The Application of Rasch measurement theory to improve the functioning of a mathematics assessment instrument

BY

HARRISON NGIRISHI

209528282

Submitted in partial fulfilment of the requirements of the Degree of

Doctor of Philosophy (Education)

(Mathematics Education)

Department of Mathematics, Science and Computer Science Education

School of Education

University of KwaZulu-Natal

Supervisor:

Prof Sarah Bansilal

2021

DECLARATION OF ORIGINALITY

I, Harrison Ngirishi, declare that this study is my own work, submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy at the University of KwaZulu-Natal and that this thesis has never been submitted at any other university or institution for any purpose, academic or otherwise.

[redacted]

Harrison Ngirishi

209528282

DATE: ....08 November 2021

Statement by Supervisor

As the candidate's supervisor, I agree to the submission of the thesis/dissertation

[redacted]

Prof Sarah Bansilal

DATE…....08 November 2021

ACKNOWLEDGEMENTS

My sincere gratitude is extended to all the people who were directly or indirectly involved in the completion of this research study and in particular:

- My supervisor and mentor, Prof Sarah Bansilal for her motivation, guidance, professional advice and support throughout the study. Prof Bansilal's support, patience, encouragement and understanding of my shortcomings were overwhelming. The expertise she generously shared with me until the completion of this dissertation is highly appreciated.
- The National Research Foundation for providing financial support throughout the study.
- My wife, Mollen for the moral support, encouragement, words of intelligence and everlasting patience.
- My son, Tinotenda and daughters Makanaka and Rumbidzwai for their patience and understanding when I had to leave them to carry out this research.
- My parents, Mr Neva Ngirishi and Mrs Loice Ngirishi for being the sources of inspiration, strength and love that pushed me to start and finish this tiring journey. They are a special gift in my life and will always play a pivotal role in my journey of life.
- My mathematics subject adviser Mr L.O.V Buthelezi for the continued support throughout the study.
- My mathematics colleagues, Mr B. Vumase, Mr M. Sosibo, Mr Zungu and Mr S. Khanyile for the constructive debates and advice throughout this study.
- My mathematics colleagues in Umlazi District for affording me the opportunity to engage them in discussions and for supporting me throughout the study.
- The Kutlwanong Centre for Maths, Science and Technology family for their unwavering support throughout the study.
- All the learners who contributed in this work, it was their participation, which allowed this study to be completed.

List of Acronyms

DoBE	Department of Basic Education
TIMSS	Trends in Mathematics and Science Study
NSC	National Senior Certificate
PIRLS	Progress in International Reading Literacy Study
ANA	Annual National Assessments
KZN	KwaZulu-Natal
RMT	Rasch Measurement Theory
SPSS	Statistical Package for the Social Sciences
ICC	Item Characteristic Curves
CPC	Category Probability Curves
DIF	Differential Item Functioning
SES	Socio-Economic Status
SAQMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
PCK	Pedagogic Content Knowledge
FET	Further Education and Training
GET	General Education and Training
NPPPPR	National Policy Pertaining to the Programme and Promotion Requirements
CAPS	Curriculum Assessment Policy Statement
SBA	School Based Assessment
WIAT	Wechsler Individual Achievement Test
CTA	Common Tasks for Assessment
CA	Continuous Accuracy
AMESA	Association of Mathematics Education South Africa
MCF	Multiplicative Conceptual Field
HOD	Head of Department
CTT	Critical Test Theory
PCA	Principal Component Analysis
ANOVA	Analysis of Variance
NCS	National Curriculum Statement

List of Figures used in the Study

Tables used in the study

TABLE OF CONTENTS

ABSTRACT

Assessment is an integral part of the teaching and learning process. Concerns about student performance in assessments often drive the teaching and learning. In South Africa there has been numerous concerns about poor learning outcomes in mathematics and this has led to calls for all stakeholders to work together to try and find solutions. This study focuses on the assessment of mathematics with particular interest in the KZN provincial Grade 12 mathematics trial examination paper 2. The study explored the use of Rasch analysis in improving the functioning of the mathematics assessment instrument. The aim of the study was to use the Rasch analysis to report on the functioning of the test instrument in measuring proficiency in mathematics, checking on the targeting and reliability of the test instrument, explain anomalies where data did not fit the Rasch model, investigate differential item functioning (DIF), response dependency and multidimensionality. The study also sought the teachers' views about the findings of the Rasch analysis.

A sequential explanatory design was used in this study, where the Rasch analysis provided the theoretical framework for the analysis of the quantitative data. The qualitative analysis of the teachers' responses helped to get more understanding of the results of the quantitative analysis of the leaners' responses.

The study found that the assessment instrument was difficult for this particular cohort, some items displayed DIF for language and response dependency due to some teachers not applying continuous accuracy marking. The study revealed that some teachers were not applying the continuous accuracy marking process. Items which carried more than two accuracy marks, showed misfit to the Rasch model. Teachers cited not applying continuous accuracy marking due to time constraints and large number of learners in classes. Teachers blamed poor performance on learners' lack of basic understanding, adequate preparation and motivation, societal influences, poor understanding of proof type questions, allocation of many accuracy marks on one item and the language barrier. The recommendations of this study if implemented may help teachers in the teaching process and examiners in producing fair assessment instruments. The recommendations may lead to improvement of mathematics results.

CHAPTER 1 INTRODUCTION

1.1 Introduction

Stake holders in the South African education system have shown great concern with regard to the education processes (teaching and learning), with many reports and studies showing that learners have not been doing well in mathematics (Department of Basic Education (DoBE), 2011, 2012, 2015; Reddy, Lee & Ginsburg, 2009). Cassim (2006) stated that learners' performances in mathematics, has been poor, a finding that was confirmed by the 2015 Trends in Mathematics and Science Study (TIMSS) report, which showed poor performance by learners in South Africa in Mathematics (Reddy, Visser, Winnaar, Arends, Juan & Prinsloo, 2016). In an earlier study, Roux (2003) reported that high school mathematics learners in South Africa have not been performing well in mathematics. Based on these and other studies, and in the interest of improving mathematics education, it is important for all stakeholders to work together to find solutions to the difficulties and challenges faced in the learning and teaching of mathematics. This study is one of the initiatives which aims at finding solutions by trying to identify ways in which the assessment instrument for mathematics can be improved.

1.2 Background and Rationale of the Study

In any education system, assessment plays a very important role regardless of which subject is under consideration. Matters (2009) maintains that assessment has a very big part to play in the teaching and learning process, but this role can only be effective if the assessments are of good quality in order to provide evidence of the ability of learners in mathematics. According to Dunne, Long, Craig and Venter (2012), a good assessment helps to give information about learners' development as individuals and about the status of the education system. The assessment process involves the theoretical exploration of the construct mathematics when one is dealing with mathematics assessment, the operationalization of the construct in questions used to measure ability, the setting of the assessment task, the process of writing by the learners and the marking process. Many studies have been carried out to identify misconceptions and challenges that students have when dealing with different mathematics concepts and on the proficiency of mathematics teachers. However very few studies have been carried out on the assessment instruments themselves, to check on their validity and whether they can be improved and made to function better. This study applied the Rasch measurement theory (RMT), and the associated Rasch analysis, to suggest ways in which the functioning of the mathematics assessment instrument could be improved.

Monitoring of quality teaching and learning is required to maintain standards and make sure that the targets of education systems are achieved, and to design intervention strategies if needed. This is especially important in South Africa, where the curriculum has been changing since the 1994 political transformation (Archer, 2011; Lemmer, Van Wyk & Berkhout, 2010). In South Africa different assessments are used to check progress, which include provincial and national assessments such as the Grade 12 National Senior Certificate (NSC), the Provincial Trial Examinations and international studies such as the Progress in International Reading Literacy Study (PIRLS) and the Trends in Mathematics and Science Study (TIMSS) (Visser, Juan & Feza, 2015). Annual National Assessments (ANAs) were used to monitor achievement in each grade (Grades 1 to 9) of the South African school population (Kanjee & Moloi, 2016). All of these assessments have revealed poor performance by South African learners, with the 2015 TIMSS report showing that South Africa was at position 38 out of 39 countries for mathematics at Grade 9 level and position 47 out of 48 countries for mathematics at Grade 5 level (Reddy et al., 2016). The 2017 diagnostic report for the NSC examinations showed that the mathematics national pass rate was 51.9% for the 2017 group, a slight improvement from the pass rate of 51.1% for the 2016 group (DoBE, 2018).

The 2017 mathematics diagnostic report identified sections of the curriculum that needed urgent focus to enhance performance, such as strengthening learners' understanding in trigonometry, and exposing learners to complex and problem-solving type questions and questions that require application of concepts learnt in all topics in the mathematics curriculum (DoBE, 2018). For the Grade 12 learners one intervention that can enhance performance at the NSC examinations is writing trial examinations that are of good quality standards and fair to the learners.

Provincial trial examinations are very important in preparing Grade 12 learners for the final NSC assessments. Each province sets its own trial examinations for their learners. The KwaZulu-Natal (KZN) trial examinations are very important assessments, which cover the same content as the final examinations. The trial examinations are used by schools to check the readiness of their learners for the NSC examinations, to identify areas of strengths and weaknesses and to plan for interventions as they prepare their learners for the final examinations. As a Grade 12 mathematics teacher, I use the KZN mathematics trial examinations to prepare my learners for their NSC examinations. The results of the trial examinations give me guidance on what I need to do to help my learners improve their proficiency and thus influence me in my teaching in the classroom before they write the NSC examinations. It is therefore undoubtedly important that the assessment and the scoring in the trial examinations are fair and valid for all students. In mathematics, the learners write Paper 1 and Paper 2, and my experience as a mathematics teacher has shown that learners struggle more with Paper 2 than Paper 1.

I decided to look at the assessment tool, to get more insight into how well it was functioning, and to examine ways in which the Rasch analysis can help to improve the item functioning. The results of the Rasch analysis can alert us to where assessment instruments are working according to good assessment principles and where there might be some problems. Furthermore, it was important to hear from teachers, their own views about what their challenges were in teaching and assessing, as well as about some of the findings of the Rasch analysis.

The rationale of the study is to contribute to the body of knowledge in mathematics education, specifically assessment. There have been limited studies, carried on the improvement of assessment instruments in mathematics. Many studies have focused on the learners' performance, factors affecting performance, challenges, misconceptions faced by learners in specific sections of mathematics and pedagogical content knowledge of mathematics teachers (Bansilal, 2015; Bansilal, Brijlal & Mkhwanazi, 2014; Chigonga, 2016; De Villiers, 2012; Dune, Long, Craig & Venter, 2012; Jacobs, Mhakure, Fray, Holtman & Julie, 2014, Long, Dunne & Craig, 2010, Luneta, 2015). Studies on assessment instruments include those on the language and readability of assessment instruments (O'Conor, 2009; Prins, 1998; Thompson, Johnstone & Thurlow, 2002; Dempster & Reddy, 2007; Khan, 2012). Other studies of assessment instruments include that by Long, Bansilal and Debba (2014) who investigated a mathematical literacy assessment instrument using the Rasch analysis. Mudaly (2006) looked at how marking matric mathematics scripts should be used as a time for teaching too. It seems that research in the area of improving the assessment instrument itself is limited. This study will add value for the improvement of mathematics assessment instruments.

The RMT has been used in the education sector South Africa in many mathematics studies (Moduka, Long & Machaba, 2019; Long et al., 2014; Bansilal, 2014; Long et al., 2010; Dunne, Long, Craig & Venter, 2012; Stols, Long & Dunne, 2015). Most of these studies focused on other areas of mathematics, with only one done on a mathematical literacy assessment instrument by Long et al., (2014). This study is therefore very relevant and useful as it will explore the use of the Rasch analysis on a mathematics assessment instrument and make recommendations to a number of stakeholders.

1.3 Research aims

The study was carried out to explore the use of the RMT in improving the functioning of a mathematics assessment instrument. The Rasch model was applied on a KZN mathematics Grade12 trial examination Paper 2 for 2017 to investigate the functioning of items, allocation of marks, the grading of proficiency along a continuum, the results of the original scoring rubric and the revisions, together with the educational considerations. The study also sought teachers' views about the findings of the Rasch analysis so as to get a deeper understanding of the phenomenon.

The following research questions will be answered in this study:

1. What are some trends in performance of learners in the Grade 12 mathematics assessment?
2. What does a Rasch analysis reveal about the targeting and functioning of the instrument as a whole?
3. How can the Rasch analysis be used to improve the functioning of the mathematics assessment instrument?
    a. To what extent are the items functioning as expected?
    b. How can the use of the Rasch analysis contribute to the improvement of the scoring rubric?
    c. To what extent do the items display differential item functioning (DIF), multidimensionality and item dependency?
4. What are the teachers' views about some of the findings of the Rasch analysis?

1.4 A brief description of the methodology

The study used a mixed-method approach, where a sequential exploratory design was used in quantitative data collection and analysis, and the results then informed the gathering and analysis of qualitative data and analysis in the second phase. The quantitative data were generated from the learner' scripts and were made up of the scores for each of the items in the mathematics trial examination Paper 2. Six hundred and four learners' scripts were collected from six Umlazi District schools in KZN. The qualitative data were generated from questionnaires which were completed by teachers from the participating schools. Note that the questionnaires were given to fourteen teachers but only seven teachers completed them.

The initial data analysis used Excel and the Statistical Package for the Social Sciences (SPSS) to analyse the data and get an overall picture of the learners' performances and at individual schools in different mathematics sections and for different questions. The next quantitative data analysis used the Rasch analysis to determine the item functioning, ordering of persons and items, and the functioning of the scoring rubric, and to check the differential item functioning, item dependency and multidimensionality. The qualitative data analysis of the teachers' responses was done by coding the teachers' responses and grouping those which were similar. Written consent was sought from all stakeholders, learners, teachers, principals, school governing bodies, the Department of Education and the university. Questionnaires were restricted to teachers only the views of the learners were not collected. By the time the quantitative data collection and analysis was finished to inform the qualitative process, the Grade 12 learners were already done with their final examinations and it was very difficult to trace them and make them fill in the questionnaires. Of the participating schools, one was a quintile 5 school, three were quintile 4 schools, one was a quintile 2 school and one was an independent school, which is not a fair representation of all the schools in South Africa.

It is hoped that this study will add to the knowledge of assessing mathematics learners, especially in designing the assessment tool and allocation of marks. The study is also expected to make a contribution on how teachers should carry out the marking process.

1.5. Overview of the study

Chapter 1: Introduction

In this chapter, the background of the study is presented with a discussion of the motivation of the study. The chapter also explains the research approach to the study, introduces the research question and provides a brief scope of the study.

Chapter 2: Literature review

In this chapter the literature relevant to the study is presented as a way of providing a strong theoretical basis for the study. Teaching, learning and assessment of mathematics and the applications of the Rasch measurement are covered in the literature provided in this chapter.

Chapter 3: Methodology

This chapter discusses in detail the context of the study, the research purpose and research questions, participants and research design. The quantitative and qualitative analysis of data are also discussed and lastly the validity issues are discussed.

Chapter 4: Initial data analysis

This chapter presents the initial data analysis using the SPSS. It presents average scores for all schools and compares them with the average percentage score for individual schools. The average percentages were calculated for different sections, for different questions and for every sub-question (item) for the individual schools. Lastly, an analysis of zeros and blank spaces is presented.

Chapter 5: Item functioning

The chapter presents the results of the Rasch analysis by discussing summary statistics and ordering of items. The item fit is also discussed by presenting and comparing Item characteristic curves (ICC), category probability curves (CPC) and fit residuals.

Chapter 6: Rescoring of items

This chapter presents the rescoring and analysis of items which showed mis-fit to the model in Chapter Five, starting with the items with fit residuals which are not within the acceptable range followed by items with disordered thresholds.

Chapter 7: Differential item functioning (DIF), multidimensionality and response dependency

This chapter presents the DIF findings from the Rasch analysis followed by a discussion on the items which showed response dependency. A discussion on multidimensionality is also presented.

Chapter 8: Qualitative analysis of the teachers' responses

The chapter presents the teachers' responses to the questionnaire about their views on the findings of the Rasch analysis.

Chapter 9: Summary, conclusion, recommendations and limitations of the study

This chapter presents a summary of the study by providing answers to the research questions of the study. The conclusion, recommendations and limitations of the study are provided.

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

There are many challenges and problems in mathematics education in South African schools as evidenced by poor performance of learners in national examinations such as the NSC examination and the ANA, as well as in international assessments like the TIMSS (Fleisch, 2008; DoBE, 2012; Reddy, 2006). However, the challenges do not start and end with mathematics results at the NSC level, but involve all of the stakeholders, including parents, learners, teachers, principals, education officials, curriculum planners and examiners. In trying to contribute to an improvement in the learner outcomes in mathematics, it is important to consider what research has identified about challenges in the learning, teaching and assessment of mathematics. In this chapter I first provide a synopsis of research on the learning of mathematics, by considering affective and socio-economic issues before focusing on particular difficulties in the various mathematics strands. The second section focuses on the teaching of mathematics and this is followed by a discussion of assessment issues in mathematics.

2.2 Research on difficulties in learning mathematics

Mathematics is seen as a challenging subject by many, and there has been much research across the world as well as in South Africa to try to understand the reasons why the subject is experienced as being difficult. In this section, I first briefly look at some factors that were identified in research as affecting mathematics achievement. I look at research about affective issues of attitude and motivation before discussing socio-economic issues. I then focus on specific difficulties related to the strands of geometry, trigonometry and statistics which are the focus of the assessment instrument in this study

2.2.1 Confidence and attitudes towards mathematics

Attitude is one of the factors that influence learners' performance and can be defined as different types of feelings towards mathematics such as love, hate or interest, and the perception of the usefulness of mathematics in one's life (McLeod, 1994; Pajares, 2002). Attitudes are moderate in duration, intensity and stability, and have emotional content (Pajares, 2002; Hannula, 2002). In mathematics, attitude can be referred to as the learners' responses to mathematics in a positive or negative way in class. For McLeod (1992), when learners continuously pass or fail a subject, the feeling that develops as a result is referred to as attitude. For McLeod (1992), attitude develops due to negative or positive experiences. A positive attitude develops from positive experiences in mathematics. It is important for teachers to help learners improve their results so that they develop positive attitudes towards mathematics (Ma & Xu, 2004). Learners with positive attitudes about mathematics and science achieved better average scores even when other factors like gender and socio-economic status were considered as revealed by the TIMSS 2011 study (Juan, Reddy, Zuze, Namome & Hannan, 2016).

One study focusing on disaffection in mathematics classrooms was conducted in three Year 9 mathematics classrooms in England (Nardi & Steward, 2003). They found that students resented mathematics learning as a rote-learning activity that required them to use formulas that cannot be manipulated. The study found that memorisation resulted in correct answers and satisfaction for the teacher, but this approach offered limited intellectual appeal to the students. The findings revealed that reducing mathematical learning to use of formulas and rules is limited to simple mathematical calculations and they cannot be applied to questions that require higher-order thinking. Learners will end up disliking the subject as they cannot make sense of the concepts or provide rational reasons as to why the steps were carried out.

