	0.14
Q_TYPE	LR	1	0.17	0.04	***	0.03	0.03	0.01
SLA LOT	LR	1.03	0.22	0.04	***	0.15	0.02	0.06
	Q_TYPE	1.03	0.29	0.03	***		0.08	
SLA HOT	Q_TYPE	1.12	0.12	0.03	***	0.12	0.02	0.05
	SLA LOT	1.12	0.28	0.03	***		0.08	
SBM LOT	LR	1	0.17	0.05	***	0.03	0.03	0.01
SBM HOT	SLA LOT	1.12	0.11	0.04	***	0.06	0.03	0.02
	SLA HOT	1.12	0.18	0.04	***		0.01	
INDI_JUD	SBM LOT	1.04	-0.12	0.04	***	0.02	0.01	0.01
	EXP_EDU	1.04	0.07	0.03	**		0.01	
MCONF	INDI_JUD	1.05	0.43	0.03	***	0.30	0.25	0.13
	SBM LOT	1.03	0.25	0.03	***		0.09	
	SBM HOT	1.04	0.22	0.03	***		0.07	
LIKE_MATH	VALUE_MATH	1.09	0.31	0.03	***	0.27	0.14	0.13
	MCONF	1.09	0.33	0.03	***		0.12	
VALUE_MATH	SBM HOT	1.01	0.24	0.03	***	0.08	0.06	0.04
	EXP_EDU	1.01	0.12	0.03	***		0.02	
MATH_LOT	IDAREA	1.26	-0.15	0.02	***	0.34	0.03	0.33
	SES	1.31	0.25	0.03	***		0.07	
	SBM LOT	1.13	-0.35	0.02	***		0.16	
	LIKE_MATH	1.31	0.18	0.03	***		0.05	
	GENSTU	1.03	-0.10	0.02	***		0.02	
MATH_HOT	IDAREA	1.28	-0.08	0.02	***	0.51	0.01	0.50
	SES	1.42	0.11	0.03	***		0.02	
	EXP_EDU	1.10	0.09	0.02	***		0.02	
	MATH_LOT	1.24	0.60	0.02	***		0.60	

Note: $^{**}p < 0.05$, $^{***}p < 0.01$

The next step in evaluating the inner model involves examining the R square (R^2) value. The R^2 coefficient is used to assess the model's predictive accuracy, obtained from the squared correlation of actual and predicted values. Consequently, R^2 also indicates the amount of variance explained in the endogenous construct. The estimated R^2 values in a specific situation are considered as substantial (0.75), moderate (0.50), and weak (0.25) (Cohen, 1988). The outcomes of the model, students' mathematics performance related to HOT (MATH_HOT) and mathematics performance related to LOT (MATH_LOT), are the two variables with the highest R^2.

In addition to evaluating the R^2 values of all endogenous latent variables, it is also important to evaluate the effect size (f^2). The f^2 indicates an exogenous latent variable's contribution to the amount of variance explained by an endogenous latent variable , with the f^2 values of 0.02, 0.15 and 0.35 indicating a small, medium and large effect size respectively (Cohen, 1988). The last step in the inner model evaluation involves assessing the model's predictive relevance (Q^2). Thus, the PLS-SEM has predictive relevance when "it accurately predicted the data points of indicators in reflective measurement models of endogenous constructs and endogenous single-item constructs" (Hair et al., 2013, p. 178). A model has predictive relevance when the Q^2 value is greater than 0 and the model is lacking in predictive relevance when the Q^2 value is less than 0 (Henseler et al., 2009).

The results of f^2 and Q^2 values for each of the endogenous variables in the model are also presented in Table 7.5. It shows that the effect sizes of the exogenous latent variables to the endogenous latent variables are varied in this study. Only one exogenous variable has a large effect size, this being the students' mathematics performance related to LOT (MATH_LOT) to the students' mathematics performance related to HOT (MATH_HOT) (f^2=0.60). Four exogenous variables have a medium effect size, these being the school location (ID_AREA) to students' socio-economic status (SES) (f^2 = 0.24), students' individual judgement of mathematics ability (INDI_JUD) to mathematics confidence (MCONF) (f^2 = 0.25), students' valuing mathematics (VALUE_MATH) to liking mathematics (LIKE_MATH) (f^2 = 0.14), and students' beliefs concerning mathematics related to LOT to the students' mathematics performance related LOT (MATH_LOT) (f^2 = 0.16). The remaining exogenous variables have a weak effect size. In addition, all endogenous variables exhibit predictive relevance indicated by the Q^2 values being greater than 0. However, some of the variables (Q_TYPE, SBM LOT and INDI_JUD) have a low predictive relevance value.

7.4.2. Relationships between Variables in Structural (Inner) Model of the Student-level Model

This section provides a detailed discussion of the relationships between the latent variables within the inner model, as they are shown in Figure 7.2. Each endogenous variable with its causal paths is discussed individually. In the path model, the three exogenous variables are (a) Gender of Student, (b) School Location and (c) Learning Resources.

Expected Education (EXP_EDU)

Gender of Student (GENSTU) and Socioeconomic Status (SES) are the two variables that have a direct effect on Expected Education (EXP_EDU). The R^2 value for Expected Education (EXP_EDU) is 0.13, indicating that Gender of Student (GENSTU) and Socioeconomic Status (SES) collectively account for the 13 percent of the variance in the Expected Education (EXP_EDU). The positive path coefficient of SES to EXP_EDU (β = 0.29, *SE* = 0.02) indicates that students with parents of a higher SES have a higher expectation for their future educational level. The negative path coefficient from GENSTU to EXP_EDU (β = -0.23, *SE* = 0.03) indicates that female students have a higher expectation for their future educational level than male students.

Socio-economic Status (SES)

School Location (IDAREA) (β = -0.44, *SE* = 0.02) is the only variable having a direct effect on the scale of SES, with the R^2 value being 0.19, indicating that School Location explains 19% of the variance of SES. The negative path coefficient indicates that schools in the urban areas have more students from higher SES backgrounds than schools in rural areas.

Question Types used in the Mathematics Classroom (Q_TYPE)

Learning Resources (LR) has a direct effect on the Question Types used in mathematics classroom (Q_TYPE) and accounts for only three per cent of the variance in question types (R^2 = 0.03). The positive path coefficient of LR (β = 0.17, *SE* = 0.03) indicates that the regularity of learning resources teachers use in the classroom (including textbooks, worksheets or workbooks, concrete objects, and computer software) have a positive effect on the regularity of four types of mathematics questions used in the classroom: (a) questions similar to what the teacher solves in the classroom; (b) questions which apply my knowledge from previous topics; (c) questions which can be solved in many ways, and (d) word problems.

Student Learning Activities related to LOT (SLA LOT)

Learning Resources (LR) (β=0.22, *SE*=0.04) and Questions Types (Q_TYPE) (β = 0.29, *SE* = 0.03) have a direct effect on Student Learning Activities related to LOT (SLA LOT). The R^2 value of 0.15 indicates that both variables explain 15% of the variance in SLA LOT. The positive path coefficients indicate that the learning resources frequently used in the mathematics classroom as well as the type of questions given to the students have a positive impact on the regularity of learning activities related to LOT that are experienced by the students in the mathematics classroom.

Student Learning Activities related to HOT (SLA HOT)

Question Types (Q_TYPE) and Student Learning Activities related to LOT (SLA_LOT) are the two variables that directly influence SLA_HOT. The two scales account for 12 per cent of the variance of the SLA HOT as indicated by the R^2 value of 0.12. The positive path coefficients of student learning activities related to LOT (β = 0.28, *SE* = 0.03) indicate that the more frequently students have learning activities related to LOT, the more they have learning activities related to HOT. Question types also have a positive impact on SLA HOT (β = 0.12, *SE* = 0.03), indicating that the types of mathematics questions given to students influence the frequency of students having learning activities related to HOT.

Students' Beliefs concerning Mathematics related to LOT (SBM LOT)

Learning Resources is the only variable with a direct effect on SBM LOT (β = 0.17, *SE* = 0.05), indicating that the frequency of using learning resources in the mathematics classroom positively influences students' beliefs concerning mathematics related to LOT. The R^2 for SBM LOT is 0.03, indicating that learning resources (LR) explains only three per cent of the variance of SBM LOT.

Students' Beliefs concerning Mathematics related to HOT (SBM HOT)

Student Learning Activities related to LOT (SLA LOT) and Student Learning Activities related to HOT (SLA HOT) have a direct effect on SBM HOT. The positive path coefficients of SLA LOT (β = 0.11, *SE* = 0.04) and SLA HOT (β = 0.18, *SE* = 0.04) indicate that the frequency of students having learning activities related to LOT as well as the frequency of students having learning activities related to HOT both positively influence the students' beliefs concerning mathematics related to HOT. The more frequently students have learning activities related to LOT and HOT, the more positive are students' beliefs concerning mathematics related to HOT, with SLA HOT having a slightly stronger impact on SBM HOT compared with SLA LOT. SLA LOT and SLA HOT together account for only six per cent of the variance in SBM HOT ($R^2 = 0.06$).

Student Individual Judgement of their Mathematical Abilities (INDI_JUD)

Student Beliefs concerning Mathematics related to LOT (SBM LOT) and Expected Education (EXP_EDU) have a direct effect on INDI_JUD. Student Beliefs concerning Mathematics related to LOT negatively influence Individual Judgement ($\beta = -0.12$, $SE = 0.04$). This indicates that students who have more positive beliefs concerning mathematics related to LOT tend to have a less positive individual judgement of their ability in mathematics. Students' expected education has a positive direct effect on their individual judgement ($\beta = 0.07$, $SE = 0.03$), indicating that students who expect to achieve a higher educational level tend to have a more positive individual judgement regarding their mathematics abilities. The R^2 value for student individual judgement of their mathematical abilities (INDI_JUD) is small ($R^2 = 0.02$). Both variables account for only two per cent of the variance in INDI_JUD.

Mathematics Confidence (MCONF)

INDI_JUD, SBM LOT and SBM HOT directly influence Mathematics Confidence (MCONF). The R^2 value for MCONF is 0.30, indicating that the INDI_JUD, SBM LOT and SBM HOT collectively account for 30 per cent of the variance in mathematics confidence (MCONF). The positive path coefficients of INDI_JUD (β =

0.43, *SE* = 0.03), SBM LOT (β = 0.25, *SE* = 0.03) and SBM HOT (β = 0.22, *SE* = 0.03) show the positive direct effects of the three scales to MCONF. This indicates that students who have a higher individual judgement of their mathematics abilities are likely to have greater mathematics confidence. Also, both SBM LOT and SBM HOT positively influence mathematics confidence, indicating that students with more positive beliefs concerning mathematics related to LOT and HOT tend to have higher mathematics confidence.

Liking of Mathematics (LIKE_MATH)

Mathematics Confidence (MCONF) and Valuing Mathematics (VALUE_MATH) are the two variables that have direct effects on Liking of Mathematics (LIKE_MATH). The R^2 value for LIKE_MATH is 0.27, indicating that the MCONF and VALUE_MATH explain 27 per cent of the variance of LIKE_MATH. The positive effect of MCONF on LIKE_MATH (β = 0.31, *SE* = 0.03) indicates that the more confident the students are about their mathematics ability, the more the students like mathematics. In other words, students who have more positive confidence in mathematics like mathematics more. VALUE_MATH also positively affects LIKE_MATH (β = 0.33, *SE* = 0.03), indicating that when students value mathematics positively, they have a more positive attitude toward liking mathematics.

Valuing Mathematics (VALUE_MATH)

SES and SBM HOT are the two variables which have direct effects on Valuing Mathematics (VALUE_MATH). The R^2 value for VALUE_MATH is 0.08, indicating that the path model explains eight per cent of the variance of valuing mathematics. The path coefficient of SES to VALUE_MATH is positive (β = 0.24, *SE* = 0.03), indicating that students coming from a higher SES tend to value mathematics more positively, seeing this subject as more highly valued for their daily life and their future education. SBM HOT also positively influences VALUE_MATH (β = 0.12, *SE* = 0.03), indicating that students who have more positive beliefs concerning mathematics related to HOT tend to place a more positive value on mathematics.

Students Mathematics Performance related to LOT (MATH_LOT)

There are five variables that have a direct effect on the Students' Mathematics Performance related to LOT (MATH_LOT), namely: (a) SBM LOT (Student Beliefs concerning Mathematics related to LOT); (b) GENSTU (Gender); (c) IDAREA (School Location); (d) SES (Socio-economic Status); and (e) LIKE_MATH (Liking of Mathematics). The R^2 value for MATH_LOT is 0.34, indicating that five variables collectively account for the 34 per cent of the variance of Students' Mathematics Performance related to LOT.

Both SES (β = 0.25, *SE* = 0.03) and LIKE_MATH (β = 0.18, *SE* = 0.03) have a positive direct effect on MATH_LOT. This indicates that students who have a more positive liking of mathematics are performing better in mathematics related to LOT. In other word, students' mathematics performance related to LOT is positively influenced by attitude in liking mathematics. The positive path coefficient of SES indicates that students from higher SES tend to perform better in mathematics related to LOT in this study.

SBM LOT (β = -0.35, *SE* = 0.02), GENSTU (β = -0.10, *SE* = 0.02) and IDAREA (β = -0.15, *SE* = 0.02) have a negative direct effect on MATH_LOT. This indicates that students who are have more positive beliefs concerning mathematics related to LOT tend to perform lower in mathematics related to LOT. This means that students who agree a little or agree a lot to most of the items in SBM LOT tend to have lower mathematics performance related to LOT, indicating that students who disagree a little or disagree a lot to those statements are performing better. These statements are itemised as: (a) mathematics is just about addition, subtraction, multiplication and division; (b) in mathematics, a correct answer is more important than the way to get it; (c) all mathematics problems can be solved in one way only; (d) mathematics is just a collection of rules and formulas; (e) the way teacher solves mathematics problem is the only correct way to solve the problem; (f) memorising is the most important thing in learning mathematics; (g) I should always follow the procedures the teacher taught in solving the mathematics problem; (h) the main goal of doing mathematics problems is to obtain a correct answer.

Female students perform better in mathematics related to LOT than do male students. Furthermore, students from urban schools have a higher achievement in mathematics related to LOT than do students from rural schools.

Students Mathematics Performance related to HOT (MATH_HOT)

Four variables have direct effects on the Students' Mathematics Performance related to HOT (MATH_HOT). The R^2 value for MATH_HOT is 0.51, indicating that the four variables collectively account for 51 per cent of the variance of students' mathematics performance related to HOT.

The results show that MATH_LOT (β = 0.60, *SE* = 0.02), EXP_EDU (β = 0.09, *SE* = 0.02), and SES (β = 0.11, *SE* = 0.03) positively affect MATH_HOT. The path coefficients show that MATH_LOT is the strongest variable, indicating that students' mathematics performance related to LOT is a strong predictor of students' mathematics performance related to HOT. In other words, students who achieve higher in mathematics related to LOT tend to achieve higher in mathematics related to HOT. Furthermore, the results also signify that students who have higher expectations for their future educational level are more likely to perform better in mathematics related to HOT. In line with the result for students' mathematics performance related to LOT, students coming from higher SES also tend to achieve higher in mathematics related to HOT.

ID_AREA has a negative effect on MATH_HOT (β = -0.08, *SE* = 0.02). This indicates that students from urban schools tends to achieve higher mathematics performance related to HOT than students attending rural schools.

Two variables are found to have relatively stronger indirect effects on MATH_HOT, namely LIKE_MATH (β = 0.11) and SBM_LOT (β= -0.20). This indicates that students' attitude of liking mathematics, mediated by students' mathematics performance related to LOT, positively influence students' mathematics performance related to HOT. In other words, the more positive students' attitude towards liking mathematics, the higher their mathematics performance related to HO. In addition, the result also indicates that student beliefs concerning mathematics related to LOT, mediated by students' mathematics performance related to LOT, negatively influence students' mathematics performance related to HOT. This can be interpreted as students who has more positive beliefs on mathematics related LOT tend to achieve lower in mathematics performance related to HOT. There are some other variables that show relatively weaker indirect effect on MATH_HOT. However, the effect size is very small and therefore they are not mentioned here.

7.5 Summary

Partial least square path analysis (PLS-PA) is a structural equation modelling (SEM) technique. The PLS-PA uses the partial least squares (PLS) procedure to examine the relationship between the latent variables and their manifest variates, as well as the interrelationships between the latent variables. The PLS-PA procedure involves modelling that does not require data to be normally distributed or based on large sample sizes. Furthermore, PLS-PA is able to incorporate both reflective and formative constructs. These characteristics of PLS-PA make it suitable for analysis in this study. It is also important to note that the PLS-PA operates very effectively for exploratory research.

In this study, the PLS-PA is carried out using SmartPLS 3.2.6 (Ringle et al., 2015), to investigate variables influencing students' mathematics performance related to LOT and HOT. The analysis is based on the theoretical framework presented in Chapter 2. More specifically, the path analysis is conducted to investigate the variables at the student-level that have a predicting relationship on students' mathematics performance related to LOT and HOT. The analysis is conducted to seek answers to the research questions: (1) What are the student-level factors influencing students' mathematics performance related to LOT and HOT? and (2) What are the interrelationships among student-level factors influencing students' mathematics performance related to LOT and HOT?

For the first research question, the analysis indicates that the following variables are directly influencing Students' Performance related to LOT: (a) Students' Beliefs concerning Mathematics related to LOT, (b) gender, (c) school location, (d) socio-economic status, and (e) liking of mathematics. Furthermore, the analysis suggests that four variables have a direct effect on students' mathematics performance related to HOT: (a) mathematics performance related to LOT, (b) school location, (c) highest educational expectation, and (d) socio-economic status.

The analysis shows that some variables have positive effects while others have negative effects on students' mathematics performance related to LOT and HOT. Students' socio-economic status and liking of mathematics have direct positive effects on the students' mathematics performance related to HOT, while students' beliefs concerning mathematics related to LOT, gender and school location have negative direct effects. The results indicate that that mathematics performance related to LOT, gender, highest educational expectation, and SES positively influence students' mathematics performance related to HOT. However, school location has a negative effect indicating students from urban schools perform better in mathematics related to HOT.

The results of the analysis indicate that students' attitude, as represented by students' liking mathematics, positively influence their mathematics performance. It is also noted that students' mathematics performance related to LOT is the strongest predictor for their performance related to HOT. The relationship between gender and mathematics performance indicates an interesting initial finding: female students are more likely to achieve higher in mathematics performance related to LOT while gender has no effect on mathematics performance related to HOT.

In regards to the second research question, the analysis shows gender and SES positively influence students expected education, indicating that female students and students with parents of a higher SES have a higher expectation for their future educational level. SES is also influenced by school location indicating that schools in the urban areas have more students from higher SES backgrounds than schools in rural areas.

The analysis also indicates that the frequency of using learning resources in the classroom have a positive effect on the regularity of the four types of mathematics questions used in the classroom. Both learning resources and the question types positively influence the regularity of learning activities related to LOT in the mathematics classroom. In addition, the question types and students learning activities also have positive effects on the frequency of students having learning activities related to HOT.

The analysis shows that learning resources also positively influences the students' beliefs concerning mathematics related to LOT. While both students learning activities related to LOT and HOT have direct effects on students' beliefs concerning mathematics related to HOT.

Students' beliefs concerning mathematics related to LOT and students' expected education influence students' individual judgement of mathematics ability. Students' individual judgement of mathematics ability, together with, students' beliefs concerning mathematics related to both LOT and HOT have positive effects on student' mathematics confidence. Furthermore, students' mathematics confidence and students' attitude of valuing mathematics positively influence their attitude of liking mathematics. While the students' attitude of liking mathematics is positively influenced by SES and their attitude of valuing mathematics.

The results in this chapter are presented as a single-level model of the student-level variables. A more rigorous investigation is required incorporating both teacher-and school-level in a multilevel analysis. The HLM analysis results are presented in Chapter 9.

Chapter 8 Single-Level Path Analysis: Teacher-level

8.1 Introduction

This chapter presents the results of the partial least square path analysis (PLS-PA) at the teacher-level, following the same procedures as the student-level analysis, carried out in Chapter 7. The analysis is based on the theoretical framework discussed in Chapter 2, examining the teacher-level variables and their possible relationships. It is hypothesised that there are teacher-level variables that influence their classroom practices in the mathematics classroom, including their practices related to higher order thinking and lower order thinking The variables at the teacher-level are: (1) teachers' background: (a) teachers' age, (b) teachers' experience, (c) teachers' certification and (d) teachers' professional development; (2) teaching resources used; (3) teachers' beliefs concerning the nature of mathematics and mathematics learning; (4) teachers' beliefs concerning mathematics teaching; (5) types of questions given to students; (6) teachers' engagement with students; and (7) instructional activities in the mathematics classroom.

The PLS-PA at the teacher-level is also carried out using the SmartPLS 3.2.6 (Ringle et al., 2015). The sequence of model building and model evaluation of the results of the PLS-PA procedure at the teacher- level follows the same process as in Chapter 7. This chapter presents the model building, including the hypothesised and final models, followed by the presentation and discussion of results for both measurement (outer) and structural (inner) models. All the analyses are based on the data set of 46 teachers.

8.2 Model Building of PLS-PA

Creating a visual representation of the model including the possible relationships between the variables is the first stage of model building for the teacher-level path analysis. The path diagram and path model consist of both latent variables and manifest variates and the relationships between the variables and variates are represented by single path arrows. The hypothesised relationships are drawn following the theoretical framework presented in Chapter 2.

The models tested in this chapter are of teacher-level factors that may be influencing the classroom practice of instructional activities used in the mathematics classroom. The hypothesised model shows that there are 15 latent variables, are presented in Table 8.1, with their acronyms, their mode (unity, formative or reflective) and description, the manifest variates that are associated with each latent variable and their description, and finally the coding employed. The hypothesised path model of factors influencing classroom practice of instructional activities is illustrated in Figure 8.1.

The analysis that follows illustrates the strategy that can be employed with PLS-PA when the sample is small (*N*=46). Such a size presents the problem that a singular matrix may occur during the estimation, especially when running the bootstrapping analysis (in order to obtain the estimates of the significance values of the path coefficients). The complexity of the model and the number of variables involved need to be considered in the analysis. Barclay, Higgins, and Thompson (1995) recommended as a guideline that the appropriate minimum sample size for a robust PLS-PA solution be ten times the largest number of indicators associated with a single variate or ten times the maximum number of path arrows projected on any variable in the outer model. In order to be able to meet this minimum requirement of sample size, and to simplify the model, four variables related to beliefs are transformed to unity mode (using a Rasch scale) namely: (a) TBMT_HOT; (b) TBMT_LOT; (c) TBNMML_HOT; and (d) TBNMML_LOT.

The final model is presented in Figure 8.2, with the variables of gender of teacher (GENTE), school location (T_AREA), teacher professional development (TPD), teachers' beliefs concerning nature of mathematics and mathematics learning related to LOT (TBNMML_LOT). Teachers' beliefs concerning the nature of mathematics and mathematics learning related to HOT (TBNMML_HOT) and teachers' beliefs concerning mathematics teaching related to HOT (TBMT_HOT) are not included in the final model as there is no significant path entering to or exiting from them.

Table 8.1

Variables at Teacher-level Model

Latent Variables	Mode	Description	Manifest Variate	Description	Coding
GENTE	Unity	Respondent’s gender	Sex	Female, Male	0=F 1=M
T_Age	Unity	Respondent’s age	Age	Years old	
T_Area	Unity	The school location of respondent is teaching	Location	Urban Area, Rural Area	1=U 2=R
TC	Unity	Respondent’s status of teacher certification	Teacher certification	Certified or Not certified	0=N 1=Y
TPD	Unity	The number of professional development programs attended by respondents in the past 2 years	Total number of Professional development programs attended	Mathematics content, (b) Mathematics pedagogy/instruction, (c) Mathematics curriculum, (d) Integrating information technology to mathematics, (e) Improving students critical thinking and problem solving, (f) Mathematics assessment, (g)Addressing individual students needs	
YT	Unity	Respondent’s years of teaching	Years	Total years of teaching	
TR	Formative	The teaching resources used in mathematics classroom	TR_A to TR_D	Frequency of the teaching resources used in mathematics classroom	0=Not used 1=Used as supplement 2=Used as basic of instructions
TBNMML_HOT	Unity	Respondent’s beliefs concerning nature of mathematics and mathematics learning related to HOT		Beliefs concerning nature of mathematics and mathematics learning related to HOT	Rasch score
TBNMML_LOT	Unity	Respondent’s beliefs concerning nature of mathematics and mathematics learning related to LOT		Beliefs concerning nature of mathematics and mathematics learning related to LOT	Rasch score
TBMT_HOT	Unity	Respondent’s beliefs concerning mathematics teaching related to HOT		Beliefs concerning mathematics teaching related to HOT	Rasch score
TBMT_LOT	Unity	Respondent’s beliefs concerning mathematics teaching related to LOT		Beliefs concerning mathematics teaching related to LOT	Rasch score
QT_Daily	Formative	The types of mathematics questions respondent used in mathematics classroom	TQ_A to TQ_B	Frequency of 6 types of mathematics questions teacher used in mathematics classroom	1=Never, 2=Seldom, 3=Sometimes, 4=Usually and 5=Always
QT_Exam	Formative	The types of mathematics questions respondent used in mathematics examination	TQE_A to TQ_F	Frequency of 6 types of mathematics questions teacher used in mathematics examination	1=Never, 2=Seldom, 3=Sometimes, 4=Usually and 5=Always
IAS	Reflective	Instructional activities respondent given to students in mathematics classroom	IAS_C to IAS_J	Frequency of instructional activities teacher given to students in mathematics classroom	1=Never, 2=Seldom, 3=Sometimes, 4=Usually and 5=Always
TES	Reflective	Respondent’s effort of engaging with students	IAT_A to IAS_F	Frequency of engaging activities teachers do	1=Never, 2=Seldom, 3=Sometimes, 4=Usually and 5=Always

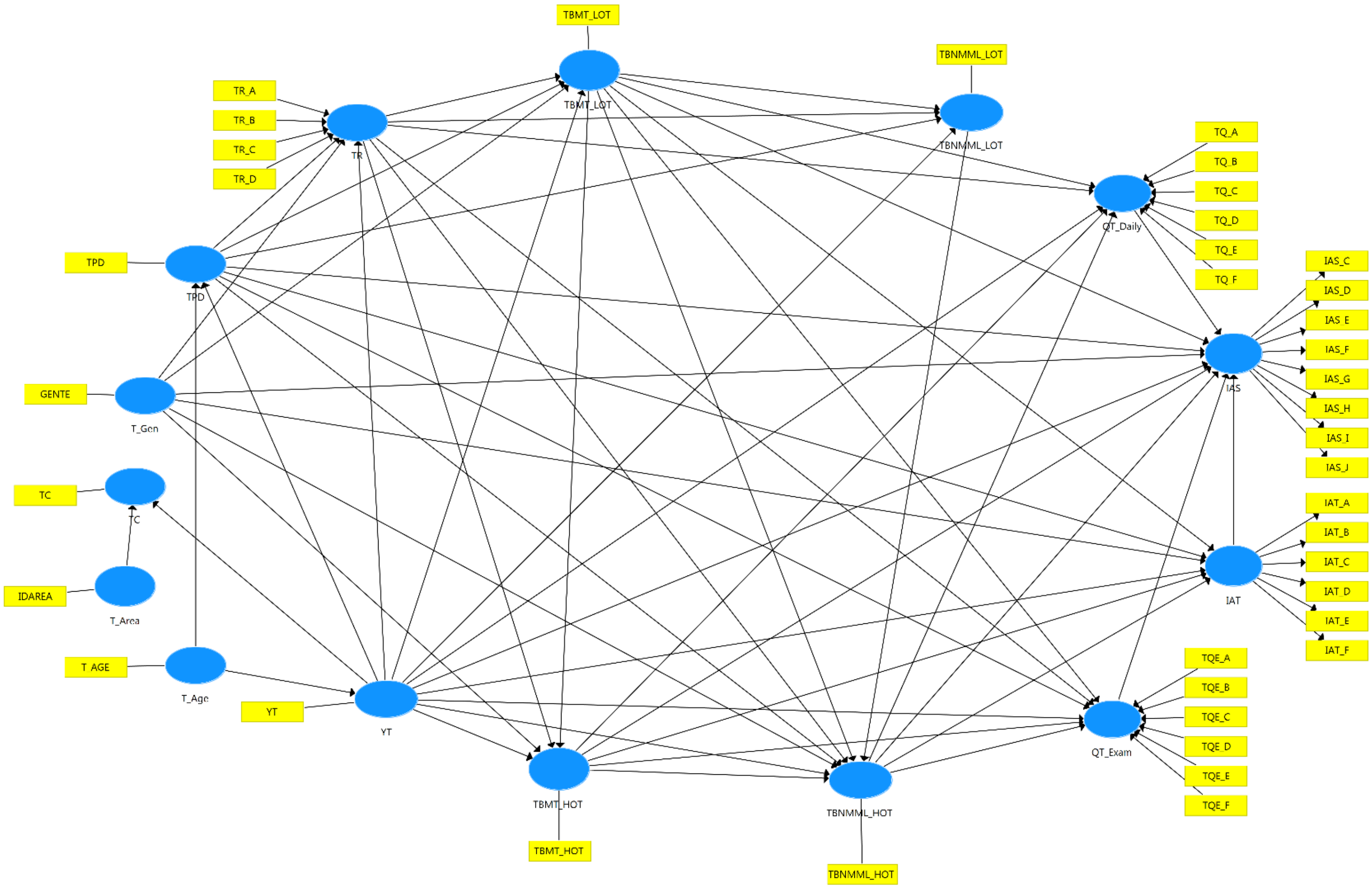


Figure 8.1 Hypothesised model for teacher-level model *(N* = 46)

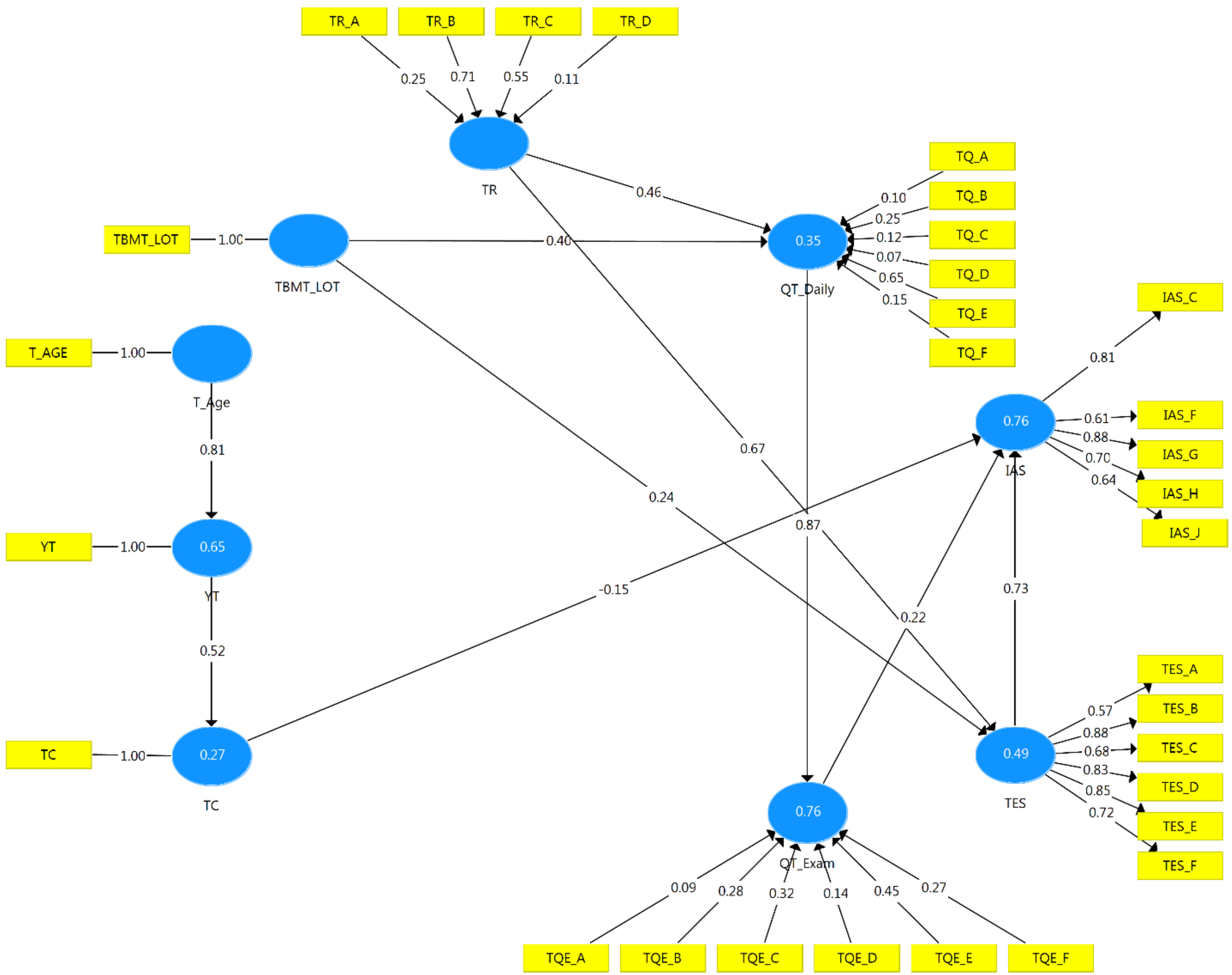


Figure 8.2 Final model for teacher-level model (N = 46)

8.3 Measurement Model (Outer Model)

The relationships between the latent variables and the observed variables are included in the measurement or outer model. The stages of partial least square path analysis for the teacher-level model are similar to those in the student-level model in the previous chapter. Hypothesising the relationships that are involved in the model is the first stage in path model building. The next stage involves estimating the measurement model by assessing its convergent and discriminant validity, followed by a comprehensive evaluation of the results and the subsequent refinement of the measurement model.

8.3.1. Reflective Constructs of Measurement Model

Table 8.2 presents the criteria reported for the evaluation of the reflective models. Indicator loadings of variables are the first criteria in assessing the measurement models. The results indicate that all indicator loadings included in the final model are above 0.50.

Table 8.2

Result Summary for Reflective Measurement (Outer) Model

Latent Variables	Indicators	Loadings	Composite Reliability	AVE
IAS	IAS_C	0.81	0.85	0.54
	IAS_D	Excluded		
	IAS_E	Excluded		
	IAS_F	0.61		
	IAS_G	0.88		
	IAS_H	0.70		
	IAS_J	0.64		
TES	TES_A	0.57	0.89	0.58
	TES_B	0.88		
	TES_C	0.68		
	TES_D	0.83		
	TES_E	0.85		
	TES_F	0.72		

Assessing the convergent validity is the next step in evaluating the measurement model by looking at the average variance extracted (AVE) as a criterion, with the recommendation that it should be greater or equal to 0.50 (Bagozzi & Yi, 1988). Both latent variables meet this requirement, namely TES and IAS. Initially, the AVE of IAS is below 0.50 (0.45). Therefore two indicators (IAS D and IAS E), with loadings below 0.60, are excluded in order to achieve the acceptable AVE.