For many learners, a negative attitude can result in mathematics anxiety which is defined as the discomfort one has, together with feelings of tension and worry that affect their ability to carry out mathematics operations and do mathematics tasks in everyday life situations and the school environment (Meece, Wigfield & Eccles, 1990; Bessant, 1995; Ma, 1999; Tobias, 1990). Mathematics anxiety comes from worry and emotions which result in one not remembering basic mathematics operations and losing self-confidence (Bessant, 1995; Tobias, 1990). Mathematics anxiety may result in stress during tests, very low self-belief and lack of interest towards mathematics (Bessant, 1995). In his study Tapia (2004) noted that mathematics anxiety is constantly related to low marks in mathematics at all ages, but that learners who have little or no fear of mathematics get high marks. According to Ashcraft and Kirk (2001), learners who are highly anxious in mathematics will avoid getting into positions where they have to perform mathematical problems. Mathematics avoidance tends to create learners who have low mathematics ability and learners who do not practice or do mathematics tasks, resulting in learners who have given up hope of achieving success in mathematics (Ashcraft & Kirk, 2001). Mathematical anxiety has an emotional impact on learners who tend to become more anxious when they face tests, assessments or examinations (Ma, 1999). Hlalele (2012) found that mathematics anxiety has adverse effect on the achievement, motivation and confidence of rural high school learners in South Africa. Learners experience mathematics anxiety at different levels in academic settings and teachers must observe its prevalence and implement strategies towards the alleviation of its effects (Hlalele, 2012). Mutodi and Ngirande (2014) found that there were high mathematics anxiety levels among mathematics students, especially among the female students at a university in South Africa. Mathematics anxiety is one psychological factor that affects students' achievements and their general practices (Mutodi & Ngirande, 2014).

Although many learners have mathematics anxiety which affects their performance negatively, others may have strong motivation which helps them to succeed at mathematics. According to Ames (1992) motivation exists as a result of one's desired results and what a person views as important, and it determines whether or not one will engage with a given task. Academic motivation in the academic settings has two types, intrinsic motivation and extrinsic motivation. Vansteenkiste, Lens and Deci (2006) defined extrinsic motivation as the desire to engage in a task for the benefit of something that is different from the task itself. With extrinsic motivation an individual engages in an activity because it leads to unrelated outcomes (Deci, 1972). A learner does not participate in an activity for the pleasure of the activity, as a result there is no genuine desire that drives learners to engage with the activity in extrinsic motivation (Vansteenkiste, et al., 2006). Intrinsic motivation is characterised by high autonomy and a sense of control (Deci & Ryan, 1985). Academic intrinsic motivation is the urge to learn and participate in learning activities because they are interesting (Deci, 1972). Learners with intrinsic motivation do tasks for the sake of enjoyment and it gives them value among other learners (Middleton, 1995). Muller and Louw (2004) found that most students at a South African university exhibited intrinsic motivation while few were extrinsically motivated. When academic requirements are transparent, motivational processes are promoted and maintained during the duration of the study (Muller & Louw, 2004). Yiga, Khoarai, Khosana, Lesupi, Mduli, Shadwell, Botes and Joubert (2019) found that teacher style, peers and academically supportive home environment were the main factors influencing academic motivation in a South African state school.

Confidence is another factor which can influence learners' achievement in mathematics positively. Research shows that the beliefs that learners have about mathematics and about themselves as mathematics learners influence their learning and their own success in mathematics (Hannula, Maijala & Pehkonen, 2004). Learners' self-efficacy can be influenced by their teachers' personal beliefs or self-efficacy (Pajares & Schunk, 2002). According to Gibbons and Borders (2010), teacher efficacy is the self-confidence that the teachers have in their capabilities to design learning strategies and carry out planned strategies for maximum learner achievement. Teachers with high self-efficacy create learners who are more successful, while teachers with low instructional efficacy can undermine learners' cognitive development and self-efficacy (Pajares & Schunk, 2002). The teachers' statements pertaining to the value of a task in the classroom influence self-efficacy and motivation (Pajares & Schunk, 2001). A study by Schunk (2001) on learning strategies found that cognitive modelling enhanced achievement for learners who scored below grade level in mathematics. Schunk (2001) further found that when learners experiencing challenges with subtraction were exposed to mastery models, they performed well. When teachers model self-efficacious behaviour by confirming the value of tasks in their daily interactions with their learners, this can yield increased levels of achievement if followed by self-monitoring, as opposed to modelling without monitoring (Schunk, 2001).

2.2.2 Socio-economic factors and mathematics achievement

Research indicates that many of the reasons for low outcomes in mathematics can be linked to inequitable access to effective learning opportunities by learners from poor socio-economic backgrounds (Reddy et al., 2016). Academic achievement has been shown to be influenced by the socio-economic status (SES) in South Africa, where the SES consists of factors such as education levels of parents, home resources, availability of books at home, the educational goals that a student sets for them self and the home environment (Zuze et al., 2017). Learners with better educated and working parents tend to perform better academically (Branson, Lam & Zuze, 2012; Case & Deaton, 1999). When socio-economic resources (education levels and income of parents, books, computer devices, stimulating toys, interactive games, internet access) are abundant at home, children will have more material resources to support schooling as compared to high poverty homes, where children have less help with school work (Zuze et al., 2017). The private resources provided by parents for the education of their children is still a big factor which continues to separate learners in South Africa (Zuze et al., 2017). According to the TIMSS 2015, children growing up in poorly resourced surroundings do not perceive the role of education as meaningful to their futures, even though their parents will be trying to give them all the support they can afford (Zuze et al., 2017). Parents and other guardians have an important job to do in the education and emotional development of their children. Parents can help to make sure that what was done in the classroom is revised through homework and other activities (Sebastian, Moon & Cunningham, 2017; Wilder, 2014). Parents can help in the education of their children by engaging with teachers, encouraging their children to have love for and to do their school work, providing children with enough time to do homework and checking the homework on a regular basis (Zuze, Reddy, Visser, Winnaar & Govender, 2017).

The TIMSS 2015 constructed an asset-based index of SES using principal components analysis, to assess the significance of SES in educational achievement using a similar one to that used by Taylor and Yu (2009). Five learner SES quintiles were created based on the presence or absence of 16 assets in a learner's home, with quintile 1 identifying the lowest SES and quintile 5 the highest (Zuze et al., 2017). Two main groups were created, schools that do not pay fees as the first group and schools that pay school fees and independent schools as the second group. In no-fee schools, the majority of the learners were in the first three SES quintiles, while in schools that were paying fees and independent schools, the majority of learners were in quintiles 4 and 5. A very small percentage (8%) of learners in public schools that were paying fees and private schools, were in SES quintiles 1,2 and 3 compared 30% of the learners in schools that were not paying fees (Zuze et al., 2017).

I will now try to gain a deeper understanding of the issues related to this study, by focusing on research about difficulties related to learning specific mathematics strands.

2.2.3 Difficulties in the learning of mathematics concepts

This section looks at what the literature says about some of the issues involved in learning of mathematics with emphasis on the strands of geometry, trigonometry and statistics, which form the central focus of this study.

Statistics is the part of mathematics that deals with collection, sorting, representation, interpretation and reporting of data (Montague-Smith & Price, 2012). Mvududu and Kanyongo (2011) reported that statistics in schools was aimed at raising statistical awareness amongst learners, so that they are able to respond to real world issues by using the experience that comes with exposure to statistical models and concepts. Ijeh and Onwu (2013) are of the opinion that learning of statistics needs different skills from those required in other areas of mathematics. The authors suggest that one of the problems may be that learners are not given opportunities by their teachers to work with experiments and simulations that will allow them to gain a deeper understanding of statistics

Teachers must have appropriate and relevant understanding of the subject in order to design learning strategies that promotes understanding (Groth, 2007). A study by Umugiraneza, Bansilal and North (2017) exploring how teachers teach mathematics and statistics in KZN schools, showed that teachers are not as confident in teaching statistics topics as they are with other more general mathematics topics.

The DoBE diagnostic report for the Grade 12 NSC examinations noted a wide range of learning challenges associated with statistical concepts (DoBE, 2020). Students displayed many misconceptions related to the concept of regression. When asked to use the least squares regression line or to predict a given variable students struggled to identify the independent and dependent variables. They displayed similar problems in trying to plot the graph of the least squares regression line (DoBE, 2020). In fact, many students resorted to guessing the answers. When asked to determine the least squares regression line in the 2017 examinations, students sometimes interchanged the values of A and B in the linear equation $y = A + Bx$, showing a fundamental misunderstanding of the linear regression relationship between the variables. It was reported that some learners cannot differentiate between reading off from the graph of the least squares regression line and using the equation to find or predict the missing values (DoBE, 2018). Many learners did not understand the impact of excluding the outlier when determining the equation of the least squares regression line (DoBE, 2018).

Learners also displayed misconceptions about cumulative frequency. When working with the frequency polygon, some learners gave the frequency corresponding to a class interval, when asked to calculate the cumulative frequency. This error shows that some learners are not able to differentiate between frequency and cumulative frequency (DoBE, 2020). Many learners struggled with basic applications. When given a frequency table, some learners could not identify the modal class. When asked to plot the ogive graph, some learners plotted the cumulative frequency against the lower limit or the midpoint of the class interval while others did not ground the ogive and some used a ruler to join the points instead of drawing smooth curves (DoBE, 2020).

Edwards, Ogun-Koca and Barr (2017) found that students had difficulties in reading and interpreting different types of data representations such as box plots. In most cases students are able to construct boxplots but are not able to interpret them correctly, failing to comment on the spread of the data using the box plots (Edwards et al., 2017).

Cooper and Shore (2008) carried out a study to identify and discuss challenges students face in making judgements of the measures of central tendency and measures of dispersion when data are represented graphically and revealed that learners had notions of variability that are tenuous. The majority of students determined variability by using the heights of the bars of the histograms instead of using the data values. The student had misconceptions about measures of dispersion from the histograms, as they measured the range of the data by subtracting the highest frequency from the lowest frequency (variability in frequency) instead of finding the difference between the maximum and minimum values. The study revealed that students may answer basic questions on histograms correctly without an understanding of how the spread of the data links the frequencies (heights of the bars) with the values on the horizontal axis. The study showed that students had challenges with computing measures of centre from histograms as they were unable to calculate the median of the data. Learners confused frequencies with data values.

I now discuss the challenges learners face in geometry. Geometry is treated as a very difficult section by many high school learners and teachers (Luneta, 2015) and Euclidean geometry was identified as the most challenging (Siyepu & Mtonjeni, 2014). Research has shown that Further Education and Training (FET) learners in South Africa find geometry extremely difficult and lack conceptual understanding, which is evident when they write examinations (Oberdorf & Taylor-Cox, 1999; Bowie, 2009; Roux, 2003; Van der Sandt, 2007). For Confrey (1990), students' difficulties and errors are a result of a lack of understanding of the strategies used by the teachers which in most cases are at a higher level than that which the learners are working at. When teachers give instructions at high geometric levels and learners are working at lower levels, misconceptions arise and concepts are not understood fully (Lim, 2011; Luneta, 2015).

Luneta (2015) studied students' misconceptions by analysing the Grade 12 final examination questions in geometry and found that the majority of the students could not solve the problem involving finding the equation of a straight line (analytical geometry) despite the fact that graphs are taught in Grade 9. The study revealed that most students were working at the second level or first level of the Van Hiele levels of geometry thinking although most of the questions required students who are working at the second level or above. Students in Grade 12 are expected to operate at levels 3 and 4 of the Van Hiele levels. The study revealed that most students were unable to answer questions on basic geometry, showing a lack of conceptual understanding. The study also revealed that when students do not have conceptual understanding they leave blank spaces during examinations (Luneta, 2015).

In their study Naidoo and Kapofu (2020) revealed that learners perceived Euclidean geometry to be difficult and confusing, with findings pointing to learners having challenges with determining angles and giving reasons for the answers. The study revealed that learners found it easier to work with geometric questions which were numeric in nature or which did not require them to give reasons to justify how they arrived at their solutions. Geometry becomes a problem only when the activity required them to provide reasons and justifications for their solutions (Naidoo & Kapofu, 2020). The learners in the study revealed that geometry became challenging when they had to remember and link different proofs and carry out proofs involving different statements and theorems. The study revealed that learners found analytical geometry questions less challenging, these requiring application of the concepts learnt and in most cases direct substitution into formulas provided on the formula sheet.

Ngirishi and Bansilal (2019) revealed that a third of Grade 10 and 11 learners were reasoning at level 1 of the Van Hiele' levels of geometry thinking while less than 40% were reasoning at the informal deduction level despite them having spent at least 10 years working with geometric figures in school. The study revealed that learners did better in geometry questions that had diagrammatic representations than those without, stressing the importance of pictures, images and diagrams in the learning of geometry. The study also revealed that when learners were given properties and definitions in the natural language, most were unable to relate them to the iconic representations. Learners struggled to work with properties of figures when diagrams were not provided. The findings also included the fact that learners struggled with proof type questions, and questions involving class inclusion.

The DoBE diagnostic report for the Grade 12 NSC examinations noted many challenges and misconceptions that learners displayed when working with Euclidean geometry questions (DoBE, 2018, 2020). When asked to find the size of angles, many learners gave incorrect or incomplete reasons or named the angles incorrectly. Many learners wrote a number of correct statements and reasons which were irrelevant and did not lead to solving the problems. When asked to prove that a given quadrilateral is cyclic, learners made several correct statements but that could not help to prove that the given quadrilateral is cyclic (DoBE, 2018). The reason for the quadrilateral being cyclic was either missing or the theorem was given instead of the converse. Confusing the theorem and its converse theorem was also noted when learners were asked to prove that a line is a tangent to a circle. After showing that the angle between a line and a chord is equal to the angle in the alternate segment, many learners then wrote that the line is a tangent to the circle because of the tangent-chord theorem instead of giving the converse as the reason (DoBE, 2018). A similar problem was displayed in the 2019 NSC examinations diagnostic report, which noted that learners were not able to tell the difference between the theorem and its converse and that these were often used interchangeably in many questions (DoBE, 2020).

Many learners assumed information that was not given on diagrams or that could not be proved, for example assuming that a given quadrilateral is a cyclic quadrilateral without being told or without proving it, or assuming that lines are parallel and that lines are perpendicular (DoBE, 2018, 2020). Learners ignored the fact that diagrams in the Grade 12 NSC mathematics examinations are not drawn to scale (stated under the instructions and information section) and so their assumptions would lead to incorrect answers.

It was also reported that learners could not differentiate between alternate and corresponding angles, and were not able to link correct angles at the centre of a circle to correct angles at the circumference of the circle (DoBE, 2020). When given parallel lines on a circle and asked to find the size of angles with reasons, many learners gave the wrong reasons for equal angles, in many instances giving those for alternating angles instead of corresponding angles (DoBE, 2018). Other challenges reported included learners not being able to differentiate between a parallelogram and a cyclic quadrilateral, including using the theorem of the exterior angle of a cyclic quadrilateral incorrectly.

When asked to find the ratio of the sides of triangles, many learners used incorrect ratios when applying the theorem that states that a line joining two points on two sides of a triangle and parallel to the third side cuts the two sides so as to divide them into equal proportions (DoBE, 2018). A number of learners confused the ratio of the sides with the actual sides of the triangles, equating the sides instead of equating the ratios of the sides.

A number of challenges and misconceptions associated with Analytical Geometry were reported by the DoBE diagnostic reports for the Grade 12 NSC examinations (DoBE, 2017, 2018, 2020). Learners displayed a number of misconceptions when asked to determine the gradient of a line under different circumstances. Learners were unable to use the gradient formula correctly, using the formula $m = \frac{x_2 - x_1}{y_2 - y_1}$ instead of $m = \frac{y_2 - y_1}{x_2 - x_1}$, while others made incorrect substitution into the correct formula, or swap the x and y coordinates in the formula (DoBE, 2018, 2020). After substituting into the correct formula, some learners were unable to simplify correctly resulting in incorrect answers. Many basic errors were displayed with signs and computation, learners often incorrectly writing formulas that were given in the information sheet: apart from the gradient formula, these included the midpoint formula and the distance formula (DoBE, 2017). When asked to calculate the gradient when given collinear points, some learners failed to realise that the gradient between any of the collinear points will be the same or equal (DoBE, 2018).

Misconceptions were also displayed when learners were asked to determine the equation of a straight line. Some learners used points which did not lie on the straight line to find the value of c in the formula $y = mx + c$ after substituting the correct gradient (DoBE, 2020). When asked to determine the equation of a line with only one point provided, learners confused perpendicular lines with parallel lines and in the end used the wrong gradient in their calculations (DoBE, 2017). Some learners assumed that parallel lines have the same equations instead of the same gradients. Many learners struggled to prove that a point lies on a straight line for which the equation was provided, with some substituting both the x and y coordinates into the equation and not knowing how to give the conclusion (DoBE, 2018, 2020).

Learners showed a lack of understanding of the properties of quadrilaterals. They were not able to use the properties of a parallelogram to find the coordinates of a fourth vertex, or give the other three vertices (DoBE, 2018). They also displayed a lack understanding of the properties of rectangles, especially the fact that opposite sides of a rectangle are parallel and hence their gradients will be equal which made it difficult to respond correctly to some questions (DoBE,2017).

When asked to calculate angles, some learners often referred to many angles as θ, without stating explicitly which angle they were referring to, resulting in a number of angles having the same name but different angle sizes (DoBE, 2017, 2020). This caused confusion and resulted in learners using incorrect angles in their calculations.

Learners also displayed a lack of knowledge of Euclidean geometry and its integration into Analytical Geometry. When asked why an angle on the circumference of a circle was a right angle, some learners provided incorrect reasons, while others used inappropriate words like angles subtended by a chord instead of diameter, and facing instead of subtended (DoBE, 2017). When asked to find the equation of a circle whose tangents were the x-axis and the y-axis, many learners were unable to interpret the diagram correctly and hence were not able to determine the coordinates of the centre of the circle (DoBE, 2020). Some learners had challenges when asked to determine the equation of a tangent to a circle, with some learners using the gradient of the radius as equal to the gradient of the tangent, instead of the product of the gradient of the tangent and radius being equal to -1 (DoBE, 2020).

Trigonometry and geometry are strongly related and an understanding of one of these, will be an advantage in understanding the other, hence the lack of understanding of geometry concepts will impact negatively on learners' success in trigonometry (Van Laren, 2012). Maor (1998) points to a strong link between achievement in trigonometry and geometry, where competency in trigonometry depends a great deal on understanding geometry. The Nuffield Foundation (Gorard & See, 2013) links success in trigonometry to strength in algebra and geometry.

Atagana, Mogari, Kriek, Ogbonnaya and Makwakwa (2009) carried out a survey which revealed that trigonometry is one of the chapters which learners find very difficult to learn. The study showed that 46% of the 222 learners who took part had challenges in learning trigonometry (Atagana et al., 2009). Learners face many challenges in trigonometry, which include answering questions on trigonometric functions (Siyepu, 2015) and solving triangles (DoBE, 2017). Developing understanding based on trigonometry links is not easy for learners and traditional ways of teaching trigonometry do not overcome students' difficulties. Demir (2012) noted that understanding trigonometry is very tough and is usually complicated further when teachers use traditional ways of teaching it.

Failure to understand trigonometric concepts leads to learners having many challenges and misconceptions (Yulandari, 2012). Learners struggle with proofs of trigonometric identities as they demand high levels of accuracy and ability to manipulate algebraically (Yulandari, 2012; Mustamir, 2019).

According to Usman and Hussaini (2017) common mistakes in trigonometry, have more to do with lack of conceptual understanding, with few process skills errors. A lack of conceptual understanding leads to learner failing to identify the correct approach to use to solve the trigonometry problem (Rohimah & Prabawanto, 2019). According to a study by Chigonga

(2016), students have difficult in trigonometry because they only have procedural knowledge and have not mastered conceptual knowledge.

De Villiers and Jugmohan (2012) showed that learners displayed no background knowledge of the Pythagoras theorem, that is required for further trigonometry engagements at grade 10, like proving/deriving the squares identity $sin^2x + cos^2x = 1$. Some learners showed misconceptions that the sine function was linear.