The discriminant validity of both models is then assessed based on the cross loadings of the indicators and the Fornell-Larcker criterion. TES meets the requirements, as the comparison of the square root of AVE is higher than any latent variables correlation with that particular construct. IAS has the square root of AVE slightly lower than the one other latent variables correlation and therefore does not meet the requirement of the Fornell-Larcker criterion. The details of cross loadings of all indicators and the calculation of the Fornell-Larcker criterion are provided in Table 8.3.

Table 8.3

Relationship between Correlations among Constructs and Square Root of AVEs (Fornell-Larcker Criterion)

	IAS	QT_Daily	QT_Exam	TBMT_LOT	TES	T_Age	TC	YT	TR
IAS	**0.74**								
QT_Daily	0.67	**FMS**							
QT_Exam	0.68	0.87	**FMS**						
TBMT_LOT	0.14	0.38	0.23	**SC**					
TES	0.84	0.66	0.60	0.20	**0.76**				
T_Age	0.02	0.27	0.19	-0.20	0.21	**SC**			
TC	-0.07	-0.04	-0.13	-0.21	0.15	0.48	**SC**		
YT	0.06	0.28	0.19	-0.16	0.17	0.81	0.52	**SC**	
TR	0.61	0.43	0.42	-0.06	0.66	0.23	0.14	0.28	**FMS**

Note. The square root of AVE values is shown on the diagonal and printed in bold (reflective construct only); nondiagonal elements are the latents variable correlations.
SC= Single-item construct
FMS=Formative measurement model

The last assessment of the reflective measurement models involves examining the composite reliability. All variables in the model satisfy the requirement of the composite reliability with the values being greater or equal to 0.70. Based on the assessment of the indicator loadings, convergent validity, discriminant validity and composite reliability, it can be concluded that all reflective variables are acceptable.

Although some variables do not satisfy all four criteria used to assess the reflective measurement models, they all meet at least two of the requirements.

8.3.2. Formative Constructs of Measurement Model

The evaluation of the formative constructs or variables in the measurement models involves collinearity assessments between the formative constructs or variables and the significance testing of the variables. The collinearity of each formative construct is assessed using SmartPLS 3.2.6 (Ringle et al., 2015). Results shown in Table 8.3 indicate that there is no indication of a collinearity problem; the tolerance is above 0.20 and the variance inflation factor (VIF) is less than 5.0 for all constructs.

The significance of weight and loading indicators are presented on Table 8.4. The significance results are obtained through the bootstrapping procedure using SmartPLS 3.2.6 (Ringle et al., 2015). In assessing the significance of formative variables, it is necessary to assess the significance of the outer weights. The significance of the outer loadings are then assessed when the outer weight is not significant. Hair et al. (2013) suggested that an indicator should be removed from the model when the weight is low and both the outer weight and loading are not significant. However, they qualify this, saying that the decision to remove indicators should also be based on the consideration of content validity, as removing the indicators means removing part of the construct (Hair et al., 2013). Table 8.4 shows that most indicators in the model have non-significant weights but mostly have significant loadings. There are three indicators that have neither significant weights nor loadings (these being TQ_D, TQE_A, TR_A and TR_D). However, none of the problematic indicators have been removed from the models as the indicators have an important role based on content validity consideration for its intended construct.

Table 8.4

Outer Weights and Loading Significance Testing Results for Formative Measurement (Outer) Model

Formative Construct	Formative Indicator	Outer Weight (Outer Loading)	Significance Level	VIF
QT_Daily	TQ_A	0.10 (0.43)	NS (*)	1.65
	TQ_B	0.25 (0.69)	NS (***)	2.21
	TQ_C	0.12 (0.72)	NS (***)	1.95
	TQ_D	0.07 (0.22)	NS (NS)	1.07
	TQ_E	0.65 (0.91)	*** (***)	1.94
	TQ_F	0.15 (0.58)	NS (***)	1.44
QT_Exam	TQE_A	0.09 (0.26)	NS (NS)	2.84
	TQE_B	0.28 (0.66)	NS (***)	3.54
	TQE_C	0.32 (0.56)	*** (***)	1.57
	TQE_D	0.14 (0.42)	NS (*)	1.56
	TQE_E	0.45 (0.74)	*** (***)	1.72
	TQE_F	0.27 (0.76)	** (***)	1.87
TR	TR_A	0.25 (0.05)	NS (NS)	1.09
	TR_B	0.71 (0.83)	*** (***)	1.07
	TR_C	0.55 (0.65)	** (***)	1.17
	TR_D	0.11 (0.33)	NS (NS)	1.06

Note: NS= Not Significant
*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

8.4 Structural Model (Inner Model)

After finalising the outer model, the examination of the structural or inner model is conducted. The examination of the outer model may result in the removal of some paths and some variables from the model. The assessment of the inner model follows the procedures outlined in the previous chapter, namely: (a) examining the collinearity between latent variables in the model; (b) examining the significance and relevance of the structural relationships; (c) examining the level of R^2; (d) examining the effect sizes (f^2) and the predictive relevance of Q^2.

Table 8.5 present the summary of final results of the structural or inner model including VIF values. The collinearity assessment of the model shows no indication of collinearity problems as the values of the variance inflation factor (VIF) are below 5, as seen in Table 8.5. Once the collinearity assessment is confirmed, the assessment of path coefficients is conducted, including its magnitude and significance. Sellin (1990) suggested that only significant path coefficients (*β)* with a value greater than 0.10 can be retained as meaningful in the final model in a small sample. As the sample size for teacher-level path analysis is small (*N*=46), the path coefficients below 0.10 are not retained. Table 8.5 also presents the summary of final results of the structural or inner model.

Table 8.5

Summary of Results of Structural (Inner) Model

Criterion	Predictor	VIF	*β*	*SE*	*p*	R^2	f^2	Q^2
YT	T_Age	1	0.81	0.09	***	0.65	1.87	0.61
TC	YT	1	0.52	0.11	***	0.27	0.37	0.25
QT_Daily	TR	1	0.46	0.15	***	0.35	0.32	0.10
	TBMT_LOT	1	0.40	0.15	***		0.25	
QT_Exam	QT_Daily	1	0.87	0.04	***	0.76	3.17	0.12
TES	TR	1	0.67	0.08	***	0.49	0.89	0.23
	TBMT_LOT	1	0.24	0.08	***		0.12	
IAS	TC	1.11	-0.15	0.08	**	0.76	0.09	0.35
	QT_Exam	1.71	0.22	0.10	**		0.27	
	TES	1.72	0.73	0.08	***		1.30	

Note: ***p* < 0.05, *** *p* < 0.01

The weakest path coefficient is the path between teacher certification (TC) and instructional activities (IAS) (β = -0.15, *SE* = 0.08). The strongest coefficient is the path between types of questions used in the exam (QT_Exam) and types of questions used in the daily mathematics classroom (QT_Daily) (β = 0.87, *SE* = 0.04). Most of the path coefficients are significant with one per cent or less probability of error ($p \leq 0.01$), with only two path coefficients being significant with five per cent probability of error ($p \leq 0.05$).

The estimation of the path coefficients is followed by the calculation of R^2 for each of the endogenous variables in the model. Table 8.5 shows that within the endogenous variables instructional activities (IAS) ($R^2 = 0.76$) and types of questions used in the examination (TQ_Exam) (R^2 = 0.76) have the highest R^2. These R^2 values are considered as substantial (Cohen, 1988).

The assessment of effect size (f^2) and predictive relevance (Q^2) are now done. The f^2 indicates an exogenous latent variable contribution to the amount of variance explained by an endogenous latent variable, with the f^2 values of 0.02, 0.15 and 0.35 indicating a small, medium and large effect size respectively (Cohen, 1988). Table 8.5 shows that all variables have either a medium or large effect size. Furthermore, all endogenous variables exhibit predictive relevance indicated by the Q^2 values being greater than 0.

8.4.1. The Relationships between Variables in the Structural (Inner) Model of Student-level Model

This section discusses in detail the relationships between the latent variables within the inner model for the final model, as shown in Figure 8.2. Each endogenous variable with its predictors is looked at individually. There are three exogenous variables in the path model namely: teacher's age (T_AGE), teaching resources (TR) and teachers' beliefs concerning mathematics teaching related to HOT (TBMT_HOT).

Teachers' Experience (YT)

The result shows that the age of teachers (T_AGE) has a direct effect on their experience (YT) (β = 0.81, SE = 0.10). The positive path coefficient of age to experience indicates that the older the teachers are the more experienced they are. This result is to be expected, as the indicator of teacher experience is the number of years teachers have been teaching. Older teachers tend to have more teaching years than younger ones. The R^2 value for YT is 0.65, which indicates that the T_Age explains 65 per cent of the variance teacher experience (YT).

Teachers' Certification (TC)

The experience of teachers (YT) directly relates to the certification of teachers (TC) with the positive path coefficient (β = 0.52, *SE* = 0.11) indicating that teachers with more experience are more likely to have passed the certification examination. The age of the teachers and years of teaching are two of the requirements for teachers to be eligible to participate in the teacher certification programme in Indonesia. The R^2 value for TC is 0.27. This means that teacher experience explains 27 per cent of the variance of teachers' certification (TC).

Types of Questions Used Daily in Mathematics Classroom (QT_Daily)

Two variables, TR (β = 0.46, *SE* = 0.13) and TBMT_LOT (β = 0.40, *SE* = 0.14), have positive direct effects on QT_Daily. This indicates that teaching resources is positively related to the types of questions used in the mathematics classroom. The positive coefficient between TBMT_LOT and QT_Daily indicates that teachers' beliefs concerning mathematics teaching related to LOT has a positive relation to the choice of the types of questions that they use frequently in the mathematics classroom. The R^2 for QT_Daily is 0.35, indicating that the path model explains 35% of the variance of QT_Daily.

Questions Types for Examination (QT_Exam)

The types of questions used daily in the mathematics classroom (QT_daily) is strongly related to the types of questions used in the mathematics examination (QT_Exam). The positive coefficient (β = 0.87, *SE* = 0.05) shows that teachers are likely to give students in examinations similar types of questions to those that they use in the mathematics classrooms. This may also be interpreted as teachers preparing their students for the types of questions that they are given in the examination. QT_Daily accounts for 76% of the variance in QT_Exam (R^2 = 0.76).

Teachers Engaging Students (TES)

Teachers engaging students (TES) covers the following ideas: (a) summarising what students should have learned from the lesson; (b) relating the lesson to students' daily lives; (c) using questioning to elicit reasons and explanation; (d) encouraging all students to improve their performance; (e) praising students for good effort; and (f) bringing interesting materials to class. There are two variables that have a direct positive effect on TES, namely, TR (β = 0.67, *SE* = 0.05) and TBMT_LOT (β = 0.24, *SE*=0.05). This indicates that the teaching resources used in the mathematics classroom have a positive relationship on the engagement of students by teachers. Furthermore, this also indicates that the teachers who have more positive beliefs concerning mathematics teaching related to LOT are recorded as having a more positive engagement with students. The R^2 value for teachers engaging students is 0.49, indicating that the path model explains 49% of the variance of TES.

Instructional Activities for Student (IAS)

Instructional activities (IAS) cover the following: (a) asking students to work on problems with peers with the teacher's guidance; (b) working on problems together with the whole class with direct guidance from the teacher; (c) encouraging students to apply facts, concepts and procedures to solve routine problems; (d) requiring students to explain their answers; (e) encouraging students to relate what they are learning in mathematics to their daily lives; (f) having students decide on their own procedures for solving complex problems; (g) having students work on problems for which there is no immediate obvious method of solution; and (h) giving students a written test or quiz.

QT_Exam (β = 0.32, *SE* = 0.07) and TES (β = 0.65, *SE* = 0.10) have positive direct effects on the instructional activities. This indicates that the types of questions given to students in the examination are positively related to the instructional activities given to students. Furthermore, teachers' efforts at engaging students in the mathematics classroom also positively relate to the instructional activities, indicating that the teachers who frequently engage the students are also more likely to engage in the instructional activities listed. The negative path coefficient between teacher certification (TC (β = -0.17, *SE* = 0.10)) and the instructional activities indicates that teachers who have not passed the certification examination are less likely to engage students in the instructional activities. The path model explains 79% of the variance of instructional activities for students in the mathematics classroom as indicated by the value of R^2 (0.79).

8.5 Summary

In this chapter partial least squares path analysis (PA), using SmartPLS 3.2.6 (Ringle et al., 2015), is employed to examine the relationships associated with variables at the teacher-level. The analysis is based on the theoretical framework presented in Chapter 2. In particular, the path analysis is carried out to investigate the variables at the teacher-level that have an explanatory relationship to classroom practices, including teachers engaging students and instructional activities given to students. The analysis is carried out to seek answers to these research question: (1) What are the teacher-level factors influencing instructional activities in the mathematics classroom? And (2) What are the interrelationships among the teacher-level factors influencing instructional activities in the mathematics classroom?

For the first research questions, the analysis of the final model indicates that the following variables directly influence classroom practices of instructional activities in the mathematics classroom: (a) teacher certification; (b) types of questions used in the examination; and (c) teachers' engagement with students. The teachers' engagement with students is directly influenced by the teaching resources used in mathematics classroom and teachers' beliefs concerning mathematics teaching related to LOT. In addition, the types of questions used in the mathematics classroom is positively related to both the teaching resources used as well as teachers' beliefs concerning mathematics teaching related to LOT. The types of mathematics questions given in the examination is greatly influenced by the choice of questions used in daily mathematics classroom.

In regard to the second research question, the analysis shows that teachers' age is positively associated with teachers' experience; and teachers' experience is positively associated teachers holding certification. The types of questions used in the mathematics classroom and teachers' engaging students are positively influenced by teaching resources and teachers' beliefs concerning mathematics teaching related to LOT.

The results reported in this chapter are concerned with a single-level model of teacher-level variables. A further examination is required using a strategy which can include both teacher-level and student-level variables. This will involve a multilevel analysis, namely hierarchical linear modelling (HLM). The HLM analysis results are presented and discussed in the next chapter.

Chapter 9 Hierarchical Linear Modelling Analysis: Student Mathematics Performance

9.1 Introduction

The data relating to the three levels of student, teacher and school require an investigation using a multilevel analysis procedure. Hierarchical linear modelling (HLM) analysis is a statistical procedure that enables an analysis to be undertaken when the variables involved operate at different levels. A three-level HLM analysis enables both the examination of significant variables associated with the outcome variable within each level as well as the investigation of how significant variables at the upper level influence the slopes of the significant variables at the lower levels toward the outcome variable (Bryk & Raudenbush, 2002).

In this study, the student data is nested within the teacher data and the teacher data is nested within the school data. When data is of a nested nature, it is likely that the relationships between variables do not occur simply at one level but also between and across the various hierarchical levels (Hofmann, 1997). When nested data is mishandled or ignored, incorrect conclusions arise concerning the situation both with respect to aggregation bias and statistical significance (Snijders, 1999). In order to obtain a better understanding of the relationship within and between the hierarchical levels, in this study the HLM 6.08 computer program developed by Raudenbush et al. (2004) is employed.

Earlier techniques that have been employed for dealing with nested data either involve disaggregating data to the lower level unit, with the analysis then being based on the lower level unit, or aggregating the lower level data to the higher level data (Hofmann, 1997; J. W. Osborne, 2000). It has long been acknowledged (Thorndike, 1917) that this generates serious bias, commonly referred as 'aggregation bias'. Thus the disaggregating and aggregating processes not only generate bias but also result in a misestimating of precision and the confounding of the unit of analysis (Bryk & Raudenbush, 2002). These imprecise techniques in the past led to incorrect conclusions as the result of the exclusion of variability at certain levels (Snijders, 1999). HLM analysis is a technique spesifically developed to handle the nested data using maximum likelihood statistical procedures. In addressing the limitations of the conventional techniques, HLM allows for the investigation of causal relationships at both individual levels and between levels as well as examining how the variables at the higher level of the multilevel structure influence the explanation of effects of the variables at the lower level.

This chapter presents the HLM analysis of a three-level model of student mathematics performance. The performance is divided into mathematics performance related to LOT and performance related to HOT. Two three-level models of mathematics performance are examined: (1) a three-level model of mathematics performance related to LOT; and (2) a three-level model of mathematics performance related to HOT. The analyses are conducted in two major stages, specifying the null model and finalising the conditional model. The results of both the models are presented and discussed. A summary of the findings are also presented to conclude the chapter. The analyses are based on the student data set (1135 students), the teacher data set (46 teachers) and the school data set (25 schools).

9.2 The Variables and their Level of Operation

The three sets of variables at the student, teacher and school levels are chosen taking into consideration the results of the analysis in the previous chapters. The variables used at the student-level (level-1), the teacher-level (level-2) and the school-level (level-3) are presented in Table 9.1. Overall there are 15 variables at level-1, 11 variables at level-2 and 11 variables at level-3. As HLM version 6.08 does not incorporate features that allow latent variables to be formed within the overall model, Rasch scores obtained using Conquest 2.0 software (Wu et al., 2007) are used for all constructs that involve composite variates. The variables with Rasch scores at level-1 and listed in Table 9.1 are: LIKE_MATH, VALUE_MATH, MCONF, INDI_JUD, SBM_LOT, SBM_HOT, SLA_LOT, and SLA_HOT; at level-2 are TBNMML LOT, TBNMML HOT, TBMT HOT, TBMT LOT, IAS, TES, T_Resource, QT_Daily and QT_Exam. None of the variables operating at level-3 require Rasch scores as most of the variables are categorical (ID_AREA) and continuous (NUM_STU, CL_SIZE, NUM_TE, NUM_T, NUM_MT, MT_C, MT_Y9, MPRO_A, MPRO_B).

One variable at level-1 is a composite variable, namely home possessions (HOMEPOS), and this composite variable is created by arithmetically summing the availability of the possessions at the student's household. One variable at level-2 is a composite variable, namely teacher professional development (TPD), and this composite variable is created by arithmetically summing the number of professional development training courses attended by each teacher in the last two years. Furthermore, one variable at level-3 is a composite variable, namely school resources (SR), and this composite variable is created by arithmetically summing the availability of the resources at school.

Table 9.1

Observed Variables in Hierarchical Linear Modelling Analyses with Mathematics Performance as the Outcome

Variable		Description	Coding
Student level :Level-1			
Students' background	GENSTU	Gender of student	0=Female, 1=Male
Students' background	H_EDU	Highest education of parent	0= Did not complete Year 6 to 7= Master's degree, doctoral degree or professional degree such as law or medicine
Students' background	EXP_EDU	Students' expected education	0= High School to 5= Doctoral degree
Students' background	HOMEPOS	Home possessions	Raw score
Students' attitudes	LIKE_MATH	Liking mathematics	Rasch score
Students' attitudes	VALUE_MATH	Valuing mathematics	Rasch score
Students' beliefs	MCONF	Mathematics confidence	Rasch score
Students' beliefs	INDI_JUD	Individual judgement of mathematics ability	Rasch score
Students' beliefs	SBM_LOT	Student belief of mathematics related to lower order thinking	Rasch score
Students' beliefs	SBM_HOT	Student belief of mathematics related to higher order thinking	Rasch score
Learning activities	SLA_LOT	Learning activities related to lower order thinking	Rasch score
Learning activities	SLA_HOT	Learning activities related to higher order thinking	Rasch score
Mathematics performance	MATH_LOT	Mathematics score related to lower order thinking	Rasch score
Mathematics performance	MATH_HOT	Mathematics score related to higher order thinking	Rasch score
Teacher level :Level-2			
Teachers' background	YT	Years of teaching	Years
Teachers' background	GENTE	Gender of teacher	0=Female, 1=Male
Teachers' background	TC	Teacher's certification	1=Yes 0=No
Teachers' background	TPD	Teacher professional development	Raw score
Teachers' beliefs	TBNMML_HOT	Teachers beliefs of nature of mathematics and mathematics learning related to HOT	Rasch score
Teachers' beliefs	TBMT_HOT	Teachers beliefs of mathematics teaching related to HOT	Rasch score
Teachers' beliefs	TBNMML_LOT	Teachers beliefs of nature of mathematics and mathematics learning related to LOT	Rasch score
Teachers' beliefs	TBMT_LOT	Teachers beliefs of mathematics teaching related to LOT	Rasch score
Instructional activities	IAS	Instructional activities	Rasch score
Instructional activities	TES	Engaging students	Rasch score
School level :Level-3			
Schools' background	IDAREA	ID area	1=Urban, 2=Rural
Schools' background	NUM_STU	Numbers of students at school	Raw score
Schools' background	CL_SIZE	Class size	Raw score
Schools' background	NUM_TE	Number of teachers at schools	Raw score
Schools' background	NUM_T	Number of teachers with certification at schools	Raw score
Schools' background	NUM_MT	Number of mathematics teachers at school	Raw score
Schools' background	MT_C	Numbers of mathematics teachers with certification	Raw score
Schools' background	MT_Y9	Number of mathematics teachers teaching in year 9	Raw score
Schools' background	MPRO_A	Additional program: Additional mathematics program after school	1=Yes 0=No
Schools' background	MPRO_B	Additional program: Mathematics Olympiad	1=Yes 0=No
School Resources	SR	School resources	Raw score
Outcome			
MATH_LOT		The outcome of Model 1	
MATH_HOT		The outcome of Model 2	

The conceptual model for the three-level HLM analyses of mathematics performance related to LOT and HOT are illustrated in Figures 9.1 and 9.2.

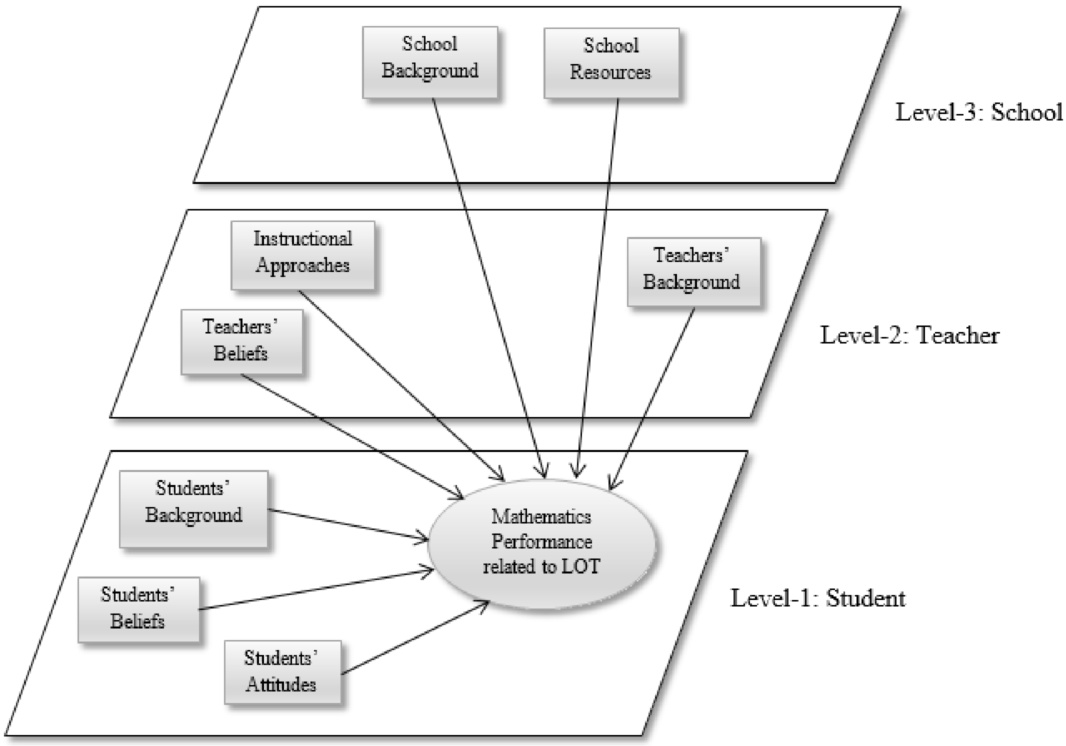


Figure 9.1 *Conceptual Three-level model of mathematics performance related to LOT*

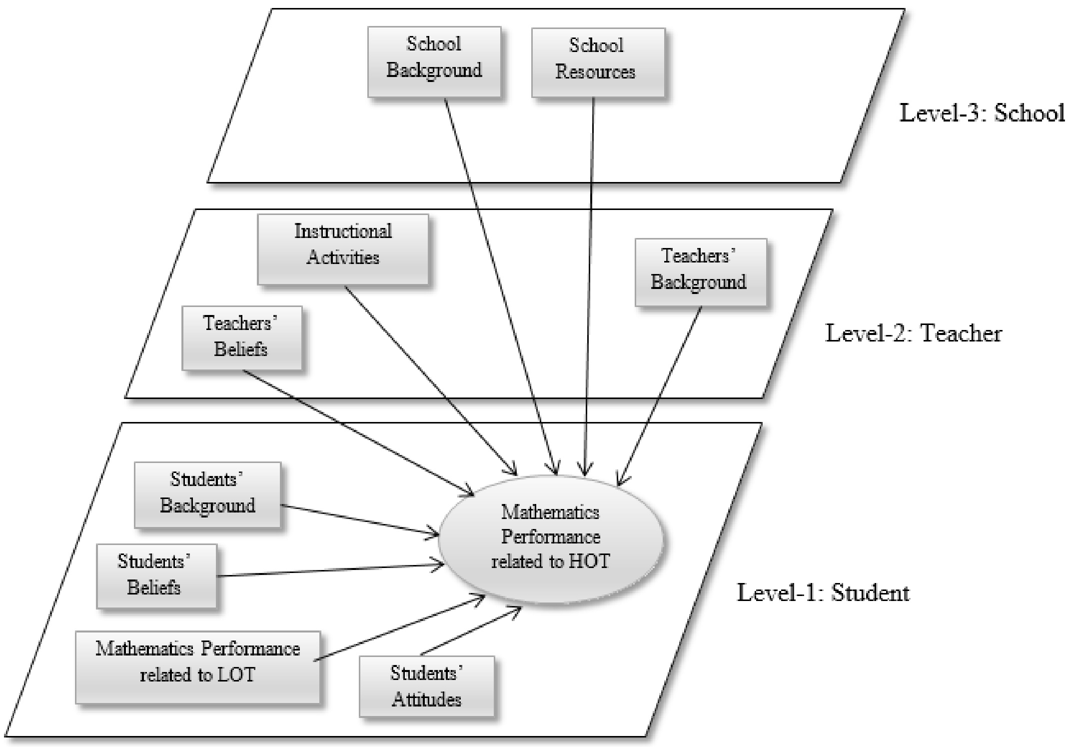


Figure 9.2 *Conceptual Three-level model of mathematics performance related to HOT*

9.3 Formulating and Testing Three-level Models

The three-level HLM model in this study is created using a single cross-sectional data set with a three-level structure consisting of students (level-1), nested within classrooms (teachers, level-2), and teachers nested within schools (level-3). There are two major stages of conducting the HLM analysis. The first stage involves examining the fully unconditional or null model (the terms mean the same and are used interchangeably) of a three-level model. This is the simplest model and consists of no explanatory variables at any level, and involves the calculation of the available variance in the outcome variable across the three-level model of students, teachers and schools (Bryk & Raudenbush, 2002). The second stage of a three-level HLM analysis specifies the conditional model, involving the examination of which of the variables at each level can explain the variability obtained across different levels of the model (Bryk & Raudenbush, 2002). The investigation in the second stage begins with examining and estimating the significant variables at level-1 (student) related to the outcome variable, followed by examining and estimating the significant variables at level-2 (teacher) related to the outcome variables as well as the estimated relationships between level-1 variables and the outcome variable, and finalised by examining and estimating the significant variables at level-3 (schools) related to the outcome variables as well as the relationships between the significant variables at level-1 and level-2 to the outcome variable. The process occurs in several steps and involves forming hypotheses related to the many possibilities of the conditional model. The final step involves the full three-level conditional model with the estimated significant variables at the appropriate level.

HLM 6.08 software is employed to run firstly for the null model, then for the many alternative models until the final conditional model is obtained with only significant effects included. At each step maximum likelihood estimation procedures are employed rather than least squares, although at the final stage of the examination of the final model it is necessary to consider the variance explained at each level when compared with the estimated variance available at each level of the initial null model.

9.4 HLM Findings for Mathematics Performance as the Outcomes Variables

This section presents and discusses the HLM analyses of mathematics performance, related to both LOT and HOT, as the outcome variables. The null model, level-1 conditional model and final three-level model for both mathematics performance outcomes are presented.

9.4.1. Three-level Model of Mathematics Performance related to LOT

The null model, final level-1 conditional model and final three-level model of HLM result for student mathematics performance related to LOT are presented in this section.

Null Model

The null model with mathematics performance as the outcome variable is presented as follows:

Level-1 model: the student level model

$$MATH\_LOT_{ijk} = \pi_{0jk} + e_{ijk} \qquad [9.1]$$

Where

$MATH\_LOT_{ijk}$ is the mathematics performance related to LOT of student *i* under teacher *j* in school *k*;

π_{0jk} is the mean of mathematics performance related to LOT under teacher *j* in school *k*; and

e_{ijk} is a random student effect

Level-2 Model: the teacher level model

$$\pi_{0jk} = \beta_{00k} + r_{0jk} \qquad [9.2]$$

Where

β_{00k} is the mean of mathematics performance related to LOT in school *k*;

r_{0jk} is a random teacher effect

Level-3 model: the school level model

$$\beta_{00k} = \gamma_{000} + u_{00k} \qquad [9.3]$$

Where

γ_{000} is the grand mean;

u_{00k} is random school effect

The results for the fully unconditional model of mathematics performance related to LOT is recorded in Table 9.2. The results enable the calculation of variance available at each level employing the values of σ^2, τ_π and τ_β obtained from the analysis. The variances available are 46%, 20% and 34% at level-1, level-2 and level-3 respectively (the detailed computation of variance available is in Table 9.3). The variance for mathematics performance is more widely spread across the levels than recorded (in Chapter 9) for the attitudes as the outcomes, where variance is strong at in level-1. The results for the fully unconditional model of mathematics performance related to LOT also indicate that the null model has high reliability at 0.91 and 0.70 for level-2 and level-3 intercepts respectively. Furthermore, the chi-square value for level-1 and level-2, χ^2 (16) = 195.75, p=0.00, demonstrates that variance is significant. The variation at level-3 is also significant which is shown by the value of chi-square χ^2 (24) = 91.48, p=0.00.

Table 9.2

Fully Conditional Model of Mathematics Performance related to LOT

Final estimation of fixed effects (with robust standard errors)					
Fixed effect	Coefficient	Standard error	t-ratio	Approx. *df*	*p*-value
For INTRCPT1, π0					
For INTRCPT2, β00					
INTRCPT3, γ000	-1.07	0.18	-5.86	24	0.00
Final estimation of level -1 and level-2 variance components					
Random effect	Standard deviation	Variance component	*df*	Chi-square (χ^2)	*p*-value
INTRCPT1, r0	0.59	0.34	16	195.75	0.00
level-1, e	0.89	0.79			
Final estimation of level -3 variance components					
INTRCPT1/INTRCPT2, u00	0.77	0.59	24	91.48	0.00
Statistics for current covariance components model					
Deviance			3081.42		
Number of estimated parameters			4		

Final Model

The final model is specified after the analysis and examination of the null model by identifying the significant variables at level-1, then level-2 and then level-3. The final three-level model of mathematics performance related to LOT is presented in the Equations 9.4 to 9.15.

Level-1 model: the student level model

$$MATH\_LOT_{ijk} = \pi_{0jk} + \pi_{1jk}(LIKE\_MATH)_{ijk} + \pi_{2jk}(SBM\_HOT)_{ijk} + \pi_{3jk}(SBM\_LOT)_{ijk} + e_{ijk} \quad [9.4]$$

Level-2 Model: the teacher level model

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(TPD)_{ijk} + \beta_{02k}(IAS)_{ijk} + \beta_{03k}(TBMT\_HOT)_{ijk} + r_{0jk} \quad [9.5]$$

$$\pi_{1jk} = \beta_{10k} + r_{1jk} \quad [9.6]$$

$$\pi_{2jk} = \beta_{20k} + r_{2jk} \quad [9.7]$$

$$\pi_{3jk} = \beta_{30k} + r_{3jk} \quad [9.8]$$

Level-3 model: the school level model

$$\beta_{0jk} = \gamma_{000} + \gamma_{001}(SR) + u_{00k} \quad [9.9]$$

$$\beta_{01k} = \gamma_{010} \quad [9.10]$$

$$\beta_{02k} = \gamma_{020} \quad [9.11]$$

$$\beta_{03k} = \gamma_{030} \quad [9.12]$$

$$\beta_{20k} = \gamma_{200} \quad [9.13]$$

$$\beta_{30k} = \gamma_{300} \quad [9.14]$$

Mixed model

$$MATH\_LOT_{ijk} = \gamma_{000} + \gamma_{001}(SR)_k + \gamma_{010}(TPD)_{ijk} + \gamma_{020}(IAS)_{ijk} + \gamma_{030}(TBMT\_HOT)_{ijk} + \gamma_{100}(LIKE\_MATH)_{ijk} + \gamma_{200}(SBM\_HOT)_{ijk} + \gamma_{300}(SBM\_LOT)_{ijk} + r_{0jk} + r_{1jk}(LIKE\_MA)_{ijk} + r_{2jk}(SBM\_HOT)_{ijk} + r_{3jk}(SBM\_LOT)_{ijk} + u_{00k} + +e_{ijk} \quad [9.15]$$

The model indicates that mathematics performance related to LOT is defined as a function with seven main effects across the levels. The three main effects at level-1 are: Students' Liking of Mathematics (LIKE_MATH); Students' Beliefs concerning Mathematics related to HOT (SBM_HOT); and Students' Beliefs concerning Mathematics related to LOT (SBM_LOT). The three main effects at level-2 are: Teachers' Professional Development (TPD); Teachers' Beliefs concerning Mathematics Teaching related to HOT (TBMT_HOT); and Instructional Activities (IAS). The only effect at level-3 is school resources (SR). There are no interactions between the explanatory variables and the slopes across the levels. The result of the final three-level model of Mathematics Performance related to LOT (MATH_LOT) is presented in Table 9.2.

The Effects of Level-1, Level-2 and Level-3 Significant Variables on Mathematics Performance related to LOT

Based on the result of the three-level HLM analysis in Table 9.3, the relationships between significant variables across the levels are shown in Figure 9.3. This indicates that three variables at level-1 have direct effects on the outcome, two with positive relationships (LIKE_MATH and SBM_HOT) and one with a negative relationship (SBM_LOT). Students' Liking of Mathematics has a positive relationship to Mathematics Performance related to LOT (LIKE_MATH) ($\gamma 100$ = 0.10, *SE*=0.02), from which it can be stated that the more students like mathematics, the higher will be their Mathematics Performance related to LOT. Students' Beliefs concerning Mathematics related to HOT also positively influence their Mathematics Performance related to LOT (SBM_HOT) ($\gamma 200$ = 0.07, *SE*=0.03). It can be argued that the more positive students' beliefs toward mathematics related to HOT, the higher be their mathematics performance related to LOT. However, students' beliefs concerning mathematics related to LOT has a negative effect on performance (SBM_LOT)($\gamma 300$ =-0.09, *SE*=0.04), which can be argued to indicate that students who have more positive beliefs concerning mathematics related to LOT tend to have lower mathematics performance related to LOT.

There are three main effects at level-2. Instructional activities (IAS, $\gamma 020$= 0.18, *SE*= 0.07) and Teacher Beliefs concerning Mathematics related to HOT (TBMT_HOT) ($\gamma 030$=0.12, *SE*=0.05) both positively influence the outcome. This indicates that the more frequently teachers conduct the instructional activities, the higher the students' perform in mathematics related to LOT. Furthermore, those students whose teachers have more positive beliefs concerning mathematics related to HOT are found to perform at a higher level. On the other hand, Teacher Professional Development (TPD) ($\gamma 010$=-0.13, *SE*=0.05) has a negative relationship to the outcome, indicating that students whose teachers are involved in fewer teacher professional development programs (TPD) tend to perform higher in mathematics related to LOT.

School resources (SR) is the only explanatory variable at level-3 having a direct effect on the outcome ($\gamma001$=0.41, *SE*=0.17). This can be interpreted as the better equipped the schools are, the better students perform in mathematics related to LOT.

Table 9.3

Final Model of Students' Mathematics Performance related to LOT

Final estimation of fixed effects					
Fixed effect	Coefficient	Standard error	*t*-ratio	Approx. *df*	*p*-value
For INTRCPT1, π0					
For INTRCPT2, β00					
INTRCPT3, γ000	-1.04	0.15	-7.02	23	0.00
SR, γ001	0.41	0.17	2.47	23	0.02
For TPD, β01					
INTRCPT3, γ010	-0.13	0.05	-2.77	37	0.01
For IAS, β02					
INTRCPT3, γ020	0.18	0.07	2.56	37	0.02
For TBMT_HOT, β03					
INTRCPT3, γ030	0.12	0.05	2.27	37	0.03
For LIKE_MA slope, π1					
For INTRCPT2, β10					
INTRCPT3, γ100	0.10	0.02	4.44	40	0.00
For SBM_HOT slope, π2					
For INTRCPT2, β20					
INTRCPT3, γ200	0.07	0.03	2.34	40	0.03
For SBM_LOT slope, π3					
For INTRCPT2, β30					
INTRCPT3, γ300	-0.09	0.04	-2.37	40	0.02
Final estimation of level-1 and level-2 variance components					
Random effect	Standard deviation	Variance component	*df*	Chi-square (χ^2)	*p*-value
INTRCPT1, r0	0.54	0.29	13	286.68	0.00
LIKE_MA slope, r1	0.10	0.01	40	84.94	0.00
SBM_LOT slope, r2	0.12	0.01	40	60.55	0.02
SBM_HOT slope, r3	0.18	0.03	40	83.29	0.00
level-1, e	0.81	0.66			
Final estimation of level-3 variance components					
INTRCPT1/INTRCPT2, u00	0.59	0.34	23	104.05	0.00
Statistics for current covariance components model					
Deviance			2934.97		
Number of estimated parameters			20		

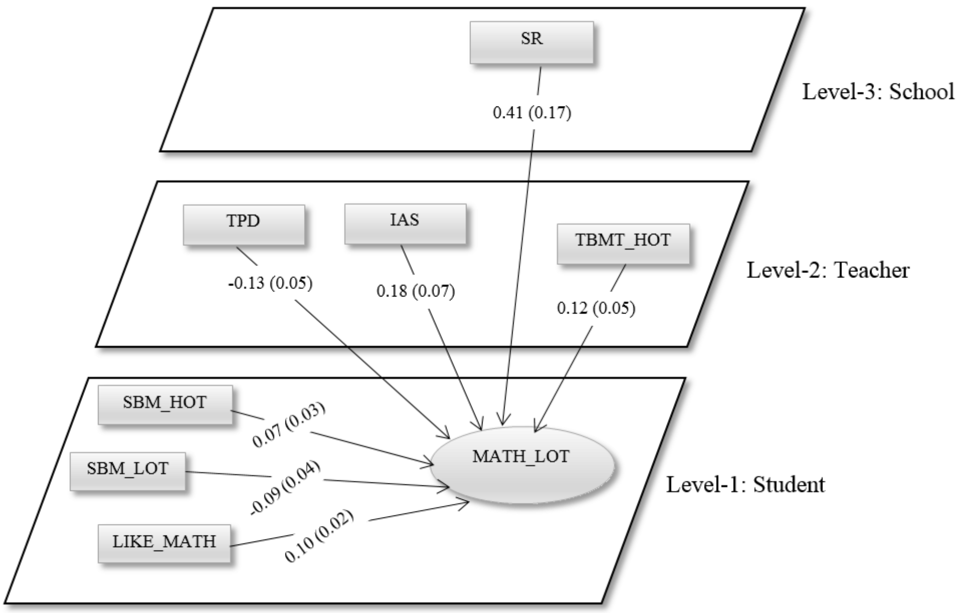


Figure 9.3 *Final three-level HLM model for mathematics achievement related to LOT*

Total Variance Extracted and Deviance Reduction

The detailed calculation of the estimated variance component at each level and variance extracted are recorded on Table 9.4.

Table 9.4

Estimation of Variance Component and Explained Variance for Mathematics Achievement related to LOT

	Estimation of variance component between:		
	Students (level-1)	Teachers (level-2)	School (level-3)
Number of cases	1135	41	25
Null Model	$\delta^2 = 0.79$	$\tau_\pi = 0.34$	$\tau_\beta = 0.59$
Final three-level HLM model	$\delta^2 = 0.66$	τ_π=0.29	$\tau_\beta = 0.34$
Variance available at each level	$\frac{\delta^2}{\delta^2 + \tau_\pi + \tau_\beta}$ $\frac{0.79}{1.72} = 0.46 = 46\ \%$	$\frac{\tau_\pi}{\delta^2 + \tau_\pi + \tau_\beta}$ $\frac{0.34}{1.72} = 0.20 = 20\ \%$	$\frac{\tau_\beta}{\delta^2 + \tau_\pi + \tau_\beta}$ $\frac{0.59}{1.72} = 0.34 = 34\ \%$
Proposition of variance explained by final three-level model	$\frac{\delta^2\ (null) - \delta^2(final\ model)}{\delta^2(null)}$ $\frac{0.79-0.66}{0.79} = 0.16 = 16\%$	$\frac{\tau_\pi\ (null) - \tau_\pi(final\ model)}{\tau_\pi(null)}$ $\frac{0.34-0.29}{0.34} = 0.15 = 15\%$	$\frac{\tau_\beta(null) - \tau_\beta(final\ model)}{\tau_\beta(null)}$ $\frac{0.59-0.34}{0.59} = 0.42 = 42\%$
Overall variance explained	$\frac{(0.79 - 0.66) + (0.34 - 0.29) + (0.59 - 0.34)}{0.79 + 0.34 + 0.59} = 0.25 = 25\ \%$		
Deviance reduction	3081.42 -2934.97 = 146.45		

The variance extracted by level-1, level-2 and level-3 are 16, 15 and 45 per cent respectively. The total variance extracted across the three levels is 25 per cent. Based on the result of the final estimation of the three-level model of mathematics performance related to LOT in Table 9.2, the deviance of the model is decreased by 146.45 points (from 3081.42 to 2934.97) with 20 parameters estimated. The large deviance reduction indicates that the final model is stronger than the null model.

9.4.2. Three-level Model of Mathematics Performance related to HOT

The null model and final model of the three-level analysis of mathematics performance related to HOT are presented in this section.

Null Model

The null model of the mathematics performance related to HOT as the outcome is as follow:

Level-1 model: the student level model

$$MATH\_HOT_{ijk} = \pi_{0jk} + e_{ijk} \quad [9.16]$$

Where

$MATH\_HOT_{ijk}$ is the mathematics performance related to HOT of student *i* under teacher *j* in school *k*;

π_{0jk} is the mean mathematics performance related to HOT under teacher *j* in school *k* and;

e_{ijk} is random student effect

Level-2 Model: the teacher level model

$$\pi_{0jk} = \beta_{00k} + r_{0jk} \quad [9.17]$$

Where

β_{00k} is the mean of mathematics performance related to HOT in school k;

r_{0jk} is a random teacher effect

Level-3 model: the school level model

$$\beta_{00k} = \gamma_{000} + u_{00k} \quad [9.18]$$

Where

γ_{000} is the grand mean;

u_{00k} is a random school effect

Table 9.5 records the results of the fully unconditional model of mathematics performance related to HOT. The variance extracted at level-2 is significant, indicated by the chi-square value, χ^2 (16) = 119.94, p= 0.00. The significance is also shown for the variance at level-3, with the chi-square value, χ^2 (24) = 102.06, p= 0.00.

Table 9.5

Fully Unconditional Model of Mathematics Achievement related to HOT

Final estimation of fixed effects (with robust standard errors)					
Fixed effect	Coefficient	Standard error	*t*-ratio	Approx. *df*	*p*-value
For INTRCPT1, π0					
For INTRCPT2, β00					
INTRCPT3, γ000	-2.56	0.26	-9.97	24	0.00
Final estimation of level -1 and level-2 variance components					
Random effect	Standard deviation	Variance component	*df*	Chi-square	*p*-value
INTRCPT1, r0	0.74	0.55	16.00	119.94	0.00
level-1, e	1.51	2.27			
Final estimation of level -3 variance components					
INTRCPT1/INTRCPT2, u00	1.10	1.21	24	102.06	0.00
Statistics for current covariance components model					
Deviance			4267.08		
Number of estimated parameters			4		

In terms of the reliability of the intercepts, both level-1 and level-2 have a high reliability of 0.86 and 0.74 respectively. The variance allocated is stated by σ^2, τ_π and τ_β with 56%, 14% and 30% of variance allocated at level-1, level-2, and level-3 respectively. Mathematics performance related to HOT show that the major components of variance are allocated at level-1 and level-3.

Final Model

There are four explanatory variables found to be significantly related to the outcome: (a) Students' Expected Education (EXP_EDU); (b) Students' Individual Judgements of their Mathematics Ability (INDI_JUD); (c) Students' Belief concerning Mathematics related to LOT (SBM_LOT); and (d) Mathematics Performance related to LOT (MATH_LOT).

The significant variables at level-2 and level-3 are then included in the final model. The final model of the three-level model of mathematics related to HOT is stated in Equations 9.19 to 9. 32.

Level-1 model: the student level model

$$MATH\_HOT_{ijk} = \pi_{0jk} + \pi_{1jk}(EXP\_EDU)_{ijk} + \pi_{2jk}(INDI\_JUD)_{ijk} + \pi_{3jk}(SBM\_LOT)_{ijk} + \pi_{4jk}(MATH\_LOT)_{ijk} + e_{ijk} \quad [9.19]$$

Level-2 model: the teacher level model

$$\pi_{0jk} = \beta_{00j} + \beta_{01k}(TC)_{jk} + \beta_{02k}(TBMT\_HOT\ )_{jk} + r_{0jk} \quad [9.20]$$

$$\pi_{1jk} = \beta_{10j} \quad [9.21]$$

$$\pi_{2jk} = \beta_{20j} \quad [9.22]$$

$$\pi_{3jk} = \beta_{30j} + r_{3jk} \quad [9.23]$$

$$\pi_{4jk} = \beta_{40j} + r_{4jk} \quad [9.24]$$

Level-3 model: the school level model

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(MPRO\_B\ )_{k} + u_{00k} \quad [9.25]$$

$$\beta_{01k} = \gamma_{010} \quad [9.26]$$

$$\beta_{02k} = \gamma_{020} \quad [9.27]$$

$$\beta_{10k} = \gamma_{100} \quad [9.28]$$

$$\beta_{20k} = \gamma_{200} + u_{20k} \quad [9.29]$$

$$\beta_{30k} = \gamma_{300} \quad [9.30]$$

$\beta_{40k} = \gamma_{400}$ [9.31]

Mixed model

$MATH\_HOT_{ijk} = \gamma_{000} + \gamma_{001}(MPRO\_B)_k + \gamma_{010}(TC)_{jk} + \gamma_{020}(TBMT\_HOT)_{ijk} + \gamma_{100}(EXP\_EDU)_{ijk} + \gamma_{200}(INDI\_JUD)_{ijk} + \gamma_{300}(SBM\_LOT)_{ijk} + \gamma_{400}(MATH\_LOT)_{ijk} + +r_{0jk} + r_{3jk}(SBM\_LOT)_{ijk} + r_{4jk}(MATH\_LOT)_{ijk} + u_{00k} + u_{20k}(INDI\_JUD)_{ijk} + e_{ijk}$ [9.32]

The final model states that Mathematics Performance related to HOT (MATH_HOT) is estimated to be a function with seven main effects and random errors. The seven main effects are: the four explanatory variables at level-1, as stated above; two variable at level-2 (Teacher Certification (TC) and Teachers' Beliefs concerning Mathematics Teaching related to HOT (TBMT_HOT)); and one variable at level-3 (the availability of a special mathematics program 'Mathematics Olympiad' (MPRO_B)).

The Effect of Level-1, Level-2 and Level-3 Explanatory Variables on Mathematics Performance related to HOT

Table 9.6 records the result of the final estimation of the final model of mathematics performance related to HOT. The relationships between significant variables and the outcome are also presented in Figure 9.4. At level-1 there are four main effects toward the outcome, three of which are positive: (a) Students' Educational Expectation (EXP_EDU) ($\gamma100 = 0.09, SE = 0.03$); (b) Students' Mathematics Performance related to LOT (MATH_LOT) ($\gamma400 = 0.60, SE = 0.08$); and (c) Students' Individual Judgement of Mathematics Ability (INDI_JUD) ($\gamma200 = 0.10, SE = 0.03$). One variable has a negative effects toward the outcome, which is student beliefs concerning mathematics related to LOT (SBM_LOT) ($\gamma300 = -0.08, SE = 0.04$).

The results can be interpreted that students' mathematics performance related to LOT influences the mathematics performance related to HOT whereby the higher they perform in mathematics related to LOT the higher is their mathematics performance related to HOT. Students' educational expectation has a positive effect on performance whereby the higher their educational expectation, the higher will be their mathematics performance related to HOT. Individual judgement also has a positive relationship to the outcome, which indicates that those students who have stronger individual judgements of their mathematics ability tend to perform higher in mathematics related to HOT. Students' beliefs concerning mathematics related to LOT influence the outcome negatively, indicating that those students with more positive beliefs concerning mathematics related to LOT tend to perform lower in mathematics related to HOT.

At level-2, two significant variables, (Teacher Certification (TC) ($\gamma 010 = 0.45, SE = 0.22$) and Teachers' Beliefs concerning Mathematics related to HOT (TBMT_HOT) ($\gamma 020 = 0.16, SE = 0.04$)) have direct positive effects on the outcome. These results indicate that students who are taught by those teachers who have already passed the certification tends to have a higher mathematics performance related to HOT. In addition, the more positive the teachers' beliefs concerning mathematics teaching related to HOT, the higher is the students' performance on mathematics related to HOT. Conversely, students whose teachers have more positive beliefs concerning mathematics teaching related to HOT tend to have higher mathematics performance related to HOT.

Only one significant variable at level-3 has a direct positive effect on the outcome, namely, the availability of 'Mathematics Olympiad' club at the school (MPRO_B) ($\gamma 001 = 0.98, SE = 0.31$), indicating that those schools with the "Mathematics Olympiad" club tend to have students with high mathematics performance related to HOT.

Table 9.6

Final Model of Students' Mathematics Performance related to HOT

Final estimation of fixed effects (with robust standard errors)					
Fixed effect	Coefficient	Standard error	*t*-ratio	Approx. *df*	*p*-value
For INTRCPT1, π0					
For INTRCPT2, β00					
INTRCPT3, γ000	-3.37	0.21	-15.72	23	0
MPRO_B, γ001	0.98	0.31	3.13	23	0.01
For TC, B01					
INTRCPT3, γ010	0.45	0.22	2.09	38	0.04
For TBMT_HOT, B02					
INTRCPT3, γ020	0.16	0.04	3.95	38	0
For EXP_EDU slope, π1					
For INTRCPT2, β10					
INTRCPT3, γ100	0.09	0.03	3.06	1127	0.00
For INDI_JUD slope, π2					
For INTRCPT2, β20					
INTRCPT3, γ200	0.10	0.03	3.28	24	0.00
For SBM_LOT slope, π3					
For INTRCPT2, β30					
INTRCPT3, γ300	-0.08	0.04	-2.49	23	0.02
For MATH_LOT slope, π4					
For INTRCPT2, β40					
INTRCPT3, γ400	0.60	0.08	7.72	24	0
Final estimation of level -1 and level-2 variance components					
Random effect	Standard deviation	Variance component	*df*	Chi-square	*p*-value
INTRCPT1, r0	0.46	0.21	12	86.62	0
SBM_LOT slope, r3	0.20	0.04	38	54.14	0.04
MATH_LOT slope, r4	0.32	0.11	38	98.23	0
level-1, e	1.33	1.77			
Final estimation of level -3 variance components					
INTRCPT1/INTRCPT2, u00	0.47	0.22	23	74.25	0
INDI_JUD/INTRCPT2, u20	0.08	0.01	24	35.76	0.06
Statistics for current covariance components model					
Deviance			3997.31		
Number of estimated parameters			18		

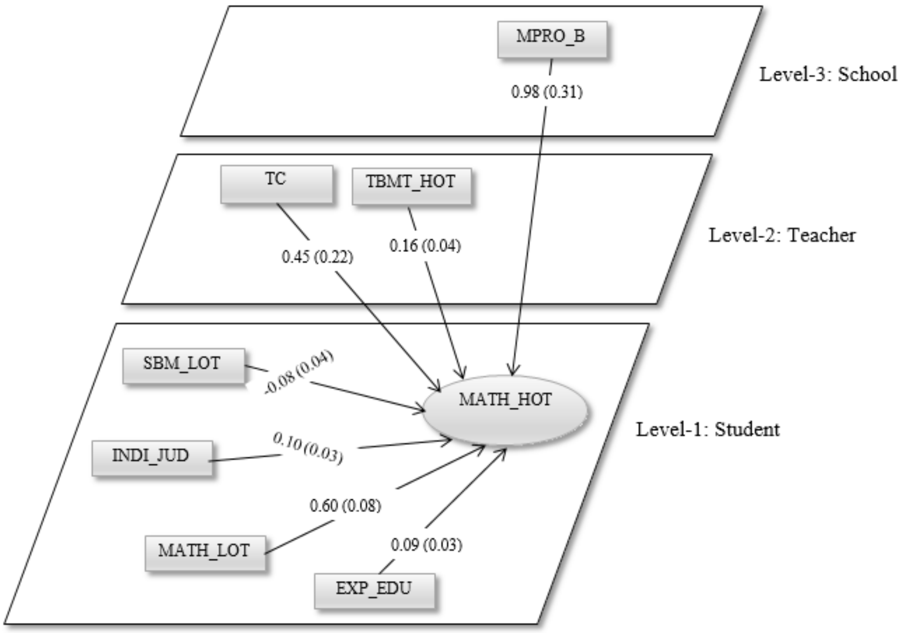


Figure 9.4 *Final three-level HLM model for mathematics performance related to HOT*

Total variance explained and deviance reduction

Table 9.7 records the detailed computation for the initial variance available at each level and variance explained across the levels as well as the total variance extracted by the three-level model.

Table 9.7

Estimation of Variance Component and Explained Variance for Mathematics Performance related to HOT

	Estimation of variance component between:		
	Students (level-1)	Teachers (level-2)	School (level-3)
Number of cases	1135	41	25
Null Model	$\delta^2 = 2.27$	$\tau_\pi = 0.55$	$\tau_\beta = 1.21$
Final three-level HLM model	$\delta^2 = 1.77$	τ_π=0.21	$\tau_\beta = 0.22$
Variance available at each level	$\frac{2.27}{4.03} = 0.56 = 56\%$	$\frac{0.55}{4.03} = 0.14 = 14\%$	$\frac{1.21}{4.03} = 0.30 = 30\ \%$
Proposition of variance explained by final three-level model	$\frac{2.27-1.77}{2.27} = 0.22 = 22\%$	$\frac{0.55-0.21}{0.55} = 0.62 = 62\%$	$= \frac{1.21-0.22}{1.21} = 0.82 = 82\%$
Overall variance explained	$\frac{(2.27-1.77)+(0.55-0.21)+(1.21-0.22)}{2.27+0.55+1.21} = 0.45 = 45\ \%$		
Deviance Reduction	4267.08-3997.31=269.77		

The initial variance available at each level are 56, 14 and 30 per cent at level-1, level-2 and level-3 respectively, and the proportion of variance explained at each level are 22%, 62%, and 82% at level-1, level-2 and level-3 respectively. The three-level model explains a high variance of 45 per cent. Furthermore, the reduction in deviance also indicates that the three-level model of mathematics performance related to HOT is stronger than the null model. The deviance is decreased by 269.77 with 18 estimated parameters.

9.5 Summary

This chapter examines the three-level model of student mathematics performance including performance related to LOT and performance related to HOT. The multilevel model is required in consideration of the nested nature of the data. The HLM analysis is employed to examine the model for significant effects at each level, the student level (level-1), the teacher level (level-2) and the school level (level-3) that are hypothesised to influence the outcome of students' mathematics performance involving students' mathematics performance related to LOT and HOT. The analyses also enable an examination of the interaction effects operating between variables and across levels. The HLM analysis is conducted in two major stages, namely, specifying the unconditional model and final conditional model.

The analyses in presented in this chapter are carried out in order to address this research questions: What are the interrelationships between student-, teacher- and school-level factors in influencing students' mathematics performance related to LOT and HOT?. The research question is addressed by conducting two HLM analyses, the first model involves mathematics performance related to LOT as the outcome and the second model involves mathematics performance related to HOT as the outcome.

The outcome the first model examined is mathematics performance related to LOT and it is directly influenced by seven variables, three of which are at level-1, three of which are at level-2 and one that operates at level-3. At level-1, student beliefs concerning mathematics related to HOT and student liking mathematics positively influence student mathematics performance related to LOT. Conversely, student beliefs concerning mathematics related to LOT has a negative effect. At level-2 the teacher instructional approach and the teacher beliefs concerning mathematics related to HOT influence the outcome positively while teacher attendance of professional development training has a negative effect. At level-3 the variable of school resources has a positive influence on student mathematics performance related to LOT. There is no interaction between the explanatory variables and the slopes across the levels for the outcome of mathematics performance to LOT.

Seven significant variables directly influence mathematics performance related to HOT as the outcome. Four of which are at level-1, two is at level-2 and one is at level-3. The following variables at level-1: (a) students' educational expectation, (b) student mathematics performance related to LOT, and (c) individual judgement have positive effects on mathematics performance related to HOT, while (d) student beliefs concerning mathematics related to LOT has a negative effect. At level-2, teacher certification and teacher beliefs concerning mathematics related to HOT has a positive effect on student mathematics performance related to HOT. At level-3 the availability of 'Mathematics Olympiad' club positively influences the outcome. Interaction occurs between one level-3 variable, school resources, and student beliefs concerning mathematics related to LOT.

Overall, the three-level models of student mathematics performance involving mathematics performance related to LOT and HOT show that mathematics performance is influenced by several explanatory variables at the three-level. Both models of students' mathematics performance related to LOT and students' performance related to HOT are influenced by some common variables. Those are teachers' beliefs concerning mathematics related to HOT and school resources that influence both outcomes positively and students' beliefs concerning mathematics related to LOT that influences both outcomes negatively. Furthermore, the available variance and variance explained of both models are not just dominated by the student level (level-1) variables. Consequently, it can be argued that the variability in mathematics performance is influenced by variables at the students, teachers and also schools level of operations, in which the students are nested within teachers and the teachers are nested within schools.

Chapter 10 Discussion and Conclusion

10.1 Introduction

The importance of education and, specifically in this study, the importance of mathematical skills, ranks highly in any measure of a country's socio-economic development and wellbeing. In Indonesia, a rapidly developing nation, mathematics education is seen as an important and integral part of the curriculum and yet Indonesian students are ranked lowly internationally in both their LOT and HOT skills. This contrasts markedly with the national testing at the end of Year 9, where more than 95% of students are passing. There is obviously a disconnection here between international and national standards. The high national pass rate, say researchers such as Hendayana, Asep, et al. (2010), is because teachers teach directly to the test questions as well as advising on strategies to answer multiple choice questions, while failing to develop an understanding of mathematical principles in their students. Unfortunately, it is also known that there is an element of cheating involved in these examinations and there is a current plan to abolish the present assessment methods (Murtiana, 2011).

This over-emphasis on rote learning and LOT that researchers have found in Indonesian classrooms could be assumed to transfer to high results in all LOT skill testing. However, neither the outcome of the PISA and TIMMS testing, nor the outcome of this study reflect this. Indeed it has been noted that an overemphasis on LOT skills can have an overall negative effect on outcomes (Soar & Soar, 1976). Previous researchers have also noted the lack of emphasis on HOT in Indonesian classrooms, the lack of problem solving and the application of mathematical skills in real life situations (Cahyono, 2010; Hadi, 2002; Saragih & Napitupulu, 2015; Sembiring et al., 2008). From these studies and the results of the Aceh mathematics test, it would seem that mathematical education is failing the majority of Indonesian students.

The current study is the first in Aceh to investigate the possible factors involved in relation to the poor outcome in mathematics performance of students in a test *not* modelled on the national assessment. This quantitative study has been specifically designed to examine the factors involved in students' mathematics performance related to both LOT and HOT. Thus it seeks to provide an analysis reflecting the findings of educational effectiveness theory, involving a multilevel analysis of interrelated factors between the schools, teachers, and students. Educational effectiveness theory brings in a fourth level, that of the national socio-cultural context. While this research is informed by this, through background reading, it has not sought to include it in the research design.

In this chapter, the summary of findings and discussion are presented. There are sections discussing the implications of the study for theory, methodology and policy (curriculum), as well as the practice of mathematics education. From the theory, it is not appropriate to make assessments of student performance in relation to LOT and HOT without engaging with a hierarchical model of education, teaching and learning. Finally, the limitations of the study and some recommendations for future research are also presented.

10.2 Discussion of Findings

This study addresses the main research question: What are the student-, teacher- and school-level factors that influence student mathematics performance related to LOT and HOT?

It follows a multilevel analysis as indicated by the theoretical and conceptual framework. A number of factors at the student-, teacher- and school-levels that may have an influence on students' mathematics performance related to LOT and HOT are identified and are listed in the conceptual framework presented in Chapter 2. The appropriate instruments for the factors were then selected carefully based on the aim of the study and the research questions. The methods of analyses were also established.

The research is primarily concerned with discovering factors influencing students' ability in relation to LOT and HOT. This findings look firstly at the Acehnese students' ability to solve mathematical problems related to LOT and HOT (10.2.1). It then focuses on the students' mathematics performance related to LOT and HOT individually, looking at the three-level factors impacting on these results (10.2.2 and 10.2.3). Performing this hierarchical analysis provide information that may be used to assess the effectiveness of education in Aceh. The interrelationships of students-level factors (10.2.4) and the interrelationship of teacher-level factors (10.2.5) follow.

10.2.1. Students' Abilities in Solving Mathematics Problems related to LOT and HOT

The descriptive statistics presented in Chapter 6 provide information of students' mathematics performance related to LOT and HOT. Overall, the students in this study have low performance with solving mathematics problems related to LOT and an even lower performance for mathematics problems related to HOT. More than 70% of the students fall into the category of low achieving (73% for mathematics performance related to LOT and 77% for the performance related to HOT). Approximately 18% of students have an average mathematics performance related to LOT and 20% of students have an average mathematics performance related to HOT. The number of students classified in the high achieving group is very small for both mathematics performance related to LOT and performance related to HOT. Eight per cent of the students scored at the high achieving level for LOT and less than three per cent of the students scored at the high achieving level for HOT.

These results from a sample of 1135 students in Aceh reflect the results obtained by PISA testing carried out throughout Indonesia, where Indonesian students have a very low performance and have made no significant improvement since testing began in 2000 (Firman, 2016). Analysis of the PISA 2009, 2012 and 2015 results show that approximately 70% of the students are low performers and less than one per cent are in the top performers internationally (OECD, 2016a). PISA categorises low performers as students who achieve a proficiency of below level 1, where students are expected to be able to answer questions involving familiar contexts, where the questions are well-defined, where all relevant information is provided and it can solved using routine procedures. These criteria are aspects of LOT, meaning that nearly three-quarters of the Indonesian students do not have a grasp of even the basics of mathematics. PISA defines top performers as students who achieve a proficiency of level 5 and 6, where students are expected to be able to evaluate and decide on appropriate problem solving strategies, to utilise well-develop thinking and reasoning, and to develop new approach to solve non-routine problems with non-standard contexts (OECD, 2013b). These are all aspects of HOT skills, meaning that less than one per cent of Indonesian students are able to perform HOT related problems.

The test devised by the researcher involved mathematics problems related to LOT and HOT appropriate for Year 9 students. The test was not as long nor as wordy as the PISA and TIMMS tests. While researchers (Firman, 2016; Saragih & Napitupulu, 2015; Siswono, 2014; Susanti & Darhim, 2014) have been emphasising the poor results of Indonesian students' mathematics performance related to HOT, little acknowledgment has been made of the poor standard of LOT (W. Widjaja, 2011, being an exception). The results of the PISA and the Aceh test strongly confirm that Indonesian students are unable to solve complex, non-routine problem requiring HOT involving a higher level of reasoning and the integration of multiple concepts (Firman, 2016). Stacey (2014), in her article concerning mathematical literacy in Indonesia, has stated that the majority of the Indonesian students are not able to comprehend the questions of PISA. Indonesian students' reading literacy is also low, as reported in PISA results, with more than 50% of students being at or below level 1 in reading proficiency (OECD, 2013c, 2016a). Shahrill, Mahalle, Matzin, Hamid, and Mundia (2013), reporting on research in which they were involved, state that "the successful students in mathematics depended largely on language-based expressive learning styles and study strategies, such as speaking, writing and reading" (p.16). The relationship between language skills and mathematical skills needs to be further investigated, especially given that Indonesian students are often not receiving school education in their mother tongue. It is noted that in Singapore, one of the highest performer in mathematics in the international testing, mathematics classes are delivered in English (Kirkpatrick, 2011).

10.2.2. Student-, Teacher- and School-Level Factors Influencing Students' Mathematics Performance related to LOT

Acehnese students, along with Indonesian students generally, are performing poorly in questions related to LOT in the test administered in this study and international examinations. The student-level PLS-PA analyses indicates that students' mathematics performance related to LOT are influenced by five variables: (a) students' beliefs concerning mathematics related to LOT; (b) gender; (c) school location; (d) socio-economic status (SES); and (e) students' attitude of liking mathematics. The multilevel analyses, using HLM, indicate that there are seven variables that directly influence the students' mathematics performance related to LOT. Three are three variables at the student-level: (a) students' liking of mathematics; (b) students' beliefs concerning mathematics related to LOT; (c) students' beliefs concerning mathematics related to HOT. At the teacher-level, there are also three variables: (a) teachers' professional development; (b) instructional activities; (c) teachers' beliefs concerning mathematics teaching related to HOT. At the school-level, school resources is the only variable.

Both the student-level and multilevel analyses indicate that a student's liking of mathematics tends to improve their LOT performance. This result is in keeping with those reported by Areepattamannil (2014) and Mohamed and Waheed (2011) who also found an influence between students' attitudes towards mathematics and their improved performance.

Also both analyses indicate that the results also indicate that students who do not have strong positive beliefs concerning mathematics related to LOT tend to perform better in mathematics related to LOT. We can interpret this to mean that students who see mathematics as something that belongs to LOT (such as: 'mathematics is just about addition, subtraction, multiplication, and division' and 'mathematics is just a collection of rules and formulas') are actually weaker mathematics students, students who are unable to see beyond the basics. This is in line with (Garofalo, 1989a) who argued that when students hold beliefs that mathematics is about memorising formula, their approach in mathematics tends to be of a mechanical nature, which then confines them to remembering, without involving creativity, which in turn impacts on their overall poor performance. These findings are also supported by Schoenfeld (1989) and Schommer-Aikins et al. (2005) in relation to the importance of students' beliefs concerning mathematics and student performance where they report that students' beliefs concerning mathematics knowledge and learning predicted students' academic performance.

Based on the three-level HLM analysis, students who have stronger positive beliefs concerning mathematics related to HOT perform higher in mathematics related to LOT. Students who agree with the statements such as 'I usually try to create my own solution for a mathematics problem' (the list of all items in the scale of students' beliefs concerning mathematics related to HOT can be found in Chapter 4), have a higher performance related to LOT. Positive beliefs concerning HOT are likely to be related to greater competency in mathematics generally and it is not surprising that students who would hold such beliefs would also perform better in LOT.

Based on the student-level model, it is concluded that gender, school location and SES have direct effects on student mathematics performance related to LOT. Female students have a higher mathematics performance related to LOT than do male students. The relationships between gender and mathematics performance have varied in previous studies. Some studies have reported that male students outperformed female students and some others reported that females outperformed males. The results in this study are consistent with the study conducted in the US reported by Gallagher and De

Lisi (1994) and the study of Malaysian students reported by Abdullah et al. (2016), in which females were reported to perform better than males in solving mathematics problems. Also, in a longitudinal study conducted by Suryadarma (2015) it was reported that 11 years old Indonesian girls outperform 11 years boys, and the gap is even wider seven years later. However, it contradicts the finding reported by Darmawan (2016), that Indonesian male students performed better than female students in mathematics performance in general. It is also contradicts the results of PISA 2012, which shows that on average females do not perform as well as males (OECD, 2013b).

Students who are enrolled in the urban schools perform better in LOT than do students enrolled in the rural schools. Higher SES is associated with urban schools. These findings are consistent with Mohammadpour and Ghafar (2014), Oh (2013), Chen (2016) and Thien and Darmawan (2016) who reported that students' SES positively influenced mathematics performance in general. In this study parents' education is a more significant indicator of SES than parents' occupation (as reported in the assessment of the formative construct in Chapter 7) which confirms the findings reported by Farooq et al. (2011) that parents' occupations were less important than parents' education in relation to mathematics performance. This suggest that parental attitude towards education, influenced by their own attainment transfers to their children ability in mathematics.

The three-level HLM analysis indicates that that teachers who have strong positive beliefs concerning mathematics teaching related to HOT and use particular instructional activities frequently have a positive influence on their students' mathematics performance related to LOT. In previous studies teachers' beliefs were found to influence their classroom practices which in turn influenced their students' outcomes (Ertmer, 2005; Stipek et al., 2001) and Polly et al. (2013) reported a direct influence between teachers' beliefs and student outcomes.