Chigonga (2016) carried out a study to investigate the errors made by learners when solving equations involving trigonometry and showed that learners have challenges when solving those that require solutions in a specific interval. The errors revealed by the study showed a lack of background knowledge on the part of the learners. The study also revealed that when solving trigonometry equations of the nature $sin\theta.tan\theta = sin\theta$, in most cases learners divide by a variable expression, in this case they divide by $sin\theta$ thereby eliminating possible solutions since $sin\theta$ equal zero. The study also revealed that some learners did not check whether their solutions worked or not, especially for trigonometry equations involving radical expressions and the lack of knowledge about periodicity of trigonometric functions. As revealed by the study, learners find it hard to determine the other solutions in the given interval after finding the reference angled. The connection of k to integers was not a common feature since learners could not link revolutions to integers. The study revealed that learners struggled with finding particular (exact) or general solutions of trigonometry equations. The findings of the study also revealed that learners misinterpreted the trigonometry ratios when their values were negative, failing to identify relevant quadrants and making invalid inferences.

Rohimah and Prabawanto (2019) carried out a study to investigate challenges faced by high school learners when solving trigonometry equations and proving identities. The study revealed that students had difficulties in solving trigonometric equation problems, namely describing the form of the problem, factoring the form of a quadratic equation in trigonometry and solving basic trigonometric equation. It also found that learners struggled with trigonometric identities and the use of basic algebra to solve trigonometry problems.

The DoBE diagnostic reports for the Grade 12 NSC examinations revealed a number misconceptions and challenges faced by learners when working with trigonometry concepts (DoBE, 2017, 2018, 2020). Learners displayed lack of knowledge of the reduction formula. When asked to simplify an expression to a single trigonometric ratio, learners struggled with the reduction formula, especially with the signs of the reduced trigonometric ratios (DoBE, 2018). Learners struggled to simplify the co-functions $cos(90° - x)$ and $sin(90° - x)$ with some opting to use the compound angle formulas unnecessarily and making mistakes in the process (DoBE, 2016, 2018).

When asked to determine certain trigonometric ratios in terms of a variable, some learners defied the instruction not to use a calculator and instead wrote the numerical values of the trigonometric ratios (DoBE, 2017). Some were incorrectly selecting and applying the reduction formulae, making errors with the signs in the final answers. When determining the missing coordinates on a Cartesian plane, some learners did not realise the quadrants in which a given angle lies, and hence were not able to write the correct sign for the missing x or y-coordinate (DoBE. 2018).

There were many challenges when working with double angles and compound angles. Many learners struggled with the simplification of double angle identities and many confused $Cos2x$ and $2Cosx$ (DoBE, 2017). Some learners wrote the expansion of $Cos2x$ and $Sin2x$ incorrectly even though these were given on the information sheet (DoBE, 2018). Learners were unable to manipulate the double angle expansion correctly (DoBE, 2020). When expanding compound angles, some learners were simplifying incorrectly. Some learners failed to realise that some angles could be expressed as compound angles that include a special angle when asked to calculate some trigonometric ratios without using calculators (DoBE, 2018). Many learners could not identify the compound angle expansion in the expression $Sin(x + 25)Cos15 - Cos(x + 25)Sin15$, and opted to expand $Sin(x + 25)$ and $Cos(x + 25)$, which made the question more complicated (DoBE, 2020).

Misconceptions were also noted when learners worked with trigonometric equations. Many learners were not able to solve trigonometric equations that required the use of the quadratic formula (DoBE, 2018). Some learners struggled with trigonometric equations that required general solutions (DoBE, 2017). When asked to solve trigonometric equations with compound angles involved, many learners used compound angle formulae to expand both sides of the equation, this made the problem far more complicated, with some learners resorting to using the calculator to solve the equation without showing key details in the working (DoBE, 2017). Other learners were selecting the wrong quadrant for the final answer, after getting the correct reference angle.

Learners also experienced a lot of challenges when asked to sketch trigonometric graphs and interpret them. Some were unable to draw the graph of $y = 2Sinx - 1$, they struggled with the shape of the graph as they were unsure of the location of the x intercept, while some learners used a ruler to join the points (DoBE, 2018). Some learners could also not sketch the graph of $y = -2Cos2x$, struggling with the shape as they were not sure of the location of the turning points (DoBE, 2017). Learners did not observe the domain of the given graph and drew arrows at the end of the graph, while some did not give the coordinates of the critical values on the graph (x-intercepts, y-intercepts, turning points) (DoBE, 2017, 2018). Some learners showed a lack of understanding of the maximum value of a graph and confused it with the range of the graph, giving the answer for maximum value as an interval instead of a single value (DoBE, 2017). When given two graphs and asked to write down the range, many learners were confused between the domain and range, with a number writing the range in terms of x, while others failed to correctly identify the graph of f(x) and hence wrote the range of an incorrect graph (DoBE, 2020). Other learners wrote the interval incorrectly as $0 \geq y \geq -2$. Some learners were unable to read off the critical values correctly and consequently were unable to give correct answers when asked to write down the intervals over which the graph was decreasing (DoBE, 2020).

When asked to determine the value(s) of x for which the distance between two graphs will be maximum, some learners calculated the points of intersection of the two graphs, not realising that they were stating that the distance between the two graphs was 0, and hence it was not the maximum distance (DoBE, 2020). Other learners were able to write down the expression for the distance between the two graphs but could not proceed any further.

Learners also displayed a lack of understanding when working with two and three-dimensional shapes. When asked to calculate the lengths of sides and size of some angles on a three-dimensional shape, some learners had difficulty in seeing the different planes in the sketch and were not able to answer questions which required them to link two triangles in different planes (DoBE, 2018). Given the three-dimensional plane, other learners were unable to identify the right angled triangles, and did not realise which rule to use when given a non-right-angled triangle (DoBE, 2017, 2018). Many learners displayed poor algebraic manipulation skills, as they failed to make one trigonometric ratio the subject of the formula or failed to square fractions and remove common fractions.

In the next section I discuss some of the findings of research about the teaching of mathematics.

2.3 Research about the teaching of mathematics

Morrow (2007) suggested that educators need to know the content of the subjects they teach, as well as the best methods to teach the subject and the challenges that learners face when dealing with the specific subject. Designing learning programmes that will facilitate learning and understanding in learners requires educators who have knowledge of the content of the subjects they teach (Morrow, 2007). In addition to knowing the content, mathematics teachers must also have an idea of which steps to take to stimulate interest in learners and enhance learning, in what Shulman (2004) referred to as pedagogical content knowledge (PCK). Shulman (2004) described PCK as the teachers' understanding of how to transform the content knowledge which they have into forms that are pedagogically powerful and yet adaptive to variations in prior knowledge, understanding levels and learning difficulties presented by individual students. PCK is further elaborated as the knowledge formed at the intersection of content and pedagogy that is demonstrated by teachers through their understanding of how particular subject matter, topics, problems, or issues are organised, represented, adapted to the diverse interests and abilities of the learners, and then used to engage learners during instruction (Shulman, 2004). More concisely, PCK is the knowledge that teachers use in transforming subject matter knowledge into forms that are easily understandable by students (Shulman, 2004).

For the Mathematics Trial paper 2, teachers need to understand and be able to teach four knowledge areas, statistics, analytical geometry, trigonometry and Euclidean geometry. In terms of having knowledge of the learners, teachers need to understand the areas where the learners usually face challenges and to plan accordingly. The NSC diagnostic reports for the 2017, 2016 and 2019 examinations reported learners as having challenges on questions which require integration of topics, realising that mathematics cannot be studied in compartments or isolated sections which are not connected (DoBE, 2017, 2018, 2020). The Curriculum and Assessment Policy Statement (CAPS) document for mathematics for Grades 10-12, stated that questions must not be compartmentalised in sections but various topics can be integrated in the same question (DoBE, 2011). According to the DoBE (2018), learners struggled with concepts in the curriculum that required deeper conceptual understanding, where they had to interpret information or provide justification. Having this knowledge about learners helps teachers during planning so that they can address some of these issues in their teaching.

An important aspect of PCK includes the instructional or teaching strategies used by the teacher. For Rosenshine (2012) instructional strategies refer to all the activities, approaches and methods that teachers use to deliver content and these strategies may vary from one lesson to another and from one group of learners to another. Teachers should spend time teaching learners to interrogate what they are doing and teaching concepts with more insight, not merely substituting values into formulas (DoBE, 2017). The methods that the teachers use to teach the learners affect how successful the learners will be in the subject (Firmender, Gavin & McCoach, 2014). The best instructional strategies need to be developed which can help to solve the problem of poor performance by learners in South Africa (Moila, 2006; Kriek & Gryson, 2009). Mji and Makgato (2006) found that poor teaching methods result in low performance by learners.

For Moss & Brookhart (2009) the success of learners in mathematics is heavily linked to the instructional strategies, where effective strategies yield better performance and non-effective methods result in poor results. Learners should be given opportunities that allows them to actively engage with the curriculum (Firmender, Gavin & McCoach, 2014). The NSC diagnostic report for 2016 suggested that teaching should be done for understanding mathematics concepts not for just passing examinations, for example teachers should avoid using the calculator as the main teaching tool for sketching graphs in trigonometry and functions, and statistics (DoBE, 2017). Instructional strategies used by the teachers should therefore encourage conceptual understanding rather than memorisation. In terms of sketching graphs, instructional strategies should therefore place emphasis on properties of functions such as intercepts with the axes, axes of symmetry, asymptotes and turning points (DoBE, 2016).

For Salako, Eze and Adu (2013), success is enhanced when cooperative learning is used as an instructional strategy by teachers to teach in the classroom. According to Qamar and Ahmad Khurram Niaz (2015), teaching methods that engage learners more are viewed as arousing the interest of learners in the subject and pushing them to reach full their potential. When learners cooperate and interact with each other during learning, they help each other in an environment that is friendlier (Reus, 2010). One of the aims of the National Curriculum Statement (NCS) Grade R-12 is to produce learners that are able to work effectively as individuals and with others as members of a team (DoBE, 2011). Effective teaching methods are characterised by proper instructional strategies that learners find easy to implement (Rahman, Khalil, Jumani, Ajmal, Malik & Sharif, 2011).

Teachers need to take into consideration what learners already know about a given concept and then connect it to the new knowledge (Baviskar, Hartle & Whitney, 2009). Building on previous knowledge should be done by asking questions and setting short formal tasks. The NSC diagnostic reports noted that many of the errors made by learners in answering the mathematics papers had their origins in poor understanding of the basics and foundational competencies taught in the earlier grades (DoBE, 2017, 2018, 2020). For Ermeling, Hiebert and Gallimore (2015) good practices must be accompanied by clear learning goals, which require the use of different learning strategies and a good link between what was previously learnt and what needs to be learnt in a particular grade.

Learners in different school environments receive a different quality of education (Hoadley, 2012; Carter, 2010; Wood, Levinson, Postlethwaite & Black, 2011). Varying results in schools might be a result of different methods and approaches by educators, some with non-effective instructional strategies leading to poor performances (Mji & Makgato, 2006). According to Mulkeen (2006), learners in rural areas receive low-quality education since few teachers are qualified to teach the subject and there are limited resources compared to schools in towns. In their study in the Western Cape Province, O'Connor and Geiger (2009) found that some mathematics teachers worked with overcrowded classes, and most schools lacked the necessary equipment for normal teaching and learning to take place. A teacher with good PCK for mathematics must be able to identify learners' challenges and address them by employing different types of instructional strategies. The CAPS document clearly states that to manage inclusivity, all relevant support structures within the school community, including teachers must identify barriers to learning and address them by using various curriculum differentiation strategies (DoBE, 2011).

Research has shown that many mathematics teachers in South Africa struggle with the content of the subjects they are teaching and this often reflects negatively on the learners as witnessed by poor results by South African learners in national and international assessments (Bansilal, 2015; Mji & Makgato, 2006; Ross, McDougall, Hogaboam-Gray & Lesage, 2003; Wilkins, 2000). Ross, et al., (2003) suggested that the interest which a learner shows in mathematics and their attitude are affected by their teacher's understanding of the subject and the teaching strategies they use.

Bansilal, Brijlall and Mkhwanazi (2014), carried out a study to investigate how mathematics teachers understood the content they were teaching their learners. In their study, 253 teachers responded to a test that had questions from one of the past Paper 1 examinations written by Grade 12 learners. The study showed that many teachers struggled with mathematics questions set at a higher cognitive level. A third of the teachers were unable to respond successfully to higher order questions and the non-routine problems. The authors questioned how the teachers would be able to design appropriate assessments, when they could not solve many of the level 3 and 4 questions. The study revealed low proficiency levels for the teachers, with the mean for the teachers' ability and item location being close to each other.

The results of the study also showed that the teachers did badly on questions requiring an object level of understanding of quadratic functions, such as completing the square for quadratic function, indicating a very low teachers' engagement with the concept of quadratic functions. Interventions are not possible if the teachers themselves are working at the action level of the concept. If the teachers can work on particular procedures only in an externally driven manner, they will not be able to recognise the demands of the questions and will not have the ability to design teaching programmes that will push the learners to their full potential and enable them to answer the higher-order questions (Bansilal et al., 2014). The teachers' poor mathematics content knowledge is a barrier to the pedagogic content strategies that they were supposed to draw upon in the class (Bansilal et al., 2014).

Hugo, Wedekind and Wilson (2010) carried out a study to explore the relationship between teachers' mathematical content knowledge, teachers' practice and learner outcomes in Grade 6 mathematics classrooms. The test comprised items from the Grades 5 and 6 mathematics curriculum as well as questions that required teachers to identify common errors made by learners in primary school mathematics.

The study revealed that when the teachers were asked to choose appropriate statistical descriptors (mean, median, mode), 38% of them thought (incorrectly) that the mean and/or the mode are appropriate descriptors to determine a score that is the most popular, showing a complete lack of conceptual understanding of the concept of mean and mode. Of the teachers who picked the mode as the statistical descriptor that describes the most popular score, many also picked an inappropriate descriptor. Teachers showed clear difficulties in deciding between representing data as bar graphs or pie charts and histograms or line graphs.

When the teachers were asked how many decimal numbers there are between 0.30 and 0.40 , 47% of the teachers said there are nine, and 24% said there are ten. The study revealed that incorrect thinking was from the natural numbers to the rational/real numbers where there is always a unique natural number following another natural umber (3 follows 2, 45 follows 44), the same assumption was made about rational numbers. Only about 29% gave the correct answer as an infinite number.

The study also showed that 57% of the teachers had problems with a ratio question which compared part to part, rather than part of a whole. Few teachers (32%) were able to determine the total number of sweets of two kinds when the total number and the ratio were given. Of the teachers 35% added the numbers in the ratio 3 : 1 to come to an answer of 4. However, the study noted that the teachers were comfortable working with simple ratio problems.

When asked to identify correct rules to predict a number in a pattern, 71% of the teachers chose an option that was not even a rule, and 15% identified the correct formula for the sum of the first n integers while 9% picked the incorrect formula which looked similar to the correct formula. The researchers suggested a lack of conceptual understanding of variables as the cause of these findings.

When it comes to geometry the study revealed that many of the teachers were operating at a low Van Hiele level. Of the respondents 71% claimed that it is impossible to construct a square which is also a rectangle (Hugo, Wedekind & Wilson, 2010). The researchers referred to this as a common misconception, which has its roots in the visual introduction of the common shapes not being corrected through later engagements with the qualities and characteristics of geometrical figures, indicating that the teachers were working at Van Hiele level 1. The same thinking was shown when 29% of the teachers indicated that it is possible to construct a rectangle that is not a parallelogram. Fifty-nine per cent (59%) of the respondents did not consider it a mistake for a learner to split a figure into two rectangles in order to find the perimeter. This reflects a confusion of area and perimeter. The method of dividing an area into smaller parts cannot be transferred to finding the perimeter, and when teachers used this approach it indicated lack of conceptual understanding (Hugo et al., 2010).

In their study, Mji and Makgato (2006) used a non-experimental, exploratory and descriptive method to establish learners' and teachers' views about reasons for poor results in mathematics and physical sciences. Purposively selected participants were chosen from seven schools with bad results in District 3 of Tshwane North in South Africa. Focus group interviews were done with ten Grade 11 learners from each school and semi-structured interviews were held with teachers from the chosen schools. The findings revealed an absence of strong mathematics understanding on the part of educators, teaching strategies used by the teachers, motivation and language as some of the factors that influenced learners' performance. The learners indicated that they did not understand the content and their teachers did not know how to assist them while the teachers revealed the sections which they were not comfortable teaching. Regarding teaching strategies, learners reflected on their inabilities and the impatience of the teachers when they asked for help. Some learners complained about others for disturbing and unwarranted behaviour. Teachers indicated a lack of interest and lack of seriousness from the learners, and learners not wanting to try. All these factors amount to teachers using the wrong teaching strategies which do not stimulate interest from learners.

2.4 Assessment

In this section I first present an overview of assessment, before looking more closely at assessment in the South African curriculum. This is followed by an overview of research about assessment in mathematics.

2.4.1 An overview of assessment

Assessment is an important part of mathematics education and a continuous planned process of identifying, gathering and interpreting information about the performance of learners, using various forms (DoBE, 2011). It involves gathering evidence of learning over a period of time. Schools use different types of assessments in order to cater for different learning styles in classrooms and those assessments need to meet the standards set by the curriculum planners for effective learning to occur and for important inferences to be made from the data collected (DoBE, 2011). Although school assessments are targeted at learners, teachers must have knowledge of the aims and objectives of the assessment process, regardless of whether the assessments are school based or external (Ainsworth, 2015). This section looks at different types of assessments and the literature done on assessment relevant to this study.

Assessments can be used by teachers in the classroom to support the teaching and learning process, and can also be used by stakeholders to inform them about how learners are progressing in a specific subject and grade (Black, 1988; Kanjee &Sayed, 2013). Depending on whether the assessment support teaching and learning or informs stake holders about the progress made by learners, it is categorised as either a formative assessment or a summative assessment. These will be discussed in detail in the sections below. In both types of assessments, it is vital that learners are given progress reports or feedback so that they are aware of areas that need improvement (DoBE, 2011).

2.4.1.1 Formative assessments

Assessment that is designed to support teaching and learning is referred to as formative assessment, with some authors calling it assessment for learning (Bennett, 2011). For Black and Wiliam (1998) all of the activities or processes carried out by teachers and/or with learners, which are used to inform and improve everyday teaching and learning can be referred to as formative assessments. These formative assessments may include questions and answers during lessons, and written work that is designed and given to all learners in a specific grade in a particular school and is of the same standard as questions asked during examinations (Fisher & Frey, 2009). For Ainsworth (2006) fairness in comparing learners' progress can be achieved through common formative assessments. The results of common assessments can be used by teachers to design learning programmes that address challenges, and can help teachers with developmental feedback.

Although formative assessments are used as tools for developmental purposes on a daily basis, summative assessments have become increasingly more important, as reports to stakeholders are based on the summative assessments (Stiggins, 2005). The next section will discuss summative assessments in detail.

2.4.1.2 Summative assessments

Assessments that give feedback to stakeholders by making judgements and grading learners' success are referred to as summative assessments. For Garrison and Ehringhaus (2013) summative assessments measures the amount of learning a learner has gained in terms of subject content. Summative assessment is used by stakeholders to check on the functionality of education institutions and also informs them about learner progress (Bennett, 2010, 2011; Sambell, McDowell & Montgomery, 2013). To IIiya (2014) summative assessment summarises and reports on the subject knowledge gained by the learners over some time.

According to Stiggins (2005) different summative assessments are used to measure the learning that has taken place in a specific subject. Common examinations, tests, practical examinations, midyear examinations, trial examinations (applicable to Grade 12) and end-of-year examinations are some of the common summative assessments used in South Africa.

Summative assessments can help to gauge the functionality of the whole education system as well as the basic functionality of a school or education institution (IIiya, 2014). Summative assessments are also used for purposes of progression from one grade to another. According to Garrison and Ehringhaus (2013), summative assessment results can be used for school evaluation and for subject improvement purposes. In general assessment in South Africa serves a number of stakeholders that include the DoBE, tertiary institutions and companies that offer various job training services and job opportunities.