The relationship between teachers' professional development and students' mathematics performance related to LOT is negative, indicating that students who are taught by teachers who have attended more professional development programmes during the last two years tend to have lower mathematics performance related to LOT. This finding requires some understanding. It contradicts that reported by Yoon, Duncan, Lee, Scarloss, and Shapley (2007) in which substantial professional development for teachers increases students' mathematics performance. The negative finding related to teachers' development programmes in this study may be related to the state of the education system and teaching profession in Indonesia, especially in outlying provinces like Aceh and rural areas. The descriptive analysis in Chapter 6 reports that most teachers said they attended development programs related to mathematics content and instruction. This is a consequence of not all teachers being professionally trained, especially those in rural areas, and the ongoing changes being made by the Ministry of Education and Culture to the mathematics curriculum that means that teachers attend seminars that simply report curriculum changes. Teachers only just become familiar with one change when another is brought in. Fewer teachers attend development programmes that focus on improving students' critical thinking and problem solving, on addressing individual student needs and teaching in a way that develops understanding and creativity. The issues of teacher training and teacher effectiveness will be discussed further in policy implications in 10.5.3.

The three-level HLM analysis also indicates that school resources directly influence students' mathematics performance related to LOT. Students who are studying in the schools with more resources perform better in mathematics related to LOT than do students enrolled in the less equipped schools. This is consistent with the results of studies by Lee and Zuze (2011) and Murillo and Román (2011) who reported that school resources influenced the students' mathematics performance in general. The descriptive analysis in this study indicates that it is rural school generally that have fewer resources and lower mathematics performance, this is an indication of the challenge face my developing countries like Indonesia that need to firstly provide adequate infrastructure to rural region in order to facilitate equality of education.

10.2.3. Student-, Teacher- and School-level Factors Influencing Students' Mathematics Performance related to HOT

The findings of the PLS-PA and HLM analyses indicate that students' mathematics performance related to HOT are directly influenced by several variables at the student, teacher- and school-level. Some of the variables are shared with the factors influencing students' performance related to LOT mentioned in the previous section.

The result of student-level model of PLS-PA records that four variables directly influence student mathematics performance related to HOT, namely: (a) student mathematics performance related to LOT, (b) students' educational expectations, (c) SES and (d) school location. The three-level model of HLM analysis indicates seven variables directly influence student mathematics performance related to HOT. There four variables at the student-level: (a) student mathematics performance related to LOT; (b) student educational expectation; (c) students' individual judgement of mathematics ability; (d) student beliefs concerning mathematics related to HOT. At the teacher-level there are two variables: teacher certification and teachers' beliefs concerning mathematics teaching related to HOT. There is one variable at the school-level: the availability of a 'Mathematics Olympiad' club.

The results of both the student-level model and the three-level model of HLM indicate that students' mathematics performance related to LOT has a strong influence on students' mathematics performance related to HOT. Students who score higher for mathematics problems related to LOT are more likely to obtain higher scores for problems related to HOT. In mathematics, for a student to be able to solve HOT problems it is necessary for them to have the basic skills; those with an ability to perform HOT will do well at LOT tasks.

Both the student-level model and the three-level model of HLM also show that a student's expectation of their highest educational qualification also has a direct effect on mathematics performance related to HOT, with students aiming higher performing better on HOT problems. This is in line with the finding reported by Hammouri (2004). It should be noted that a very high number (58%) of the Year 9 students in this study indicated that they intend to complete PhDs. This figure is well be beyond the realm of the general Indonesian population, where approximately only 11.48 of 270 thousands of academics in Indonesia complete doctoral programmes (Satrio, 2017), and is likely to be the result of the impact of the researcher telling the students that she was a PhD candidate.

The results of the student-level model also indicates that students who are from a higher SES background have a higher mathematics performance related to HOT. This is consistent the study of the US students reported by Namok Choi and Mido Chang (2011). In addition, school location also influences students mathematics performance related to HOT, with urban students doing better. HOT skills then are associated with urban living and higher SES. This important finding provides an explanation of Indonesian students' poor performance in PISA and TIMMS and Aceh's students in this research. It contradicts an early assessment of the role of SES in education that was subsequently disproved by many researchers (Chen, 2016; Lindberg et al., 2010; Petty et al., 2013; Steinmayr & Spinath, 2008; Tsui, 2007).

The important contribution of this finding is to emphasise that there are factors beyond the school-, teacher- and student-levels that impact on student mathematic performance. Developing nations, such as Indonesia, are less likely to improve their performance in HOT-related testing while a large proportion of the population lives in rural areas. Equality of educational opportunity is, as previously stated, a challenge to developing nation. However, it should also be noted that Vietnam, a developing country, is able to perform very well in HOT-related testing (OECD, 2016a).

The results of the three-level HLM model indicate that students' beliefs related to LOT are inversely related to students' performance in HOT. Students who see mathematics education to be concerned primarily with LOT skills are likely to have poor HOT skills. As was said in the previous section, mathematics performance is influenced by students' beliefs about mathematics. This finding supports the studies by Schoenfeld (1989) and Schommer-Aikins et al. (2005) who said that that students' beliefs concerning mathematics knowledge and learning predicted students' academic performance. The HLM model also indicate that students' individual judgement of mathematics ability is positively associated with their mathematic performance related to HOT. Students who judge higher of their mathematics ability are having higher mathematics performance related to HOT.

The findings of the HLM analysis also indicate that teacher certification has a positive influence on students' mathematics performance related to HOT. Moss (2012) also reported that more qualified teachers have students who are able to perform better. Again, given that the number of teachers with certification is lower in rural areas, students' mathematics performance related to HOT is compromised. A further finding of the HLM analysis shows that students whose teachers have stronger positive beliefs concerning mathematics related to HOT have a better mathematics performance related to HOT. This is consistent with the findings of the study reported by Tschannen-Moran and Barr (2004) regarding the positive relationship between teachers' beliefs and students' mathematics performance. However, it should be noted that the teacher questionnaire was self-reporting in relation to a teacher's beliefs towards HOT. While the teachers generally had a positive belief towards HOT their understanding of it and their ability to transfer this understanding was not tested. It could be assumed that most teachers who had even a slight knowledge of the Bloom taxonomy, would agree that HOT skills were important to mathematics education. Recent research in the US (T. Thompson, 2008) discovered that teachers were often unable to accurately define LOT and HOT and consistently underestimated the nature of a HOT test question when asked to design one. This suggests that a teacher giving a positive response to a statement concerning the value of HOT does not necessarily translate to a teacher's of understanding of HOT or their capacity to facilitate their students' performance in mathematics related to HOT. A further investigation of relationships between teachers' beliefs concerning mathematics and their classroom practices transferring to better students' outcome is needed.

The result of the HLM analysis, apart from also indicating the importance of location, highlights a further school-level variable. The presence of a 'Mathematics Olympiad' club, significantly influences students' mathematics performance related to HOT. This indicates that students are more likely to be able perform better in HOT mathematics problems in the schools where clubs are available. These clubs are organised by the school as an additional lesson for students interested in competing in the Olympiad, a prestigious, national competition. This club attracts high achieving students and provides them with the opportunity to develop their interest in mathematics beyond the classroom. From the descriptive analysis, the schools that organised this club are predominantly in urban areas.

The results presented in section relating to factors influencing students' mathematics performance related to LOT and HOT can be summarised to provide a comparison of factors. Some factors are influence both LOT and HOT performance as well as being present in both single-level and multilevel models.

From the results of the single-level analysis, mathematics performance related to LOT is higher in urban areas with a higher SES. It is also higher when students like mathematics and feel strongly about beliefs related to LOT. Female students outperformed males. The multilevel analysis further reveals that performances related to LOT are higher when students have a stronger beliefs concerning mathematics related to LOT and HOT. This analysis was also able to capture a link between students' performance and teacher-level factors concerning beliefs of mathematics teaching related to HOT. The unexpected negative influence of teachers' higher involvement in professional development programmes results in poorer performance related to LOT (as has already been explained). The school-level brings in the variable of schools resources, where being in a better equipped urban school predicts a higher LOT performance.

Looking at the mathematics performance related to HOT, a higher students' mathematics performance related to HOT is obviously influenced by performance related to LOT in both single-level and multilevel models. At the single-level model, the significant variables (as in LOT) are higher educational expectation, higher SES and an urban location. The additional factor to influence performance related to HOT is students' high educational expectation. This indicating that students who are better at mathematics performance related to HOT also have high educational expectation. At the multilevel analysis, performance related to HOT is negatively associated with student beliefs related to LOT (as previously explained). A higher educational expectation is also showing here. The additional factor discovered to be important is students' high judgement of their mathematical ability. At the teacher-level, performance related to HOT is related to teachers having certification and stronger beliefs of mathematics related to HOT. The presence of the 'Mathematics Olympiad club' predominantly in urban schools, is the additional school factor.

10.2.4. Interrelationships between the Student-level Factors

The student-level model of PLS-PA shows that there are multiple interrelationships within the student-level factors that may influence their relationship with mathematics performance related to LOT and HOT.

Students' individual judgement of their mathematical abilities is influenced by two variables, namely: (a) their expectation of their highest educational level; and (b) their beliefs concerning mathematics related to LOT. Students who expect to achieve higher in their educational level have more positive individual judgements of their mathematics ability. Here at the student-level, there is a further link with putting a high beliefs concerning mathematics related to LOT: in this case it is linked with less positive individual judgements of their mathematics abilities. This can be somewhat explained by the low standard of students' general mathematical ability; a student who judges themselves as having low mathematical abilities will generally believe they need to focus on LOT skills.

Students' mathematics confidence is positively influenced by students' beliefs concerning mathematics related to LOT; students' beliefs concerning mathematics related to HOT; and their individual judgement of their mathematics ability. In relation to their confidence, students' individual judgements have the strongest influence; when students judge their capacity higher, they are more likely to be confident about learning mathematics. This result is expected as the validation of the scale in Chapter 5 reports that both students' individual judgements and mathematics confidence are subscales of students' self-efficacy and a moderate correlation is found between the subscales. Furthermore, students who have stronger positive beliefs concerning mathematics related LOT and HOT also tend to be more confident in learning mathematics. This is consistent with Kloosterman (2002) who reported the secondary school students' beliefs of mathematics and mathematics learning have an impact on students' interest in mathematics. In addition, students' attitude towards liking mathematics is influenced by their mathematical confidence and the value they place in mathematics; strong confidence and placing a high value on mathematics inspire students to like mathematics. Further, with regard to students' beliefs concerning mathematics related to LOT and HOT, the results of the student-level model indicate that learning resources used in the mathematics classroom have a positive effect on students' beliefs concerning mathematics related to LOT. Students learning activities related to both LOT and HOT positively influence students' beliefs concerning mathematics related to HOT.

In relation to, SES and student educational expectation, the student-level model results also report that SES is strongly related to school location, with students who are enrolled in the urban schools are those with parents of a higher SES. Student educational expectation is influenced by gender and SES, with female students and students from higher SES have a higher expectation on their future educational level.

In relation to students' perceptions of the types of mathematics questions used in their classroom and their learning activities, the student-level model indicates that learning resources positively influence the frequency of the type of mathematics questions used in the classroom. It also shows that the learning resources and the frequency of the types of questions used in the classroom have a positive influence with the learning activities related to LOT, while the frequency of the types of mathematics questions used in the classroom and the learning activities related to LOT positively influence students learning activities related to HOT.

10.2.5. The Interrelationships between the Teacher-level Factors

The teacher-level path analysis is conducted based on the sample of 46 teachers. Despite this small sample size, SmartPLS 3.2.6 (Ringle et al., 2015) enables the analysis of the model and a few relationships can be determined. The teacher-level model is conducted to identify the interrelationships among the teacher-level factors including classroom practices and teacher background.

The results show that the teacher's instructional activities conducted in the classrooms are positively influenced by the frequency of the types of questions used in mathematics examinations; the teachers' methods of engaging the students; and whether or not they are certified. Generally, the tendency of teaching for examinations in Indonesian mathematics classrooms is well documented (Hendayana, Supriatna, & Imansyah, 2010), where understanding and creativity are passed over in order for the school to achieve the maximum percentage of students passing the examination. There is also a link between the frequency of instructional activities and the frequency of teachers' effort in engaging with students, indicating that the teacher's frequency of interaction in one area reinforces the other. Finally, a teacher with certification is more likely to involve their students in the instructional activities.

The results of the teacher-level model also indicate that both teachers' effort in engaging with students and the types of questions teachers used in mathematics classrooms are positively influenced by the teaching resources used and their beliefs concerning mathematics teaching related to LOT. In regard to the influence of beliefs on teachers' engagement with students, this is in line with Nisbet and Warren (2000) who reported that teachers' beliefs concerning mathematics influence their approach to teaching. In term of teaching resources that are associated with the types of questions used, in this study the most commonly used teaching resources are textbooks, with more than 70% of teachers using them as the basic means of instruction; the most common type of mathematics problem given both in the classrooms and examinations are word problems, and the least common type of questions given to students is unfamiliar problem (non-routine problem). The known importance of the link between teachers' activities and beliefs and student learning and performance requires some further comments and contextualisation.

10.3 Implications of the Study

The focus of this study has been an analysis of the factors impacting on Acehnese students' mathematic performance in relation to LOT and HOT. Aceh is a less-developed province within a developing nation. Education, particularly mathematics education, despite the government's best intentions, needs great improvement for Indonesia to meet the demands of a technologically sophisticated, globalised world. Why are Indonesian students performing so poorly in international assessments? The answer is not simple and involves an assessment of all levels of society and education. The blame cannot be placed on resources, as education funding throughout Indonesia is adequate. Nor can schools, or teachers or indeed the students shoulder the blame. As educational effectiveness theory has shown, the answer is in a complex dynamic that is forever shifting (Creemers & Kyriakides, 2008). The findings of this study provide a basis for an understanding of students' mathematics performance in both LOT and HOT as an outcome of their society, their schools, their teachers and their own background and capacities. At the heart of this is mathematics teaching and learning, the future of Indonesia. The implications of the study include theoretical, methodological and practical significance for the improvement of mathematics teaching and learning in Aceh, Indonesia.

10.3.1. Theoretical Implication

Many researchers have examined the numerous and multilevel factors that might contribute to students' mathematics performance, with the factors including student attitudes and beliefs, student background, classroom practices, school background and school climate and resources among others. These factors have been found to have significant relationships to student mathematics performance. While the findings in this study themselves do not highlight a great difference between factors relating to LOT and HOT, the decision to test for this remains significant as it throws light on the general state of mathematics education in Indonesia and the struggle students are having in reaching satisfactory mathematical outcomes for both LOT and HOT.

There are few studies related to these topics conducted in the Asian context, in particular in the Indonesian context. Given the student outcomes and Indonesia's on-going economic development, there is an obvious need to develop multilevel analyses of the situation. The framework of the present study, as presented in Chapter 2, in seeing the effectiveness of education as stemming from many variables and all the levels involved in an education system and impacting on a student, builds on the research established in the field. The results of the analysis of the data sets confirm some of the hypothesised relationships in the conceptual framework.

Analyses of Indonesian mathematics teaching have tended to focus on its lack of emphasis on HOT and its over-emphasis on LOT. Yet PISA, TIMMS and Aceh data all point to a lack of both LOT and HOT skills. This study provides a theoretical contribution by examining the factors independently related to LOT and HOT in the context of Aceh, Indonesia. The fact that this study is conducted in a non-western or developing country context enriches the existing research that has mainly focused on a Western, or developed country context.

10.3.2. Methodological Implication

The review of research provides the theoretical basis and conceptual framework (Chapter 2) for the quantitative methods that are employed (Chapter 3) in order to obtain answers for the research questions (Chapter 1). The analyses of the data collected initially involve single-level analyses using partial least squares path analysis (PLS-PA), with formative variables included in the model. The subsequent HLM analyses allow for a more accurate interpretation of the nested levels.

As is indicated above, the inclusion of some formative variables results in the use of path analysis with partial least squares (PLS) instead of a maximum likelihood (ML) approach. The partial least squares path analysis (PLS-PA) is employed to examine procedure relationships between variables involving the student- and teacher-level. The PLS-PA employs a variance-based structural equation modelling (SEM) approach for estimating the path coefficients by maximizing the variance explained by the dependent variables. The PLS-PA has the flexibility of being broadly applicable to work with a diversity of data (data could be nominal, ordinal or categorical), it has the ability to work with both large and small sample sizes and can include both formative and reflective latent variables. The feature of formative latent variables being included in the PLS-PA is one of the reason why the technique is chosen for single-level analysis at the student- and teacher-level. More specifically, the small sample size is an advantage of PLS-PA since it is beneficial in the analysis of the teacher data due to the low number of cases involved at the teacher level. In addition, the lack of studies focusing on factors influencing student mathematics performance that are related to LOT and HOT and involve students' and teachers' beliefs results in this examination being classified as an exploratory research. This also suits the background of the partial least squares approach that aims for exploratory research. In this study, SmartPLS 3.2.6 (Ringle et al., 2015) is used to carry out the partial least squares path analysis.

Due to the hierarchical nature of the data collected, a single-level analysis is not sufficient as the disaggregating and aggregating process can generate bias, resulting in a mis-estimation of the precision of the effect and can also lead to the neglect of important and meaningful variance at other levels. Therefore, a multilevel analysis technique is employed with models that go beyond the single-level analysis. The multilevel analysis technique employed is hierarchical linear modelling (HLM) using HLM 6. The technique overcomes or reduces the limitations of a single-level analysis, taking into account all possible variance at each level as well as enabling interactions between variables and cross level effects.

In addition, acknowledging the limitations of the raw score of classical test theory (CTT), this study also employs Rasch analysis and item response theory (IRT). The

Rasch model is employed in this study for validating the scales and transforming the raw scores of both the mathematics test and the attitudinal questionnaire items using interval scale measures. The Rasch model is used for measurement in the scales and a complimentary techniques to establish the validity and reliability of the scales. The Rasch model involves item response theory (IRT) that provides more rigorous measurement by taking into account both item and person data. The Rasch technique is also used to transform the raw scores using a weighted likelihood estimation (WLE). The WLE scores for the mathematics test and attitudinal questionnaires items are obtained by employing Conquest 2.0 software (Wu et al., 2007).

10.3.3. Policy and Practice Implication

In a practical sense, this study potentially benefits the students, the teachers, the schools and the policy makers. The fact that this study reports that Acehnese students have a low performance for both mathematics problems related to LOT and HOT indicates that the mathematics performance of Acehnese students needs attention. The findings of this study may also begin to map the strengths and weaknesses of current mathematics classroom practices in Aceh. Despite most students reporting that they have positive attitudes of liking mathematics, their individual judgement and mathematics confidence are less positive. Furthermore, textbooks are still used as the basis of instruction in most schools and the use of technology in mathematics classrooms is still very limited.

Perhaps the most important finding of this study is that Indonesian students in Year 9 have a very low performance in mathematical testing that includes both LOT and HOT. This means that there is a problem somewhere within the education system. In the past the blame has been laid at the feet of the education ministry, the curriculum and a lack of resources. Today, education in Indonesia is not suffering through a lack of financial resources. Both the central government and the provincial governments are generally providing adequate finances. In some provinces a lack of infrastructure, including roads, transport and internet availability, continues to be a problem, particularly in more remote provinces like Aceh. The curriculum has also been blamed. However, the national curriculum is explicit in its incorporation of both LOT and HOT.

After 15 years of carrying out research on education in Indonesia, Bjork (2013) concluded that there is a problem within the system. He acknowledged the enormous challenges that the nation faced after Independence and its setting up, very quickly, an enormous network of schools, one of the largest in the world. He has reviewed the curriculum and government policies and found them in line with current international benchmark in education. Rather he sees the problem in 'the culture of teaching' where teachers are seen and see themselves primarily as civil servants rather than as innovators of education. Consequently, Ministry of Education reforms do not result in change in the classrooms and increases in student performance. Following research carried by Hattie (2003), which synthesised over 500,000 studies, he concluded that the most direct way of improving student outcome is through improving teacher quality. But in the case of Indonesia this will require a change in the status of teachers:

> Developing an infrastructure that treats teachers as professionals and gives them the support necessary to act autonomously is an essential antecedent to fundamental reform. Once this foundation has been laid, the outcomes of policy should more closely match the ministry's predictions of change (Bjork, 2013, p. 66).

The foundation for this change lies in teachers recruitment, training and in-service professional development. It is hoped that this study may contribute to Indonesian students receiving a mathematics education that will serve them well in their future.

10.4 Limitations of the Study and Recommendations for Future Research

This study has resulted in some preliminary and significant findings concerning mathematics performance of Acehnese students; however, its design is not without some shortcomings. The timeframe, cost and human resources are some of the reasons for the limitation of the study. The short period available for data collection has limited the area covered and the instruments used in the study. First, the data collection was conducted in 25 schools from two regions in Aceh, an urban and a rural area. It would have been preferable to choose more than one city and district for the study. However, considering the timeframe, cost efficiency and the lack of human resources to carry out the study, the choice of the regions included is optimum. Second, the classroom practice data is obtained through questionnaires, where teachers self-reported on their teaching methodologies and beliefs in relation to mathematics. It would have been preferable to have some classroom observation to gain more detailed information and provide a reference for self-reported data. Third, the number of mathematics questions included in the test is limited to eight questions which covered topics covered the topic from Year 7 and Year 8. This was due to the fact that the data collection was conducted at the beginning of Year 9. A greater number of mathematics questions may also have be preferable to be able to cover more subtopics of mathematics content. Fourth, prior performance is not included in the study. Ideally, prior performance is a good predictor of the current performance. However, due to the limited resources available for this study, only a cross sectional study can be conducted, not a longitudinal study. Standardised data on prior performance are also unavailable.

The limitations concerning the time frame, cost and human resources resulted in a compromise for the generalisation of the findings. While the findings cannot be generalised to reflect the Aceh province or the Indonesian population, it should be noted that the student outcomes in the testing reflect the data gathered through PISA and TIMMS. The findings and conclusion from the study can be used for the basis of future studies whether on a smaller or larger scale and within the same or other contexts.

10.5 Concluding remarks

Educators have a responsibility to ensure the best possible outcomes for their students. There is much to be gained from broad scale international comparative studies in establishing contributing factors to student outcomes. The contributing factors require understanding and reform. The findings of this study show that mathematics teaching and learning in Aceh is in a similar state to that throughout Indonesia. Indonesian mathematics students are underperforming. Teachers are the bridge between the national system and local school policies and the curriculum; it is they who work with their students to achieve their potential. While internationally educators are trying to find the most effective methods for teaching HOT, Indonesia must face the reality that its system of mathematics teaching must be rigorously reviewed.

Appendices

Appendix A Ethics Approval from the University of Adelaide

THE UNIVERSITY of ADELAIDE

RESEARCH BRANCH
OFFICE OF RESEARCH ETHICS, COMPLIANCE AND INTEGRITY

BEVERLEY DOBBS
EXECUTIVE OFFICER
LOW RISK HUMAN RESEARCH ETHICS REVIEW GROUP (FACULTY OF HUMANITIES AND SOCIAL SCIENCES AND FACULTY OF THE PROFESSIONS)
THE UNIVERSITY OF ADELAIDE
SA 5005
AUSTRALIA
TELEPHONE +61 8 8313 4725
FACSIMILE +61 8 8313 7325
email: beverley.dobbs@adelaide.edu.au

21 May 2013

Dr I Darmawan
School of Education

Dear Dr Darmawan

ETHICS APPROVAL No: HP-2013-034
PROJECT TITLE: Higher order thinking skills in mathematics classroom: the nexus between teacher's beliefs, classroom practices and student's achievement

I write to advise that the Low Risk Human Research Ethics Review Group (Faculty of Health Sciences) has approved the above project. The ethics expiry date for this project is **31 May 2016**.

Ethics approval is granted for three years subject to satisfactory annual progress and completion reporting. The form titled *Project Status Report* is to be used when reporting annual progress and project completion and can be downloaded at http://www.adelaide.edu.au/ethics/human/guidelines/reporting. On expiry, ethics approval may be extended for a further period.

Participants in the study are to be given a copy of the Information Sheet and the signed Consent Form to retain. It is also a condition of approval that you **immediately report** anything which might warrant review of ethical approval including:

- serious or unexpected adverse effects on participants,
- previously unforseen events which might affect continued ethical acceptability of the project,
- proposed changes to the protocol; and
- the project is discontinued before the expected date of completion.

Please refer to the following ethics approval document for any additional conditions that may apply to this project.

Yours sincerely

ASSOCIATE PROFESSOR PAUL BABIE
Convenor
Low Risk Human Research Ethics Review Group (Faculty of Humanities and Social Sciences and Faculty of the Professions)

THE UNIVERSITY of ADELAIDE

RESEARCH BRANCH
OFFICE OF RESEARCH ETHICS, COMPLIANCE AND INTEGRITY

BEVERLEY DOBBS
EXECUTIVE OFFICER
LOW RISK HUMAN RESEARCH ETHICS REVIEW GROUP (FACULTY OF HUMANITIES AND SOCIAL SCIENCES AND FACULTY OF THE PROFESSIONS)
THE UNIVERSITY OF ADELAIDE
SA 5005
AUSTRALIA
TELEPHONE +61 8 8313 4725
FACSIMILE +61 8 8313 7325
email: beverley.dobbs@adelaide.edu.au

Applicant:	Dr I Darmawan
School:	Education
Application/RM No:	16416
Project Title:	**Higher order thinking skills in mathematics classroom: the nexus between teacher's beliefs, classroom practices and student's achievement**

Low Risk Human Research Ethics Review Group (Faculty of Health Sciences)

ETHICS APPROVAL No: **HP-2013-034**

APPROVED for the period: **14 May 2013 to 31 May 2016**

This study is to be conducted by Ms Elizar, PhD Candidate.

ASSOCIATE PROFESSOR PAUL BABIE
Convenor
Low Risk Human Research Ethics Review Group (Faculty of Humanities and Social Sciences and Faculty of the Professions)

Appendix B Ethics Approval from the Indonesian government

Approval from the Ministry of Education, Banda Aceh

PEMERINTAH KOTA BANDA ACEH
DINAS PENDIDIKAN, PEMUDA DAN OLAHRAGA
JALAN. P. NYAK MAKAM No. 23 GP. KOTA BARU TELP/FAX. (0651) 7555136,7555137
E-mail:disdikporabna@gmail.com Website: www.disdikporabna.com
Kode Pos: 23125

IZIN PENELITIAN
NOMOR : 074/A.2/ 5929 /2013

TENTANG
PENGUMPULAN DATA THESIS

Dasar : Surat School of Education, the University Of Adelaide, Adelaide, South Australia Nomor : HP-2013-034, Tanggal 3 Juli 2013 Hal Izin Mengumpulkan Data Thesis.

MEMBERI IZIN :

Kepada :

Nama : ELIZAR
Nomor Mahasiswa : 1195731
Prodi : S3 Pendidikan Matematika

Untuk : Mengumpulkan data dalam rangka penyusunan Thesis, dengan judul " **KETERAMPILAN BERPIKIR TINGKAT TINGGI DALAM PEMBELAJARAN MATEMATIKA : HUBUNGAN ANTARA PENDAPAT GURU, PRAKTEK KELAS DAN PRESTASI SISWA".**

Adapun sekolah-sekolah pengambilan data penelitian antara lain :
1. SMP N [redacted]
2. SMP N [redacted]
3. SMP N [redacted]
4. SMP N [redacted]
5. SMP [redacted]
6. SMP [redacted]
7. SMP [redacted]

Dengan ketentuan sebagai berikut:
1. Harus berkonsultasi langsung dengan kepala sekolah yang bersangkutan dan sepanjang tidak mengganggu proses belajar mengajar.
2. Bagi Mahasiswa yang bersangkutan supaya menyampaikan foto copy hasil Penelitian sebanyak 1 (satu) eksemplar ke Dinas Pendidikan Pemuda dan Olahraga Kota Banda Aceh.
3. Surat ini berlaku sejak tanggal 19 Agustus 2013 s/d 19 September 2013

Demikianlah surat ini disampaikan untuk dapat dipergunakan semestinya.

Ditetapkan di Banda Aceh
Pada tanggal 19 Agustus 2013

An. KEPALA DINAS PENDIDIKAN PEMUDA DAN OLAHRAGA KOTA BANDA ACEH,
Kasi Pengembangan Kurikulum
Bidang Pendidikan Dasar dan Lanjutan

MARWAN, S. Ag. M. Pd
Pembina
NIP. 19730410 1999903 1 009

TEMBUSAN :
1. Koordinator Pasca Sarjana School of Education The University of Adelaide
2. Mahasiswa/i yang bersangkutan
3. Pertinggal

wonderful indonesia
Visit Aceh

PEMERINTAH KOTA BANDA ACEH
DINAS PENDIDIKAN, PEMUDA DAN OLAHRAGA
JALAN P. MYAK MAKAM No. 23 GP. KOTA BARU
TELP/FAX. (0651) 7555136, 7555137
E-mail: disdikporabna@gmail.com Website: www.disdikporabna.com

RESEARCH PERMIT NUMBER 074/A.2/ 5929 /2013
WITH REGARDS TO DATA COLLECTION FOR THESIS

Basic : A letter of School of Education, the University Of Adelaide, and Adelaide, South Australia Number: HP-2013-034, date 3 July 2013, Permit of Data Collection for Thesis.

The permit is granted to:

Name	: Elizar Ali
Student identification number	: 11957311
Program Study	: Doctor of Mathematic Education

For purpose of : Gathering data in order to complete the requirement for thesis writing with the title **'Higher Order Thinking (Hot) Skills: Investigating The Nexus Between Teachers' Beliefs, Classroom Practices, Student Beliefs And Mathematics Performance In Aceh, Indonesia'**
The schools that being the sources of data collection are:

1. SMP N [redacted]
2. SMP N [redacted]
3. SMP N [redacted]
4. SMP N [redacted]
5. SMP [redacted]
6. SMP [redacted]
7. SMP [redacted]

With the following conditions

1. It is required to consult directly to the School's Principals and it should not interfere the learning process.
2. The person is required to submit a copy of the final results of the research of 1 (one) copy to the Department of Education, Youth and Sports Banda Aceh.
3. The letter is valid from 19 August 2013 until 19 September 2013

This statement has been issued for use as intended.

Issued in Banda Aceh
Dated 19 August 2013
On behalf of Head of Education, Youth and Sport, City of Banda Aceh
Head of Curriculum Development in Elementary and Secondary Education

MARWAN, S.Ag, M,Pd
Supervisor
NIP. 19730410 1999903 1 009

Cc:

1. *Coordinator of postgraduate School of Education, the University of Adelaide, Adelaide, South Australia*
2. *The student*
3. *Archives*

Approval from the Ministry of Religious Affairs, Banda Aceh

KEMENTERIAN AGAMA
KANTOR KOTA BANDA ACEH
Jln. Mohd. Jam No.29 Telp. 27959 – 22907 Fax. 22907
BANDA ACEH (Kode Pos 23242)

Nomor : Kd.01.07/2/TL.00./ 606 /2013 — Banda Aceh, 10 Juli 2013
Lampiran : -
Perihal : **Rekomendasi Melakukan Penelitian**

Kepada
Yth, 1. MTsN [redacted]
2. MTs[redacted]
3. MTsN [redacted]
4. MTsS [redacted]
5. MTs[redacted]

Assalāmu'alaikum Wr. Wb.

Sehubungan dengan surat Koordinator Pascasarjana School Of Education, The University Of Adelaide, South Australia Nomor : HP-2013-034 tanggal Juli 2013, perihal sebagaimana tersebut dipokok surat, maka dengan ini kami mohon bantuan Saudara untuk dapat memberikan data maupun informasi lainnya yang dibutuhkan dalam rangka memenuhi persyaratan bahan penulisan ***Thesis,*** dengan judul **"Keterampilan Berpikir Tingkat Tinggi Dalam Pembelajaran Matematika: Hubungan Antara Pendapat Guru, Praktek Kelas Dan Prestasi Siswa "** kepada saudara :

Nama	:	**Elizar**
NIM	:	1195731
Prodi	:	Pendidikan Matematika

Dengan ketentuan sebagai berikut :

1. Harus berkosultasi langsung dengan Kepala Madrasah yang bersangkutan dan sepanjang tidak mengganggu proses belajar mengajar.
2. Tidak memberatkan Madrasah.
3. Tidak menimbulkan keresahan-keresahan lainnya di Madrasah.
4. Bagi yang bersangkutan supaya menyampaikan foto copy hasil penelitian sebanyak 1 (satu) eksemplar ke Kantor Kementerian Agama Kota Banda Aceh.

Demikian rekomendasi ini kami keluarkan, atas perhatian dan kerja sama yang baik kami ucapkan terima kasih.

An. Kepala,
Kepala Seksi Pend. Madrasah,

Drs. Aiyub, MA
NIP. 119680414 199905 1 001

Tembusan :

1. Kepala Kantor Wilayah Kementerian Agama Provinsi Aceh.
2. Direktur Koordinator Pascasarjana School Of Education, The University Of Adelaide, South Australia
3. Mahasiswa yang bersangkutan.

KEMENTERIAN AGAMA
KANTOR KOTA BANDA ACEH
Jln. Mohd. Jam No. 29 Telp. 27959 – 22907 Fax. 22907
BANDA ACEH (Kode Pos 23242)

Banda Aceh, 10 July 2013

Number : Kd.0 I.07/2/TL.00./606/2013
Appendix : -
Subject : Recommendation to Conduct Research

To:
1. MTsN [redacted]
2. MTsN [redacted]
3. MTsN [redacted]
4. MTsS [redacted]
5. MTsS [redacted]

Assalamu 'alaikum Wr. Wb.

Pursuant to a letter from Coordinator of Postgraduate School of Education, the University of Adelaide, Adelaide, South Australia, number HP-2013-034 on July 2013. As mentioned in the letter, it is hereby requested to you to give data and information needed in order to complete the requirement for writing thesis materials with the title **'Higher order thinking (HOT) skills: Investigating the nexus between teachers' beliefs, classroom practices, student beliefs and mathematics performance in aceh, indonesia'** to:

Name : Elizar
Student identification number : 11957311
Program Study : Doctor of Mathematic Education

With the following conditions:

1. It is required to consult directly to the School's Principals and as far as it has no interference with the learning process.
2. It will not disturb the schools.
3. It will not cause any unrest in the schools.
4. The person is required to submit a copy of the final results of the research of 1 (one) copy to the Ministry of Religion Banda Aceh.

This recommendation has been issued, thank you for your attention and your cooperation.

On behalf of theHead of Madrasah Education

Drs, Aiyub, MA
NIP. 119680414 199905 1 001

Cc:

1. *Head Office of the Ministry of Religion of Aceh.*
2. *Coordinator of postgraduate School of Education, the University of Adelaide, Adelaide, South Australia*
3. *The student*

Approval from the Ministry of Education, Aceh Besar

PEMERINTAH KABUPATEN ACEH BESAR

DINAS PENDIDIKAN

Jln. T. Bachtiar Panglima Polem, SH. Kota Jantho (23918) Telp. (0651) 92156, 92547 Fax. (0651) 92389
Email;dinas pendidikanacehbesar@gmail.com website www.disdikacehbesar.org

Nomor : 070/ 8756 /2013
Lamp : -
Hal : **Izin Pengumpulan Data**

Kota Jantho, 15 Agustus 2013
Kepada Yth,
Kepala Sekolah :

1. **SMP Negeri 1** [redacted]
2. **SMP Negeri 1** [redacted]
3. **SMP Negeri 2** [redacted]
4. **SMP Negeri 2** [redacted]
5. **SMP Negeri 1** [redacted]
6. **SMP** [redacted]
7. **SMP** [redacted]
8. **SMP** [redacted]

Kabupaten Aceh Besar
di –
Tempat

Dengan hormat,

Sehubungan dengan Surat Koordinator Pasca Sarjana School of Education, the University of Adelaide Nomor : HP.2013-034 tanggal Juli 2013, Kepala Dinas Pendidikan Kabupaten Aceh Besar memberi izin kepada :

Nama	: **Elizar**
Nomor Mahasiswa	: **1195731**
Program Studi	: **S3 Pendidikan Matematika**

Untuk melakukan penelitian dan mengumpulkan data untuk keperluan penyusunan Thesis yang berjudul :

"KETRAMPILAN BERPIKIR TINGKAT TINGGI DALAM PEMBELAJARAN MATEMATIKA : HUBUNGAN ANTARA PENDAPAT GURU, PRAKTEK KELAS DAN PRESTASI SISWA"

Setelah mengadakan penelitian 1 (satu) eks laporan dikirim ke Dinas Pendidikan Kabupaten Aceh Besar.

An. Kepala Dinas Pendidikan
Kabupaten Aceh Besar
Kabid. Pendidikan Pra sekolah
Dikdas dan PLB

Amri, S.Pd, M. Pd
NIP. 19660728 199003 1 008

Tembusan :
1. Koordinator Pasca Sarjana School of Education, the University of Adelaide
2. Arsip.

PEMERINTAH KABUPATEN ACEH BESAR
DINAS PENDIDIKAN
Jl. T. Bachtiar Panglima Polem, SH Kota Jantho (23918)
Telp. (0651) 92156, 92547 Fax. (0651) 92389
Email: dinas pendidikanacehbesar@gmail.com
website: www.disdikacehbesar.org

City of Jantho, 15 August 2013

Number : 070/2756/2013
Appendix : -
Subject : Permit for Data Collection

To:
1. **SMP** [redacted]
2. **SMP** [redacted]
3. **SMP** [redacted]
4. **SMP** [redacted]
5. **SMP** [redacted]
6. **SMP** [redacted]
7. **SMP** [redacted]
8. **SMP** [redacted]

With all due respect,
Pursuant to a letter from Coordinator of Postgraduate School of Education, the University of Adelaide, Adelaide, South Australia, number: HP.2013-034 on July 2013, Head of the Education Office in Aceh Besar District has given permission to:

Name	: **Elizar Ali**
Student identification number	: **11957311**
Program Study	: **Doctor of Mathematic Education**

To conduct research and to collect data for the purpose of Thesis with the title of

'HIGHER ORDER THINKING (HOT) SKILLS: INVESTIGATING THE NEXUS BETWEEN TEACHERS' BELIEFS, CLASSROOM PRACTICES, STUDENT BELIEFS AND MATHEMATICS PERFORMANCE IN ACEH, INDONESIA'

After conducting the research, a copy of the report should be sent to the Education Office of Aceh Besar district.

On behalf of Head of Education Agency Aceh Besar District
Head of Preschool, Elementary school and Special School.

Amiri, S.Pd, M.Pd
NIP. 19660728 199003 1 008

Cc:
1. Coordinator of postgraduate School of Education, the University of Adelaide, Adelaide, South Australia
2. Archives

Approval from the Ministry of Religious Affairs, Aceh Besar

KEMENTERIAN AGAMA REPUBLIK INDONESIA
KEMENTERIAN AGAMA
KANTOR KABUPATEN ACEH BESAR
Jl. Bupati T. Bachtiar Panglima Polem, SH Telp. 92174 Fax. 0651 - 23745
KOTA JANTHO. 23911

Nomor : Kd. 01.04/I/Kp.06/ 629 / 2013
Lampiran : -
Perihal : Mohon Bantuan dan Izin Mengumpulkan Data

Kepada Yth.
Kepala MTsN [redacted]
Kepala MTsN [redacted]
Kepala MTsN [redacted]
Kepala MTsN [redacted]
Kepala MTs [redacted]
Kepala MTs [redacted]
Kepala MTs [redacted]
Kepala MTs [redacted]

Di – Tempat

Sehubungan dengan surat kordinator Pascasarjana School of Education, the University of Adelaide,Adelaide,South Australia, tanggal Juni 2013. Perihal sebagaimana tersebut dipokok surat, maka dengan ini dimohonkan kepada saudara memberikan bantuan kepada mahasiswa yang tersebut namanya dibawah ini:

Nama	: Elizar
Nim	: 11957311
Pogram Studi	: S3 Pendidikan Matematika

Untuk melakukan pengumpulan data dalam rangka penyusunan Tesis untuk meyelesaikan studinya pada Prograram S3 Pendidikan Matematika di Pascasarjana School of Education, the University of Adelaide,Adelaide,South Australia, Adapun judul Tesis :

" KETERAMPILAN BERFIKIR TINGKAT TINGGI DALAM PEMBELAJARAN MATEMATIKA:HUBUNGAN ANTARA PENDAPAT GURU, PRAKTEK KELAS DAN PRESTASI SISWA ".

Demikian surat ini dibuat atas bantuannya kami ucapkan terima kasih.

Kota Jantho, 15 Agustus 2013
Ka.Sub.Bag. TU

Nashrullah, S.Ag
NIP.196209271992031003

Tembusan :
1. *Direktur Koorditor pascasarjana School of Education, the University of Adelaide,Adelaide,South Australia*
2. *Mahasiswa/i ybs*

KEMENTERIAN AGAMA REPUBLIK INDONESIA
KEMENTERIAN AGAMA
KANTOR KABUPATEN ACEH BESAR
Jl. Bupati T. Bachtiar Panglima Polem, SH
Telp. 92174 Fax. 0651 – 23745
KOTA JANTHO, 23911

Number	: Kd. 01.04/I/Kp.06/ 629 / 2013
Appendix	: -
Subject	: Appeal for Assistance and Permission for Data Collection

To:
Head of [redacted]
Head of [redacted]
Head of [redacted]
Head of [redacted]
Head of [redacted]
Head of [redacted]
Head of [redacted]
Head of [redacted]

Pursuant to a letter from Coordinator of Postgraduate School of Education, the University of Adelaide, Adelaide, South Australia, on June 2013. As mentioned in the letter, it is hereby requested to you to grants to student who is the name below:

Name	: Elizar
Student identification number	: 11957311
Program of Study	: Doctor of Mathematic Education

To conduct data collections for the proposed thesis to complete her program of study in Doctorate in Postgraduate School of Education the University of Adelaide, Adelaide, South Australia. The title of the thesis is

'HIGHER ORDER THINKING (HOT) SKILLS: INVESTIGATING THE NEXUS BETWEEN TEACHERS' BELIEFS, CLASSROOM PRACTICES, STUDENT BELIEFS AND MATHEMATICS PERFORMANCE IN ACEH, INDONESIA'

This statement has been issued for use as stated and intended and thank you in advance.

City of Jantho, 15 August 2013
Head of Administration

Nashrullah, S.Ag
NIP. 196209271992031003

Cc:
1. *Coordinator of postgraduate School of Education, the University of Adelaide, Adelaide, South Australia*
2. *Intended student*

Appendix C School Questionnaire

School Questionnaire (English)

THE UNIVERSITY *of* ADELAIDE

SCHOOL OF EDUCATION

Level 8, 10 Pulteney St, The University of Adelaide, Adelaide SA 5005; Tel: (+618) 8303 5628, Fax: (+618) 8303 3604

SCHOOL QUESTIONNAIRE

The aim of this questionnaire is to gather information about your school. Researcher will keep the answers **confidential**. Names of schools **will not** be used in the report of the study. Please take about **10 minutes or so** to complete this questionnaire and return it.

Instructions:
Please tick (✓) the appropriate box or complete the answer.

1. What is the type of your school?
   a. Public school ☐
   b. Islamic public school ☐
   c. Private school ☐
   d. Islamic private school ☐

2. As at July 2013, what was the total enrolment (number of students) at your school?
   ____

3. Which of the following factors are considered when students are admitted to your school? (You can choose more than one answer)
   a. Residence in particular area ☐
   b. Student's record of academic performance ☐
   c. Placement test ☐

4. Some schools organise instruction differently for students with differing abilities. Which of the following best describe policies in your school?
   a. Students are grouped by ability into different classes ☐
   b. Students are grouped by ability within their classes ☐
   c. Students are not grouped according to ability into classes or within classes ☐

5. What is the average class size in your school?
   ____

6. How many teachers are in your school?
   ____

7. How many of them have passed the certification exam?
   ____

8. How many mathematics teachers are in your school?
   ____

9. How many of the mathematics teachers have passed the certification exam?
   ____

10. How many mathematics teachers teach year 9 classes?
    ____

1

11. Which of these facilities are available in your school?
(You can choose more than one answer)

a. Computers for mathematics instruction ☐
b. Computers with internet connection for mathematics instruction ☐
c. Concrete material for mathematics teaching ☐
d. Computer software for mathematics instruction ☐
e. Library material relevant to mathematics instruction ☐
f. Mathematics text books ☐

12. What is the approximate number of each of these facilities available in your school?

a. Computers for mathematics instruction ________
b. Computers with internet connection for mathematics instruction ________
c. Concrete material for mathematics teaching ________
d. Computer software for mathematics instruction ________
e. Library material relevant to mathematics instruction ________
f. Mathematics text book ________

13. Does your school provide the following?
(You can choose more than one answer)

a. Additional mathematics programs after school hours ☐
b. Mathematics Olympiad club ☐

14. What are the approximate hours per week for the additional mathematics programs?

	≤ 2 hours	*3 hours*	*4 hours*	*≥ 4 hours*
a. Additional mathematics programs after school hours	☐	☐	☐	☐
b. Mathematics Olympiad club	☐	☐	☐	☐

Thank you for participating in this questionnaire!
Please return the completed form!

Only for the researcher

School Code	:	....
Date	:	....

School Questionnaire (Bahasa)

THE UNIVERSITY *of* ADELAIDE

SCHOOL OF EDUCATION

Level 8, 10 Pulteney St, The University of Adelaide, Adelaide SA 5005; Tel: (+618) 8303 5628, Fax: (+618) 8303 3604

KUESIONER UNTUK SEKOLAH

Tujuan dari kuesioner ini adalah untuk mengumpulkan informasi tentang sekolah Anda. Peneliti akan berusaha sebaik mungkin agar jawaban dalam kuesioner ini bersifat rahasia. Nama sekolah tidak akan digunakan dalam laporan penelitian. Pengisian kuesioner ini akan membutuhkan waktu sekitar 10 menit. Silakan mengisi kuesioner ini dan mengembalikannya

Petunjuk
Berilah tanda centang (✓) pada kotak yang disediakan atau lengkapilah.

1. Jenis sekolah:
   a. SMPN ☐
   b. MTSN ☐
   c. SMP swasta ☐
   d. MTS swasta ☐

2. Per Juli 2013, berapakah jumlah total siswa di sekolah anda?
   ____

3. Manakah dari faktor-faktor berikut ini yang dijadikan bahan pertimbangan saat penerimaan siswa baru di sekolah anda? (Jawaban boleh lebih dari satu)
   a. Tempat tinggal siswa/Rayon ☐
   b. Nilai siswa dari jenjang pendidikan sebelumnya ☐
   c. Tes ujian masuk ☐

4. Sebagian sekolah membedakan pembelajaran berdasarkan kemampuan siswa. Manakah dari kebijakan berikut ini yang mengambarkan kebijakan di sekolah anda?
   a. Kelas disusun berdasarkan ranking ☐
   b. Kelas tidak disusun berdasarkan ranking, tetapi komposisi kelas diatur agar siswa-siswa di kelas tersebut heterogen ☐
   c. Kelas di susun secara acak, tanpa memperhatikan ranking ataupun komposisi kelas. ☐

5. Berapakah rata-rata jumlah siswa dalam satu kelas di sekolah anda?
   ____

6. Berapakah jumlah guru di sekolah anda?
   ____

7. Berapakah jumlah guru yang sudah lulus sertifikasi?
   ____

8. Berapakah jumlah guru matematika di sekolah anda?
   ____

9. Berapakah jumlah guru matematika yang sudah lulus sertifikasi?
   ____

10. Berapakah jumlah guru yang mengajar di kelas 9?
   ____

11. Manakah dari fasilitas berikut ini yang tersedia di sekolah anda? (Jawaban boleh lebih dari satu)

a. Komputer yang bisa digunakan untuk pembelajaran matematika ☐
b. Komputer (dengan koneksi internet) yang bisa digunakan untuk pembelajaran matematika ☐
c. Alat peraga untuk pembelajaran matematika ☐
d. Program komputer untuk pembelajaran matematika ☐
e. Bahan pustaka yang relevan untuk pembelajaran matematika ☐
f. Buku cetak matematika ☐

12. Berapakah jumlah dari setiap fasilitas berikut yang tersedia di sekolah anda?

a. Komputer yang bisa digunakan untuk pembelajaran matematika ________
b. Komputer (dengan koneksi internet) yang bisa digunakan untuk ________
c. Alat peraga untuk pembelajaran matematika ________
d. Program computer untuk pembelajaran matematika ________
e. Bahan pustaka yang relevan untuk pembelajaran matematika ________
f. Buku cetak matematika ________

13. Apakah sekolah anda memiliki program berikut ini?
(Jawaban boleh lebih dari satu)

a. Kelas matematika tambahan setelah jam sekolah ☐
b. Klub olimpiade matematika ☐

14. Dalam satu minggu, berapa durasi waktu program-program berikut ini di sekolah anda?

	≤ 2 jam	*3 jam*	*4 jam*	*≥ 4 jam*
a. Kelas matematika tambahan setelah jam sekolah	☐	☐	☐	☐
b. Klub olimpiade matematika	☐	☐	☐	☐

Terima kasih atas partisipasi anda dalam kuesioner ini!

Hanya untuk peneliti

Kode sekolah	:	....
Tanggal	:	....

2

Appendix D Teacher questionnaire

Teacher Questionnaire (English)

THE UNIVERSITY of ADELAIDE

SCHOOL OF EDUCATION

Level 8, 10 Pulteney St, The University of Adelaide, Adelaide SA 5005; Tel: (+618) 8303 5628, Fax: (+618) 8303 3604

TEACHER QUESTIONNAIRE

Researchers will be careful to keep your answer to this questionnaire **confidential**. Reports of findings **will not use** names of respondents or schools.
Please take about **20 minutes or so** to complete this questionnaire and return it.

Section 1. Background

Instruction:
Please tick (✓) the appropriate box or complete the answer.

1. By the end of this school year, how many years will you have been teaching altogether?
   ____________ years (please round to the nearest whole number)

2. Your gender: Female ☐ Male ☐

3. Your age: __________ years old (please round to the nearest whole number)

4. What is the highest level of formal education you have completed?

   a. Completed a 2-year college or university degree (i.e. Diploma) ☐
   b. Completed a 4-year college or university degree (i.e. Bachelor) ☐
   c. Completed a master's degree ☐
   d. Completed a doctorate (Ph.D or Ed.D) ☐

5. During your college or university education, what was your major or main area(s) of study?

   a. Mathematics Education ☐
   b. Mathematics (Non Education) – Pure mathematics ☐
   c. Mathematics (Non Education) – Informatics ☐
   d. Mathematics (Non Education) – Statistics ☐
   e. Other (Please specify) ______________________________

6. If your degree is not mathematics education or mathematics (non education), how many mathematics subjects did you complete during the course of your degree?

   ____

7. Have you passed the teaching certification exam?
   Yes ☐ No ☐

8. In the past two years, have you participated in professional development in any of the following?

   a. Mathematics content ☐
   b. Mathematics pedagogy/instruction ☐
   c. Mathematics curriculum ☐
   d. Integrating information technology to mathematics ☐
   e. Improving students critical thinking and problem solving ☐
   f. Mathematics assessment ☐
   g. Addressing individual students needs ☐

9. How often do you attend professional development programmes for mathematics teachers?

a. Every 2 weeks	☐	d. Every 3 months	☐
b. Every month	☐	e. Every 6 months	☐
c. Every 2 months	☐	f. Less regularly	☐

Section 2. Teacher's beliefs

Instruction:
Please tick (✓) the appropriate box.
There is no right or wrong answer. Please choose the answer which represents ***your opinion***

10. How much do you agree with these statements about the nature of mathematics?

	Disagree a lot	*Disagree a little*	*Agree a little*	*Agree a lot*
a. Mathematics is not a flexible process where accuracy, speed, and memory is important	☐	☐	☐	☐
b. Mathematics is just addition, subtraction, multiplication and division				
c. Mathematics problems given to students should be quickly solvable in a few steps	☐	☐	☐	☐
d. Mathematics is the dynamic searching for order and pattern in the learner's environment	☐	☐	☐	☐
e. Mathematics is a creative human endeavour that is both a way of knowing and a way of thinking	☐	☐	☐	☐
f. Right answers are more important in mathematics than the ways in which you get them	☐	☐	☐	☐
g. Mathematics is about remembering the rules	☐	☐	☐	☐
h. Analysis is important in solving mathematics problems	☐	☐	☐	☐
i. Mathematics problems could be solved in various ways	☐	☐	☐	☐

11. How much do you agree with these statements about mathematics learning?

a. Mathematics learning is enhanced by activities which build upon students' thinking skills	☐	☐	☐	☐
b. Mathematics learning is enhanced by challenges within a supportive environment	☐	☐	☐	☐
c. Students' critical thinking skills are enhanced when they work on an open-ended mathematics problem	☐	☐	☐	☐
d. Mathematics learning is being able to transfer the skills to a new unfamiliar problem	☐	☐	☐	☐
e. Mathematics learning is enhanced by exposing students to repetitive routine problems only	☐	☐	☐	☐
f. A demonstration of good reasoning should be seen as more important than a student's ability to find the correct answer	☐	☐	☐	☐
g. Mathematics learning is enhanced when students are given enough opportunities to discover their own solutions	☐	☐	☐	☐
h. Being able to memorise facts is very critical in mathematics learning	☐	☐	☐	☐
i. Mathematics learning is being able to get the right answer quickly	☐	☐	☐	☐

12. How much do you agree with these statements about mathematics teaching?

	Disagree a lot	*Disagree a little*	*Agree a little*	*Agree a lot*
a. Teachers should provide instructional activities which result in challenging situations for learners	☐	☐	☐	☐
b. Teachers should allow students to work in a cooperative learning environment with their peers	☐	☐	☐	☐
c. Teachers should allow students the opportunity to analyse problems before they try to find solutions	☐	☐	☐	☐
d. Teachers should not tell students if their answers are correct or incorrect. Rather, they should challenge them to explain their strategies	☐	☐	☐	☐
e. Teachers should always demonstrate how to solve simple problems before students are allowed to solve problems	☐	☐	☐	☐
f. Students should be encouraged to justify their solutions and reasoning	☐	☐	☐	☐
g. Teachers should teach students the exact procedures for solving problem	☐	☐	☐	☐
h. Mathematics should be taught as a collection of rules, procedures and algorithms	☐	☐	☐	☐
i. Mathematics should be taught using a combination of routine and non-routine problems to develop students' thinking skills	☐	☐	☐	☐
j. Higher order thinking tasks should not be given to low achieving students	☐	☐	☐	☐
k. The role of the mathematics teacher is to transmit mathematical knowledge and to verify that learners have received this knowledge	☐	☐	☐	☐

Section 3. Classroom practices

Instruction:
Please tick (✓) the appropriate box.

13. How often do you include the following types of questions in your daily mathematics teaching?

	Never				Always
	1	2	3	4	5
a. Questions similar to what you have solved in the classroom	☐	☐	☐	☐	☐
b. Questions involving application of mathematical procedures	☐	☐	☐	☐	☐
c. Question which can be solved in many ways	☐	☐	☐	☐	☐
d. Question involving unfamiliar problems	☐	☐	☐	☐	☐
e. Open-ended questions	☐	☐	☐	☐	☐
f. Word problems	☐	☐	☐	☐	☐

14. How often do you include the following types of questions in your mathematics tests examination?

a. Questions similar to what you have solved in the classroom	☐	☐	☐	☐	☐
b. Questions involving application of mathematical procedures	☐	☐	☐	☐	☐
c. Question which can be solved in many ways	☐	☐	☐	☐	☐
d. Question involving unfamiliar problems	☐	☐	☐	☐	☐
e. Open-ended questions	☐	☐	☐	☐	☐
f. Word problems	☐	☐	☐	☐	☐

15. In teaching mathematics to this class, how often do you usually ask students to do the following?

	Never 1	2	3	4	Always 5
a. Listen to me explaining how to solve problems	☐	☐	☐	☐	☐
b. Memorise rules, procedures and facts	☐	☐	☐	☐	☐
c. Work problems with peers with my guidance	☐	☐	☐	☐	☐
d. Work problems together in whole class with direct guidance from me	☐	☐	☐	☐	☐
e. Apply fact, concepts and procedures to solve routine problem	☐	☐	☐	☐	☐
f. Explain their answer	☐	☐	☐	☐	☐
g. Relate what they learning in mathematics to their daily lives	☐	☐	☐	☐	☐
h. Decide on their own procedures for solving complex problems	☐	☐	☐	☐	☐
i. Work on problems for which there is no immediate obvious method of solution	☐	☐	☐	☐	☐
j. Take a written test or quiz	☐	☐	☐	☐	☐

16. How often do you do the following in teaching this class?

a. Summarize what students should have learned from the lesson	☐	☐	☐	☐	☐
b. Relate the lesson to student's daily life	☐	☐	☐	☐	☐
c. Use questioning to elicit reasons and explanation	☐	☐	☐	☐	☐
d. Encourage all students to improve their performance	☐	☐	☐	☐	☐
e. Praise students for good effort	☐	☐	☐	☐	☐
f. Bring interesting materials to class	☐	☐	☐	☐	☐

17. When you teach mathematics to this class, how do you use the following resources?

	Not used	*Supplement*	*Basic of instruction*
a. Textbooks	☐	☐	☐
b. Worksheets or workbooks	☐	☐	☐
c. Concrete objects/materials to help students understand the quantities and procedures	☐	☐	☐
d. Computer softwares for mathematics instructions	☐	☐	☐

Thank you for your participation in this survey!
Please return the completed form!

Researcher use only

School Code	:	....
Class Code	:	....
Teacher Code	:	....
Date	:	....

Teacher Questionnaire (Bahasa)

THE UNIVERSITY of ADELAIDE

SCHOOL OF EDUCATION

Level 8, 10 Pulteney St, The University of Adelaide, Adelaide SA 5005; Tel: (+618) 8303 5628, Fax: (+618) 8303 3604

KUESIONER UNTUK GURU

Peneliti akan berusaha sebaik mungkin untuk menjaga agar jawaban dalam kuesioner ini bersifat **rahasia**.
Laporan penelitian ini **tidak akan** menggunakan nama guru maupun nama sekolah.

Bagian 1. Data guru

Petunjuk: *Berilah tanda centang (✓) pada kotak yang disediakan atau lengkapilah.*

1. Terhitung pada akhir tahun ini, sudah berapa lamakah anda mengajar matematika?
   ______________ tahun (bulatkan ke bilangan bulat terdekat)

2. Jenis kelamin: Perempuan ☐ Laki-laki ☐

3. Usia: ___________ tahun (bulatkan ke bilangan bulat terdekat)

4. Apa pendidikan terakhir anda?

   a. Diploma (D1, D2 atau D3) ☐
   b. Sarjana (S1) ☐
   c. Magister (S2) ☐
   d. Doktoral (S3) ☐

5. Apakah jurusan anda saat di universitas?

   a. Pendidikan matematika ☐
   b. Matematika (non pendidikan) – Matematika murni ☐
   c. Matematika (non pendidikan) – Informatika ☐
   d. Matematika (non pendidikan) – Statistik ☐
   e. Lainnya (sebutkan) ____________________________

6. Jika jurusan anda bukan jurusan pendidikan matematika ataupun matematika (non pendidikan), berapa mata kuliah yang berhubungan dengan matematika yang anda ambil selama kuliah?

   ____

7. Apakah anda sudah lulus sertifikasi guru? Ya ☐ Tidak ☐

8. Dalam kurun waktu 2 tahun ini, pernahkah anda mengikuti program pengembangan profesionalisme (pelatihan) guru yang berkaitan dengan topik berikut ini?

   a. Materi (bahan ajar) matematika ☐
   b. Pedagogi atau pengajaran matematika ☐
   c. Kurikulum matematika ☐
   d. Penggunaan teknologi dalam pengajaran matematika ☐
   e. Peningkatan kemampuan berpikir kritis dan pemecahan masalah untuk siswa ☐
   f. Penilaian matematika ☐
   g. Pemenuhan kebutuhan individu siswa ☐

9. Seberapa seringkah anda mengikuti program Musyawarah Guru Mata Pelajaran (MGMP) matematika?

a. Setiap 2 minggu	☐	d. Setiap 3 bulan	☐
b. Setiap bulan	☐	e. Setiap 6 bulan	☐
c. Setiap 2 bulan	☐	f. Sangat jarang / Tidak pernah	☐

Bagian 2. Pendapat guru tentang matematika dan pembelajaran matematika

Petunjuk: *Berilah tanda centang (✓) pada kotak yang disediakan. Tidak ada jawaban benar atau salah, pilihlah jawaban berdasarkan* ***pendapat anda.***

10. Apa **pendapat** anda terhadap pernyataan-pernyataan berikut ini?

	Sangat tidak setuju	*Tidak setuju*	*Setuju*	*Sangat setuju*
a. Matematika adalah suatu sistem kaku di mana ketepatan, kecepatan dan daya ingat sangat penting	☐	☐	☐	☐
b. Matematika hanya merupakan penjumlahan, pengurangan, perkalian dan pembagian	☐	☐	☐	☐
c. Soal matematika untuk siswa harus dapat diselesaikan dengan cepat dalam beberapa langkah	☐	☐	☐	☐
d. Matematika merupakan pencarian yang dinamis akan urutan dan pola dalam kehidupan siswa	☐	☐	☐	☐
e. Matematika adalah usaha kreatif manusia yang merupakan suatu cara untuk mengetahui dan juga cara untuk berpikir	☐	☐	☐	☐
f. Dalam matematika, jawaban yang benar lebih penting dari pada strategi yang digunakan untuk mendapatkan jawaban tersebut	☐	☐	☐	☐
g. Matematika adalah ilmu tentang mengingat aturan dan rumus	☐	☐	☐	☐
h. Analisis adalah bagian penting dalam pemecahan masalah matematika	☐	☐	☐	☐
i. Soal matematika dapat diselesaikan dengan berbagai cara	☐	☐	☐	☐

11. Apa **pendapat** anda terhadap pernyataan-pernyataan berikut ini?

a. Pembelajaran matematika dapat di tingkatkan melalui kegiatan pembelajaran yang mengembangkan kemampuan berpikir siswa	☐	☐	☐	☐
b. Pembelajaran matematika dapat di tingkatkan melalui tantangan dalam lingkungan belajar yang mendukung	☐	☐	☐	☐
c. Kemampuan siswa berpikir kritis akan meningkat ketika mereka mengerjakan soal matematika terbuka	☐	☐	☐	☐
d. Pembelajaran matematika adalah kemampuan mentransfer keahlian untuk menyelesaikan masalah baru yang belum di pahami atau yang belum pernah di selesaikan sebelumnya	☐	☐	☐	☐
e. Pembelajaran matematika dapat ditingkatkan dengan hanya memberikan masalah-masalah rutin kepada siswa	☐	☐	☐	☐
f. Penalaran yang baik seharusnya dilihat sebagai sesuatu yang lebih penting dari pada kemampuan siswa menemukan jawaban yang benar	☐	☐	☐	☐
g. Pembelajaran matematika akan meningkat ketika siswa di berikan kesempatan yang cukup untuk menemukan solusinya sendiri	☐	☐	☐	☐
h. Kemampuan menghafal merupakan hal penting dalam pembelajaran matematika	☐	☐	☐	☐
i. Pembelajaran matematika adalah bagaimana mendapatkan jawaban yang benar dengan cepat	☐	☐	☐	☐

12. Apa **pendapat** anda terhadap pernyataan–pernyataan berikut ini?

	Sangat tidak setuju	*Tidak setuju*	*Setuju*	*Sangat setuju*
a. Guru harus memberikan kegiatan pembelajaran yang menantang bagi siswa	☐	☐	☐	☐
b. Guru harus memberikan siswa kesempatan untuk bekerja sama dengan siswa lain dalam lingkungan belajar kooperatif	☐	☐	☐	☐
c. Guru harus memberikan siswa kesempatan untuk menganalisa masalah sebelum mereka mencari solusi	☐	☐	☐	☐
d. Guru tidak seharusnya memberitahu siswa apakah jawaban mereka benar atau salah, tetapi guru harus menantang siswa menjelaskan strategi yang mereka gunakan	☐	☐	☐	☐
e. Guru harus selalu menunjukkan cara memecahkan masalah sebelum siswa diperbolehkan untuk memecahkan masalah tersebut	☐	☐	☐	☐
f. Siswa harus didorong untuk menjelaskan jawaban dan penalaran mereka	☐	☐	☐	☐
g. Guru harus selalu mengajarkan prosedur yang pasti untuk memecahkan masalah	☐	☐	☐	☐
h. Matematika harus diajarkan sebagai kumpulan aturan, prosedur dan algoritma	☐	☐	☐	☐
i. Matematika harus diajarkan menggunakan kombinasi masalah rutin (tidak asing) dan non-rutin (asing) untuk mengembangkan kemampuan berpikir siswa	☐	☐	☐	☐
j. Soal yang membutuhkan penalaran tingkat tinggi sebaiknya tidak di berikan kepada siswa yang berkemampuan rendah	☐	☐	☐	☐
k. Peran guru matematika adalah menyampaikan pengetahuan matematika dan membuktikan bahwa siswa telah menerima pengetahuan tersebut	☐	☐	☐	☐

Bagian 3. Kegiatan di kelas matematika

Petunjuk: *Berilah tanda centang (✓) pada kotak yang disediakan.*

13. Seberapa **sering** anda menggunakan pertanyaan matematika berikut ini dalam mengajar sehari-hari?

	Tidak pernah				*Selalu*
	1	2	3	4	5
a. Soal yang mirip dengan yang saya dijelaskan di depan kelas	☐	☐	☐	☐	☐
b. Soal yang melibatkan penerapan prosedur matematika	☐	☐	☐	☐	☐
c. Soal yang dapat diselesaikan dengan banyak cara	☐	☐	☐	☐	☐
d. Pertanyaan yang melibatkan masalah non-rutin (masalah yang asing bagi siswa)	☐	☐	☐	☐	☐
e. Pertanyaan terbuka	☐	☐	☐	☐	☐
f. Soal cerita					

14. Seberapa **sering** anda menggunakan pertanyaan matematika berikut ini dalam tes?

a. Pertanyaan yang berdasarkan pada mengingat fakta dan prosedur	☐	☐	☐	☐	☐
b. Pertanyaan yang melibatkan penerapan prosedur matematika	☐	☐	☐	☐	☐
c. Pertanyaan yang melibatkan pencarian pola dan hubungan	☐	☐	☐	☐	☐
d. Pertanyaan yang melibatkan masalah non-rutin (masalah yang asing bagi siswa)	☐	☐	☐	☐	☐
e. Pertanyaan terbuka	☐	☐	☐	☐	☐
f. Soal cerita	☐	☐	☐	☐	☐

15. Ketika mengajar matematika di kelas, seberapa sering biasanya anda meminta siswa untuk melakukan hal-hal berikut?

	Tidak pernah				*Selalu*
	1	2	3	4	5
a. Mendengarkan saya menjelaskan cara menyelesaikan soal	☐	☐	☐	☐	☐
b. Menghafal aturan, prosedur dan rumus	☐	☐	☐	☐	☐
c. Menyelesaikan soal secara berkelompok dengan bimbingan dari saya	☐	☐	☐	☐	☐
d. Menyelesaikan soal bersama-sama di depan kelas dengan bimbingan langsung dari saya	☐	☐	☐	☐	☐
e. Mengaplikasikan konsep, prosedur dan rumus untuk menyelesaikan masalah rutin	☐	☐	☐	☐	☐
f. Menjelaskan jawaban dari soal yang mereka kerjakan	☐	☐	☐	☐	☐
g. Menghubungkan apa yang mereka pelajari dengan kehidupan sehari-hari	☐	☐	☐	☐	☐
h. Menentukan sendiri prosedur untuk menyelesaikan masalah yang rumit	☐	☐	☐	☐	☐
i. Menyelesaikan masalah yang solusinya tidak terlihat jelas	☐	☐	☐	☐	☐
j. Mengikuti tes tertulis atau kuis	☐	☐	☐	☐	☐

16. Seberapa sering Anda melakukan hal berikut ketika mengajar di kelas?

a. Menyimpulkan apa yang harus di pelajari siswa di kelas	☐	☐	☐	☐	☐
b. Menghubungkan pelajaran dengan kehidupan sehari-hari siswa	☐	☐	☐	☐	☐
c. Menggunakan pertanyaan untuk memperoleh alasan dan penjelasan	☐	☐	☐	☐	☐
d. Mendorong semua siswa untuk meningkatkan prestasinya	☐	☐	☐	☐	☐
e. Memuji siswa atas usaha yang baik	☐	☐	☐	☐	☐
f. Membawa bahan menarik untuk kelas	☐	☐	☐	☐	☐

17. Ketika anda mengajar matematika di kelas, bagaimana anda menggunakan sumber belajar berikut?

	Tidak menggunakan	*Sebagai suplemen*	*Sumber utama*
a. Buku cetak	☐	☐	☐
b. Lembar kerja siswa	☐	☐	☐
c. Alat peraga atau objek nyata untuk membantu pemahaman siswa	☐	☐	☐
d. Program komputer untuk pembelajaran matematika	☐	☐	☐

Terima kasih atas partisipasi bapak/ibu dalam kuesioner ini!
Silakan mengembalikan kuesioner yang sudah disi!

Hanya untuk peneliti

Kode Sekolah	:	....
Kode Kelas	:	....
Kode Guru	:	....
Tanggal	:	....

Appendix E Student questionnaire

Student Questionnaire (English)

THE UNIVERSITY *of* ADELAIDE

SCHOOL OF EDUCATION

Level 8, 10 Pulteney St, The University of Adelaide, Adelaide SA 5005; Tel: (+618) 8303 5628, Fax: (+618) 8303 3604

STUDENT QUESTIONNAIRE

Researcher will keep your answers **confidential**. Names of students or schools **will not** be used in the report of the study. Please take about **20 minutes or so** to complete this questionnaire and return it.