Among the stakeholders in the education system, different groups view assessments in different ways. The learners see assessments as tools they must use to progress to the next grade. Teachers use assessment results to improve their teaching while measuring progress made by learners (Kanjee & Sayed, 2013; Roux, 2014). Principals use assessment results as an indication of the effectiveness of the teacher, and to check on the school functionality in terms of meeting the goals of the education system (Reynolds, Livingston & Willson, 2010).

In the South African school context, assessments such as those administered at NSC level are used to grade and progress learners as well as to give access to tertiary education. Similarly, classroom assessments (school examinations which are summative) can also be used as formative assessments. Benchmark or provincial assessments are designed for formative goals such as designing interventions, and also to benchmark school, district and provincial performance (Roux, 2014). The provincial trial examinations are good examples of such benchmark assessments.

In global studies such as TIMSS and PIRLS, summative assessments are used to measure how a country is performing educationally and to rate the assessment results of South Africa with those of other countries (Mullies, Martin, Foy & Drucker, 2012).

All the assessment tasks that contribute to the final year mark, are moderated at school level and at district level for quality assurance purposes. The formal programme of assessment constitutes all these tasks and helps teachers in deciding whether a learner will progress to the next grade. Formal assessments provide feedback to stakeholders about how well learners are progressing in a grade and/or in a particular subject (DoBE, 2011).

According to the DoBE (2011), the formal assessments must have a variety of activities and tasks that will help learners to grow and to develop an understanding of the content of the subjects they are learning. The formal assessments are supposed to be designed in a way that caters for different types of learners, from high-ability to low-ability learners.

All assessment tasks should be measured against the basic education principles of content, learning and equity, if they are to draw the best outcome from the learners (Alberts & White, 1993; Messick, 1989). Educational principles may appear to work against the traditional technical and practical principles that were used to evaluate the advantages of assessments but rather they evolve from the traditional principles.

For a mathematics assessment to meet the principle of fostering an understanding of content, it must reflect the key mathematics concepts that are crucial for learners to learn (Messick, 1989). Besides the content principles, mathematics assessments should also be measured against the extent to which they reflect the learning principles, based on how learning and instruction are improved. The product of education principles should be a high-quality education system.

Assessment should cater for all groups of learners in terms of supporting their learning process, allowing them to engage with mathematical tasks even when they have very low proficiency in mathematics. The equity principles are based on the question of whether an assessment favours one group over another for different reasons that have nothing to do with the aims of the assessment, whether comparisons with performance standards are justifiable and whether the tasks are accessible to students (Frederiksen & Collins, 1989). In the next section I discuss the assessment landscape in South Africa.

2.4.2 Assessment landscape in the South African Curriculum

Various stakeholders (learners, teachers, education officials, parents, political players, tertiary institutions) have an interest in the outcomes of the South African schooling system, especially the results of the final Grade 12 examinations. Learners' performance in South Africa is therefore always being monitored due to the demands for improvement in the performance of learners in Grades 10-12 in critical subjects which includes mathematics. The mathematics national pass rate (at 30% level) for the Grade 12 National Examinations for the past five years were 49.1% in 2015; 51.1% in 2016; 51.9% in 2017; 58.0% in 2018 and 54.6% in 2019. These national pass rates are a bit worrying as there was a steady increase from year 2015 to 2018 and then a sudden drop in 2019 (DoBE, 2020). Over the past few years, learners have written different assessments to measure progress and achievement in critical subjects like mathematics and these include National Examinations, School-Based Assessments, ANA, National, Provincial and District common examinations and international assessments like the TIMSS.

In order to maintain standards across the country, the NCS for Grades R-12 outlines the knowledge, skills and values that are worth learning in South African schools (DoBE, 2011). Two phases exist within the South African schooling system, the first being the General Education and Training (GET) phase which is from Grades 0 to 9 while the second is the FET phase, from Grades 10 to 12.

The GET phase has three stages namely, the Foundation Phase, Intermediate Phase and Senior Phase, with each stage preparing learners with the skills, knowledge and values required for the next phase. The Foundation Phase ranges from Grade R to Grade 3 and four subjects are taught. At the foundation phase formative assessments are used to progress learners. The Intermediate Phase is from Grade 4 to Grade 7 and 6 subjects are covered. School-based assessments have a weighting of 75%, while the final-year examination contributes 25% to the final year mark. The Senior Phase is from Grade 8 to Grade 9 and 9 subjects are covered. School based assessment contributes 40% and the end of year examinations contribute 60% of a learner’s final year mark in a subject.

The FET phase is from Grade 10 to Grade 12 after which a learner obtains an NSC. Learners have choices of subjects even though some are limited to one or two choices, e.g. a learner has to study a home language and has to choose between mathematics and mathematical literacy. Subject choices are made at Grade 10 level as stipulated in the NCS regarding subject choices and requirements for progressing to the next grade. Below is a summary of the South African assessment system for Grades R-12.

Table 2.1

The South African Assessment Landscape (Adapted from DoBE (2018)

General Education and Training (GET) phase (Grades R - 9)	Further Education and Training (FET) Phase (Grade 10 - 12)
School based assessments (formal and summative assessment tasks) • Formal assessments such as tests, projects, assignments etc • Internal assessment tasks for Foundation Phase in Grade R to Grade 3 in numeracy and language	**School based assessments (formal and summative assessment tasks)** • Formal assessments such as tests, assignments, research etc • Language Oral Assessment tasks • Practical Assessment Tasks
Daily assessment tasks • Classwork, homework, group work etc	**Daily Assessment Tasks** • Classwork, homework, group work etc
Summative examinations (June and November) • Internally set formal examination in Grades 4, 5, 7 and 8 for all subjects • District and provincial common examinations in Grades 6 and 9 in subjects such as Natural Sciences, Mathematics, Economics and Management Sciences	**Summative examinations (June and November for Grades 10 - 12 plus trial examination for Grade 12)** • National and Provincial Common examinations in Grade 10 - 11 in Mathematics, Physical Sciences, Economics, Accounting and English First Additional Language • National Senior Certificate Examinations in Grade 12 for all subjects

Table 2.1 shows that the South African education system uses both formative and summative assessments, but for promotional purposes at the FET phase, the greater percentage of the marks comes from the summative assessment, especially the end of year examinations.

At the FET phase provincial examinations are aimed at critical learning areas like mathematics in Grades 10 and 11, and at all subjects in Grade 12. The common examinations ensure standardised assessments across the province and across the country. Learners should not be disadvantaged by virtue of writing these common examinations (DoBE, 2016).

Assessment processes in South Africa are controlled by a number of policies aimed at providing assistance for the smooth running of school assessments including examinations. The National Protocol for Assessment (NPA) was put into effect in 2011 to manage the administration of assessment tasks. Apart from the NPA, the National Policy Pertaining to the Programme and Promotion Requirements (NPPPPR) of the NSC Grades R-12 and CAPS was put in place for standardisation of the NCS for Grades R - 12 (DoBE, 2011). The NPPPPR controls all programmes of assessment up to the stage of issuing reports for all grades.

The NSC takes place over three years from Grade 10 to Grade 12 in the FET phase, where a learner has to opt for mathematics or mathematical literacy as one of the seven subjects required to achieve the NSC qualification. For mathematics formal tasks are outlined in the formal programme of assessment which includes tests, the June examination, projects, investigations, end-of-year examinations and the Grade 12 trial examinations. According to the DoBE (2011), formal tasks are administered and managed at school level for progress and certification purposes.

The provincial DoBE sets the formal assessments, such as the mid-year and trial examinations for Grades 10 - 12. The end-of-year examinations in Mathematics and Physical Sciences for Grades 10 and 11 are set by the national DoBE, while teachers mark internally at school level. The trial examination for mathematics at Grade 12 level covers all aspects that the learners will write on for the final examination; it thus helps learners to prepare for the final examination.

All formal tasks that are considered for school-based assessment are part of the School programme of assessment. The results of the assessment tasks, are recorded and reported at the end of every term. School-based assessment contributes 25% of a learner’s final year mark and it includes the June examination and trial examination for Grade 12. The end of year examination contributes 75% of a learner’s year mark. Table 2.2 below shows the programme of assessment for Grades 10 - 12 as shown in the CAPS document.

Table 2.2

Number of Assessment Tasks and Weighting for Grade 10-12 (DoBE, 2011)

		GRADE 10		GRADE 11		GRADE 12	
		TASKS	WEIGHT (%)	TASKS	WEIGHT (%)	TASKS	WEIGHT (%)
School-based Assessment	Term 1	Project /Investigation Test	20 10	Project /Investigation Test	20 10	Test Project /Investigation Assignment	10 20 10
	Term 2	Assignment/Test Mid-Year Examination	10 30	Assignment/Test Mid-Year Examination	10 30	Test Mid-Year Examination	10 15
	Term 3	Test Test	10 10	Test Test	10 10	Test Trial Examination	10 25
	Term 4	Test	10	Test	10		
School-based Assessment mark			100		100		100
School-based Assessment mark (as % of promotion mark)			25%		25%		25%
End-of-year examinations			75%		75%		
Promotion mark			100%		100%		

Table 2.2 clearly indicates which type of assessment is required for each term and the weighting for each, that is, its percentage contribution towards the final school-based assessment mark for Grades 10 - 12.

2.4.3 Research about Mathematics Assessments

Producing a test for basic mathematical skills that can be used in different cultures is a hard task. The TIMSS, ANA and other psychology tests such as Wechsler Individual Achievement Test (WIAT), have all been used to assess learners' proficiency in different countries. In South Africa before Grade 12 learners write their final examinations, they write provincially set trial examinations. This helps to check the readiness of the learners, and it is important that these trial papers meet the basic standards of fundamental measurement.

Modzuka, Long and Machaba (2019) in their study 'Rasch analysis of South Africa's Grade 6 Annual National Assessment', investigated the ANA Grade 6 Mathematics test in 2012 to observe whether the assessment tool provided accurate information about learner proficiency. The aim was to investigate whether the learners were performing poorly because of the ANA assessment instrument, which was poorly structured (Modzuka et al., 2019). In this study 546 learners' scripts were used from one district of Gauteng Province in South Africa, where 29 questions with 56 items were responded to. The study revealed that the ANA was not well targeted as the item mean and the person mean were not close to each other (Modzuka et al., 2019). I also showed that the ANA Grade 6 Mathematics instrument was not a good assessment tool as it did not meet most of the standards and principles of measurement. However, the assessment tool addressed validity issues although the reliability was very low. Some items did not meet the requirements of the Rasch model. For other items, learners of low proficiency performed better than predicted by the Rasch model and learners of high ability performed lower than predicted, while for some items learners with low proficiency performed at a lower level than expected while those of high ability performed better than expected. The ANA Grade 6 Mathematics test failed to differentiate learners well according to their different abilities, and this turns out to be very challenging for the learners. Some items were above the ability levels of all of the learners, and hence these items could not accurately measure their proficiency. The study recommended that the quality of assessment tools should be guaranteed first before conclusions can be made based on the assessment process.

In a related study focusing on a Grade 9 assessment, Bansilal and Wallace (2008) studied one teacher and her mathematics class in Durban, South African. Interviews and classroom observations were used as the methods of collecting data from the learners and the teacher. Some of the common tasks for assessments (CTA) were analysed as part of collecting data for the study. The study found challenges with learner attributes and proficiencies, the quality of tasks given to learners, and other factors to do with classroom management. The learners showed a lack of understanding in most of the common tasks for assessment (CTA) given during lessons, and had difficulties in understanding the questions because of the terms used to give instruction. Other issues were information overload, messy numbers, lexical density and the lack of relevant real-life contexts when the CTA were created. These will now be discussed in detail.

The study revealed that learners struggled with mathematics concepts involved, what the teacher referred to as learners having 'gaps' in learning (Bansilal & Wallace, 2008). The study showed that the teacher struggled with mediation of a summative assessment task as the learners were showing a lot of knowledge gaps especially where conceptual understanding was involved. The teacher's intervention through direct instruction affected the validity of the CTA programme. The study showed that the teacher tried to compensate for learners' problems but ended up teaching a lot of information in a short period of time. At one point many topics were taught in one lesson, a process not normally considered a sound pedagogical practice. The teacher's use of interventions which differed from those of other teachers further influenced the validity of comparison of results across class, school, or provinces.

An important finding revealed in the study by Bansilal and Wallace (2008) was regarding the intersection between language and assessment. They found that language was an issue, where learners' poor command of English prevented them from accessing the CTA problem, with some learners guessing the meaning of words like 'dormitory', resulting in them visualising a completely different scenario and context from that given in the question therefore affecting their response to the task. The study revealed two dimensions of language issues, poor language skills which prevented learners from understanding the context of the task, and poor language skills which prevented learners from understanding the instructions

The study also found that information overload affected learners' success in the CTA. The use of Robben Island as a place where an event happened, contained many explanations and descriptions which were grouped into context information. This covered the information that was used to paint a picture about the context, and crucial information that was necessary to solve the task. Some learners in the study were unable to extract important data which were necessary to solve the task, some used instructions which were not required or which they were discouraged from using, while others created their own information. Learners were unable to extract details about the task which were contained in the paragraph which also contained the contextual information.

The other issue revealed by the study was that of lexical density (Bansilal & Wallace, 2008). The study showed that reading the instructions was not an easy exercise for the learners. In one instance the learners determined the area instead of the dimensions of the cells and they indicated that they saw the instruction on the board but did not know what to do. The study revealed that the learners' struggles and the context make up lexical density. Halliday (1993) suggested that texts used in science and mathematics have many terms that are not used in everyday language, a situation referred to as having a very high lexical density. It was not easy for learners to understand the instructions and the questions, since they contained many content words. The instructions in the CTA task had a very high lexical density of approximately 13 content words per clause.

The last concern raised in the study by Bansilal and Wallace (2008) was the assumption made about situations that learners are used to in their daily lives. In the CTA task, the context was Nelson Mandela's imprisonment on Robben Island. The study observed that while the learners were familiar with some aspects of the context, many others were not part the learners' daily lives. The context used in the CTA task did not help to give learners a clue of what approach to use to answer the question. Learners failed to use two graphs, a scatter plot representing the mean monthly temperatures and a bar graph showing the monthly rainfall patterns at Robben Island. Some learners used the weather experience in Durban, to respond to a question about which months the weather is favourable for tourism. The environment in which the event took place prevented learners from applying the mathematics concepts they learnt in class.

Mudaly (2006) in his study, 'Marking matric mathematics scripts should be used as a time for teaching too', investigated the way in which the marking process impacts on mathematics education. The study involved 32 educators who were either experienced markers or had limited experience in marking, who completed a questionnaire and thereafter participated in a discussion. The study was done with educators from Pietermaritzburg and Durban, in the KwaZulu-Natal (KZN) province in South Africa. The study revealed that for questions which require consistent accuracy in marking, no other educator apart from the Chief Marker, allocated marks in a way that was identical to that espoused in the marking memorandum, and the experienced educators allocated the majority of marks to pure accuracy. This means that a candidate making a manipulation error in the first line will not obtain any marks at all. The study also revealed that the entire marking process is an empowering exercise, allowing educators to revisit, revise and rework their teaching strategies and content in mathematics classrooms. The study showed that marking experience broadens the mathematics knowledge of educators as new methods and strategies may be introduced during marking and teachers will apply these in their classrooms. It also revealed that most educators scrutinise the mathematics paper only during the 10 days of marking to the extent that they could now predict which sections of the syllabus should be emphasized, notwithstanding the fact that marking enables educators to easily establish common errors and misconceptions. The study revealed that very few or no workshops were conducted regarding the marking process. Very few educators stated that they attended workshops on marking of examinations scripts, and the few who did so indicated that the workshops were not conducted by Department of Education officials, but by independent bodies like the Association of Mathematics Education of South Africa (AMESA).

Performance in mathematics assessments is greatly affected by the language used in the test and the degree of familiarity that the test taker has with the test language. In many cases the language complexity or readability of the tests may contribute to outcomes which are lower than expected (Khan & Bansilal, 2012; Abedi & Lord, 2001; Prins & Ulijn, 1998).

Researchers Prins and Ulijn (1998) administered three English versions of nine mathematics tasks original, adapted and non-verbal to 108 students in South Africa. The adapted version was reconstructed to make it easier to understand, while the non-verbal versions did not have any references to context. The students were made up of three groups: those who speak English as their first language, first language Afrikaans speakers and those whose first language was an African language. The researchers found that the average score on the adapted versions were statistically significantly higher than on the original versions for students from all language groups. However, second-language English speakers were more favoured by the modifications to the mathematics test items than the first-language English speakers. The study also revealed that students whose first language was an African language had the lowest score on the original and adapted versions, but performed equally well as their counterparts on the non-verbal version of the task. For the authors this result showed that the group had the same algebraic proficiency as the other groups but factors related to language and culture acted as barriers to their performance on the non-verbal versions

Regardless of whether the students are first- or second-language speakers of the test language, they will still have to deal with the difficulties in the mathematics register, although the difficulties will be increased for the second-language speakers because the mathematics register is used simultaneously with a natural language register that they are not experts in (Prins & Ulijn, 1998). Hence success in mathematics assessments presented in textual form is necessarily linked to language fluency since the solution of the problems requires the learner to move between the mathematical symbolic register and the natural language register. Understanding the natural language is necessary to first decode the instruction, and then mathematics skills can be used to solve the problem. Students who have a deeper understanding of the relevant mathematics concept will more easily recognise the information they need to extract from the natural language. This is because a few words may trigger the activation of wanted information that they may need. Those who do not understand the relevant concepts will therefore be at a disadvantage.

A study by Abedi and Lord (2001) yielded similar results to the study by Prins and Ulijn (1998) described above. Abedi and Lord (2001) carried out their study with 1174 Grade 8 learners in the United States of America to investigate the effects of language on learners' performance in mathematics. The study revealed that the linguistic modification of some items benefited students with a poor mathematics foundation more than those who had been doing well in mathematics. However, simplifying the questions to make them more understandable benefited all of the groups, and scores were then slightly higher.

Bansilal and Khan (2012) carried out a study to identify the kind of challenges that learners have with the instructions for a task for Grade 9 mathematics students. Students who participated in the study used English as their second language of communication. The CTA was created such that a series of tasks was set using an extended context. The study involved 44 English second language-speaking learners in an English-medium high school. The study revealed that many of the instructions and passages were very hard for an average Grade 9 learner to understand. The learners could not even attempt to answer the questions because they failed to understand and interpret the questions of the CTA. With high lexical density, the questions were not accessible to the average learners. Some studies in the South African context which involved the use of the Rasch analysis will be discussed in the paragraphs that follow.

Long et al., (2014) carried out an investigation of a Mathematical Literacy (ML) assessment by making use of the Rasch measurement. The investigation was carried out on the 2009 KZN provincial paper, which had 51 items, with maximum possible score of 150 marks. The participants were 73 Grade 12 ML learners. The study illustrated how the Rasch model could be used in conjunction with professional judgement to check the validity of the assessment, identifying anomalies and inconsistences and providing educational reasons that may warrant rescoring. The study revealed cases where the method and accuracy system did not work. The allocation of method and accuracy marks unduly disadvantaged those who did not address the question correctly, as most learners who identified the method were able to obtain the correct answers because it involved just entering the numbers into the calculator and reading off answers (Long et al., 2014). The study revealed that very few learners achieved the method mark without achieving the accuracy mark. Long et at., (2014) resolved the method and accuracy issue by rescoring the respective items with such problems. Their study also revealed cases where some scores were considered as redundant, that is where the scoring rubric allocated 2 marks per reason. In these cases, no learner attained the first mark without getting the second and similarly no learners attained 3 marks without getting the fourth (Long et al., 2014). The issue of redundant marks was also resolved by rescoring. The study also revealed cases where the answer to a previous item influenced the probability of success of the learner in a following item, a situation referred to as item response dependency (Long, et al., 2014). The dependency of a subsequent item on an earlier item is regarded as an unfavourable test practice, it is the independency of items that offers greater precision (Lee, 2004).