Section 1. About you

Instructions: *Please tick (✓) the appropriate box or complete the answer.*

1. Your gender:

   Female ☐ Male ☐

2. Who usually lives at home with you?

a. Mother (including stepmother or foster mother)	☐
b. Father (including stepfather or foster father)	☐
c. Brother(s) (including stepbrothers and foster brothers)	☐
d. Sister(s) (including stepsisters and fosters sisters)	☐
e. Grandparent(s)	☐
f. Others (e.g., cousin)	☐

3. What is your mother's main job? ____________________

4. What is the highest level of schooling completed by your mother?

a. Master's, doctoral, or professional degree such as medicine or law	☐
b. Bachelor's degree (4 year college degree)	☐
c. Diploma (1, 2 or 3 year college degree)	☐
d. Vocational or technical certificate after high school (such as cosmetology or auto mechanic)	☐
e. She completed grade 12	☐
f. He completed grade 9	☐
g. He completed grade 6	☐
h. She did not complete grade 6	☐

5. What is your father's main job? ____________________

6. What is the highest level of schooling completed by your father?

a. Master's, doctoral, or professional degree such as medicine or law	☐
b. Bachelor's degree (4 year college degree)	☐
c. Diploma (1, 2 or 3 year college degree)	☐
d. Vocational or technical certificate after high school (such as cosmetology or auto mechanic)	☐
e. He completed grade 12	☐
f. He completed grade 9	☐
g. He completed grade 6	☐
h. He did not completed grade 6	☐

7. Which of the following are present in your home?

a. A desk to study at	☐
b. A room of your own	☐
c. A computer you can use for school work	☐
d. A mobile phone of your own	☐
e. Dictionary	☐
f. Books to help you with your homework	☐
g. Encyclopaedia	☐
h. Internet connection	☐
i. A guest room	☐
j. A DVD player	☐
k. Playstation®, Xbox®, Wii® or other TV/video game systems.	☐
l. TV (cable/pay)	☐
m. TV (regular free-to-air channels)	☐

8. How many of these are at home?

	0	1	2	≥ 3
a. Mobile phone	☐	☐	☐	☐
b. Televisions	☐	☐	☐	☐
c. Computers	☐	☐	☐	☐
d. Cars	☐	☐	☐	☐
e. motorcycles	☐	☐	☐	☐

9. How far in school do you expect to go?

a. High school	☐
b. Vocational or technical education after high school	☐
c. Diploma (2 years or 3 years degree)	☐
d. Bachelor's degree (4 year degree)	☐
e. Master's degree	☐
f. Doctoral degree	☐

Section 2. Mathematics in school

Instructions:
Please tick (✓) the appropriate box.
*There is **no right or wrong answer**. Please choose the answer which represents **<u>your opinion</u>**.*

10. How much do you agree with these statements?

	Disagree a lot	Disagree a little	Agree a little	Agree a lot
a. I enjoy learning mathematics	☐	☐	☐	☐
b. I wish I do not have to study mathematics	☐	☐	☐	☐
c. Mathematics is boring	☐	☐	☐	☐
d. I learn many interesting things in mathematics	☐	☐	☐	☐
e. I like mathematics	☐	☐	☐	☐

11. How much do you agree with these statements?

	Disagree a lot	Disagree a little	Agree a little	Agree a lot
a. I think learning mathematics will help me in my daily life	☐	☐	☐	☐
b. I need mathematics to learn other school subjects	☐	☐	☐	☐
c. I need to do well in mathematics to get into the university or college of my choice	☐	☐	☐	☐
d. I need to do well in mathematics to get the job I want	☐	☐	☐	☐
e. I would like a job that involves mathematics	☐	☐	☐	☐

12. How much do you agree with these statements?

a. I usually do well in mathematics	☐	☐	☐	☐
b. Mathematics is not one of my strengths	☐	☐	☐	☐
c. Mathematics is more difficult for me than for many of my classmates	☐	☐	☐	☐
d. I learn things quickly in mathematics	☐	☐	☐	☐
e. Mathematics makes me confused and nervous	☐	☐	☐	☐
f. I am good at working out difficult mathematics problem	☐	☐	☐	☐
g. My teacher thinks I can do well in mathematics classes with difficult materials	☐	☐	☐	☐
h. My teacher tells me I am good at mathematics	☐	☐	☐	☐
i. Mathematics is harder for me than any other subject	☐	☐	☐	☐

13. How much do you agree with these statements?

a. Mathematics is just about addition, subtraction, multiplication and division	☐	☐	☐	☐
b. Mathematics problems should be quickly solvable in a few steps	☐	☐	☐	☐
c. There are several ways to solve a mathematics problem	☐	☐	☐	☐
d. In mathematics, a correct answer is more important than the way to get it	☐	☐	☐	☐
e. All mathematics problems can be solved in one way only	☐	☐	☐	☐
f. Mathematics is just a collection of rules and formulas	☐	☐	☐	☐

14. How much do you agree with these statements?

a. The way teacher solves mathematics problem is the only correct way to solve the problem	☐	☐	☐	☐
b. Memorising is the most important thing in learning mathematics	☐	☐	☐	☐
c. I should always follow the procedures the teacher taught in solving the mathematics problem	☐	☐	☐	☐
d. The main goal of doing mathematics problems is to obtain a correct answer	☐	☐	☐	☐
e. I usually try to create my own solution for mathematics problem	☐	☐	☐	☐
f. I wish my teacher only provided me with those problems which I am familiar with	☐	☐	☐	☐
g. I think my teacher should always show how to solve a problem before she/he asks me to do the task	☐	☐	☐	☐
h. I like working on a mathematics problem where the solution is not obvious	☐	☐	☐	☐

Section 3. Classroom practices

Instructions:
Please tick the appropriate box.
There is ***no right or wrong answer****. Please choose the answer which represents* ***your opinion****.*

15. How often do you work on the following types of questions in your mathematics classroom?

	Never or almost never	Sometimes	Always or almost always
a. Questions similar to what the teacher solves in the classroom	☐	☐	☐
b. Questions which apply my knowledge from previous topics	☐	☐	☐
c. Questions which can be solved in many ways	☐	☐	☐
d. Word problems	☐	☐	☐

16. How often do you use the following in mathematics classroom?

a. Textbooks	☐	☐	☐
b. Worksheets or workbooks	☐	☐	☐
c. Concrete objects	☐	☐	☐
d. Computer software to understand mathematics concept	☐	☐	☐

17. How often do you do these in your mathematics classroom?

a. Listening to the teacher explaining how to solve problems	☐	☐	☐
b. Memorising formulas	☐	☐	☐
c. Working on problems with your classmates	☐	☐	☐
d. Using your knowledge from previous topics to solve problems	☐	☐	☐
e. Explaining your answer	☐	☐	☐
f. Relating what you learn in mathematics to your daily life	☐	☐	☐
g. Using your own way to solve difficult problems	☐	☐	☐
h. Working on problems which you cannot find the solution straight away	☐	☐	☐

Thank you for your participation in this survey!
Please return the completed form!

Researcher use only

School Code	:	....
Class Code	:	....
Teacher Code	:	....
Student Code	:	....
Date	:	....

Student Questionnaire (Bahasa)

THE UNIVERSITY of ADELAIDE

SCHOOL OF EDUCATION

Level 8, 10 Pulteney St, The University of Adelaide, Adelaide SA 5005; Tel: (+618) 8303 5628, Fax: (+618) 8303 3604

KUESIONER UNTUK SISWA

Jawaban dalam kuesioner ini bersifat **rahasia**. Laporan penelitian ini **tidak akan** menggunakan nama siswa maupun nama sekolah.

Bagian 1. Data pribadi siswa

Petunjuk: *Berilah tanda centang (✓) pada kotak yang disediakan atau lengkapilah.*

1. Jenis kelamin:

   Perempuan ☐ Laki-laki ☐

2. Dengan siapa kamu tinggal di rumah?

a. Ibu (termasuk ibu tiri atau ibu angkat)	☐
b. Ayah (termasuk ayah tiri atau ayah angkat)	☐
c. Saudara laki-laki (termasuk saudara tiri dan saudara angkat)	☐
d. Saudara perempuan (termasuk saudara tiri dan saudara angkat)	☐
e. Kakek atau Nenek	☐
f. Lainnya (misal: sepupu)	☐

3. Apakah pekerjaan ibumu? ______________________

4. Apakah pendidikan terakhir ibumu?

a. Magister (S2), doktoral (S3) atau pendidikan profesi (contoh: dokter atau pengacara)	☐
b. Sarjana (S1)	☐
c. Diploma (D1, D2, atau D3)	☐
d. Kursus bersertifikat setelah SMA (contoh: kursus kecantikan, ahli mekanik, dll)	☐
e. Tamat SMA	☐
f. Tamat SMP	☐
g. Tamat SD	☐
h. Tidak Tamat SD	☐

5. Apakah pekerjaan ayahmu? ______________________

6. Apakah pendidikan terakhir ayahmu?

a. Magister (S2), doktoral (S3) atau pendidikan profesi (contoh: dokter atau pengacara)	☐
b. Sarjana (S1)	☐
c. Diploma (D1, D2, atau D3)	☐
d. Kursus bersertifikat setelah SMA (contoh: kursus kecantikan, ahli mekanik, dll)	☐
e. Tamat SMA	☐
f. Tamat SMP	☐
g. Tamat SD	☐
h. Tidak Tamat SD	☐

7. Apakah kamu mempunyai benda-benda berikut ini di rumah?

a. Meja belajar	☐
b. Kamar sendiri	☐
c. Komputer/ laptop yang bisa kamu pakai untuk mengerjakan PR	☐
d. Telepon gengam / *handphone* sendiri	☐
e. Kamus	☐
f. Buku yang bisa kamu pakai untuk mengerjakan PR	☐
g. Ensiklopedia	☐
h. Koneksi internet	☐
i. Kamar tamu	☐
j. DVD player	☐
k. *Playstation*®, Xbox®, Wii® atau permainan untuk TV lainnya.	☐
l. TV berlangganan (contoh: Indovision®)	☐
m. TV biasa	☐

8. Ada berapa banyakkah benda-benda berikut ini di rumahmu?

	0	1	2	≥ 3
a. Telepon gengam / *handphone*	☐	☐	☐	☐
b. TV	☐	☐	☐	☐
c. Komputer/ laptop	☐	☐	☐	☐
d. Mobil	☐	☐	☐	☐
e. Sepeda motor	☐	☐	☐	☐

9. Apakah jenjang pendidikan tertinggi yang ingin kamu tempuh?

a. SMA	☐
b. Kursus bersertifikat setelah SMA (contoh: kecantikan, ahli mekanik, dll)	☐
c. Diploma (D1, D2, atau D3)	☐
d. Sarjana (S1)	☐
e. Magister (S2)	☐
f. Doktoral (S3)	☐

Bagian 2. Pendapat siswa tentang pelajaran matematika

Petunjuk:

Berilah tanda centang (✓) pada kotak yang disediakan.
Tidak ada jawaban benar atau salah. Pilihlah jawaban berdasarkan ***pendapatmu***

10. Apa pendapatmu terhadap pernyataan-pernyataan berikut ini?

	Sangat tidak setuju	*Tidak setuju*	*Setuju*	*Sangat setuju*
a. Saya senang belajar matematika	☐	☐	☐	☐
b. Saya berharap saya tidak harus belajar matematika	☐	☐	☐	☐
c. Matematika membosankan	☐	☐	☐	☐
d. Saya belajar banyak hal menarik dalam matematika	☐	☐	☐	☐
e. Saya suka matematika	☐	☐	☐	☐

11. Apa pendapatmu terhadap pernyataan–pernyataan berikut ini?

	Sangat tidak setuju	*Tidak setuju*	*Setuju*	*Sangat setuju*
a. Belajar matematika akan membantu saya dalam kehidupan sehari-hari	☐	☐	☐	☐
b. Saya membutuhkan matematika untuk mempelajari mata pelajaran lain	☐	☐	☐	☐
c. Saya harus mendapatkan nilai matematika yang bagus agar bisa diterima di universitas yang saya inginkan	☐	☐	☐	☐
d. Saya harus mendapatkan nilai matematika yang bagus untuk mendapatkan pekerjaan yang saya inginkan	☐	☐	☐	☐

12. Apa pendapatmu terhadap pernyataan–pernyataan berikut ini?

a. Saya biasanya mendapatkan nilai bagus untuk pelajaran matematika	☐	☐	☐	☐
b. Saya tidak mahir matematika	☐	☐	☐	☐
c. Matematika lebih sulit bagi saya dibandingkan bagi teman-teman sekelas saya	☐	☐	☐	☐
d. Saya belajar matematika dengan cepat	☐	☐	☐	☐
e. Matematika membuat saya bingung dan gugup	☐	☐	☐	☐
f. Saya mahir dalam menyelesaikan soal matematika yang sulit	☐	☐	☐	☐
g. Guru saya berpikir bahwa saya dapat menyelesaikan soal-soal matematika yang sulit	☐	☐	☐	☐
h. Menurut guru saya, saya mahir dalam matematika	☐	☐	☐	☐
i. Menurut saya, matematika lebih sulit daripada pelajaran lain	☐	☐	☐	☐

13. Apa pendapatmu terhadap pernyataan–pernyataan berikut ini?

a. Matematika hanya tentang penjumlahan, pengurangan, perkalian dan pembagian.	☐	☐	☐	☐
b. Soal matematika harus dapat diselesaikan dengan cepat dalam beberapa langkah	☐	☐	☐	☐
c. Ada banyak cara menyelesaikan soal matematika	☐	☐	☐	☐
d. Dalam matematika, jawaban yang benar lebih penting daripada cara mendapatkan jawaban tersebut	☐	☐	☐	☐
e. Semua soal matematika hanya dapat diselesaikan dengan satu cara	☐	☐	☐	☐
f. Matematika hanya merupakan kumpulan aturan dan rumus	☐	☐	☐	☐

14. Apa pendapatmu terhadap pernyataan–pernyataan berikut ini?

a. Cara yang diajarkan guru di kelas adalah satu-satunya cara yang benar untuk menyelesaikan soal matematika	☐	☐	☐	☐
b. Menghafal adalah suatu hal yang paling penting dalam belajar matematika	☐	☐	☐	☐
c. Saya harus selalu mengikuti cara yang diajarkan guru dalam menyelesaikan soal matematika	☐	☐	☐	☐
d. Tujuan utama mengerjakan soal matematika adalah untuk mendapatkan jawaban yang benar	☐	☐	☐	☐
e. Saya biasanya mencoba menciptakan solusi sendiri untuk menyelesaikan soal matematika	☐	☐	☐	☐
f. Saya berharap guru saya hanya memberikan soal yang tidak asing bagi saya	☐	☐	☐	☐
g. Menurut saya, guru harus selalu menunjukkan cara menyelesaikan soal sebelum meminta siswa mengerjakan soal tersebut	☐	☐	☐	☐
h. Saya tertantang untuk mengerjakan soal matematika yang solusinya tidak terlihat jelas	☐	☐	☐	☐

Bagian 3. Tentang kelas matematika

Petunjuk:
Berilah tanda centang (✓) pada kotak yang disediakan.
Tidak ada jawaban benar atau salah. Pilihlah jawaban berdasarkan ***pendapatmu***

15. Seberapa sering kamu mengerjakan jenis-jenis pertanyaan berikut ini saat pelajaran matematika?

	Tidak pernah atau hampir tidak pernah	*Kadang-kadang*	*Selalu atau hampir selalu*
a. Soal yang mirip dengan yang dijelaskan guru di depan kelas	☐	☐	☐
c. Soal yang memerlukan pengetahuan tentang topik sebelumnya	☐	☐	☐
d. Soal yang dapat diselesaikan dengan banyak cara	☐	☐	☐
e. Soal cerita	☐	☐	☐

16. Seberapa sering kamu menggunakan sumber belajar berikut ini saat pelajaran matematika?

a. Buku cetak	☐	☐	☐
b. Lembar Kerja Siswa (LKS)	☐	☐	☐
c. Alat peraga	☐	☐	☐
d. Program komputer untuk memahami konsep matematika	☐	☐	☐

17. Seberapa sering kamu melakukan hal-hal berikut ini saat pelajaran matematika?

a. Mendengarkan guru menjelaskan cara menyelesaikan soal	☐	☐	☐
b. Menghafal rumus dan konsep	☐	☐	☐
c. Menyelesaikan soal secara berkelompok	☐	☐	☐
d. Menggunakan pengetahuan dari topik sebelumnya untuk menyelesaikan soal	☐	☐	☐
e. Menjelaskan jawaban dari soal yang saya kerjakan	☐	☐	☐
f. Menghubungkan apa yang saya pelajari dengan kehidupan sehari-hari	☐	☐	☐
g. Menentukan sendiri cara untuk menyelesaikan soal yang rumit	☐	☐	☐
h. Menyelesaikan soal yang solusinya tidak terlihat jelas	☐	☐	☐

Terima kasih atas partisipasinya dalam kuesioner ini!
Silakan mengumpulkan kuesioner yang sudah diisi!

Hanya untuk peneliti

Kode Sekolah	:	....
Kode Kelas	:	....
Kode Guru	:	....
Kode Siswa	:	....
Tanggal	:	....

Appendix F Mathematics test

Mathematics Test (English)

Mathematics Test

Instructions

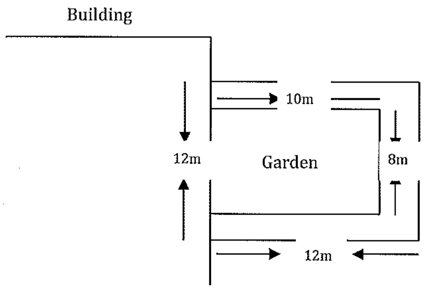
- You are given **60 minutes** to do the test
- The test consists of **8 questions**
- Please read each questions carefully before answering
- You are allowed to start with any items you find easier
- Please provide you answer in the answer sheet provided
- You **must** provide the way you get the answer for each question
- Scratch papers are provided; please let us know if you need more
- You **are not allowed** to communicate with others at all

THE UNIVERSITY *of* ADELAIDE

School of Education
The University of Adelaide
Adelaide, South Australia
www.adelaide.edu.au

Time: 60 Minutes

1. What fraction is equal to 0.6? (Please put it in the simplest fractions)
2. Simplify this equation: $2(x + y) - (2x - y)$
3. A rectangular garden which is next to a building has a path around the other three sides, as shown below.



What is the area of the path?

4. Use the formula $y = 100 - \frac{100}{1+t}$ to find the value of y when t=9.
5. Fadli from Aceh was preparing to go to Australia for 3 months as an exchange student. He needed to change some Indonesian Rupiah (IDR) to Australian Dollars (AUD).

Fadli found out that the exchange rate between Australian dollars (AUD) and Indonesian rupiahs (IDR) was:
1 AUD = 10,000 IDR

a) Fadli changed 5,000,000 Indonesian rupiahs into Australian dollars at this rate. How much money in Australian dollars did Fadli get?
b) On Returning to Aceh after 3 months, Fadli had 150 AUD left. He changed this back to Indonesian rupiah, noting that the exchange rate had changed to:
1 AUD = 9,500 IDR
How much money in Indonesian rupiah did Fadli get?
c) During these months the exchange rate has changed from 10,000 to 9,500 IDR per AUD.
Was it in Fadli's favour that the exchange rate now was 9,500 instead of 10,000 IDR, when he changed his Australian dollars back to Indonesian rupiah? Give an explanation to support your answer.

6. If $x - y = 5$ and $\frac{x}{2} = 3$
   What is the value of y?

7. The table shows a relation between x and y.

x	2	3	4	5
y	7	10	13	16

   What is the equation to express the relation between x and y. Show your work.

8. A pizzeria serves two round pizzas same thickness in different in different sizes. The smaller one has a diameter of 30 cm and cost 30,000 rupiah. The larger one has a diameter of 40 cm and costs 40,000 rupiah.
   Which Pizza is better value for money? Show your work/reasoning.

End of the test

Mathematics Test (Bahasa)

Tes Matematika

Petunjuk

- Tersedia waktu **60 menit** untuk mengerjakan **8 soal**
- Bacalah soal-soal sebelum menjawabnya
- Diperbolehkan untuk menjawab soal yang dianggap mudah terlebih dahulu
- Tulislah jawaban pada lembar jawaban yang disediakan
- Setiap jawaban **harus** disertai **cara**
- Kertas buram di sediakan, jika dibutuhkan
- Tidak diizinkan berkomunikasi dengan siswa lain selama tes berlangsung

THE UNIVERSITY *of* ADELAIDE

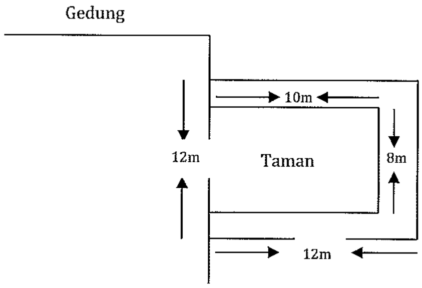
School of Education
The University of Adelaide
Adelaide, South Australia
www.adelaide.edu.au

Waktu: 60 Menit

1. Ubahlah bentuyk decimal, 0.6 ke dalam pecahan biasa paling sederhana.

2. Sederhanakanlah persamaan berikut ini: $2(x + y) - (2x - y)$

3. Sebuah taman berbentuk persegi panjang terletak tepat di sebelah gedung, taman tersebut di kelilingi oleh trotoar di ketiga sisinya (perhatikan gambar di bawah ini). Tentukanlah luas trotoar tersebut!



4. Diketahui $y = 100 - \frac{100}{1+t}$. Jika $t = 9$, tentukanlah nilai y.

5. Fadli adalah siswa yang berasal dari Aceh yang akan berangkat ke Australia untuk pertukaran pelajar selama 3 bulan. Dia harus menukarkan Rupiah (IDR) ke Dolar Australia (AUD).

   Berdasarkan informasi dari bank, nilai tukar dollar Australia dan rupiah adalah sebagai berikut.

   1 AUD= 10. 000 IDR

   a) Jika Fadli menukarkan 5.000.000 rupiah ke dollar Australia dengan nilai tukar di atas, berapa dolar uang yang Fadli dapatkan?
   b) Fadli kembali ke Aceh setelah 3 bulan. Uangnya masih tersisa 150 dolar. Dia menukarkan kembali uangnya ke rupiah. Ternyata saat dia menukar uang ke bank, nilai tukar sudah berubah menjadi:
      1 AUD = 9.500 IDR
      Berapa rupiah uang yang Fadli dapatkan?
   c) Selama 3 bulan Fadli berada di Australia, nilai tukar berubah dari 10.000 menjadi 9.500 rupiah per dolar Australia.
      Apakah Fadli mendapatkan keuntungan ketika dia menukarkan kembali uangnya ke rupiah? (nilai tukar yang sebelumnya 10.000 rupiah sekarang menjadi 9.500 rupiah per dolar Australia).

6. Jika $x - y = 5$ dan $\frac{x}{2} = 3$
   Tentukanlah nilai y?

7. Tabel di bawah ini menunjukkan hubungan antara x dan y.

x	2	3	4	5
y	7	10	13	16

   Tentukanlah persamaan yang menunjukkan hubungan antara x dan y!

8. Toko ' Pizzamania' menjual 2 jenis pizza. Kedua jenis pizza tersebut memiliki ketebalan yang sama, namun ukurannya berbeda. Pizza yang kecil berdiameter 30 cm dijual seharga 30.000 rupiah. Pizza besar berdiameter 40 cm di jual seharga 40.000 rupiah. Pizza manakah yang lebih murah dan mengapa?

Appendix G Mathematics Test and Marking Scheme Format

The mathematics test consists of eight questions adapted from *Trends in International Mathematics and Science Study (*TIMMS) and *Programme for International Student Assessment (*PISA) items. The purpose of the test is to examine students' lower order thinking skills and higher order thinking skills in mathematics, so the items have been chosen specifically to allow this assessment. There are six 'stages' of the questions involving LOT and four involving HOT.

The classification of the test items as LOT or HOT is based mainly on Bloom's revised taxonomy, in which the cognitive processes of remembering, understanding, and applying are categorised as LOT and analysing, evaluating and creating are categorised as HOT. The detailed cognitive processes of Bloom's taxonomy are presented in Table I. Table II lists the items, the content domain and the cognitive domain according to LOT and HOT.

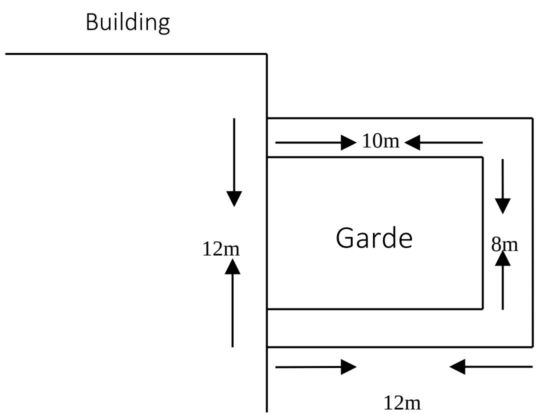
Table I

Structure of the Cognitive Process Dimension of the Revised Bloom's Taxonomy

Cognitive processes	**Description**
1. Remembering	Retrieving knowledge from long-term memory (recognising, recalling)
2. Understanding	Determining the meaning of instructional messages, including oral, written, and graphic communication (interpreting, exemplifying, classifying, summarising, inferring, comparing, and explaining)
3. Applying	Carrying out or using a familiar procedure in a given situation (executing, implementing)
4. Analysing	Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose (differentiating, organizing, attributing)
5. Evaluating	Making judgments based on criteria and standards (checking, critiquing)
6. Creating	Putting elements together to form a novel, coherent whole or make an original product (generating, planning, producing)

Table II

Items, Content and Cognitive Domains

No	Items	Content Domain	Cognitive Domain
1	What fraction is equal to 0.6? (Please put it in the simplest fractions) (TIMSS 2011)	Number (fractions and decimals)	Remembering, understanding (L)
2	Simplify this equation: $2(x + y) - (2x - y)$ (TIMSS 2007)	Algebra	Remembering, understanding (L)
3	A rectangular garden which is next to a building has a path around the other three sides, as shown below.  What is the area of the path? (TIMSS 1999)	Measurement	Analysing, evaluating (H)
4	Use the formula $y = 100 - \frac{100}{1+t}$ to find the value of y when t=9. (TIMSS 2011)	Algebra	Remembering, understanding (L)
5	Fadli from Aceh was preparing to go to Australia for 3 months as an exchange student. He needed to change some Indonesian Rupiah (IDR) to Australian Dollars (AUD). Fadli found out that the exchange rate between Australian dollars (AUD) and Indonesian rupiahs (IDR) was: 1 AUD = 10,000 IDR a) Fadli changed 5,000,000 Indonesian rupiahs into Australian dollars at this rate. How much money in Australian dollars did Fadli get? b) On Returning to Aceh after 3 months, Fadli had 150 AUD left. He changed this back to Indonesian rupiah, noting that the exchange rate had changed to: 1 AUD = 9,500 IDR How much money in Indonesian rupiah did Fadli get? c) During these months the exchange rate has changed from 10,000 to 9,500 IDR per AUD.	Algebra	Understanding (a,b) (L) Analysing, evaluating (c) (H)

No	Items	Content Domain	Cognitive Domain
	Was it in Fadli's favour that the exchange rate now was 9,500 instead of 10,000 IDR, when he changed his Australian dollars back to Indonesian rupiah? Give an explanation to support your answer. (PISA)		
6	If $x - y = 5$ and $\frac{x}{2} = 3$ What is the value of y? (TIMSS 2003)	Algebra	Understanding, applying (L)
7	The table shows a relation between x and y. x | 2 | 3 | 4 | 5 y | 7 | 10 | 13 | 16 What is the equation to express the relation between x and y. Show your work. (TIMSS 1999)	Algebra	Analysing, evaluating (H)
8	A pizzeria serves two round pizzas same thickness in different in different sizes. The smaller one has a diameter of 30 cm and cost 30,000 rupiah. The larger one has a diameter of 40 cm and costs 40,000 rupiah. Which pizza is better value for money? Show your work/reasoning. (PISA)	Geometry	Evaluating, creating (H)

Items 1, 2, 4 and 6 are categorised as lower order thinking questions, while the rest of the items are categorised as higher order thinking. Partial credit model will be employed to score students' responses. Table III presents in details the marking scheme for each item.

Table III

Mathematics Test Marking Scheme

Item	Response	Code
1	*Correct response (2 steps)* Step 1: Knowing how to convert decimal to fraction. Step 2: Knowing how to simplifying the fraction. $0.6 = \frac{6}{10} = \frac{3}{5}$ or $0.6 = \frac{60}{100} = \frac{30}{50} = \frac{3}{5}$ *Partially correct response (only 1 step correct)* $0.6 = \frac{6}{10} = \frac{3}{10}$ or $0.6 = \frac{6}{10}$ (without simplifying) *Incorrect response* None of the steps is correct.	Correct response = 2 Partially correct response =1 Incorrect response=0 Note: LOT (1-2)
2	*Correct response (2 steps)* Step 1: Knowing the hierarchical rules of simplifying. Step 2: Knowing the classification of like terms.	Correct response = 2 Partially correct response =1 Incorrect response=0

Item	Response	Code
	$2x + 2y - 2x + y$ $3y$ *Partially correct response (only 1 step correct)* $2x + 2y - 2x + y$ $4x + 4y$ *Incorrect response* None of the steps are correct or just the answer without the calculation steps.	Note:LOT (1-2)
3	*Correct response (3 steps)* Step 1: Analysing to measure the area of the path. Step 2: Knowing the formula for measuring area of square. Step 3: Knowing the formula for measuring area of rectangle. The area of garden + path (the square) $= side \times side = 12 \times 12 = 144\ m^2$ The area of garden only (the rectangle) $= length\ x\ width = 10 \times 8 = 80\ m^2$ The area of the path only $= the\ area\ of\ the\ square - the\ area\ of\ rectangle$ $= 144\ m^2 - 80\ m^2 = 64\ m^2$ *Partially correct response #1 (only 2 steps correct)* *Partially correct response #2 (only 1 step correct)* *Incorrect response* None of the steps are correct or just the answer without the calculation steps.	Correct response = 3 Partially correct response 1 =2 Partially correct response 2 =1 Incorrect response=0 Note: HOT (1-5)
4	*Correct response (2 steps)* Step 1: Knowing how to use the formula. Step 2: Order of operation. $y = 100 - \frac{100}{1+9}$ $y = 90$ *Partially correct response (only 1 step correct)* $y = 100 - \frac{100}{1+9}$ $y = \frac{0}{10} = 0$ *Incorrect response* None of the steps are correct or just the answer without the calculation steps.	Correct response = 2 Partially correct response =1 Incorrect response=0 Note: LOT (1-2)
5	*Correct response a (2 steps)* Step 1: Comprehension of exchange rates (units)	a. Correct response = 2 Incorrect response=0

Item	Response	Code
	Step 2: Computation $\frac{5{,}000{,}000}{10{,}000} = 500$, Faisal got 500 AUD *Incorrect response a* Student makes a wrong computation *Correct response b (2 steps)* Step 1: Comprehension of exchange rates (units) Step 2: Computation $150 \times 9{,}500 = 1{,}425{,}000$ Faisal got 1,425,000 rupiah *Incorrect response b* Student makes an incorrect computation. *Correct response c (2 steps)* Step 1: Analysing the change of the currency rate. Step 2: Evaluating the case to make conclusion. *Correct response c #1* 3 months ago 10,000 IDR = 1 AUD Today 9,500 IDR = 1 AUD That is today you receive less IDR for one AUD than 3 months ago. So it is not in Faisal's favour. *Correct response c #2* 1:10,000 before 1:9,500 after It was not in his favour to change the money. *Correct response c #3* No, since he paid 10,000 for $1before which is more expensive than 9,500. He lost another 5% when he changed the currency back. *Correct response c #4* No, because the value of the Indonesian currency is dropped in value. *Correct response c #5* The exchange rate was not in his favour as he received less Indonesian money at the end than at the beginning. *Correct response c #6* If the exchange rate was still 1 AUD= 10,000 IDR then he would get 1,500,000. Thus, it was not in his favour as he would have made more with the original exchange rate. *Correct response c #7* No, if the exchange rate is lower he gets less money back from Australian dollars. *Correct response c #8* No, it is his loss. Otherwise it would be 1,500,000 IDR.	Note: LOT (1-3) b. Correct response = 2 Incorrect response=0 Note: LOT (1-3) c. Correct response = 2 Partially correct response = 1 Incorrect response=0 Note: HOT (1-5)

Item	Response	Code
	Correct response c #9 Exchange rate is not in favour of Faisal because he lost 75,000 IDR. *Correct response c #10* No, he had to spend 10,000 to change \$1 but he only got 9,500 back and that is 500 IDR short. *Partially correct response* Student said 'No' but she/he gives wrong reasoning. *Incorrect response* Student said 'Yes' or said 'No' without any reasoning at all	
6	*Correct response (2 steps)* Step 1: Knowing the fraction Step 2: Knowing the concept of substitution $x - y = 5$ Equation (1) $\frac{x}{2} = 3$ Equation (2) $x = 6$ Substitute $x = 6$, to Equation (1) $x - y = 5$ $6 - y = 5$ $-y = 5 - 6$ $-y = -1$ $y = 1$ *Partially correct response* Only one step correct or both steps correct but student makes an incorrect computation. *Incorrect response* None of the steps is correct or just the answer without the calculation.	Correct response = 2 Partially correct response =1 Incorrect response=0 Note: LOT (1-3)
7	*Correct response (2 steps or more)* Step 1: Analysing the possible relation of x and y. Step 2: Evaluating the possible relation. Step 3: Concluding the general equation for x and y. Correct response #1 Using the change $3 - 2 = 1 = x$ $10 - 7 = 3 = y$ So, $y = 3x + 1$ Correct response #2 $x = 2, y = 7$ $x = 3, y = 10$ $x = 4, y = 13$ $x = 5, y = 16$ $y = x + 5$ $y = x + 7$ $y = 3x + 1$ $y = 3x + 1$	Correct response = 3 Partially correct response 1 = 2 Partially correct response 2 = 1 Incorrect response=0 Note: HOT (1-5)

Item	Response	Code

or, $y = x^2 + 3$ or, $y = x^2 + 1$

or, $y = 3x + 1$ or, $y = 3x + 1$

Correct response #3

$$+5\begin{cases}7\\2\end{cases}\overset{+3}{\underset{+1}{}} \quad +7\begin{cases}10\\3\end{cases}\overset{+3}{\underset{+1}{}} \quad +9\begin{cases}13\\4\end{cases}\overset{+3}{\underset{+1}{}} \quad +11\begin{cases}16\\5\end{cases}$$

$y = 3x + 1$

Correct response #4

$y = x$?