Long et al., (2010) carried out a study on proficiency in the multiplicative conceptual field: using the Rasch measurement to identify levels of competence. The study reported on the findings of a study in which 35 items selected from TIMSS (2003), were administered to 330 Grades 7, 8 and 9 learners at two schools. The Rasch analysis was used to compare the difficulty of the mathematics problems located within the multiplicative conceptual field, while locating the degree to which learners had mastered the necessary skills set on the same scale (Long, et al., 2010). Such location of items and learners on the same uni-dimensional scale allowed for analysis of which aspects of the problems, make one problem more difficult than another. Simultaneously the scale gave clear evidence of which students had mastered which concepts and skills and which students had not, thereby allowing more targeted assistance to the class and individual learners (Long, et al., 2010). The study suggested that the implementation of the Rasch analysis within the school classroom on appropriately designed assessments instruments would provide clarity for the teachers on the difficulties within the problems used in the assessment, and the relative degree to which individual learners were achieving success in mastering the targeted concepts.

Bansilal (2015) carried out a study on the Rasch analysis of a Grade 12 test written by mathematics teachers. The purpose of that study was to the explore mathematics teachers' proficiency in the mathematics that they teach (Bansilal, 2015). The study used a sample of 253 teachers' responses to a shortened Grade 12 examination which was analysed using the Rasch model. The study revealed that the teachers' proficiency was located close to the mean of the item locations, with the proficiency levels of almost one-third of the group lying below the difficulty level of all the level 3 and level 4 items in the test (Bansilal, 2015). The study also illustrated how the application of the Rasch model can be used to contribute to a more informative and fair assessment. In the study and in line with RMT, the test was subjected to various analyses and the results were used to improve the fit of the items and the test (Bansilal, 2015).

Stols et al., (2015) carried out a study on the application of the RMT to an assessment of geometric thinking levels. The aim of the study was to apply the Rasch model to investigate both the Van Hiele theory for geometric development and an associated test, where the objective was to investigate the functioning of a classic 25-item instrument designed to identify levels of geometric proficiency (Stols et al., 2015). The study used a dataset of responses from 244 students (106 for a pre-test and 138 for a post-test), of whom 76 sat both the pre-test and post-test. The results from that study revealed summary statistics which did not show statistically discernible differences between observed and expected scores under the Rasch model (chi-square statistic), but the Rasch analysis strongly confirmed the Van Hiele theory of geometric development (Stols et al., 2015). The study used the Rasch analysis to identify some problematic test items as they only required knowledge of a specific aspect of geometry instead of testing geometric reasoning. In terms of the Van Hiele theory, the Rasch analysis identified as problematic some items about class inclusion, an issue that has also been raised in other studies (Stols et al., 2015). Class inclusion happens when classification of items with common characteristics or properties takes place. For example, a square is a rectangle, since a square has all the properties of a rectangle. However not all rectangles are squares, as some rectangles do not have all the properties of a square. In accordance with the researchers' findings, learners in this study found such questions very difficult.

CHAPTER 3 RESEARCH METHODOLOGY

3.1 Introduction

This chapter outlines the methodology that was used in this research study. It begins with an explanation of the context within which the study was undertaken and the process by which the participants for the research were selected. Thereafter, the aim of the research based on the critical questions as well as the mixed methods research approach used to address these critical questions are outlined. The methods used to collect data for the research are also described. The validity and trustworthiness criteria, as suggested by Lincoln and Guba (1985) are also explained in this section. The chapter ends with an explanation of the ethical issues involved and the limitations of the study.

3.2 Context of the Study

This study was motivated by my own experience as a teacher teaching mathematics in a rural school in KZN for many years. A large proportion of the KZN population live in rural areas, although there are also well-developed urban areas. The Grade 12 learners in KZN account for 23.19% (116937 out of 504303) of the total number of learners who wrote the NSC examinations in 2019 (DoBE, 2020). Hence, improvement in learner performance in KZN will impact on the education system as a whole.

There is a great deal of pressure in Grade 12 for learners to do well, since it is the final year of schooling, after which they look for jobs or enter the tertiary sector. Teachers, learners and parents want the learners to do their best, especially in the subject of mathematics, which is the gateway to many science and mathematics-related careers. One intervention that has been in place for many years is that of the preparatory examination where learners write examinations based on all the content that they will be assessed on in the final national examinations.

Each province sets the trial examination for their learners, it is usually set by a panel of experienced teachers and is expected to undergo moderation processes. However, it is not clear whether the moderation processes are as rigorous as they are for the national final examinations. Some small-scale research conducted on the provincial examination in Mathematics Literacy raises concerns about the design of these KZN provincial assessments. Long et al., (2014) in a study conducted on a KZN Grade 12 examination for ML, found indiscriminate mark allocations in some instances, as well as unclear instructions, which affected the fairness of the paper. Although not all schools are compelled to write the provincial trial examination papers, the underperforming schools have no choice but to do so. Schools which are identified as having performed poorly in the Grade 12 national examinations in the previous year, are compelled to administer these examinations on a quarterly basis. Some schools opt to write the paper for convenience or opt to provide their learners with the experience of writing an external paper.

The trial examination is an important instrument for helping learners prepare for their final examinations, it is therefore very important that the instrument is well designed and provides a fair and reliable assessment of learners' proficiency at that stage of the year. Based on these considerations, I decided to embark on this study to examine the trial examination using the Rasch analysis as a tool. In addition, I sought teachers' views about the paper as a whole and about particular topics and items that were identified by the Rasch model as having issues or being problematic. The schools that were approached to participate in the study were from Umlazi district in the province of KZN.

3.3 Research Purpose

The Grade 12 KZN trial mathematics examinations aim to help Grade 12 learners prepare for their final examinations. The mathematics trial paper must fulfil most of the assessment standards if it is to serve its purpose of preparing learners for the final examination. The study aimed to explore the use of the Rasch analysis in improving the functioning of a mathematics assessment instrument. In order to achieve this purpose a number of objectives were fulfilled which include but not limited to, reporting on the functioning of the test instrument for measuring proficiency in mathematics and checking on the targeting and reliability of the test instrument. The extent to which the data fit the model was also investigated and where there were anomalies explanations were given. The study also investigated if the items were working in different ways for learners who have the same characteristics, a situation referred to as differential item functioning (DIF). In order to reach these objectives, I formulated some research questions, which were answered by the research study. These are presented in the section that follows.

3.4 Research questions

The research study aimed to answer the following research questions:

1. What are some trends in performance of learners in the Grade 12 mathematics assessment instrument?
2. What does a Rasch analysis reveal about the targeting and functioning of the instrument as a whole?
3. How can the Rasch analysis be used to improve the functioning of the mathematics assessment instrument?
    a. To what extent are the items functioning as expected?
    b. How can the use of the Rasch analysis contribute to the improvement of the scoring rubric?
    c. To what extent do the items display DIF, multidimensionality and item dependency?
4. What are the teachers' views about the findings of the Rasch analysis?

The first research question helped us to understand general performance of the learners per school and for all of the participating schools combined. The general trends in learners' performance were also sought. The second research question helped to understand what the Rasch analysis revealed in terms of how the learners performed in the mathematics trial paper. The Rasch analysis revealed items which were too hard or too easy for the learners instead of depending on the cognitive levels that are stipulated by the Department of Education. The third research question helped to identify items which are working well, in line with measurement expectations, and those which are not working as expected. The third research question also helped to explore how the Rasch analysis would contribute to the improvement of the scoring rubric. Lastly, research question three helped to check the extent to which the items displayed DIF, multidimensionality and item dependency, as these violated the principal of independency. The last research question helped to give an explanation of what the teachers believed to be the reasons behind the findings of the Rasch analysis.

3.5 Participants

The participants in this study were Grade 12 mathematics learners from six schools in Umlazi district and 10 teachers from the feeder schools and 5 from outside the feeder schools. Purposive sampling was used to select the six schools, to achieve representativeness from township schools, former Model C schools, semi-urban schools, urban schools and rural schools. Initially 10 schools were targeted, but permission was denied by the school governing bodies of four schools, for various reasons ranging from the need for long procedures before release of the scripts to fears about the safety and privacy of the scripts once they are handed over to the researcher (although the researcher tried in vain to guarantee the safety of the scripts). The initial 10 schools were selected because they were easily accessible for the researcher.

Six hundred and four (604) learners gave consent and their scripts were used in the study, which involved six schools. Two hundred and eleven (211) learners did not consent so their scripts were not used in the study. The schools were coded as AD, SS, KS, LM, SB and FF. For the learners who did not give consent, 49 were from school AD, 61 from school SS, 38 from school KS, 27 from School LM, 16 from school SB and 20 were from school FF. School KS is a former Model C school, whereas LM is located in a township, and AD and SB are schools in a semi-urban area in KZN. School SS is located in a rural area, whereas school FF is located in the central district of a major town in KZN. Table 3.1 gives a full description of the participating schools.

Table 3.1

Description of the schools

School	Quintile	Description of the School
AD	4	Located in a semi-urban area Well-resourced It is a boarding school, although some of the learners are day scholars
KS	5	Located in a low-density suburb It is a former Model C school and is well resourced
SS	4	Located in a rural area It has a huge enrolment in terms of number of learners
LM	4	Located in a high-density suburb Not well resourced
SB	2	Located in a semi-urban area Not well resourced
FF	Independent	Located in the central business district (CBD) Well-resourced and privately owned

The number of learners, the gender totals and the total number of learners per school are given in Table 3.2.

Table 3.2

Gender totals and total number of learners in the six schools

School	Home language		Gender		
	English	Others	Males	Females	Totals
AD	0	76	27	49	**76**
SS	0	131	57	74	**131**
KS	64	61	54	71	**125**
LM	0	9	43	46	**89**
SB	0	9	26	53	**79**
FF	0	104	32	72	**104**
Total	**64**	**540**	**239**	**365**	**604**

Out of the 604 learners whose scripts were used in the study, 239 were males and 365 were females. The highest number of scripts were from school SS, with 131 scripts, followed by school KS with 125 scripts, school FF with 104 scripts, school LM with 89 scripts and school SB with 79 scripts. The smallest number of scripts (76) came from School AD. School KS had 64 English home language speakers and 61 who speak other languages as their home language. All of the other schools had learners speaking other languages as their home language.

3.6 Research design

I used the mixed methods research design for this study. The aim of the mixed methods was to maximise on the strength of the qualitative and quantitative research and minimise the weaknesses of the two approaches, in a single research study (Creswell & Plano Clark, 2011; Morse & Niehaus, 2009). Mixed methods research is any type of research that makes use of two or more research approaches in one study (Creswell, 2011). This definition is supported by Shanon-Baker (2016) who saw mixing of two research approaches in one study as useful in enquiry-based research that is philosophically grounded. The mixing of the methods can take place at different stages of the study, which might be at the theoretical framework stage, data collection and analysis stage, or at the overall research design or discussion of the research stage. Mixed methods research provides opportunities to explore situations where the use of one approach is not going to yield maximum results (Creswell & Plano Clark, 2011; Morse & Niehaus, 2009).

There are six different types of mixed methods as described by Creswell (2009), sequential exploratory design, sequential transformative design, concurrent triangulation design, concurrent embedded design, concurrent transformative design and sequential explanatory design. Table 3.3 discusses the six mixed research types.

Table 3.3

The mixed method designs

Mixed method design	Purposes	Order of the phases
1.Sequential Exploratory Design	To explore a phenomenon in a selected population. To develop an instrument or typology that is not available. To assess whether qualitative themes generalise to a population.	Two phases. Starts with qualitative data collection and analysis, which then informs the quantitative data collection and analysis in second phase.
2. Sequential Transformative Design	To investigate a problem, creating sensitivity to data collection from marginalised or disadvantaged groups and to call for action.	Two phases, which might be either qualitative first or quantitative first, with a theoretical lens overlying the sequential procedures.
3. Concurrent Triangulation Design	To obtain a more complete understanding from two data bases and corroborate results from two different methods. To compare multiple levels within a system.	Quantitative and qualitative data are collected concurrently. The two data sets are then compared for any convergence, differences or combinations. Quantitative data and qualitative data are used separately, and equal weight is placed on both approaches, even though in some cases, priority may be given to one or the other of the two approaches. Mixing of the two approaches happens at the discussion or interpretation section.
4. Concurrent Embedded Design	To address different questions that call for different methods. To enhance an experiment such as by improving recruitment procedures, examining interventions and explaining reactions to participation.	Both quantitative and qualitative data are collected simultaneously. One method guides the research and one secondary database plays the supporting role. The secondary method which can either be quantitative or qualitative is embedded or nested within the predominant method, which can be either quantitative or qualitative.
5. Concurrent Transformative Design	To use specific theoretical perspectives as well as the concurrent collection of both quantitative and qualitative data	Researchers use a specific theoretical perspective together with the simultaneous collection of quantitative and qualitative data where priority maybe equal or unequal. Data are mixed by merging, connecting or embedding them.
6. Sequential Explanatory Design	To use qualitative data to help explain quantitative data results that need further exploration. To use quantitative results to purposefully select best participants for qualitative study.	Starts with quantitative data collection and analysis, which then informs the qualitative data collection and analysis in the second phase.

The sequential explanatory design was used in this study. This design will now be discussed in detail.

When a researcher collects quantitative data and analyze it, then uses the results of the quantitative approach to design ways to collect qualitative data and analyze, the design is referred to as the sequential explanatory design (Creswell, 2009). For Creswell (2009), the mixing of the data occurs when the initial quantitative results inform the secondary qualitative data collection. According to Morse (1991), the sequential explanatory design is useful when the data is hugely numerical and unexpected results arise from the quantitative data. The advantages of using the sequential explanatory design are that, it is straightforward in nature, it is easy to implement and it is easy to present and analyze the findings. However, the sequential explanatory design needs a length time to be spent in data collection for the two phases.

The sequential explanatory design is the mixed method used in this research. In this design I collected and analyzed quantitative data, then used the outcomes from the quantitative analysis to collect qualitative data in a second phase and lastly connected the phases by using the outcomes of the first phase of quantitative approach to design how the qualitative approach will be used.

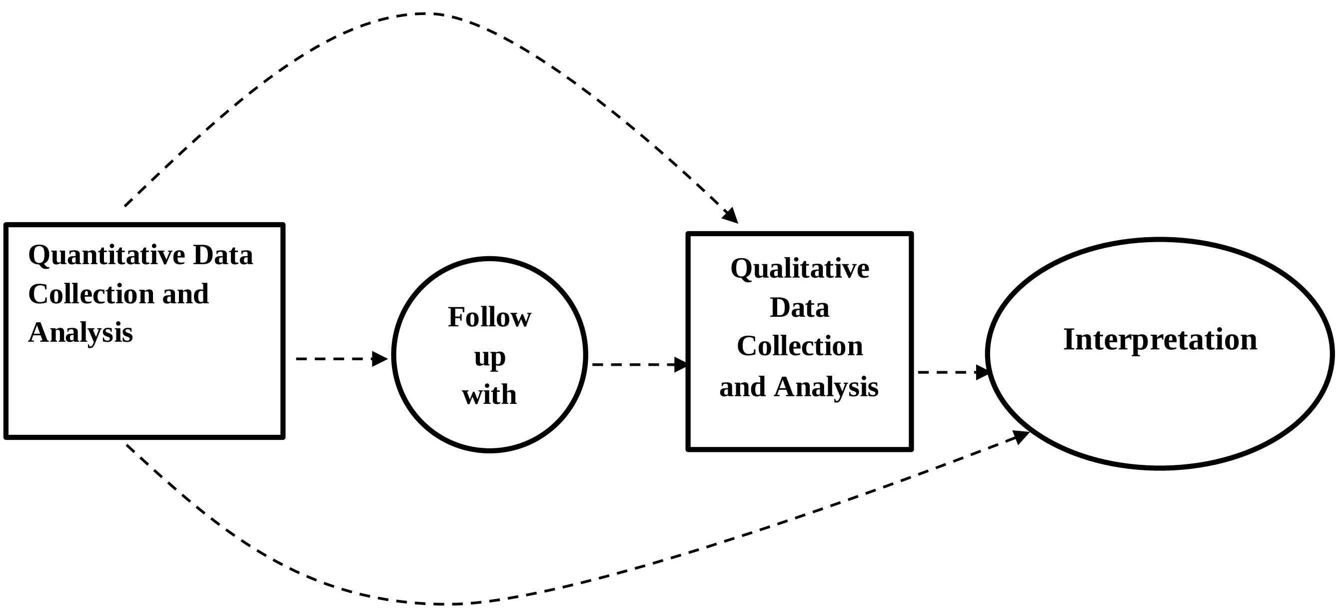


Figure 3.1. Flow Chart Showing Sequential Explanatory Design

The sequential explanatory design was used to help describe and investigate quantitative data that required further investigation and to use quantitative results to purposefully select the best participants for the qualitative study. The quantitative results were used to design a questionnaire that was used to collect qualitative data. I placed more weight on the quantitative data, as two of the three research questions were answered using the quantitative data. The initial quantitative results therefore informed the secondary qualitative data collection.

I chose this design because the research problem was quantitatively oriented. Quantitative data, which consisted of 604 scripts from 604 learners from six schools were recorded. The marks obtained by the learners were recorded item by item (question by question). The assessment tool (question paper) had 40 items (sub-questions) and these were all recorded on a spread sheet per learner and then loaded on to RUMM 2030.

This study fitted in well with the sequential explanatory design. Six hundred and four scripts with 40 individual responses for 40 items was going to be difficult to analyse using qualitative means first. The quantitative analysis using the Rasch method made it easier to take into consideration all 40 responses for each of the 604 scripts. The design made it easier for the researcher to identify trends, which then informed the drafting of questions for the qualitative data. The data collection methods are discussed in the section that follows.

3.7 Methods of Data Collection

The first set of data was the trial examinations scripts collected from six schools in Umlazi district in KZN. The scripts were collected from the six schools after they had been marked and moderated at district level. A questionnaire was designed to collect data from the teacher participants in the second phase of the study. Fifteen teachers, 10 from the participating schools and five from outside the participating schools agreed in principle to take part in the research and were given the questionnaires to fill in. The features of the design, design limitations and affordability of these data collection methods are discussed in the sections that follow.

3.7.1 The KZN Mathematics Trial Examination Paper 2

The trial examinations form part of formal assessment tasks that make up the formal program of assessment for the year. All of the formal assessment tasks are administered and marked by the subject teacher, so that progress reports are given to relevant stakeholders. According to the CAPS document (DoBE, 2011), all formal assessment tasks are subject to moderation for the purposes of quality assurance. The Grade 12 KZN provincial mathematics papers, which include the Mathematics Paper 2 and Mathematics Paper 1, help Grade 12 learners in their preparation for the final examination. Learners can use the trial examinations to identify sections which they need to put more effort into and sections where they need more assistance from the teacher. These trial examinations are used to give feedback to different stakeholders, such as parents, teachers, school management, DoBE and the learners themselves, on the preparedness of the learners for the final examination. If need be, intervention programs will be carried out depending on the outcome of these results. A team of examiners selected by the DoBE sets the mathematics trial papers and they are set for use under examination conditions.

I used the Mathematics Paper 2 scripts as the first set of data for the study. The schools wrote their trial examinations under examination conditions. Marking was done as per the rules of the schools and DoBE and moderation of the scripts was done at school level, cluster level and district level. As per the department protocol, at school level the head of department, is supposed to moderate 10% of the scripts. However, due to the large volumes of learners in most schools, the head of department ends up not moderating 10 % of the scripts. Instead, they focus on recording of marks and correct conversion of marks on the mark sheet. Also as part of the examinations protocols for the provincial examinations, the marking is supposed to be moderated at cluster level and district level. At cluster level teachers in the same cluster, teaching the same subject are supposed to moderate for each other and the cluster coordinator will then sign. Also 10% of the scripts are supposed to be moderated at this stage, but mistakes observed in the marking showed that the moderation was not focused on the correct marking but more on recording of marks and conversions, as the marking was not consistent with the marking memorandum. At district level, other teachers also moderate 10% of the scripts and the subject advisor will then sign to confirm compliance.

After the marks were verified and sent to the DoBE, I then borrowed the scripts for Mathematics Paper 2, photocopied them and returned the original scripts to the respective schools. During this process, I ensured that I did not interfere with the processes of the schools and the DoBE, in terms of how the trial examinations were conducted, or how the scripts were marked, moderated, verified and sent to the DoBE. The scripts were collected from schools after district moderations were done, to check whether the marking process was in line with the marking guideline that was supplied by the KZN education department.

The Mathematics Paper 2 question paper (instrument) had 40 items, each with marks ranging from 1 to 9. The paper had a total of 150 marks. The breakdown for the different sections is provided in Table 3.4.

Table 3.4

Break down of marks per section for the 2017 KZN Mathematics Trial Paper 2

Section	Question	Sub-questions	Marks	Total Per Section	Percentage of the paper	Marks as per CAPS document
Statistics	1	1.1	2	19	13%	20±3
		1.2	4			
		1.3.1	2			
		1.3.2	3			
	2	2.1	1			
		2.2.1	3			
		2.2.2	2			
		2.2.3	2			
Analytical Geometry	3	3.1	2	37	25%	40±3
		3.2	2			
		3.3	2			
		3.4	3			
		3.5	4			
		3.6	2			
	4	4.1.1	5			
		4.1.2	2			
		4.1.3	4			
		4.1.4	4			
		4.2.1	5			
		4.2.2	2			
Trigonometry	5	5.1	5	41	27%	40±3
		5.2	7			
		5.3	5			
		5.4	3			
	6	6.1.1	3			
		6.1.2	7			
		6.2	6			
		6.3	5			
Euclidean Geometry	7	7	9	53	35%	50±3
	8	8.1	3			
		8.2	3			
		8.3	4			
		8.4	3			
	9	9.1	4			
		9.2	5			
	10	10.1	7			
		10.2.1	4			
		10.2.2	2			
		10.2.3	4			
		10.2.4	5			

The KZN mathematics trial paper 2 covered four sections as shown in Table 3.4, statistics, analytical geometry, trigonometry and Euclidean geometry. The statistics section had two questions with eight sub-questions, contributing a total of 19 marks to the paper. The statistics section contributed 13% of the paper. The second section in the paper is analytical geometry, which had two questions, with 12 sub-questions, contributing a total of 37 marks, which corresponded to 25% of the paper. The third section was trigonometry, which had two questions, with 8 sub-questions and a total of 41 marks. The trigonometry section constituted 27% of the total marks. The last section covered in this trial question paper was Euclidean geometry, which had four questions and 12 sub-questions. The Euclidean geometry section contributed a total of 53 marks, which constitute 35% of the paper.

The CAPS document showed the breakdown of marks as recommended by the DoBE as that, statistics must contribute 20 ± 3 marks, analytical geometry must contribute 40 ± 3, trigonometry must contribute 40 ± 3 and Euclidean geometry must contribute 50 ± 3 (DoBE, 2011). A comparison of the number of marks per section in the trial paper and those recommended by the DoBE showed that the marks allocation was in line with the expectations of the department.

The question paper accommodated a range of cognitive levels, an aspect that will be discussed in detail in later chapters.

3.7.2 Questionnaire

Questionnaires are research tools that are widely used for data collection processes. A questionnaire for teachers was designed using some of the trends and responses observed in the learners' scripts. The questionnaire was used to capture the teachers' views on the assessment tool which was the KZN Mathematics Grade 12 preparatory examination and to get their views on the learners' responses as accurately and reliably as possible.

Open-ended questions were used as they provide more information which is impossible to obtain with closed-ended questions (Kronsnick & Presser, 2010). Teachers were given the freedom to answer the questionnaire in their own time.

A number of strategies were used to try and ensure a high response rate. My participation in mathematics content workshops in the Umlazi district in KZN helped me to be familiar with the teachers. My involvement in mathematics intervention programs also helped as some of the teachers completed the questionnaire soon after our regular weekly planning sessions.

One of the limitations of using a questionnaire is coverage error (Visser & Visser, 2000), since a sample can never be fully representative of the population it represents and sample bias will always be present to some extent. Effort was put into obtaining responses from all of the teachers from the participating schools. A variety of teachers with a range of teaching experience participated.

Whenever questionnaires are used there is always the challenge of non-responses, which impacts the reliability of the data (Visser & Visser, 2000). When teachers complete the questionnaires unattended non-responses becomes more common. Teachers were encouraged to respond to all questions and the completing of the questionnaires was normally done soon after the content workshop.

Whenever one is designing a questionnaire, consideration needs to be given to measurement error (Visser Visser, 2000). In questionnaire development, pre-testing the questionnaire is always vital in order to reduce measurement error. To try to overcome this problem I first gave the questionnaire to four teachers in one of the schools to respond to some of the questions. According to Krosnick and Presser (2010), the first step in questionnaire development must be to conduct an expert review to screen items and ensure unambiguous wording, after which a field test or pilot test is done. Research methods that do not use pilot testing can only provide insight into the possible problems, whereas methods that include pilot testing provide information about actual problems. The questionnaire used in this study was piloted with a group of four teachers who were experts in mathematics education. The questions were fine-tuned, and rephrased to remove ambiguity and to allow the teachers to give their views without narrowing the scope.

The questionnaire was given to 14 teachers, 10 from the participating schools and 4 from non-participating schools. All of the teachers who were given the questionnaire were teaching Grade 12 learners and most of them had more than five years teaching experience. Ten of the teachers were males and four were females. The questionnaire consisted of nine questions and teachers were required to answer all of them, however some teachers only responded to a number of questions leaving out others. Of the 14 questionnaires given to 14 teachers, only seven were returned. The other seven teachers cited different reasons for not completing the questionnaire, such as not getting time to complete the questionnaire, or being too busy and some even said they had misplaced the questionnaire and were not willing to complete another one.

The following questions were asked in the questionnaire:

1. From the analysis done, learners found the trial examination to be generally difficult. Can you please comment on this and give reasons why it might be so?
2. Despite the statistics questions (Items/Questions 1.1, 1.3.1 and 1.3.2) being placed among the easy questions (level 1 and 2 on the cognitive levels), the results showed that they were among the top most difficult questions. Why do you think these statistics questions were harder than expected?
3. The results also showed that the six easiest items/questions included analytical geometry questions. Can you give reasons on what might have been done that made learners perform better in these questions?
4. In general, the analysis showed that many items (questions) which were at the cognitive levels 1 and 2 were not always found to be easy by the learners. What are the possible causes of this?
5. The analysis showed that many items which were at the cognitive levels 3 and 4 were not always found to be difficult by the learners (for example Questions 3.3). What do you think are some of the reasons for this?
6. There were many instances where marking was not consistent with the marking guideline, for example, very few teachers in their marking were using the continuous accuracy (CA) factor hence depriving learners from getting potential marks. Can you give your views on the issue and why teachers sometimes do not adhere to this aspect?
7. In cases where an item carries 4 marks or has 4 categories and all the marks were accuracy marks (A), for example Question 4.1.4, results showed that the majority of learners were either getting a zero or were getting all marks. Very few learners got marks in between. Do you think this is a fair practice? Please explain.
8. The analysis showed that for some questions, learners who had the same total scores, from different schools or from same school, did not perform in the same way on most of the items. For example, learners from school AD did better in particular items than those from school SS, while for some other items it was the other way around. What are some of the contributing factors for these observations?
9. The analysis indicated items where learners who used English as the first home language, performed better than the English second home language speakers (Item

2.2.1). However, a different scenario was also observed for other questions like Question 5.1 where English second language speakers seemed to be performing better than the English first language speakers. What are some of the factors contributing to these findings?

3.8 Quantitative analysis

The 604 learners' scripts that were collected from six schools were captured on an Excel spreadsheet. The Mathematics Trial Paper 2 for 2017 had 40 items from the four sections on statistics, analytical geometry, trigonometry and Euclidean geometry. Table 3.4 shows the breakdown for the sections and the total number of marks per section.

The statistics section had 8 items with a total of 19 marks. The analytical geometry section had 12 items, contributing a total of 37 marks. The trigonometry section had 8 items with a total of 41 marks. The Euclidean geometry section had 12 items with a total of 53 marks. The items carried marks ranging from 1 to 9. However, some learners did not attempt some items and these were coded as 999. The code 999 for missing responses was used so that there would be a difference between a learner who attempted the item but scored a zero and a learner who did not attempt the question at all.

The data were then sorted and cleaned for use in the initial analysis and then for the Rasch analysis. The cleaning of the data involved making sure that the marks were captured properly and that there were no blank places for certain items. The data were checked to make sure that there were no scores above the maximum possible scores as these would be rejected during Rasch analysis in RUMM. The data were then run in RUMM 2030 and analysed using the Rasch analysis to investigate the preparatory paper as a whole, to investigate how the individual questions were working and to investigate the responses of the learners.
The Rasch model has different statistics that help to detect where data do not meet the expectations of the Rasch model. A number of analysis methods are used to allow the researcher to make informed conclusions.

The quantitative data analysis was carried out in two stages, the first of which was the initial data analysis followed by the Rasch analysis. The sections below describe the initial data analysis, and then provide a description of the Rasch analysis and how it was used to analyse the data and help answer the research questions.

3.8.1 Initial data analysis

The initial analysis of data helped to understand the general trends in the performances of the learners in the assessment tool. The Statistical Package for the Social Sciences (SPSS) was used to get the overall average scores and the graphs and tables for comparisons purposes and to get a clearer picture of learners' performance in different sections.

The overall average scores for all schools were generated and these were compared against the overall average scores for the individual schools. The overall average percentage scores per section were also generated for all schools together and for individual schools and were presented both graphically and in table form. Four sections were covered in the assessment tool; statistics, analytical geometry, trigonometry and Euclidean geometry.

The SPSS was also used to obtain the overall average percentage scores for each and every question for all of the schools and for individual schools, and also for each sub-question (item) for all schools and for individual schools. The number of zeroes and blank spaces were also determined for all schools and for individual schools.

3.8.2 Validity

Messick (1989) defined validity as the overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment. According to Winter (2000), validity refers to the extent to which an account contains the characteristics of the concept that it aims to describe or theorise. For McMillan and Schumacher (2006), validity refers to the extent to which the test instrument is serving the purpose it was supposed to serve. Validity is a meaning of the test scores not a property of the test or assessment, where the scores are a function of not only the items or stimulus conditions, but also of the person responding as well as the context of the item (Messick, 1989). The meaning or interpretation of the scores as well as any implications for actions that are needed, entails what needs to be valid (Cronbach, 1971). The Rasch analysis was used in this study to check whether the test scores were functioning as intended.

For Bush (2007) internal validity is the degree to which the research results cover the phenomenon being researched, and external validity refers to the extent to which the outcomes of the study can be applied in similar contexts. Cohen, Manion and Morrison (2011) described a number of different types of validity, where each type can be evaluated through expert judgement or statistical methods. For this study three main types of validity will be discussed; construct validity, content validity and citerion validity.

According to Messick (1989) the basis for construct validity lies in the integration of any evidence that comes from the meaning of the results. In assessments, construct validity can be threatened by not representing the construct well enough and by construct-irrelevant variance (Messick, 1989). Construct underrepresentation in a test occurs when important concepts of a chapter or subject are not included or covered in an assessment. Construct irrelevant variance is a result of a test having too much reliable variance, making the test easier or harder for the learners.

Messick (1989) explained content validity as the extent to which a teacher views the test as representing the concepts that are part of the work or domain of interest. Content validity does not consider how learners will respond to the test questions, but rather considers the representativeness of the content of the assessment tool. The Mathematics Trial Paper 2 covered all of the work prescribed in the NCS.

Criterion validity considers the degree to which the results obtained in a test correspond to other valid results in a similar context. Messick (1989) described criterion-related validity as based on the degree of empirical correlation between the test scores and criterion scores hence it relies on selected parts of the test's external structure. With criterion-based validity, interest is restricted to selected relationships, with measures deemed criterial for a particular applied purpose in a specific applied setting (Messick, 1989). The results of this study will be compared to results of other studies with a similar context as explained in the literature review section.

According to Cohen et al., (2011) validity is ensured in assessment at the design stage, at the data gathering stage, at the data analysis stage and at the data reporting stage. While designing the instrument it is important that assessment is relevant to the content covered in the classroom, that the assessment is understandable by the learners in terms of the language used and that the instructions and questions cannot be interpreted in many ways that can confuse the learners. It is also important that the assessment must not be too easy or too difficult for the learners. The assessment items with little discrimination of learner abilities must be avoided, and they must be not too long or too short. The assessment must not have many items covering the same concept.

At the data gathering stage, validity can be ensured by subjecting participants to the same conditions when administering tests, reducing withdrawal of participants from the study, encouraging participants to return completed questionnaires, and addressing other situational factors (Cohen et al., 2011). Learners were subjected to similar examination conditions when they wrote the Mathematics Trial Paper 2.

A number of measures can be followed to ensure validity at the data analysis stage. It is crucial that mark allocation is standard across all the scripts, and is done according to the marking guideline. The scripts must be moderated to ensure standardisation. When analysing the data, appropriate models must be used and researchers must avoid making conclusions and generalisations that cannot be justified from the results. The scripts used in this study were moderated at school level, cluster level and district level before I borrowed them for use in this study.

Validity at the data reporting stage was ensured by showing the context and parameters of the study during the time the data were gathered and sorted. It is also important for the researcher to specify the extent to which one can have confidence in the data, claims and conclusions made. Inaccurate reporting and misrepresentation of data must be avoided. It is also crucial for the researcher to make sure that the research questions are fully answered and results are released at the appropriate time (Cohen et al., 2011).

3.8.3 Reliability

Reliability refers to the degree to which an assessment will repeatedly give similar outcomes if it is used repeatedly in similar situations. According to Cohen et al., (2011) sampling is one of the key issues in considering the reliability of a questionnaire, especially a sample that does not represent the whole population, a biased sample and a sample that is not big enough. The reliability of the questionnaire was enhanced by making sure that at least one teacher from each of the participating schools filled in the questionnaire.

According to Cohen et al., (2011) reliability is a synonym for dependability, consistency and replicability over time, over instruments and over groups of respondents and is mainly concerned with precision and accuracy. Reliable research or assessment instruments would yield similar results if used with a similar group of learners in a similar environment. Although both qualitative and quantitative research have to have some degree of reliability, Guba and Lincoln (1994) suggests that the concept of reliability is largely positivist.

In assessment a reliable instrument/assessment tool can be determined in terms of its stability over a similar sample. If a test is simultaneously administered to groups of learners who have the same ability, then similar results should be obtained on the test (Cohen et al., 2011). An assessment can also be reliable in terms of its consistency over time. This means a test and a re-test will give the same results if they are administered within reasonable a period.

Reliability can also be considered in terms of equivalence, when the same type of test or instrument is constructed that will give the same results (Cohen et al., 2011). Pre-test and post-test are examples of this type of reliability, as they are similar types of instrument designed to measure the same concepts. Reliability as equivalence can also be shown when similar tests given to learners with the same characteristics will yield the same results when administered at the same time.

A number of factors can affect the reliability of assessment tools (tests and examinations). Fieldt and Brennam (1989) suggested four types of threat to reliability which include the individual's motivation, forgetfulness, carelessness, reading ability and exposure to solving the problem set which also has to do with the effects of practice. The test maker's idiosyncrasy and subjectivity, as well as instrument factors such as poor domain sampling, errors in sampling tasks, poor question items, the assumption or extent of unidimensionality in item response theory, and scoring errors, can all affect the reliability of an assessment tool (Fieldt & Brennam, 1989).

Airasian (2001) mentioned many threats to reliability in assessment tasks with respect to the examiners and markers. Reliability can be affected when errors are made in marking, such as attributing, adding and transferring of marks. Inter-rater reliability where different teachers allocate marks which are not the same for the same work, and the variations in awarding of marks for work that is close to grade boundaries can also affect the reliability of an assessment instrument.

Several sources of unreliability are associated with students and the teachers themselves. For Harlen (1994), a relationship between the marker and the learner, either positive or negative, can influence the results of an assessment. Other factors include students not always being clear on what is being asked in the question and students performing differently in questions which tested the same concepts (Black, 1988).

When it comes to the test items, reliability can be affected when the task is multidimensional (Cohen *et al*., 2011), as well as by the language used in the assessment and the assessor exerting influence on the learner being tested (Haladyna, 1997). The readability level of the task, the size and complexity of numbers or operations in a test, and tests must not favour one group more than another (Haladyna, 1997).

3.8.4 The background of the Rasch model

A Danish mathematician called George Rasch, responded to an education dilemma regarding equating tests by developing the RMT. The application of the RMT is increasing in psychometrics and in educational studies. The RMT implements addition measurement; that is, adding one more unit means the same amount extra, no matter how much there is already (Linacre, 2016).

The RMT of testing is based on the relationship between the performance of individuals on a test item and the individual's level of performance as an overall measure of the ability that the item was designed to measure (Bond & Fox, 2012). Ability is defined as the level of successful performance of the objects of measurement on the variable. The RMT is an example of a statistical model of item response theory that can be used to represent both the item's characteristics and the individual's characteristics (Linacre, 2016).

In simplified terms, the RMT sheds light on which items the learners actually experience as difficult, rather than what examiners expect learners to find difficult. The RMT offers the possibility of checking every item requirement, their collective functioning and various independence requirements (Dunne et al., 2012). The power of the RMT is harnessed when the item and independence requirements are each found to be reasonably satisfied by the test item data.

Applying the RMT when analysing tests helps to identify items that the examinees found easy and items they found difficult (Jacobs et al., 2014). Rasch (1980) provides the simple logistic (symbolic logic) model for dichotomous items, where learner ability is denoted by βi and the item difficulty is denoted by δi. These two constructs may be represented on the same scale. The equations from the RMT which are used to calculate the probability of a correct and incorrect response to a given item are outlined a little later in this chapter.

The RMT extracts the principles of measurement from the natural sciences and holds the measurement in the social sciences to the same standards (Bond & Fox, 2015; Boone, Staver & Yale, 2014; Long, 2011). Holmes (2005) argued that if psychology is the science of mental and behavioural processes, then the measurement of those processes should be scientific and should be guided by the principles of measurement.

What it means to measure is clearly put into use by examining or observing the principles of measurement (Combrinck, 2018). Measures should be equally ordered, additive and meaningful approximations of the construct under consideration (Engelhard, 2013; Massof, 2011; Wilson, 2005). Pupils who are able to answer more difficult items correctly must also be able to answer the easier items correctly. In the same way those who fails to answer the easier items correctly, will not be able to answer the difficult items. When the items and persons are perfectly ordered in this structured way, it is called the Guttman pattern. Guttman (1950) clearly defined the requirements that a set of responses should meet before a single meaningful score can arise from addition. These were if person A scores better than person B on the test, then person A should have scored all the items correct that person B has correct, and in addition some items that are more difficult. In a set of responses to more typical items in a mathematics test, it is less likely that the perfect Guttman pattern will found, but if the items are ordered according to their relative difficulty inferred from their total scores, then the Guttmen pattern is more likely (Andrich, 1985). The Guttman pattern is an important structure in the theorisation of the Rasch model (Cavanagh & Waugh, 2011; Engelhard, 2013). The Guttman pattern may not be evident for two reasons, which are if items are not assessing the same variable as expected, which implies that the scores on the items should not be summed to give a meaningful total score, and even if the items do assess the same variable, the items may be very close in difficulty and the persons may all be close in proficiency (Andrich, 1985). The probability of success on an item forms the basis of the Rasch model, on the assumption that the Guttman pattern will not appear and if it does, it will be an indication of extreme over-fit (Linacre, 2013; Wright, 1977).