Closest multiple 3

$3x2 = 6$ (7)

$3x3 = 9$ (10)

$3x4 = 12$ (13)

$3x4 = 15$ (16)

In all cases, only 1 needs to be added. So, $y = 3x + 1$

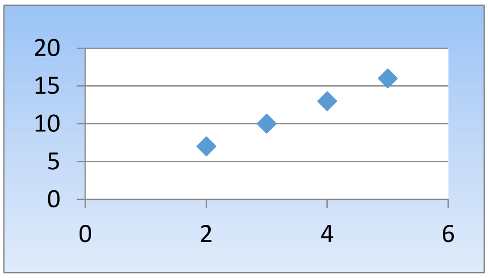
Correct response #5

Straight line

$(y - 7) = 3(x - 2)$

$y = 3x + 1$

Correct response #6



$y = 3x +$
1

Correct response #7

$$Slope = m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{16 - 13}{5 - 4} = \frac{3}{1} = 3$$

$y = mx + c$

$y = 3x + c$

$when\ y = 7, x = 2$

$7 = 3(2) + c$

$c = 1$ So, $y = 3x + 1$

Correct response #8

$y_1 = 7$, $y_2 = 10$

$x_1 = 2$, $x_2 = 3$

$y = mx + c$

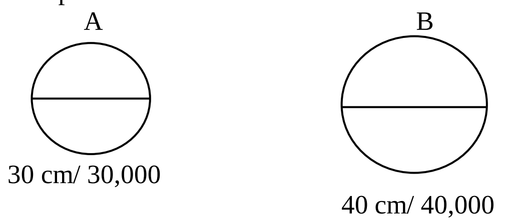
$7 = m(2) + b \rightarrow b = 7 - 2m$

$10 = m(3) + b$

$10 = 3m + 7 - 2m$

$m = 3$

$b = 7 - 2m$

Item	Response	Code
	$b = 1$ $\quad\quad\quad\quad y = 3x + 1$ *Partially correct response #1* Student shows most relevant and correct steps but she/he draw the incorrect conclusion. *Partially correct response #2* Student shows some relevant and correct steps but she/he doesn't come to any conclusion. Incorrect response. Students shows no relevant or correct steps.	
8	Correct response (3 steps) Step 1: Knowing the concept of area. Step 2: Applying the concept of area to the problem. Step 3: Doing computation to prove reasoning. Step 4: Comparing and concluding. Correct response #1 Area of smaller one: $\pi r^2 = \pi 15^2 = 225\pi$ Area of larger one: $\pi r^2 = \pi 22^2 = 400\pi$ So the larger one has better value Correct response #2  Amount of pizza is the area Area of A: $\pi r^2 = \pi 15^2 = 225\pi = 706.5$ cm^2 Area of B: $\pi r^2 = \pi 20^2 = 400\pi = 1256$ cm^2 Cost per cm^2 $A = \frac{30,000}{706.5} = 42.463$ $\quad\quad B = \frac{40000}{1256} = 31.847$ So the larger pizza has better value. Correct response #3 Area of A: $\pi r^2 = \pi 15^2 = 225\pi = 706.5$ cm^2 Area of B: $\pi r^2 = \pi 20^2 = 400\pi = 1256$ cm^2 $\frac{706.5}{1256} \times 100 = 56.25$ % $\quad \frac{30,000}{40,000} \times 100 = 75$ % The area of small pizza is only 56.25 % of the large one but the price is 75 % of the large one. Thus, by comparing to the price difference and the amount of pizza, the larger one is of better value. Partially correct response #1. Student shows relevant and correct strategies. However she/he used the wrong formula for area of circle. Partially correct response #2 Student shows relevant and correct strategies but she/he withdraw the wrong conclusion	Correct response = 4 Partially correct response #1 = 2 Partially correct response #2 = 2 Incorrect response=0 Note: HOT (1-6)

Item	Response	Code
	Incorrect response Student shows no relevant nor correct strategies	
	Total score is 24	

Appendix H Model Fit Indices

Among the fit indices, Chi square (χ^2) is the most conventional. The χ^2 works based on the null hypothesis that 'predicted covariance matrix Σ is equivalent to the observed sample covariance matrix *S*, Σ=S' (Albright & Park, 2009). If the number χ^2 is large and the null hypothesis is rejected, it indicates that the model does not fit the data well (Albright & Park, 2009). χ^2 has significant drawbacks. It is sensitive to sample size as the larger number of sample size impact on the difficulty to accept the null hypothesis (Albright & Park, 2009). Furthermore, χ^2 works with the assumption of multivariate normality which allows the possibility of rejecting the hypothesis when the data does not meet the normality assumption (Hooper et al., 2008). Consequently, the χ^2 index is rarely used as a stand-alone criteria for model fit.

Generally, the fit indices are either absolute or incremental. Absolute fit indices employ the characteristic of evaluation of 'how well a priori model reproduces the sample data' while incremental fit indices perform the evaluation of model fit by 'comparing a target model with a more restricted, nested baseline model' (Hu & Bentler, 1998, p. 426). Among the model fit indices RMSEA, AIC, RMR, SRMR, GFI, and AGFI belong to the dimension of absolute fit indices while NFI, TLI and CFI are part of the incremental fit indices dimension. RMSEA is a commonly used index as it is not sensitive to sample size but sensitive to the estimated parameters (Albright & Park, 2009). The range of acceptable value of RMSEA is less than or equal to 0.05 (Schumacker & Lomax, 2012) or less than 0.06 to 0.08 (Schermelleh-Engel, Moosbrugger, & Müller, 2003). AIC also works within the same framework as RMSEA is (Akaike, 1987). GFI and AGFI have a value range between 0 to 1, where 0 indicates no fit and 1 indicates the perfect fit. A value of more than 0.90 reflects an adequate model fit (Schumacker & Lomax, 2012). These two fit indices were originally created for the 'LISREL' SEM computer program developed by Jöreskog and Sörbom (Smith & McMillan, 2001). RMR and the SRMR are both 'the square root of the difference between the residuals of the sample covariance matrix and the hypothesised covariance model' (Hooper et al., 2008, p. 54). However, RMR is the more useful for a comparison of model fit between two model derived from the same data with the acceptable range of value decision is on the researcher (Schumacker & Lomax, 2012). SRMR was also developed by Jöreskog and Sörbom with the value less than 0.05 indicate a good model fit (Schumacker & Lomax, 2012).

Table IV.

The summary of model fit indices and acceptable range of interpretation (Schumacker & Lomax, 2012, p.76)

Model fit indices	Acceptable range	Interpretation
χ^2	Based on the χ^2 table value	Compares the obtain χ^2 value with tabled value for given *df*
GFI	0 (no fit) to 1 (perfect fit)	Value close to 0.90 or 0.95 indicates a good model fit
AGFI	0 (no fit) to 1 (perfect fit)	Value adjusted for *df* Value close to 0.90 or 0.95 indicates a good model fit
RMR	Depending on the researcher	Indicates the closeness of $\sum$ to S matrices
SRMR	< 0.05	Value less than 0.05 indicates a good model fit
RMSEA	0.05 to 0.08	Value close to 0.5 to 0.08 indicate close fit
TLI	0 (no fit) to 1 (perfect fit)	Value close to 0.90 or 0.95 indicates a good model fit
NFI	0 (no fit) to 1 (perfect fit)	Value close to 0.90 or 0.95 indicates a good model fit
PNFI	0 (no fit) to 1 (perfect fit)	Compares values in alternative models
AIC	0 (perfect fit) to positive value (poor fit)	Compares values in alternative models

Bentler and Bonnet (1980) developed NFI whose value also ranges from 0 to 1 with the cut off value for reflecting a good model fit being 0.9. CFI is another similar fit index with the value closest to 1 showing the best fit (Albright & Park, 2009). TLI (which is also known as NNFI) also ranges from 0 to 1 where 0.90 or 0.95 is considered as a good model fit (Schumacker & Lomax, 2012). While there are further model fit indices, considering the benefits and drawbacks of those described, it is recommended by Hooper, Coughlan et al (2008) to report the Chi-square, RMSEA, SRMR, CFI and one parsimony fit index, such as PNFI, as the criteria for model fit. Iacobucci (2010) suggests that CFI and TLI be used because of their power and robustness. The summary of model fit indices and the acceptable range of interpretation is presented in the table above.

Appendix I Independent *t*-test Results

Table V

School Location Differences for Teachers' Beliefs concerning Mathematics related to HOT (TBM HOT)

Levene's test for equality of variances		t-test for equality of means						
F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. error difference	95% confidence interval of the difference Lower	Upper
2.43	0.13	0.77	44.00	0.45	0.09	0.11	-0.14	0.32

Table VI

School Location Differences for Types of Questions used in Mathematics Classroom (TQE)

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
TQE_A	.02	.89	-	44.00	.03	-.91	.41	-1.75	-.08
TQE_B	1.81	.19	-.88	44.00	.38	-.30	.35	-1.00	.39
TQE_C	.15	.70	.25	44.00	.80	.09	.35	-.62	.79
TQE_D	.74	.39	-.46	44.00	.65	-.17	.38	-.94	.59
TQE_E	.75	.39	.00	44.00	1.00	.00	.43	-.86	.86
TQE_F	.93	.34	.42	44.00	.67	.13	.31	-.49	.75

Table VII

School Location Differences for Instructional Activities Approach for Students (IAS)

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
IAS_C	2.06	0.16	-0.34	44.0	.74	-.09	.26	-.61	.43
IAS_D	0.74	0.40	0.97	44.0	.34	.26	.27	-.28	.80
IAS_E	3.34	0.07	-0.86	41.9	.40	-.22	.25	-.73	.29
IAS_F	0.07	0.80	-1.75	44.0	.09	-.52	.30	-1.12	.08
IAS_G	0.48	0.49	-1.42	44.0	.16	-.35	.25	-.84	.15
IAS_H	4.82	0.03	-0.39	37.8	.70	-.13	.33	-.81	.55
IAS_I	0.36	0.55	-1.27	44.0	.21	-.52	.41	-1.35	.30
IAS_J	0.12	0.73	-1.73	44.0	.09	-.48	.28	-1.03	.08

Table VIII

School Location Differences for Teaching Resources

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Textbooks	1.75	0.19	-0.66	44.00	0.51	-0.09	0.13	-0.35	0.18
Worksheets	0.00	0.95	-0.25	44.00	0.80	-0.04	0.17	-0.39	0.31
Concrete objects	1.76	0.19	-1.08	44.00	0.29	-0.17	0.16	-0.50	0.15
Computer software	3.52	0.07	2.58	41.63	0.01	0.39	0.15	0.08	0.70

Table IX

School Location Differences for Student Liking of Mathematics

Levene's Test for Equality of Variances		t-test for Equality of Means						
F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
30.08	0.00	-2.95	1121.41	0.00	-0.08	0.03	-0.14	-0.03

Table X

Gender Differences for Student Liking of Mathematics

Levene's Test for Equality of Variances		t-test for Equality of Means						
F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
34.200	.00	2.88	962.42	.00	.09	.03	.03	.14

Table XI

Gender Differences for Valuing Mathematics

Levene's Test for Equality of Variances		t-test for Equality of Means						
F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
38.46	0.00	3.23	989.16	0.00	0.08	0.02	0.03	0.12

Table XII

School Location Differences for Student Beliefs concerning Mathematics related to LOT (SBM LOT)

Levene's Test for Equality of Variances		t-test for Equality of Means						
F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
25.60	0.00	-5.80	1129.15	0.00	-0.17	0.03	-0.23	-0.11

Table XIII

School Location Differences for Types of Questions in Mathematics Classroom

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
SFWQ_A	2.54	0.11	2.59	1133	0.01	0.08	0.03	0.02	0.14
SFWQ_B	5.82	0.02	2.64	1130.90	0.01	0.09	0.03	0.02	0.15
SFWQ_C	0.07	0.79	2.44	1133	0.01	0.08	0.03	0.02	0.15

Table XIV

School Location Differences for Types of Learning Resources

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
SFLR_C	1.99	0.16	-3.99	1133	0.00	-0.15	0.04	-0.22	-0.07
SFLR_D	210.87	0.00	9.53	1040.73	0.00	0.30	0.03	0.24	0.37

Table XV

School Location Differences for Types of Learning Resources SBM LOT

Levene's Test for Equality of Variances		t-test for Equality of Means						
F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
25.60	0.00	-5.80	1129.15	0.00	-0.17	0.03	-0.23	-0.11

References

Abbott, C., & Wilks, S. (2005). Developing an effective classroom climate for higher order thinking. In S. Wilks (Ed.), *Designing a thinking curriculum* (pp. 12-28). Australia: ACER press.

Abdul, R., Muhammad, M., Syahrullah, A., & Ikhbariaty, Q. (2014). *Teaching problem solving in mathematics learning: Reflection from PISA AND TIMSS results of the students of Indonesia*. Paper presented at the International Conference On Research, Implementation And Education Of Mathematics And Sciences 2014, Yogyakarta.

Abdullah, A. T. S., Zain, M. Z. M., Nair, S. G., Abdullah, R., & Ismail, I. (2016). PISA: Malaysia's wake up call for a more balanced approach to educational delivery and attainment. In L. M. Thien, N. A. Razak, J. Keeves, & I. G. N. Darmawan (Eds.), *What can PISA 2012 data tell us?: Performance and challenges in five participating Southeast Asian countries* (pp. 1). Rotterdam: Sense publishers.

Adams, R. J., & Khoo, S. T. (1993). Quest: The Interactive Test Analysis System. Melbourne: Australian Council for Educational Research.

Afrassa, T. M. (2005). Monitoring mathematics achievement over time a secondary analysis Of FIMS, SIMS And TIMS: a Rasch analysis. In S. Alagumalai, D. D. Curtis, & N. Hungi (Eds.), *Applied Rasch measurement: A Book of exemplars* (pp. 61-77). The Netherlands: Springer.

Aiken, L. R. (1970). Attitudes toward mathematics. *Review of Educational Research, 40*, 551-596.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*(3), 317-332.

Alagumalai, S., & Curtis, D. D. (2005). Classical Test Theory. In S. Alagumalai, D. D. Curtis, & N. Hungi (Eds.), *Applied Rasch measurement: A Book of exemplars*. The Netherlands: Springer.

Albright, J. J., & Park, H. M. (2009). Confirmatory factor analysis using Amos, LISREL, Mplus, and SAS/STAT CALIS. *The Trustees of Indiana University, 1*, 1-85.

Alexander, P., Dinsmore, D., Fox, E., Grossnickle, E., Loughlin, S., Maggioni, L., . . . Winters, F. (2011). Higher-order thinking and knowledge: Domain-general and domain-specific trends and future directions. In G. Schraw & D. H. Robinson (Eds.), *Assessment of higher order thinking skills* (pp. 47-88). Charlotte, NC: Information Age Publishers.

Anderson, L., Krathwohl, D., Airiasian, W., Cruikshank, K., Mayer, R., & Pintrich, P. (2001). A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational outcomes: Complete edition. New York: Longman.

Anderson, L., & Sosniak, L. A. (Eds.). (1994). *Bloom's taxonomy: A forty year retrospective. 93rd NSSE yearbook.* Chicago: University of Chicago Press.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573.

Areepattamannil, S. (2014). International Note: What factors are associated with reading, mathematics, and science literacy of Indian adolescents? A multilevel examination. *Journal of Adolescence, 37*(4), 367-372.

Arsaythamby, V., & Zubainur, C. M. (2014). How A Realistic Mathematics Educational Approach Affect Students' Activities In Primary Schools? *Procedia-Social and Behavioral Sciences, 159*, 309-313.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 16*(1), 74-94.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (1986). *Social foundations of thought and action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Fajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307-337). Greenwich: Information Age Publishing.

Barclay, D., Higgins, C., & Thompson, R. (1995). The partial least squares (PLS) approach to causal modeling: Personal computer adoption and use as an illustration. *Technology Studies, 2*(2), 285-309.

Barkatsas, A., & Malone, J. (2005). A typology of mathematics teachers' beliefs about teaching and learning mathematics and instructional practices. *Mathematics Education Research Journal, 17*(2), 69-90.

Barrera-Osorio, F., Garcia-Moreno, V. A., Patrinos, H. A., & Porta, E. E. (2011). Using the Oaxaca-Blinder decomposition technique to analyze learning outcomes changes over time: an application to Indonesia's results in PISA mathematics. *World Bank Policy Research Working Paper*(5584).

Battista, M. T. (1994). Teacher beliefs and the reform movement in mathematics education. *The Phi Delta Kappan, 75*(6), 462-470.

Bentler, P. M., & Bonnet, G., D. (1980). Significance test and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 9.

Bidwell, C. E., & Kasarda, J. D. (1980). Conceptualizing and measuring the effects of school and schooling. *American Journal of Education, 88*(4), 401-430.

Bigge, M. L., & Shermis, S. S. (1992). *Learning theories for teachers*. New York: HarperCollins.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning*. New York: Academic Press

Biggs, J. B., & Moore, P. J. (1993). *The process of learning*. Melbourne: Prentice Hall.

Bjork, C. (2013). Teacher training, school norms and teacher effectiveness in Indonesia. In D. Suryadarma & G. W. Jones (Eds.), *Education in Indonesia* (pp. 53-67). Singapore: Institute of Southeast Asian Studies.

Blank, R. K., & de las Alas, N. (2009). Effects of teacher professional development on gains in student achievement. *Peabody Journal of Education, 77*(4), 59-85.

Bloom, B. S., Engelhart, M., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Handbook I: Cognitive domain* (Vol. 19). New York: Longman.

Bobis, J., Anderson, J., Martin, A., & Way, J. (2011). A model for mathematics instruction to enhance student motivation and engagement. In D. Brahier & W. Speer (Eds.), *Motivation and Disposition: Pathways to Learning Mathematics* (Vol. 73, pp. 31-42). Reston, VA: National Council of Teachers of Mathematics (NCTM).

Bollen, K. A. (1989). *Structural equations with latent variables*: John Wiley & Sons.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York: Routledge.

Bonk, C. J., & Cunningham, D. J. (1998). Searching for learner-centered, constructivist, and sociocultural components of collaborative educational learning tools. In C. J. Bonk & K. S. King (Eds.), *Electronic collaborators: Learner-centered technologies for literacy, apprenticeship, and discourse* (pp. 25-50). New Jersey: Erlbaum.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht: Springer.

Bosker, R. J., & Dekkers, H. P. (1994). School differences in producing gender-related subject choices. *School Effectiveness and School Improvement, 5*(2), 178-195.

Brackett, M. A., & Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin, 29*(9), 1147-1158.

Brookhart, S. M. (2010). How to assess higher-order thinking skills in your classroom. Virginia USA: ASCD.

Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice, 16*(3), 347-363.

Bruner, J. S. (1977). *The process of education*. Cambridge: Harvard University Press.

Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Newbury Park, CA: SAGE.

Bryman, A. (2012). *Social research methods*. Oxford: UK.

Byrne, B. M. (2013). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York: Routledge.

Cahyono, Y. (2010). *Upaya mengatasi kesulitan belajar matematika melalui pendekatan pembelajaran aktif inovatif kreatif efektif dan menyenangkan (paikem) pada siswa kelas VII SMP negeri 2 Gemolong* [overcoming mathematics learning difficulty through active, innovative, creative, effecctive and joyful learning in Year 7 students in public junior high school 2 Gemolong]. (Undergraduate thesis), Universitas Muhammadiyah Surakarta.

Cardelle-Elawar, M. (1992). Effects of teaching metacognitive skills to students with low mathematics ability. *Teaching and Teacher Education, 8*(2), 109-121.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*(4), 499-531.

Caygill, R., & Eley, L. (2001). *Evidence about the effects of assessment task format on student achievement*. Paper presented at the Annual Conference of the British Educational Research Association, University of Leeds, England.

Chen, Q. (2016). A Multilevel Analysis of Singaporean Students' Mathematics Performance in PISA 2012. In L. M. Thien, N. A. Razak, J. Keeves, & I. G. N. Darmawan (Eds.), *What can PISA 2012 data tell us?: Performance and challenges in five participating Southeast Asian countries* (pp. 17-33). Rotterdam: Sense Publishers.

Cheung, K., & Keeves, J. (1990). Hierarchical linear modelling. *International Journal of Educational Research, 14*(3), 289-297.

Chin, W. W. (1998a). Commentary: Issues and opinion on structural equation modeling. *MIS Quarterly, 1*(22), 7–16.

Chin, W. W. (1998b). The partial least squares approach to structural equation modeling. *Modern Methods for Business Research, 295*(2), 295-336.

Choi, N., & Chang, M. (2011). Interplay among school climate, gender, attitude toward mathematics, and mathematics performance of middle school students. *Middle Grades Research Journal, 6*(1), 15-29.

Choi, N., & Chang, M. (2011). Interplay among School Climate, Gender, Attitude toward Mathematics, and Mathematics Performance of Middle School Students. *Middle Grades Research Journal, 6*(1), 15-28.

Chudgar, A., & Sankar, V. (2008). The relationship between teacher gender and student achievement: evidence from five Indian states. *Compare: A Journal of Comparative and International Education, 38*(5), 627-642.

Cicconi, M. (2014). Vygotsky Meets Technology: A Reinvention of Collaboration in the Early Childhood Mathematics Classroom. *Early Childhood Education Journal, 42*(1), 57-65.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources, 45*(3), 655-681.

Cohen, J. (1988). *Statistical power analysis fort he behavioral sciences*. Hillsdale, NJ: Lawrence.

Coleman, J. S. (1968). Equality of educational opportunity. *Integrated Education, 6*(5), 19-28.

Crawford, K., Gordon, S., Nicholas, J., & Prosser, M. (1998). University mathematics students' conceptions of mathematics. *Studies in Higher Education, 23*(1), 87-94.

Creemers, B. P. (1994). *The effective classroom*. London: Cassell

Creemers, B. P., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 347-366.

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

Creemers, B. P., & 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.

Creemers, B. P., Kyriakides, L., & Sammons, P. (2010). *Methodological advances in educational effectiveness research*. New York: Routledge.

Creemers, B. P., & Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement, 7*(3), 197-228.

Creemers, B. P., & Reezigt, G. J. (1999). The concept of vision in educational effectiveness theory and research. *Learning Environments Research, 2*(2), 107-135.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. California: SAGE.

Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement, 6*(4), 475-494.

Cronbach, L. J. (1947). Test “reliability”: Its meaning and determination. *Psychometrika, 12*(1), 1-16.

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

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*(3), 391-418.

Croninger, R. G., Rice, J. K., Rathbun, A., & Nishio, M. (2007). Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review, 26*(3), 312-324.

Cross, D. I. (2009). Alignment, cohesion, and change: Examining mathematics teachers' belief structures and their influence on instructional practices. *Journal of Mathematics Teacher Education*, *12*(5), 325-346.

Curtis, D. D. (2005). Comparing classical and contemporary analyses and Rasch measurement. In S. Alagumalai, D. D. Curtis, & N. Hungi (Eds.), *Applied Rasch measurement: A book of exemplars* (pp. 179-195). the Netherlands: Springer.

Da Ponte, J. P., & Chapman, O. (2006). Mathematics teachers' knowledge and practices. In A. Gutiérrez & P. Boero (Eds.), *Handbook of research on the psychology of mathematics education: Past, present and future* (pp. 461-512). Rotterdam: Sense Publisher.

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

Darmawan, I. G. N. (2003). *Implementation of information technology in local government in Bali, Indonesia*. Adelaide: Flinders University.

Darmawan, I. G. N. (2016). Assessing the Quality and Equity of Student Performance in Five Southeast Asian Countries. In L. M. Thien, N. A. Razak, J. Keeves, & I. G. N. Darmawan (Eds.), *What can PISA 2012 data tell us?: Performance and challenges in five participating Southeast Asian countries* (pp. 159-180). Rotterdam: Sense Publisher.

Darmawan, I. G. N. (2017). *Confirmatory Factor Analysis* [Power point slide].

De Vaus, D. A., & de Vaus, D. (2001). *Research design in social research*. London: SAGE.

Dechsri, P. (2016). Students' Performance in PISA and the Adequacy of Teaching and Learning. In L. M. Thien, N. A. Razak, J. Keeves, & I. G. N. Darmawan

(Eds.), *What can PISA 2012 data tell us?: Performance and challenges in five participating Southeast Asian countries* (pp. 51-62): Springer.

Demir, I., & Kilic, S. (2010). Using PISA 2003, examining the factors affecting students' mathematics achievement. *Hacettepe University Journal of Education, 38*, 44-54.

Dewi, N. R., & Kusumah, Y. (2014). Developing test of high order mathematical thinking ability in integral calculus subject. *International Journal of Education and Research, 2*(12), 101-108.

Dodeen, H., Abdelfattah, F., Shumrani, S., & Hilal, M. A. (2012). The effects of teachers' qualifications, practices, and perceptions on student achievement in TIMSS mathematics: A comparison of two countries. *International Journal of Testing, 12*(1), 61-77.

Edens, K. M. (2000). Preparing problem solvers for the 21st century through problem-based learning. *College Teaching, 48*(2), 55-60.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. United States: CRC press.

Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review, 57*(1), 72-84.

Ertmer, P. A. (2005). Teacher pedagogical beliefs: The final frontier in our quest for technology integration? *Educational Technology Research and Development, 53*(4), 25-39.

Farooq, M. S., Chaudhry, A. H., Shafiq, M., & Berhanu, G. (2011). Factors affecting students' quality of academic performance: a case of secondary school level. *Journal of Quality and Technology Management, 7*(2), 1-14.

Fauzan, A. (2002). *Applying Realistic Mathematics Education (RME) in teaching geometry in Indonesian primary schools.* (Doctoral thesis), University of Twente, Enschede. Available from University of Twente

Fennema, E. (1977). Influences of Selected Cognitive, Affective and Educational Variables on Sex-related Differences in Mathematics Learning and Studying.

In L. H. Fox, E. Fennema, & J. A. Sherman (Eds.), *Women and mathematics: Research perspectives for change*. Washington, D.C: National Institute of Education.

Fennema, E., & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization and affective factors. *American Educational Research Journal, 14*(1), 51-71.

Fernández, M., Wegerif, R., Mercer, N., & Rojas-Drummond, S. (2002). Re-conceptualizing" scaffolding" and the zone of proximal development in the context of symmetrical collaborative learning. *Journal of Classroom Interaction, 36*(2/1), 40-54.

Ferry, T. R., Fouad, N. A., & Smith, P. L. (2000). The role of family context in a social cognitive model for career-related choice behavior: A math and science perspective. *Journal of Vocational Behavior, 57*(3), 348-364.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: SAGE.

Firman, H. (2016). Diagnosing Weaknesses of Indonesian Students' Learning. In L. M. Thien, N. A. Razak, J. Keeves, & I. G. N. Darmawan (Eds.), *What can PISA 2012 data tell us?: Performance and challenges in five participating Southeast Asian countries* (pp. 63-80). Rotterdam: Sense Publisher.

Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., & Shelley, B. E. (2010). *Highlights from PISA 2009: Performance of US 15-year-old students in reading, mathematics, and science literacy in an international context. NCES 2011-004*. Washington, DC: U.S. Government Printing Office.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing research*, *18*(February), 39-50.

Forster, M. (2004). Higher order thinking skills. *Research Developments, 11*(11), 1-6.

Foy, P., Arora, A., & Stanco, G. M. (2013). *TIMSS 2011 User Guide for the International Database. Supplement 1: International Version of the TIMSS*

2011 Background and Curriculum Questionnaires. Boston: TIMSS & PIRLS International Study Center.

Franke, M. L., Kazemi, E., & Battey, D. (2007). Mathematics teaching and classroom practice. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 225-256). Charlotte, NC: Information Age Publishing.

Fujii, T. (2015). The critical role of task design in lesson study. In A. Watson & M. Ohtani (Eds.), *Task Design In Mathematics Education* (pp. 273-286). Switzerland: Springer.

Fullan, M., & Watson, N. (2011). The Slow Road to Higher Order Skills. Retrieved September 30, 2013 from: teacher.righthere.com.cn/UEditor/net/upload/file/.pdf

Fuller, B. (1987). What school factors raise achievement in the Third World? *Review of educational research, 57*(3), 255-292.

Furinghetti, F., & Morselli, F. (2009). Leading beliefs in the teaching of proof. In W. Schlöglmann & J. Maass (Eds.), *Beliefs and attitudes in mathematics education: New research results* (pp. 59-74). Rotterdam: Sense Publishers.

Gallagher, A. M., & De Lisi, R. (1994). Gender differences in Scholastic Aptitude Test: Mathematics problem solving among high-ability students. *Journal of Educational Psychology, 86*(2), 204-211.

Garofalo, J. (1989a). Beliefs and their influence on mathematical performance. *The Mathematics Teacher, 82*(7), 502-505.

Garofalo, J. (1989b). Beliefs and their influence on mathematical performance. *The Mathematics Teacher, 82*, 502-505.

Goldin, G., Rösken, B., & Törner, G. (2009). Beliefs–No longer a hidden variable in mathematical teaching and learning processes. In W. Schlöglmann & J. Maass (Eds.), *Beliefs and attitudes in mathematics education: New research results* (pp. 1-18). Rotterdam: Sense Publishers.

Gowers, W. T. (2000). The importance of mathematics. Retrieved September 30, 2012, from course.zjnu.cn/hnc/zhiliao/importance.pdf website

Grace, J. B. (2006). *Structural equation modeling and natural systems*. UK: Cambridge University Press.

Graham, J. W., & Coffman, D. L. (2012). Structural equation modeling with missing data. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 227-295). New York: Guilford press.

Graham, S. E., & Provost, L. E. (2012). Mathematics achievement gaps between suburban students and their rural and urban peers increase over time. *The Carsey Institute, Brief* (52), 1-8.

Grégoire, Y., & Fisher, R. J. (2006). The effects of relationship quality on customer retaliation. *Marketing Letters, 17*(1), 31-46.

Guay, F., Chanal, J., Ratelle, C. F., Marsh, H. W., Larose, S., & Boivin, M. (2010). Intrinsic, identified, and controlled types of motivation for school subjects in young elementary school children. *British Journal of Educational Psychology, 80*(4), 711-735.

Guion, R. M. (2002). Validity and reliability. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 57-76). Oxford: Blackwell Publishing.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science, 320*(208), 1164-1165.

Ha, L. T. M. (2016). Education Assessment System and PISA 2012 in Vietnam. In L. M. Thien, N. A. Razak, J. Keeves, & I. G. N. Darmawan (Eds.), *What can PISA 2012 data tell us? Performance and challenges in five participating Southeast Asian countries* (pp. 35-49). Rotterdam: Sense Publisher.

Hadi, S. (2002). Effective teacher professional development for the implementation of realistic mathematics education in Indonesia (Unpublished doctoral dissertation). University of Twente, Enschede, The Netherlands.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). Multivariate Data Analysis-Pearson New International Edition, New Jersey, US: Pearson.

Hair, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2013). *A primer on partial least squares structural equation modeling (PLS-SEM)*, Thousand Oaks, CA: SAGE.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, *19*(2), 139-152.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2012). Editorial-Partial Least Squares: The Better Approach to Structural Equation Modeling? *Long Range Planning*, *45*(5-6), 312-319.

Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science, 40*(3), 414-433.

Halpern, D., Wai, J., & Saw, A. (2005). A psychobiosocial model: Why females are sometimes greater than and sometimes less than males in math achievement. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics* (pp. 48-72). New York: Cambridge University Press.

Hammouri, H. (2004). Attitudinal and motivational variables related to mathematics achievement in Jordan: Findings from the Third International Mathematics and Science Study (TIMSS). *Educational Research, 46*(3), 241-257.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24*(3), 1141-1177.

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., & Hyde, J. S. (2012). Helping parents to motivate adolescents in mathematics and science an experimental test of a utility-value intervention. *Psychological Science, 23*(8), 899-906.

Harrington, D. (2008). *Confirmatory factor analysis*, Oxford: Oxford University Press.

Hart, L. C. (2002). Preservice teachers' beliefs and practice after participating in an integrated content/methods course. *School Science and Mathematics*, *102*(1), 4-14.

Hattie, J. (2003). *Teachers Make a Difference, What is the research evidence?* Paper presented at the Building Teacher Quality: What does the research tell us?, Melbourne. Retrieved May 30, 2012, from:http://research.acer.edu.au/cgi/viewcontent.cgi?article=1003&context=research_conference_2003

Hayden, M., & Martin, R. (2014). ASEAN State of Education Report 2013. Jakarta, Indonesia: Association of Southeast Asian Nations (ASEAN).

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence Based Nursing, 3*(18), 66-67.

Heise, D. R. (1975). *Causal analysis*. Oxford: John Wiley & Sons.

Hendayana, S., Asep, S., & Imansyah, H. (2010). Indonesia's issues and challenges on quality improvement of mathematics and science education. *Journal of International Cooperation in Education*, 41-51.

Hendayana, S., Supriatna, A., & Imansyah, H. (2010). Indonesia's Issues and Challenges on Quality Improvement of Mathematics and Science Education. *Journal of International Cooperation in Education*, 41-51.

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing, 20*(1), 277-319.

Henseler, J., & Sarstedt, M. (2013). Goodness-of-fit indices for partial least squares path modeling. *Computational Statistics, 28*(2), 565-580.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371-406.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23*(6), 723-744.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: guidelines for determining model fit. *Articles, 6*(1), 53-60.

Hoskins, J. (2005). Valuing and assessing higher order thinking in mathematics. In S. Wilks (Ed.), *Designing a thinking curriculum* (pp. 109-117). Australia: ACER press.

Howie, S. (2002). *English language proficiency and contextual factors influencing mathematics achievement of secondary school pupils in South Africa.* (Unpublished doctoral dissertation), University of Twente, Enschede, the Netherlands

Hox, J. J., & Roberts, J. K. (2011). Multilevel analysis: Where we were and where we are. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 1-11). New York: Routledge.

Hoyle, R. H. (2012). *Handbook of structural equation modeling*, New York: Guilford Press.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453.

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal, 20*(2), 195-204.

Hunt, A., Walton, F., Martin, S., Haigh, M., & Irving, E. (2015). Moving a school: Higher order thinking through SOLO and e-Learning: Teaching and learning research initiatives. New Zealand: Teaching and Learning Research Iniatiative.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin, 107*(2), 139-155.

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology, 20*(1), 90-98.

IEA. (2001). *TIMSS 1999 assessment TIMSS 1999 mathematics items.* Retrieved August 30, 2012, from http://timssandpirls.bc.edu/timss1999i/study.html

IEA. (2005). *TIMSS 2003 assessment TIMSS 2003 mathematics items.* Retrieved from http://timss.bc.edu/timss2003i/released.html

IEA. (2009). *TIMSS 2007 assessment TIMSS 2007 mathematics items* (pp. 92). Retrieved August 30, 2014, from http://timss.bc.edu/timss2007/items.html

IEA. (2012). Trends in international mathematics and science study 2011. Retrieved July 30, 2015 from IEA website: http://www.iea.nl/timss_2011.html

IEA. (2013). *TIMSS 2011 assessment TIMSS 2011 mathematics items* (pp. 92). Retrieved August 30, 2012, from http://timssandpirls.bc.edu/timss2011/international-released-items.html

Iryanti, P. (2010). *The expectation of implementing lesson study in mathematics education in Indonesia.* Paper presented at the the fourth APEC - Tsukuba International Conference, Tokyo, Japan.