Hubbard (2010) argued that in science, anything that can be observed either directly or indirectly, can be measured. Hence, it is possible to distinguish real science from pseudoscience by making use of measurable observations, leading to the realisation of obvious consequences for the construct, phenomenon or claim (Tal, 2017). For Hubbard (2010) if observations cannot be measured then they cannot fall under science and if it is, then it can be observed and detected in quantifiable amounts. In general, it is very challenging to obtain measurement in social science, as in most cases the constraints are latent. However, these measurements can be obtained by making use of the principles of measurement described in the section below.

3.8.5 Rasch analysis

The fundamentals of the RMT are covered in many publications (Andrich, 1988; Rasch, 1960/1980; Wright & Stone, 1979, 1999). The Rasch measurement model is a psychometric model that is used for analysing categorical data, such as questionnaire responses, as a function of the trade-off between a person's ability and item difficulty (Rasch, 1960/1980).
With RMT there is an assumption that for the construct of interest there exists a latent trait in the learner that may be gauged through operationalisation of the construct through the items (Bansilal, 2015). The latent trait is understood as a single dimension or scale along which the items can be located in terms of their difficulty. In the RMT, the learner ability, denoted by β_n, and item difficulty, denoted δ_i, may be represented on the same scale. Rasch analysis is then the formal testing of an outcome scale against a mathematical model developed by Rasch.

The term 'measurement' is often used loosely in the assessment of social and educational models. Considering measurement in the physical sciences, and a classical theory of measurement, we note that the property of invariance across the scale of measurement is required (Wright & Stone, 1999). For example, the measure of the height of a population at two different sites should not differ, nor should the means of measuring (system of units) change for different objects. Rasch analysis is the process of examining the extent to which the responses from a scale approach the pattern required to satisfy axioms of measurement in order to construct measurement. In RMT, and in conformity with classical measurement theory, the requirement is that the data must fit the model, rather than adapting the statistical model to fit the data. With RMT, the first step in approximating measurement is to define the construct being measured. The next step is to invoke a probabilistic process, a transformation that constructs natural units of measurement that are independent of both the construct and the persons being measured. This procedure involves converting a raw score percentage into its success-to-failure odds and then to its natural logarithm. Similarly, for items, the percentage of correct responses for the item is calculated and converted to a logarithm of the correct-to-incorrect odds for the item. This log-odds transformation of raw data is a first approximation of the Rasch measurement scale. Thereafter these estimations are then subjected to a series of iterations by the computer, allowing the student ability and item difficulty to be located on a common continuum so that a genuine interval scale using logits is produced.

The Rasch simple logistic model for dichotomous items is given in Equation 1. In RMT, the equation which relates the ability of learners and the difficulty of items is given by the logistic function:

$P\{X_{vi} = 1\} = \frac{e^{\beta v - \delta_i}}{1+e^{\beta v - \delta_i}}$ ....Equation 1

This function expresses the probability of a person *v*, with ability β_v responding successfully on a dichotomous item *i*, with two ordered categories, designated as 0 and 1. Here P is the probability of a correct answer; X_{vi} is the item score variable allocated to a response of person *v*, on dichotomous item *i*; and δ_i is the difficulty of item *i* (Dunne, Long, Craig & Venter, 2012).

Applying Equation 1, we can see that if a person *v* is placed at the same location on the scale as an item labelled *i*, then $\beta v = \delta_i$, that is, $\beta v - \delta_i = 0$, and the probability in Equation 1 is thus equal to 0.5 or 50%. Thus, any person will have a 50% chance of achieving a correct response to an item whose difficulty level is at the same location as the person's ability level. If an item's difficulty is above a person's ability location, then the person has a less than 50% chance of obtaining a correct response on that item, while for an item whose difficulty level is below that of the person's ability, the person would have a greater than 50% chance of producing the correct response (Dunne et al., 2012). Figure 3.2 illustrates this relationship for a single item.

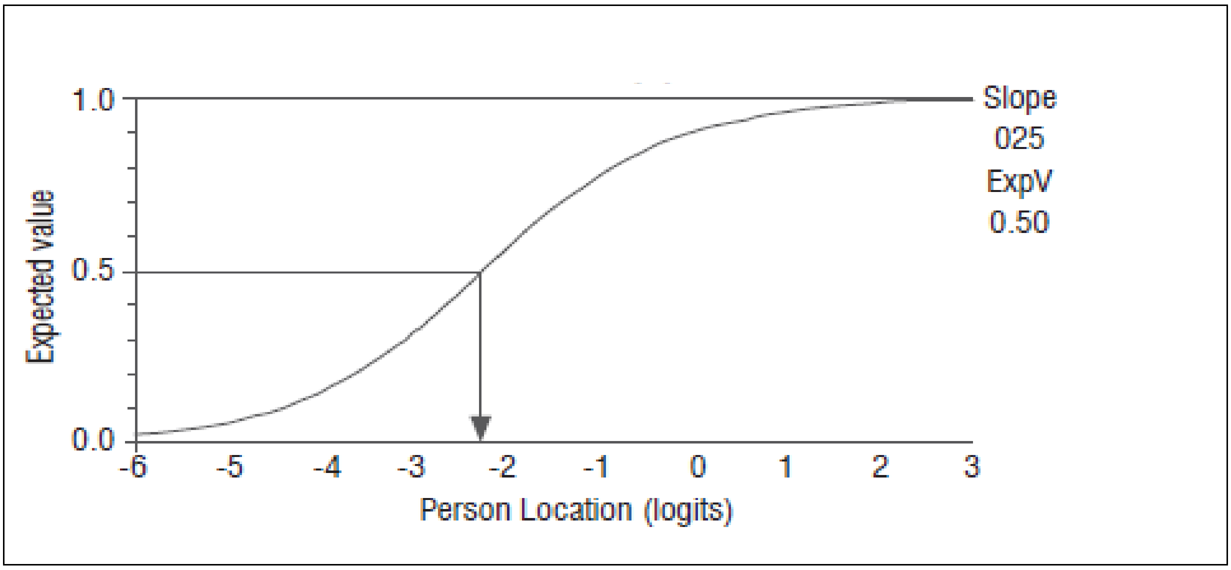


Figure 3.2. Item characteristic curve for an item located at -2.266 logits, representing a 0.5 probability of a correct answer (Bansilal, 2015)

The item characteristic curve (ICC) shown in Figure 3.2 represents the alignment of item difficulty and person ability. Learners are represented on the horizontal axis from low proficiency (to the left, towards -6) to high proficiency (to the right towards +3). The probability of a correct response is represented by the vertical axis (from 0 to 1). The person located at -2.266 logits has a 0.5 probability of answering this item correctly. Extrapolating from this notion means that out of 100 learners located at this point, the probability is that 50 will get the item correct.

With the ability of a person changing, the probability of a correct response to the item also varies. The probability that a person with low proficiency will respond correctly is correspondingly low, approaching 0 asymptotically as ability decreases. Symmetrically, the probability that a person with high proficiency will respond correctly is correspondingly high, and approaches 1 asymptotically as proficiency increases.

Equation 2 shows both the probability of responding correctly and incorrectly, where:

$$P\{X_{vi} = 0\} = 1 - \frac{e^{\beta v - \delta_i}}{1+e^{\beta v - \delta_i}} = \frac{1}{1+e^{\beta v - \delta_i}} \qquad \text{Equation 2}$$

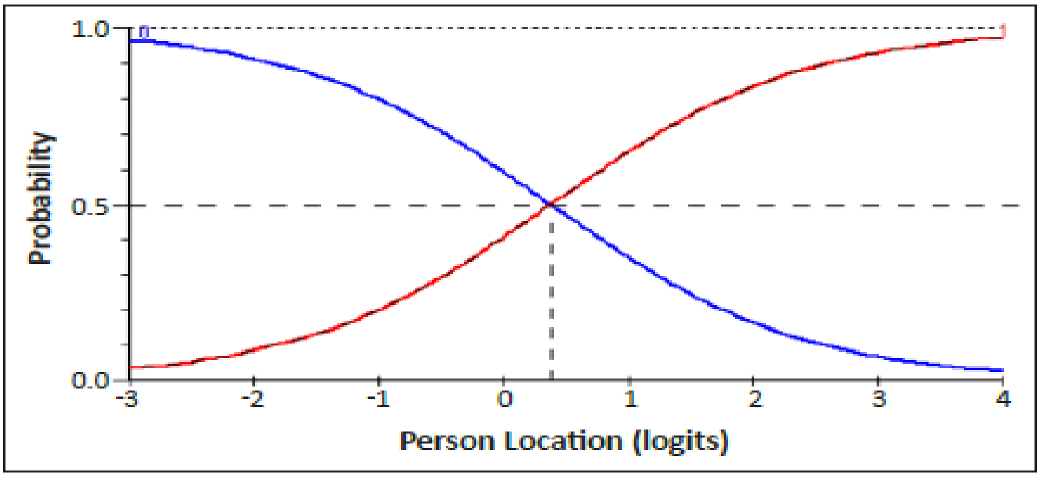


Figure 3.3. Category probability curve for an item showing the probabilities of scores 0 and 1 as a function of proficiency (Debba, 2018).

In Figure 3.3 the curve for the score of zero (0), represents the probability of a learner getting a score of zero at different proficiency levels along the ability continuum. At lower ability levels (around -3 logits), the probability of getting a score of 0 is very high but as the ability levels increases, the probability of getting a 0 decreases and hence the curve for the score of 0 also decreases. At high ability/proficiency levels (around 4 logits) the probability of getting a 0 is very low and so the curve for the score of 0 is approaching 0 (approaching the x-axis). The curve for the score of one (1) represents the probability of a learner getting a score of 1 at different ability levels. At low ability levels the probability of getting a score of 1 is very low, close to zero. As the ability levels increases, the probability of getting a score of 1 also increases as shown by the curve for the score of 1.

The location of the item is identified as the point on the ability scale where the probability curves of 0 and 1 intersect. At this point of intersection, there is an equal probability of either getting a 0 or a 1 (the probability of a response of 0 or 1 is equally likely). The item in Figure 3.3 has a location of 0.4 logits. The item is a dichotomous item; therefore, there is a probability of 0.5 of either response. As the proficiency of an individual decreases, the probability of a correct response decreases, and the probability of correct response increases as the proficiency increases, around this point of intersection of the probability curves of getting a 0 or a 1.

Rasch (1960) developed the simple logistic model for the analysis of dichotomously scored test items. Many assessment programs require greater precision or more information than a simple right/wrong scoring system. In these cases, polytomously scored items with several levels of performance may be required. Rasch's formulation of the model for polytomously scored items is an extension of the simple logistic model. Instead of dealing with dichotomous items with two response categories and possible scores of 0 and 1 only, it provides a model for test items with more than two response categories, with possible scores of 0,1,2,……..,m.. Andrich (1988) derived a model which gives the probability of a person with ability β_v being classified in a category x in a test item of difficulty δ_i, with m+1 ordered categories as:

$$P\{X_{vi} = 1\} = \frac{e^{(X\beta_v - \delta_i) - \sum_{k=1}^{x} \tau_k}}{\sum_{x=0}^{m} e^{(X\beta_v - \delta_i) - \sum_{k=1}^{x} \tau_k}}$$ Equation 3

where $x \in$ {1,2,......,m} and τ_k are the thresholds. In Equation 3 the threshold parameters are not subscripted by i, indicating that they are assumed to be identical across items, making it possible to estimate one of the thresholds which hold for all the items. If the thresholds are different across items, the model takes the form of Equation 4, shown below:

$$P\{X_{vi} = 1\} = \frac{e^{(X\beta_v - \delta_i) - \sum_{k=1}^{x} \tau_{ki}}}{\sum_{x=0}^{mi} e^{(X\beta_v - \delta_i) - \sum_{k=1}^{x} \tau_{ki}}}$$ Equation 4

The model of Equation 3 has become known as the ratings scale model and the model for Equation 4 has become known as the partial credit model. In this study partial credit model is used because it is less restrictive and allows the distances between the response categories to emerge from the data rather than being imposed on the data and also because each of the items had a different number of categories. The word 'threshold' defines the transition between two adjacent categories, for example between scoring 0 and 1 (τ_1), or scoring between 1 and 2 (τ_2). Figure 3.4 shows the category probability curves (CPC) for an item with more than two categories.

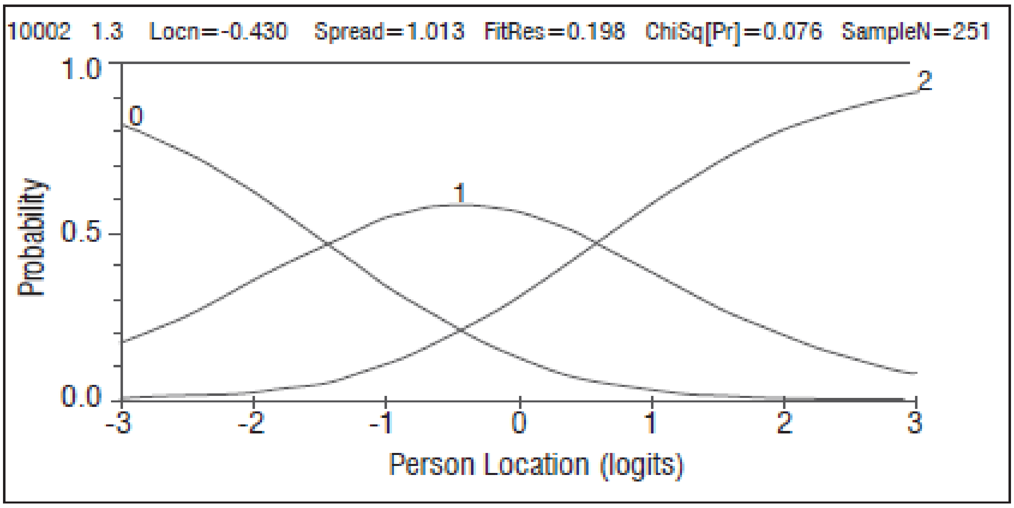


Figure 3.4. CPCs for an item with three categories (Bansilal, 2015)

In Figure 3.4, there are three category curves, corresponding to the probabilities of obtaining a score of 0, 1 or 2. The thresholds, and the categories they define, are naturally ordered in the sense that the threshold defining the two higher categories of achievement is of greater difficulty than the threshold defining the two lower categories of achievement. The first threshold(τ_1), which represents the point where a score of 1 is equally likely as a score of 0, is at a lower difficult level than the second threshold(τ_2), where a score of 2 is equally likely as a score of 1.

As seen in Figure 3.4, the two thresholds (point of intersections between the categories 0 and 1; 1 and 2) are naturally ordered, show that progressively more proficiency is required to score a 1 than a 0 and more proficiency is required to score a 2 than a 1. The curves are ordered and show that learners whose proficiency location is between 0 and approximately -1.5 logits are more likely to score a zero, while those between -1.5 and 0.6 logits are more likely to score a 1 while those who are located above 0.6 logits are most likely to score a 2. Compare this to Figure 3. 5 which shows the CPC for an item with disordered thresholds.

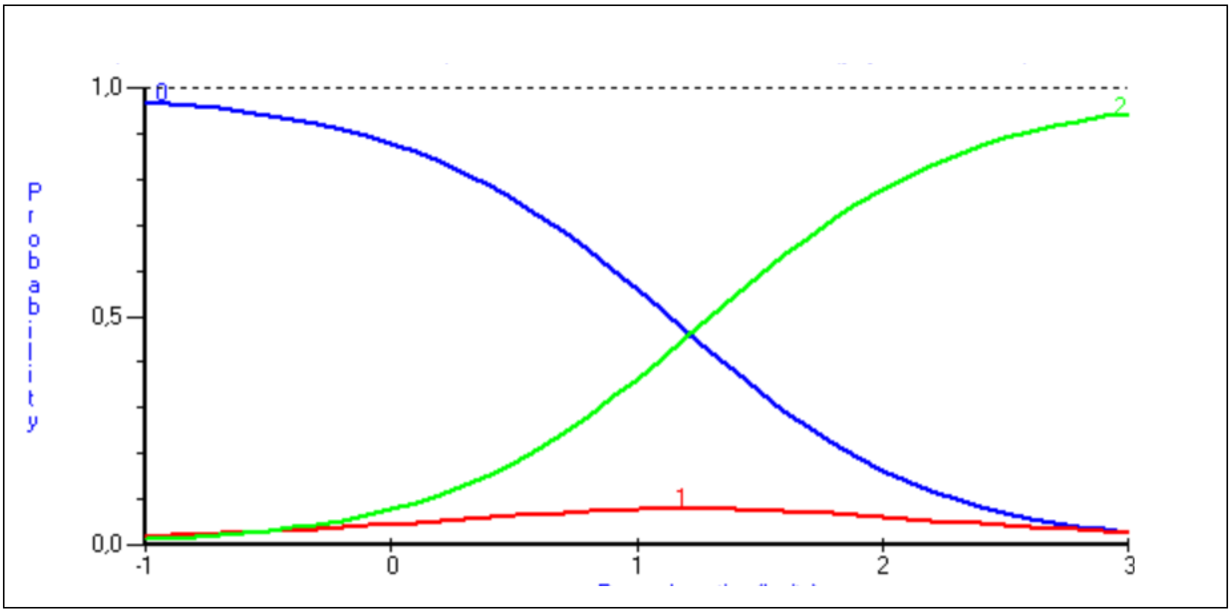


Figure 3.5. CPCs for an item showing disordered thresholds (Debba, 2018)

The CPCs in Fig 3.5 show that there is no interval in the proficiency continuum where the category curve for 1 is most likely, showing that the rubric is not working to distinguish between learners of different proficiency. Note too that the threshold between the curves for 0 and 1 are at a higher location than the threshold between 1 and 2, which does not make sense in this context.

3.8.6 Requirements of the Rasch model

The Rasch model claims that how a person answers a question depends on the proficiency of the learner and the difficulty of the question. The chances of a learner answering a question correctly, depend completely on the difference between the ability of the learner and the difficulty of the question.

The basic rule controlling the Rasch model is that a person who is more-able than another person should have a greater chance of successfully responding to a question than the person who is less able. In the same way, a question that is more challenging than the other implies that any individual has a higher chance of responding positively to the less challenging question than the more challenging question (Rasch, 1960/1980).

In RMT, the data must fulfil the requirements of the Rasch model. According to Douglas (1982), the data fitting the model is described as the correspondence between a data set and a statistical model. The model gives signs whenever the data fail to meet the principle of invariance, which may result in the data failing to fulfil the requirements of the Rasch model (Long et al., 2014). The extent of the difference between a person's answer to a question and the response predicted by the model is referred to as the fit residual. Analysing the fit of the data to the model leads to greater understanding of the construct in question or under study.

The data fitting the Rasch model requirements implies that the answers from the learners gives an invariant comparison and the sum of the scores characterise the persons, and the questions must work together and reinforce the evidence from each other (Andrich, 2011; Rasch, 1960/1980, Gigerenzer, 1993). Items do not fit the model in isolation, they work together consistently in the data set and hence an item fitting the model is the extent to which the item is working consistently with other items in the data set. Andrich (2011) suggests that the fit to the model need to be qualified, so that it indicates a degree of accuracy. High numbers of persons, items and categories and a better alignment between the persons and items, all work together to increase the potential precision (Andrich, 2011). Good precision of estimates, increases the power of the test and these will increase the chances of identifying any deviation from the model.