Jähnig, C. C. (2013). *Assessing Business Knowledge of Students in German Higher Education.* Paper presented at the Jahrbuch der berufs-und wirtschaftspädagogischen Forschung 2013.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(2), 199-218.

Jaworski, B. (1994). *Investigating mathematics teaching: A constructivist enquiry*, London: Falmer Press.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., . . . Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America*, New York: Basic Books.

Johar, R., & Afrina, M. (2011). *The Teachers' Efforts to Encourage the Students' Strategies to Find the Solution of Fraction Problem in Banda Aceh.* Paper presented at the Proceeding of 24th International Congress for School Effectiveness and Improvement Cyprus.

Jupri, A., Drijvers, P., & van den Heuvel-Panhuizen, M. (2014a). Difficulties in initial algebra learning in Indonesia. *Mathematics Education Research Journal*, *26*(4), 683-710.

Jupri, A., Drijvers, P., & van den Heuvel-Panhuizen, M. (2014b). Student difficulties in solving equations from an operational and a structural perspective. *International Electronic Journal of Mathematics Education*, *9*(1-2), 39-55.

Kagan, D. M. (1992). Implication of research on teacher belief. *Educational Psychologist, 27*(1), 65-90.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2010). Identifying effective classroom practices using student achievement data. *Journal of Human Resources, 46*(3), 587-613.

Kanyongo, G. Y., Schreiber, J. B., & Brown, L. I. (2007). Factors affecting mathematics achievement among 6th graders in three sub-Saharan African countries: The use of hierarchical linear models (HLM). *African Journal of Research in Mathematics, Science and Technology Education, 11*(1), 37-46.

Kelly, D., Nord, C. W., Jenkins, F., Chan, J. Y., & Kastberg, D. (2013). Performance of US 15-Year-Old Students in Mathematics, Science, and Reading Literacy in an International Context. First Look at PISA 2012 (NCES 2014-024). Retrieved May 15, 2017, from http://nces.ed.gov/pubsearch

King, F. J., Goodson, L., & Rohani, F. (1998). Higher order thinking skills. Retrieved August 30, 2012, from http://www.cala.fsu.edu/files/higher_order_thinking_skills.pdf

Kini, T., & Podolsky, A. (2016). Does teaching experience increase teacher effectiveness: A Review of the Research. Palo Alto: CA: Learning Policy Institute.

Kirkpatrick, A. (2011). English as an Asian lingua franca and the multilingual model of ELT. *Language Teaching, 44*(2), 212-224.

Kline, R. B. (1998). Software review: Software programs for structural equation modeling: Amos, EQS, and LISREL. *Journal of psychoeducational assessment, 16*(4), 343-364.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*, New York: Guilford press.

Kloosterman, P. (2002). Beliefs about mathematics and mathematics learning in the secondary school: Measurement and implications for motivation. In G. C. Leder, E. Pehkonen, & G. Torner (Eds.), *Beliefs: A hidden variable in* *mathematics education?* (pp. 247-270). the Netherlands: Kluwer Academic Publishers.

Kolovou, A., Van den Heuvel-Panhuizen, M., & Bakker, A. (2009). Non-routine problem solving tasks in primary school mathematics textbooks-A needle in a haystack. *Mediterranean Journal for Research in Mathematics Education, 8*(2), 31-67.

Kozma, R. B. (2003). Technology and classroom practices: An international study. *Journal of Research on Technology in Education, 36*(1), 1-14.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212-218.

Kumar, M. (2011). Conceptions of mathematics to mathematics education research. *International Journal of Mathematics Trends and Technology, 2*(3), 7-10.

Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school *Education Finance and Policy*. Washington, D.C: American Institutes for Research.

Lamb, S., & Fullarton, S. (2002). Classroom and school factors affecting mathematics achievement: A comparative study of Australia and the United States using TIMSS. *Australian Journal of Eucation*, *46*(2), 154-171.

Laura, B., McMeeking, S., Orsi, R., & Cobb, R. B. (2012). Effects of a Teacher Professional Development Program on the Mathematics Achievement of Middle School Students. *Journal for Research in Mathematics Education, 43*(2), 159-181.

Lay, Y. F., Ng, K. T., & Chong, P. S. (2015). Analyzing Affective Factors Related to Eighth Grade Learners' Science and Mathematics Achievement in TIMSS 2007. *The Asia-Pacific Education Researcher, 24*(1), 103-110.

Leder, G., Pehkonen, E., & Törner, G. (2003). Setting the scene. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A hidden variable in mathematics education?* (pp. 1-10). Dordrecht: Kluwer Academic publishers.

Lee, V. E., & Zuze, T. L. (2011). School resources and academic performance in Sub-Saharan Africa. *Comparative Education Review, 55*(3), 369-397.

Leonardo, Z., & Manning, L. (2017). White historical activity theory: toward a critical understanding of white zones of proximal development. *Race Ethnicity and Education, 20*(1), 15-29.

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into Practice, 32*(3), 131-137.

Li, Y., & Lappan, G. (2014). Mathematics curriculum in school education: Advancing research and practice from an international perspective. In Y. Li & G. Lappan (Eds.), *Mathematics curriculum in school education* (pp. 3-12). Dordrecht: Springer.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychological Bulletin, 136*(6), 1123.

Liu, O. L. (2009). An investigation of factors affecting gender differences in standardized math performance: Results from US and Hong Kong 15 year olds. *International Journal of Testing, 9*(3), 215-237.

Lloyd, J. E., Walsh, J., & Yailagh, M. S. (2005). Sex differences in performance attributions, self-efficacy, and achievement in mathematics: If I'm so smart, why don't I know It? *Canadian Journal of Education, 28*(3), 384-408.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis*, New Jersey: Lawrence Erlbaum Associates Publishers.

Lohmöller, J.-B. (2013). *Latent variable path modeling with partial least squares*, Berlin: Springer Science & Business Media.

MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In R. H. Hoyle (Ed.), *Structural equations modelling concepts issues and application* (pp. 16-36). Thousand Oaks, London and New Delhi: SAGE Publication, Inc.

Mailizar, M., Alafaleq, M., & Fan, L. (2014). A historical overview of mathematics curriculum reform and development in modern Indonesia. *Teaching Innovations, 27*(3), 58-68.

Marshall, J. C., & Horton, R. M. (2011). The relationship of teacher-facilitated, inquiry-based instruction to student higher-order thinking. *School Science and Mathematics, 111*(3), 93-101.

Martínez, J. F., Stecher, B., & Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence from the ECLS. *Educational Assessment*, *14*(2), 78-102.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.

Masters, G. N., & Keeves, J. P. (1999). Issues in educational measurement. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and assessment*. Amsterdam: Pergamon.

McCoach, D. B., Goldstein, J., Behuniak, P., Reis, S. M., Black, A. C., Sullivan, E. E., & Rambo, K. (2010). Examining the unexpected: Outlier analyses of factors affecting student achievement. *Journal of Advanced Academics, 21*(3), 426-468.

McConney, A., & Perry, L. B. (2010). Socioeconomic status, self-efficacy, and mathematics achievement in Australia: a secondary analysis. *Educational Research for Policy and Practice, 9*(2), 77-91.

McCormack-Larkin, M. (1985). Ingredients of a Successful School Effectiveness Project. *Educational leadership, 42*(6), 31-37.

McCormack-Larkin, M., & Kritek, W. J. (1982). Milwaukee's Project RISE. *Educational Leadership, 40*(3), 16-21.

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575-596). New York: Macmillan.

Menard, S. (2002). *Applied logistic regression analysis*, Thousands Oaks: SAGE.

Messick, S. (1987). Validity. *ETS Research Report Series, 1987*(2), i-208.

Mevarech, Z., & Kramarski, B. (2014). Educational Research and Innovation Critical Maths for Innovative Societies: The Role of Metacognitive Pedagogies. Paris: OECD Publishing.

Miri, B., David, B. C., & Uri, Z. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in Science education, 37*(4), 353-369.

Mji, A., & Makgato, M. (2006). Factors associated with high school learners' poor performance: a spotlight on mathematics and physical science. *South African Journal of Education, 26*(2), 253-266.

Mohamed, L., & Waheed, H. (2011). Secondary students' attitude towards mathematics in a selected school of Maldives. *International Journal of Humanities and Social Science, 1*(15), 277-281.

Mohammadpour, E. (2012). Factors accounting for mathematics achievement of Singaporean eighth-graders. *The Asia-Pacific Education Researcher, 21*(3), 507-518.

Mohammadpour, E., & Ghafar, M. N. A. (2014). Mathematics Achievement as a Function of Within-and Between-School Differences. *Scandinavian Journal of Educational Research, 58*(2), 1-33.

Moss, P. L. (2012). *Teacher certification and student achievement.* (Doctor of Philosophy), The University of Southern Mississippi. Retrieved August 30, 2014, from http://aquila.usm.edu/dissertations/814

Muijs, D., & Reynolds, D. (2002). Teachers' beliefs and behaviors: What really matters? *Journal of Classroom Interaction, 37*(2), 3-15.

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., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Boston: TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Smith, T. A., Garden, R. A., Gregory, K. D., Gonzalez, E. J., . . . O'Connor, K. M. (2003). *TIMSS assessment frameworks and specifications 2003*. Boston: TIMSS & PIRLS International Study Center.

Murillo, F. J., & Román, M. (2011). School infrastructure and resources do matter: analysis of the incidence of school resources on the performance of Latin American students. *School Effectiveness and School Improvement*, *22*(1), 29-50.

Murphy, P. E. (2017). Student Approaches to Learning, Conceptions of Mathematics, and Successful Outcomes in Learning Mathematics. In L. N. Wood & Y. A. Breyer (Eds.), *Success in Higher Education* (pp. 75-93). Sydney, Australia: Springer.

Murtiana, R. (2011). *Rethinking the National Examination: Is a uniform assessment effective for diverse students in Indonesia?* Paper presented at the Indonesian Student International Conference: Thinking of Home While Away.

Nathan, M. J., & Koedinger, K. R. (2000). An investigation of teachers' beliefs of students' algebra development. *Cognition and Instruction*, *18*(2), 209-237.

Newmann, F. M. (1988). Higher order thinking in the high school curriculum. *NASSP Bulletin*, *72*(508), 58-64.

Newmann, F. M. (1990). Higher order thinking in teaching social studies: A rationale for the assessment of classroom thoughtfulness. *Journal of Curriculum Studies, 22*(1), 41-56.

Nisbet, S., & Warren, E. (2000). Primary school teachers' beliefs relating to mathematics, teaching and assessing mathematics and factors that influence these beliefs. *Mathematics Education Research Journal*, *13*(2), 34-47.

Nyikos, M., & Hashimoto, R. (1997). Constructivist theory applied to collaborative learning in teacher education: In search of ZPD. *The Modern Language Journal, 81*(4), 506-517.

OECD. (2004). First result from PISA 2003: executive summary. Retrieved August 30, 2012, from http://www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/34002454.pdf

OECD. (2006). *PISA Releseased Items-Mathematics* (pp. 136). Retrieved August 30, 2012, from www.oecd.org/pisa/38709418.pdf

OECD. (2007). Executive summary PISA 2006: Science competencies for tomorrow's world. 56. Retrieved August 30, 2012, from: http://www.oecd.org/pisa/pisaproducts/pisa2006/39725224.pdf

OECD. (2010). *PISA 2009 Results: What Students Know and Can Do* (Vol. 1). Paris: OECD Publishing.

OECD. (2012). About PISA. Retrieved September 30 2012, from http://www.oecd.org/pisa/aboutpisa/

OECD. (2013a). PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy: OECD.

OECD. (2013b). *PISA 2012 Results in Focus. What 15-year-olds know and what they can do with what they know* (Revised Edition February 2014 ed. Vol. I). Paris: OECD Publishing.

OECD. (2013c). PISA 2012 Results: What Student Can Know and Can Do - Student Performance in Mathematics, Reading and Science (Vol. I). PISA.

OECD. (2015). *Education in Indonesia: Rising to the Challenge.* Retrieved October 10, 2012, from http://dx.doi.org/10.1787/9789264230750-en

OECD. (2016a). *PISA 2015 Results (Volume I): Excellence and Equity in Education* (Vol. I). Paris: OECD Publishing.

OECD. (2016b). Results from PISA 2015: Indonesia. Retrieve from https://www.oecd.org/pisa/PISA-2015-Indonesia.pdf

Oh, Y. (2013). Factors Affecting Mathematics Achievement Gaps in Korea. *Journal of Mathematics Education at Teachers College, 3*(2), 63-67.

Oktiningrum, W., Zulkardi, Z., & Hartono, Y. (2015). *Developing PISA-like mathematics task with Indonesia natural and cultural heritage as context to promote reasoning skills of students*. Paper presented at the The Third South

East Asia Design/Development Research International Conference, Palembang.

Op't Eynde, P., & De Corte, E. (2003). *Students' Mathematics-Related Belief Systems: Design and Analysis of a Questionnaire*. Paper presented at the the symposium "The relationship between students' epistemologicalbeliefs, cognition, and learning", Chicago.

Op’t Eynde, P., De Corte, E., & Verschaffel, L. (2003). Framing students’ mathematics-related beliefs. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A hidden variable in mathematics education?* (pp. 13-37). Dordrecht: Kluwer Academic Publishers.

Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity, 10*, 265-279.

Osborne, J. W. (2000). Advantages of hierarchical linear modeling. *Practical Assessment, Research & Evaluation*, *7*(1), 1-3.

Pajares, M. F. (1992). Teachers’ beliefs and educational research: Cleaning up a messy construct. *Review of educational research*, *62*(3), 307-332.

Pajares, M. F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary educational psychology, 24*(2), 124-139.

Palardy, G. J., & Rumberger, R. W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis, 30*(2), 111-140.

Papanastasiou, C. (2000). Internal and external factors affecting achievement in mathematics: Some findings from TIMSS. *Studies in Educational Evaluation, 26*(1), 1-7.

Pearl, J. (2010). The foundations of causal inference. *Sociological Methodology, 40*(1), 75-149.

Pegg, J. (2010). *Promoting the acquisition of higher order skills and understandings in primary and secondary mathematics.* Paper presented at the Teaching

Mathematics? Make it count: What research tells us about effective teaching and learning of mathematics, Melbourne.

Perry, B., Tracey, D., & Howard, P. (1999). Head mathematics teachers' beliefs about the learning and teaching of mathematics. *Mathematics Education Research Journal, 11*(1), 39-53.

Peterson, P. L., Fennema, E., Carpenter, T. P., & Loef, M. (1989). Teacher's pedagogical content beliefs in mathematics. *Cognition and Instruction, 6*(1), 1-40.

Petty, T., Wang, C., & Harbaugh, A. P. (2013). Relationships between student, teacher, and school characteristics and mathematics achievement. *School Science and Mathematics, 113*(7), 333-344.

Philipp, R. A. (2007). Mathematics Teachers beliefs and affect. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 257-315). Charlotte, NC: Information Age Publishing.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31*(6), 459-470.

Pogrow, S. (2005). HOTS revisited: A thinking development approach to reducing the learning gap after grade 3. *Phi Delta Kappan, 87*(1), 64-75.

Polly, D., McGee, J. R., Wang, C., Lambert, R. G., Pugalee, D. K., & Johnson, S. (2013). The association between teachers' beliefs, enacted practices, and student learning in mathematics. *Mathematics Educator, 22*(2), 11-30.

Presmeg, N. (2002). Beliefs about the nature of mathematics in the bridging of everyday and school mathematical practices. In G. C. Leder, E. Pehkonen, & G. Torner (Eds.), *Beliefs: A Hidden Variable in Mathematics Education?* (pp. 293-312). The Netherlands: Kluwer Academic Publishers.

Puchner, L. D., & Taylor, A. R. (2006). Lesson study, collaboration and teacher efficacy: Stories from two school-based math lesson study groups. *Teaching and Teacher Education, 22*(7), 922-934.

Rakhmani, I., & Siregar, F. (2016). Reforming Research in Indonesia: policies and practice *Working Paper*: GDN

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests*. Copenhagen: Danish Institute for Educational Research.

Raudenbush, S. W. (1993). Hierarchical linear models and experimental design. In L. K. Edwards (Ed.), *Applied analysis of variance in behavioral science* (Vol. 137, pp. 459-496). New York: Marrel Dekker.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). HLM 6 for Windows. Lincolnwood, IL: Scientific Software International.

Razak, N. A., & Shafaei, A. (2016). The Variation in Teaching and Learning Practices and their Contribution to Mathematics Performance in PISA 2012. In L. M. Thien, N. A. Razak, J. Keeves, & I. G. N. Darmawan (Eds.), *What can PISA 2012 data tell us?: Performance and challenges in five participating Southeast Asian countries* (pp. 123-157). Rotterdam: Sense Publisher.

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research, 75*(2), 211-246.

Reyes, L. H. (1984). Affective Variables and Mathematics Education. *The Elementary School Journal, 84*, 558-581.

Reyna, V. F., & Brainerd, C. J. (2007). The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. *Learning and Individual Differences, 17*(2), 147-159.

Reynolds, D., Teddlie, C., Creemers, B. P., Scheerens, J., & Townsend, T. (2000). An introduction to school effectiveness research. In D. Reynolds & C. Teddlie (Eds.), *The international handbook of school effectiveness research* (pp. 3-25). London: Falmer Press.

Ringle, C. M., Wende, S., & Will, A. (2015). SmartPLS 3.2.6. Hamburg: SmartPLS.

Roberts, P., Priest, H., & Traynor, M. (2006). Reliability and validity in research. *Nursing standard*, *20*(44), 41-45.

Rodriguez, M. C. (2004). The role of classroom assessment in student performance on TIMSS. *Applied Measurement in Education, 17*(1), 1-24.

Rooney, C. (2012). How am I using inquiry-based learning to improve my practice and to encourage higher order thinking among my students of mathematics. *Educational Journal of Living Theories, 5*(2), 99-127.

Ross, K. N. (1978). *Sample design for educational survey research*, Oxford: Pergamon Press.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Sandoval-Hernandez, A. (2008). School effectiveness research: a review of criticisms and some proposals to address them. *Educate, Special Issue, 1*(1), 31-44.

Sangwin, C. (2017). Practice and Practise in University: What Defines Success and How Does Online Assessment Support Achieving This? In L. N. Wood & Y. A. Breyer (Eds.), *Success in Higher Education* (pp. 111-130). Sydney, Australia: Springer.

Saragih, S., & Napitupulu, E. (2015). Developing Student-Centered Learning Model to Improve High Order Mathematical Thinking Ability. *International Education Studies, 8*(6), 104-112.

Satrio, F. A. (2017, 29 March). Walah, Jumlah Doktor di Indonesia Baru 11,48 Persen [the number of PhD graduates in Indonesia is only 11,48 per cent]. Retrieved August 20, 2017, from https://www.timesindonesia.co.id/read/145163/20170329/212406/walah-jumlah-doktor-di-indonesia-baru-1148-persen/

Savery, J. R. (2006). Overview of problem-based learning: Definitions and distinctions. *The Interdisciplinary Journal of Problem-based Learning, 1*(1), 5-15.

Saxe, G. B., Gearhart, M., & Seltzer, M. (1999). Relations between classroom practices and student learning in the domain of fractions. *Cognition and Instruction, 17*(1), 1-24.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23-74.

Schoenfeld, A. H. (1989). Explorations of students' mathematical beliefs and behavior. *Journal for Research in Mathematics Education, 20*, 338-355.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-370). New York: Macmillan.

Schoenfeld, A. H. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational researcher, 31*(1), 13-25.

Schommer-Aikins, M., Duell, O. K., & Hutter, R. (2005). Epistemological beliefs, mathematical problem-solving beliefs, and academic performance of middle school students. *The elementary school journal, 105*(3), 289-304.

Schreiber, J. B. (2002). Institutional and student factors and their influence on advanced mathematics achievement. *The journal of educational research, 95*(5), 274-286.

Schulz, H., & FitzPatrick, B. (2016). Teachers' Understandings of Critical and Higher Order Thinking and What This Means for Their Teaching and Assessments. *Alberta Journal of Educational Research*, *62*(1).

Schumacker, R. E., & Lomax, R. G. (2012). *A beginner's guide to structural equation modeling*. New Jersey: Routledge Academic.

Seddon, G. M. (1978). The properties of Bloom's taxonomy of educational objectives for the cognitive domain. *Review of educational research, 48*(2), 303-323.

Sellin, N. (1990). *PLS path version 3.01 program manual*. Hamburg: Hamburg University.

Sembiring, R. K., Hadi, S., & Dolk, M. (2008). Reforming mathematics learning in Indonesian classrooms through RME. *ZDM, 40*(6), 927-939.

Sendag, S., & Ferhan Odabasi, H. (2009). Effects of an online problem based learning course on content knowledge acquisition and critical thinking skills. *Computers & Education*, *53*(1), 132-141.

Shahrill, M., Mahalle, S., Matzin, R., Hamid, M. H. S., & Mundia, L. (2013). A comparison of learning styles and study strategies used by low and high math achieving Brunei secondary school students: Implications for teaching. *International Education Studies, 6*(10), 39.

Shahrill, M., & Mundia, L. (2014). The use of low-order and higher-order questions in mathematics teaching: Video analyses case study. *Journal of Studies in Education, 4*(2), 15-34.

Sheffield, C. C. (2007). Technology and the gifted adolescent: Higher order thinking, 21st century literacy, and the digital native. *Meridian Middle School Computer Technologies Journal*, *10*(2), 1-5.

Shin, J., Lee, H., & Kim, Y. (2009). Student and school factors affecting mathematics achievement international comparisons between Korea, Japan and the USA. *School Psychology International*, *30*(5), 520-537.

Shrigley, R. L., Koballa, T. R., & Simpson, R. D. (1988). Defining attitude for science educators. *Journal of research in science teaching*, *25*(8), 659-678.

Silva, E. (2008). Measuring skills for the 21st century *Education Sector Reports* (Vol. 11). Washington, D.C: Education Sector.

Siswono, T. Y. E. (2014). Levelling students' creative thinking in solving and posing mathematical problem. *Journal on Mathematics Education, 1*(1), 17-40.

Smith, T. D., & McMillan, B. F. (2001). *A Primer of Model Fit Indices in Structural Equation Modeling*. New Orleans: South Educational Research Association.

Snijders, T. A. (1999). *Multilevel analysis*. Great Britain: Springer.

Snijders, T. A., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*, London: SAGE.

Soar, R. S., & Soar, R. M. (1976). An attempt to identify measures of teacher effectiveness from four studies. In D. R. Cruickshank & J. J. Kennedy (Eds.), *Research in Teacher Education* (Vol. 27, pp. 261-267).

Stacey, K. (2014). The PISA view of mathematical literacy in Indonesia. *Journal on Mathematics Education, 2*(02), 95-126.

Stacey, K., & Steinle, V. (2006). A case of the inapplicability of the Rasch model: Mapping conceptual learning. *Mathematics Education Research Journal, 18*(2), 77-92.

Staples, M. E., & Truxaw, M. P. (2010). The mathematics learning Discourse project: fostering higher order thinking and academic language in urban mathematics classrooms. *Journal of Urban Mathematics Education, 3*(1), 27-56.

Staub, F. C., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology, 94*(2), 344-355.

Steinmayr, R., & Spinath, B. (2008). Sex differences in school achievement: What are the roles of personality and achievement motivation? *European Journal of Personality, 22*(3), 185-209.

Stillman, G., Cheung, K., Mason, R., Sheffield, L., Sriraman, B., & Ueno, K. (2009). Challenging mathematics: Classroom practices. In P. J. Taylor & E. J. Barbeau (Eds.), Challenging Mathematics In and Beyond the ClassroomThe 16th ICMI Study (Vol. 12, pp. 243-283). New York: Springer.

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education, 17*(2), 213-226.

Stuart, C., & Thurlow, D. (2000). Making it their own: Preservice teachers' experiences, beliefs, and classroom practices. *Journal of Teacher Educatio, 51*(2), 113-121.

Suhendra. (2015). *Reforming mathematics education in Indonesia using the productive pedagogies framework.* (Doctor of Philosopy), Curtin University.

Sullivan, P., & Lilburn, P. (2002). *Good questions for math teaching: Why ask them and what to ask*. Sausalito, CA: Math Solutions.

Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching, 37*(9), 963-980.

Suryadarma, D. (2015). Gender differences in numeracy in Indonesia: evidence from a longitudinal dataset. *Education Economics, 23*(2), 180-198.

Suryadarma, D., & Jones, G. W. (Eds.). (2013). *Education in Indonesia*. Singapore: Institute of Southeast Asian Studies.

Susanti, E., & Darhim, Y. S. K. J. S. (2014). Computer-Assisted Realistic Mathematics Education for Enhancing Students' Higher-Order Thinking Skills (Experimental Study in Junior High School in Palembang, Indonesia). *Computer, 5*(18).

Sutton, A., & Soderstrom, I. (1999). Predicting elementary and secondary school achievement with school-related and demographic factors. *The Journal of Educational Research, 92*(6), 330-338.

Syah, S. M., Fitri, Y., Yani, B., Qurnati, T., & Idris, T. (2011). Action Research on the Implementation of Teaching for Active Learning in Two Elementary Madrasahs in Aceh. *Excellence in Higher Education, 2*(2), 79-89.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.), USA: Pearson Education, Inc.

Tan, M., & Saw Lan, O. (2011). Teaching mathematics and science in English in Malaysian classrooms: The impact of teacher beliefs on classroom practices and student learning. *Journal of English for Academic Purposes, 10*(1), 5-18.

Tassell, J. L., McDaniel, K., Johnson, H., Norman, A., Pankratz, R., & Tyler, R. (2012). EXPLORE Performance in Mathematics and Science: Why are Middle School Students Unprepared for Success in Mathematics and Science? *International Journal of Innovation in Science and Mathematics Education (formerly CAL-laborate International), 20*(1).

Taufik, T. (2014). Increase Student Achievement Through Realistic Mathematics Education in Materials Association in Junior High School. *Jurnal Pendidikan Sains (JPS), 1*(4), 404-412.

Taylor, B. O. (1990). *Case Studies in Effective Schools Research*. Madison, Winconsin: National centre for effective schools and development.

Teddlie, C., Reynolds, D., & Pol, S. (2000). Current topics and approaches in school effectiveness research: The contemporary field. In D. Reynolds & C. Teddlie (Eds.), *The international handbook of school effectiveness research* (pp. 26-51). London: Falmer Press.

Tenenhaus, M., Amato, S., & Vinzi, V. E. (2004). *A global goodness-of-fit index for PLS structural equation modelling.* Paper presented at the the XLII SIS scientific meeting, Padova.

Thien, L. M., & Darmawan, I. G. N. (2016). Factors Associated with Malaysian Mathematics Performance in PISA 2012. In L. M. Thien, N. A. Razak, J. Keeves, & I. G. N. Darmawan (Eds.), *What can PISA 2012 data tell us?: Performance and challenges in five participating Southeast Asian countries* (pp. 81-105). Rotterdam: Sense Publisher.

Thompson, A. G. (1984). The relationship of teachers' conceptions of mathematics and mathematics teaching to instructional practice. *Educational Studies in Mathematics, 15*(2), 105-127.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). New York: McMillan.

Thompson, T. (2008). Mathematics teachers' interpretation of higher order thinking in Bloom's taxonomy. *International Electronic Journal of Mathematics Education, 3*(2).

Thompson, T. (2011). An analysis of higher-order thinking on algebra I end-of-course tests. *International Journal for Mathematics Teaching and Learning, 2011*. Retrieved June 30, 2017, from: https://www.researchgate.net

Thomson, S., De Bortoli, L., & Buckley, S. (2013). *PISA 2012: How Australia measures up: the PISA 2012 assessment of students' mathematical, scientific and reading literacy*. Victoria, Australia: Australian Council for Educational Research

Thomson, S., De Bortoli, L., Nicholas, M., Hilman, K., & Buckley, S. (2013). PISA in brief: Highlights from the full Australian report: PISA 2012: How Australia measures up. Victoria, Australia: Australian Council for Educational Report.

Thorndike, E. L. (1917). Reading as reasoning: A study of mistakes in paragraph reading. *Journal of Educational Psychology, 8*(6), 323-332.

Tikhonova, E., & Kudinova, N. (2015). *Sophisticated thinking: higher order thinking skills*. Paper presented at the SGEM 2015 International Multidisciplinary Scientific Conferences on Social Sciences and Arts, Bulgaria.

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times* (Vol. 1), San Francisco: Jossey-Bass.

Tschannen-Moran, M., & Barr, M. (2004). Fostering student learning: The relationship of collective teacher efficacy and student achievement. *Leadership and Policy in Schools, 3*(3), 189-209.

Tsui, M. (2007). Gender and mathematics achievement in China and the United States. *Gender Issues, 24*(3), 1-11.

Turner, J. C., Christensen, A., & Meyer, D. K. (2009). Teachers' beliefs about student learning and motivation. In L. J. Saha & A. G. Dworkin (Eds.), *International handbook of research on teachers and teaching* (pp. 361-371). New York: Springer.

Uher, R., Farmer, A., Maier, W., Rietschel, M., Hauser, J., Marusic, A., . . . Schmael, C. (2008). Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychological Medicine, 38*(2), 289-300.

Vinzi, V. E., Trinchera, L., & Amato, S. (2010). PLS path modeling: from foundations to recent developments and open issues for model assessment and improvement *Handbook of partial least squares* (pp. 47-82). New York: Springer Verlag.

Vriens, M., & Melton, E. (2002). Managing missing data. *Marketing Research, 14*(3), 12-17.

Vygotsky, L. (1978). *Mind in society: The development of higher psycho- logical process*. Cambridge, MA: Harvard University Press.

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership, 41*(8), 19-27.

Wallace, M. R. (2009). Making sense of the links: Professional development, teacher practices, and student achievement. *The Teachers College Record, 111*(2), 573-596.

Wang, J., Wildman, L., & Calhoun, G. (1996). The relationships between parental influence and student achievement in seventh grade mathematics. *School Science and Mathematics, 96*(8), 395-399.

Wardhani, S., Anggraena, Y., & Marfuah. (2015). *Materi pelatihan guru implementasi kurikulum 2013 SMP pelajaran matematika [teachers' training materials of the implementation of 2013 junior high school mathematics curriculum]*. Jakarta: Badan Pengembangan Sumber Daya Manusia Pendidikan dan Kebudayaan.

Warwick, D. P., & Jatoi, H. (1994). Teacher gender and student achievement in Pakistan. *Comparative Education Review, 38*, 377-399.

Watson, J. M., Collis, K. F., Callingham, R. A., & Moritz, J. B. (1995). A model for assessing higher order thinking in statistics. *Educational Research and Evaluation, 1*(3), 247-275.

Weiss, R. E. (2003). Designing Problems to Promote Higher-Order Thinking. *New Directions for Teaching and Learning, 2003*(95), 25-31.

Wenglinsky, H. (2001). Teacher classroom practices and student performance: How schools can make a difference *Research Report-Educational Testing Service Princeton* (Vol. 2001, pp. i-37). Princeton, NJ: ETS Research Report Series.

Wenglinsky, H. (2002). The Link between Teacher Classroom Practices and Student Academic Performance. *Education Policy Analysis Archives, 10*(12), 1-30.

Widjaja, W. (2011). Towards mathematical literacy in the 21st century: perspectives from Indonesia. *Southeast Asian Mathematics Education Journal, 1*(1), 75-84.

Widjaja, Y. B., & Heck, A. (2003). How a realistic mathematics education approach and microcomputer-based laboratory worked in lessons on graphing at an Indonesian junior high school. *Journal of Science and Mathematics Education in Southeast Asia, 26*(2), 1-51.

Wijaya, A. (2015). *Context-based mathematics tasks in Indonesia: Toward better practice and achievement.* (Doctoral thesis), Utrecht University.

Wilks, S. (2005). *Designing a thinking curriculum*. Camberwell, Victoria: ACER Press.

Wold, H. (2006). Partial least squares *Encyclopedia of statistical sciences* (Vol. 9). New York: John Wiley & Sons.

Wolf, C., & Best, H. (2015). Linear Regression. In C. Wolf & H. Best (Eds.), *The SAGE handbook of regression analysis and causal inference* (pp. 55-81). London: SAGE.

Woodcock, R. W. (1999). What Can Rasch-Based Scores Convey About a Person's Test Performance. 71. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 105-127). Mahwah, NJ: Erlbaum.

World Bank. (2006). Aceh Public Expenditure Analysis – Spending for Reconstruction and Poverty Reduction. Jakarta: The World Bank.

World Bank. (2008). The Impact of the Conflict, the Tsunami and Reconstruction on Poverty In Aceh *Aceh Poverty Assessment 2008*. Jakarta: World Bank.

Wright, B., & Linacre, M. (1994). Reasonable mean-square fit values. Rasch Measurement Transactions 8: 3. *Rasch Measurement Transactions, 8*, 370. Retrieved August 30, 2012, from: www.rasch.org/rmt/rmt83b.htm

Wright, B., & Stone, M. (1979). *Best Test Design*. Chicago: Mesa Press.

Wright, B., & Stone, M. (1999). Measurement essentials (2nd ed.). Wilmington, DE: Wide Range, Inc.

Wu, M., & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Melbourne: Educational Measurement Solutions

Wu, M., Adams, R., Wilson, M., & Haldane, S. (2007). ACER ConQuest (Version 2.0). Camberwell, Australia: ACER Press

Yoon, K. S., Duncan, T., Lee, S. W.Y., Scarloss, B., & Shapley, K. L. (2007). Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement. Issues & Answers. REL 2007-No. 033. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Zakaria, E., & Musiran, N. (2010). Beliefs about the Nature of Mathematics, Mathematics Teaching and Learning Among Trainee Teachers. *The Social Sciences, 5*(4), 346-351.

Zakaria, E., Solfitri, T., Daud, Y., & Abidin, Z. Z. (2013). Effect of cooperative learning on secondary school students' mathematics achievement. *Creative Education, 4*(2), 98-100.

Zamri, S. N. A. S. (2016). Problem-Solving Skills among Malaysian Students. In L. M. Thien, N. A. Razak, J. Keeves, & I. G. N. Darmawan (Eds.), *What can PISA 2012 data tell us?: Performance and challenges in five participating Southeast Asian countries* (pp. 107-121). Rotterdam: Sense Publisher.

Zhang, Q., & Stephens, M. (2013). Utilising a construct of teacher capacity to examine national curriculum reform in mathematics. *Mathematics Education Research Journal, 25*(4), 481.

Zohar, A., Degani, A., & Vaaknin, E. (2001). Teachers' beliefs about low-achieving students and higher order thinking. *Teaching and Teacher Education, 17*(4), 469-485.

Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive?. *The Journal of the Learning Sciences, 12*(2), 145-181.

Zoller, U. (2002). Algorithmic, LOCS and HOCS (chemistry) exam questions: performance and attitudes of college students. *International Journal of Science Education, 24*(2), 185-203.