The size of the sample must not be too small, as any data can fit any model statistically if the sample is too small. There is no model that can describe a particular data set to an infinitely high level of precision, hence a large enough sample should be found to show that the data do not fit the model (Andrich, 2011; Gigerenzer, 1993). Andrich (2011) recommended a sample size of approximately 10 - 20 persons for each threshold for the set of items. Thus for 10 items (questions) each with 3 categories (2 thresholds each), and therefore a total of 20 thresholds, a sample of 15 times 20, which is equal to 300 persons can be used.

A good spread of persons in relation to the thresholds of the items is important for the power in the test of fit to be achieved. Responses will not have a range of probabilities if all persons have similar value and the persons were aligned to the items. Similar probabilities of responses occur when persons are similar in value and well aligned to the difficulty of an item with two categories, the probability of success is close to 0.5 and that of no success is also 0.5 and it is impossible to decide if a response in unlikely and therefore misfits in some sense.

The person separation index is used with respect to the distribution of the persons and the test of fit. In RUMM 2030, the person separation index is a scale used to measure the internal consistency reliability. When the person distributions are well aligned to the threshold distribution, it is similar in construction to the Critical Test Theory index of reliability (Andrich, 2011). Traditional reliability is estimated by the coefficient alpha, and if the persons are well aligned to the thresholds of the items, then the Person Separation Index is similar in value to that of the Coefficient Alpha, and it increases systematically with the number of item thresholds (Andrich, 2011). When the thresholds and persons are well aligned and the items are covering the range of person locations, a Coefficient Alpha index of 0.8 or greater can be obtained for 20 or more thresholds (Gigerenzer, 1993; Andrich, 2011). An index value of 0.85 gives an excellent power for the test of fit and an index value of 0.5 gives a very weak power in detecting misfit (Tennant & Conaghan, 2007).

3.8.7 Applications of the Rasch model

The Rasch model is a useful psychometric model used to analyse categorical data. The usefulness of the model comes from the fact that the analysis is a function of the trade-off between the respondent's abilities and the item's difficulty (Bond & Fox, 2012). The Rasch model is used in psychometrics, educational research, the health profession as well as market research due to its general applicability. RMT is a useful statistical tool for the improvement of tests.

The application of the RMT to mathematics assessment tests can be particularly helpful (Edwards & Alcock, 2010). Rasch-based research combines the rigorous measurement demands of the model with qualitative distinctions demanded by researchers in the mathematics education field (Callingham & Bond, 2006).

When tests are in the process of being designed, RMT can help identify questions that should be excluded or included in an assessment based on the functioning of the item (Anderson, Lai, Alonzo & Tindal, 2011). The technical adequacy of the measures can be increased while at the same time increasing the accessibility of the items, especially for those learners who struggle with mathematics (Anderson et al., 2011).

The process of assessment informed by RMT has the potential to aid both in classroom-based assessment and systematic assessment (Dunne, Long, Craig & Venter, 2012). This is because the requirements of RMT echo the requirements of good educational practice (Dunne et al., 2012). A well-designed assessment instrument/tool is able to provide detailed information of individual learners and at the same time inform external stakeholders on how healthy the education system actually is.

RMT has increasingly been used in mathematics education research in recent years, both in South Africa and internationally (Callingham & Bond, 2006). Recent research using RMT in South Africa include the analysis of assessments of teachers' ability (Bansilal, 2015), sequences and series in a high stakes examination (Jacobs et al., 2014), and assessments of geometry thinking levels (Stols et al., 2015).

Internationally RMT is used in the analysis of TIMMS, (Mullis, Martin, Foy & Drucker, 2012), measuring the mathematics readiness of US middle school students (Ketterlin-Geller, Gifford & Perry, 2015) and algebra reasoning of lower secondary students in the United Kingdom (Hodge, Brown, Coe & Kuchemann, 2012).

Many researchers have used RMT when investigating the mathematics attainment of learners as well as analysing how accurately an assessment measures this attainment. RMT can be used to determine whether an instrument is unidimensional, to determine whether items are anomalous and why, to develop a ranking of items in order of difficulty and to measure the mathematics attainment of the students within the tested population (Craig & Campbell, 2013). RMT theory can also be used to analyse the validity and reliability of algebra of mathematics items (Nopiah, Osman, Razali, Ariff & Asshaari, 2010). RMT can aid in increasing the quality and reliability of mathematics tests by reducing item gaps and identifying items that need to be rephrased or replaced (Nopiah et al., 2010). It is possible and beneficial to make use of RMT to improve questions and to test if questions are functioning differently in a cohort (for example success or failure on an item differs according to gender, language or location) (Bond & Fox, 2012). Hence the potential for RMT in mathematics education research is considerable (Dunne, Long, Craig & Venter, 2012).

3.8.7.1 Item misfit

When data are analysed using the Rasch analysis, the items may be shown to be working well as good indicators of learners' abilities or it may be observed that some of the items are highlighted as not functioning as expected and are thus considered as problematic. The Rasch model can also confirm that a scoring rubric (marking memorandum) is functioning as expected by the model or is failing to meet the model requirements.

The Rasch analysis test of fit places learners into groups of nearly equal sizes. In this study I used five groups (intervals) The average of each of the five groups becomes the horizontal coordinate (x-coordinate) of the points when plotting the ICC, representing the probability of answering correctly (Andrich, 1988). When the data fits the model, the expected curve (predicted by the model) and the actual proportions (the calculated average of the actual item scores in the five chosen groups) are very close to each other. Where the predicted curve and the actual proportions are not close to each other, and deviate substantially, it is an indication that the data do not fit the model. Observed proportions and the theoretical expectation can be related in four different ways; fairly good fit, under-discrimination, over-discrimination and haphazard misfit (Andrich, 1988).

When the observed proportions are flatter than the predicted curve, it shows that the item does not discriminate well enough, that is the item does not separate learner abilities well. This type of category is known as under-discrimination or under-fit. Under-discrimination shows that learners with low ability seem to achieve better than predicted, while the learners with higher ability are falsely estimated to respond to the item as if it was much more difficult than it really was.

The third category happens when the calculated proportions are steeper than the predicted curves, which result in discrimination that is more than predicted. Over-discrimination in items may benefit learners of high ability and disadvantage lower proficiency learners. In RMT, items which discriminate more may result in response dependence in one way or the other (Long et al., 2014).

The fourth category occurs when the calculated proportions are haphazardly but substantially different from the theoretical requirement. This pattern will require specific investigations of the construct, with a study of the relevance/suitability of the items.

Apart from checking whether the data fitted the model or not, I also checked to see whether the data showed differential item functioning (DIF), which will be explained in the next section.

3.8.7.2 Differential item functioning (DIF)

DIF occurs when different groups of people within the same sample who are at the same ability level overall, do not respond in the same manner to an individual item (Andrich & Hagquist, 2004). When questions work differently for two or more groups of people, the difference is a violation of the property of invariance and is also considered as a bias.

DIF does not necessarily mean that a certain group of leaners is performing better on an item as compared to another group of learners (Andrich & Hagquist, 2015). When one develops an assessment task, it is important to make sure that the task does not favor one group of learners and become too difficult to another group. The items have to work in the same way across different groups of learners who have the same academic characteristics.

According to Andrich and Hagquist (2004), there are two ways of identifying DIF. The first approach involves the estimations of a single set of parameters for each item and then studying the residuals identified by the different groups. For example, if the groups were men and women, one would analyse men's and women's responses in the same analysis and compare a mean residual for men with a mean residual for women. The second approach would be to estimate parameters in different groups and then compare the estimates. In other words, one would analyse the men's responses separately from the women's responses and then compare the estimated item locations for the men with the estimated locations for the women.

DIF can be detected graphically by means of ICC or it can be confirmed statistically.

3.8.7.2.1 Identifying DIF graphically

Within the tradition of the Latent Trait Theory, the fundamental idea of no DIF among groups is that for the same values of the latent trait, the expected value of a member from any group of individuals is identical (Andrich & Hagquist, 2004). The expected values are displayed in an item's ICC. Three different kinds of DIF can be identified in ICCs. In the first type of DIF, the curves are parallel, which is referred to as uniform DIF (Andrich & Hagquist, 2015). The second type of DIF is where the locations are the same but their slopes are different, that is referred to as non-uniform DIF and has non-parallel slopes. The third type of DIF is where both their slopes and their locations are different and is also referred to as non-uniform DIF since the slopes are not parallel.

To check whether an item has different meaning for females and males, I looked at the ICC for the item and display the observed values separately for males and females. I also checked for

DIF for the language factor, to see if for the same person location, those who use other languages as their home language have systematically higher observed mean values than the English home language speakers for this item.

3.8.7.2.2 Identifying DIF statistically

While the graphical displays of ICC, give the visual orientation of the data, DIF needs to be confirmed statistically through an analysis of the residuals.

The standardised residual of each person *n* to each item *i* is given by:

$$z_{ni} = \frac{x_{ni} - E[x_{ni}]}{\sqrt{V[x_{ni}]}}$$

For the purposes of the detailed analysis, each person is identified by the gender group, g, and by the class interval ***c***. This gives the residuals

$$Z_{n_{cg}i} = \frac{Z_{n_{cg}i} - E[x_{n_{cg}i}]}{\sqrt{V[x_{n_{cg}i}]}}$$

The residuals are analysed according to a standard analysis of variance (ANOVA). ANOVA is used statistically to determine whether there is a significant difference between the means of two or more groups of people. In identifying DIF, ANOVA is used to determine whether there is a significant difference among the mean residuals for the groups of interest (Andrich & Hagquist, 2004).

In ANOVA the F-ratios are constructed, which is a ratio of the estimated variance of residuals among groups and the estimated variance of residuals within the groups. With the assumption that the means come from a single random set of residuals from within the groups, then this ratio should be 1.0. An F-ratio bigger than 1.0 could indicate that there is a real difference between the group means, while a smaller than 1.0 F-ratio will lead to one concluding that the difference is not statistically significant.

Since we are working with estimates of variances in ANOVA, one cannot say with 100% certainty that an observed difference is real, since it maybe a difference occurring by chance, a peculiarity of the particular sample of residuals. In an ANOVA output, both an F-ratio and a probability are given. If the probability is less than a certain chosen criterion one can conclude that the difference between the means is statistically significant. That means the F-ratio of this magnitude would occur by chance less often than indicated by the probability (Andrich & Hagquist, 2004). If the probability is greater than the chosen criterion, one can conclude that the difference is not statistically significant. The statistical significance is usually determined using the criterion of 0.01 or 0.05. In this analysis, the statistical significance for the DIF criterion was set at the 0.05 level.

Apart from the data fitting the model or not and DIF, I also checked the data for any violations of the assumptions of local independence by checking for multidimensionality and response dependence which are discussed in the section below.

3.8.7.2.3 Multidimensionality and response dependency

Within the Rasch model, measurement is assumed to be unidimensional so that measuring a single construct can be built up to prevent varying difficulty along that measurement scale. According to Tennant and Conaghan (2007), when the construct of the assessment is unidimensional, the scores of all the items can be summed together in order to underlay a single construct. Validity and reliability are believed to be enhanced if the responses to the items are added up rather than taking the response to one item only. Each of these items needs to measure the same latent trait as the other items in the scale and provide some unique information that is not given by the other items in the scale or test. The Rasch analysis indicates as an anomaly those items that do not provide related information or do not perform independently.

In Rasch analysis, local independence refers to the notion that the differences in responses to an item are accounted for by the person parameter β, and for the same value β, there is no further relationship among responses (Marais & Andrich, 2008). Correlations between items not captured by the latent trait are indicative of local dependence. Local independence can be violated by multidimensionality and response dependency (Marais & Andrich, 2008). Christensen, Makransky and Horton (2017) studied how local dependence in the Rasch model can be identified using residual correlations. Christensen et al., (2017) concluded that within the parameter ranges that were tested, any residual correlation greater than 0.2 above the average correlation would appear to indicate local dependence and any residual correlation of independent items at a value greater than 0.3 above the average would seem unlikely.

With response dependency, the answer to an item influences the chances of success of the learner in the question that follows. The learners who answer the first item correctly have higher chances of answering the second item correctly and learners who answered the first question incorrectly are less likely to answer the second item correctly. The dependence of subsequent items on earlier items is regarded as unfair test practice (Marais & Andrich, 2008). Splitting an item into two or more groups depending on how the groups fared in the previous item can help to resolve response dependency.

With multidimensionality, there may be person parameters other than β that are involved in the response. Multidimensionality acts as an extra source of variation and the responses are less Guttman-like than they would be in the absence of dependence. High correlations between standardised residuals indicate a violation of the assumption of independence. Multidimensionality can also be observed by examining the principal component analysis (PCA), for any meaningful pattern which might suggest that the scale or test is not unidimensional.

Rasch programs generally run a principal component analysis of the residuals for detecting a sign of multidimensionality. According to McGill (2009), PCA procedure groups items into sets that correlate with one another but are relatively independent of other internally consistent subsets of items. The PCA explains the variability between items using a linear combination of the item scores using data reduction. Studies report that if there is no meaningful pattern on the residuals, the assumption of unidimensionality is achieved (Dawit, et al., 2014; Tennant & Conaghan, 2007).

3.9 Qualitative analysis

The teachers' responses to the questionnaire were analysed to check for similar observations and trends. Simon (2011) described qualitative data analysis as processing data, arranging the data, organising the data into units that are easy to work with, putting codes to the data, synthesising the data and observing similar patterns. Logical interpretation and analysis of the collected information is crucial in qualitative research. For Khan (2014), qualitative research is an organised way of showing and explaining participants' experiences and making sense of them. In this study, the data collected using questionnaires were organised and interpreted using the qualitative means.

According to Creswell (2009) data organisation includes attaching codes and themes to the data and considering how often the themes appear in the whole data set. For Charmaz (2015 coding of data involves categorising information using short descriptions as a way of summarising the collected data. I took all of the teachers' responses to the questionnaire and first organised the responses into main points per teacher. The main points from all of the teachers were then compared to check for similarities. The transcripts were studied and summarised many times to identify themes and categories. Themes were then created for common responses and for those responses with some similarities.

When the data collected from the questionnaires were organised into themes and codes, it was important to avoid being biased towards certain findings. Themes were used to distinguish elements for analysis, for example, CA marking, lexical density of questions and instruction, redundant marks, language factors, content knowledge of teachers and response dependency of items.

The teachers' responses helped the researcher in understanding some of the causes of the differences and similarities among the learners' responses both within the same school and between different schools. The similarities and differences between the teachers' responses meant that some observations were common between the schools and therefore needed a broader approach to solve, while some challenges were restricted to individual schools and then needed specialised and customised attention and solutions.

Trustworthiness was introduced for the purposes of representing many of the quality measurements in qualitative research and key criteria of validity in qualitative research are credibility, transferability, dependability and conformability (Lincoln & Guba, 1985). According to Cohen et al., (2011) the validity of a questionnaire can be achieved when participants complete questionnaires to the best of their abilities. Validity in a questionnaire can also be achieved when a researcher makes multiple follow-up rounds to request returns of questionnaires. The researcher must stress the importance and benefits of the questionnaire, follow up questionnaires with personal telephone calls, maintain certain features of the questionnaire and encourage participation by a third party (Hudson & Miller, 1997). For this study I explained at length the importance of the questionnaires and their benefits to the teachers, who were the respondents. I also begged the teachers to be honest and accurate when completing the questionnaire. Telephone calls were made to teachers to try and ensure the questionnaires were returned.

3.9.1 Credibility

Credibility is the extent to which findings are congruent with the data presented (Lincoln & Guba, 1985). Human beings are the primary instruments of data analysis in qualitative research and carry inherent biases which make them unable to capture the objective truth (Merriam, 2009). However, researchers can use a number of measures to increase the credibility of their findings.

One of the most common strategies used by researchers is triangulation, where one uses many data collection methods, many sources of data, many investigators and many theories (Cohen et al., 2011). The use of different data collection methods and different data sources strengthens the credibility of the findings as those from one data source can be validated by those from a different source. This study used quantitative (mathematics trial examination) and qualitative (teacher questionnaires) data collection tools. Data were collected from different schools in different settings to ensure that deductions were not based on one type of setting.

The credibility criterion in this study was also addressed by keeping to the learners' written answers in the scripts and comments given by the teachers when responding to questions in the questionnaire. Some of the learners' work is presented without alterations to ensure credibility. The same data (learners' scripts) were analysed using different methods (Rasch analysis and item analysis) to triangulate the findings.

Credibility is also strengthened when the participants are given the freedom to participate (Shenton, 2004). Teachers who filled in the questionnaire were given the choice to participate or not. Some of the teachers from the participating schools chose not to participate. When teachers were responding to questions on the questionnaire I made it clear that there was no right or wrong answer, and that they were participants in a research project and not subjects of an experiment.

3.9.2 Transferability

In qualitative studies, transferability can be compared to external validity in quantitative research, and it is defined as the degree to which the outcomes of one study can be applied to other situations (Merriam, 2009). According to Lincoln and Guba (1985), in comparison to how a quantitative researcher establishes external validity by determining statistical confidence limits, the qualitative researcher can only provide thick descriptions of the context, making it possible for researchers wishing to replicate findings to make judgements about the extent to which the study is transferable.

Case studies tend to seek analytical generalizability rather than statistical generalisability (Yin, 2003). Analytical generalisability refers to the case's ability to contribute to the expansion and generalisability of the theory. One case can assist the researcher to understand similar cases or situations. It therefore provides the opportunity to test a theory in more than one empirical case rather than generalise the findings of a few cases to others in general. In this study the learners could represent many other learners in similar learning environments in Grade 12, and the teachers could represent many other teachers with similar teaching experience and qualifications, teaching mathematics and working in similar circumstances.

I have provided a detailed description of the settings and made the data and statistical reports available to allow other researchers to apply the findings to similar settings.

3.9.3 Dependability

Dependability in qualitative research is analogous to reliability in quantitative research. Lincoln and Guba (1985) argue that credibility cannot be shown without showing dependability. However, researchers should be explicit about how they have established both aspects of trustworthiness in their studies (Shenton, 2004). The inclusion of the process of how data were collected and analysed can enable other researchers to repeat the work if necessary. Dependability considers whether the research was done in a traceable manner and presented in clear and consistent format. In this study a detailed description of the research design and sample selection have been provided. The design and final versions of all the data collection tools were included, and the data analysis procedures were also described in detail. The details of the transcription of the marks as well as the cleaning of the data are included in Chapter 4. The results of the Rasch analysis and the interpretation of statistics, appear in Chapter 5. Chapter 6 details the conditions under which post-hoc rescoring of items was carried out and the results of the rescoring. The details of the results for the DIF and ways in which response dependency was identified appear in Chapter 7. The discussion of the teachers' responses to the questionnaire is provided in Chapter 8.

3.9.4 Confirmability

The concept of confirmability is comparable to objectivity in quantitative research (Shenton, 2004). Confirmability is the objectivity or neutrality and control of research bias. Confirmability should be maintained through the data transcripts and data analysis as well as the description of the findings and relating them to the literature. The findings of the study must be supported by the data collected. Triangulation can help reduce researcher bias. As mentioned previously, multiple data collection methods and data sources as well as triangulation were used to reduce researcher bias and ensure that findings and deductions came from the collected data. All pertinent details about the analysis of the data from the transcription to the final results are provided in the results sections. At any point, any researcher could use the same data and carry out a similar analysis and would arrive at similar conclusions.

3.10 Ethical issues

An application was made to the University of KwaZulu-Natal and KZN Department of Education stakeholders in respect of ethical clearances required. The ethical clearance from the University of KwaZulu-Natal is attached in Appendix B and the ethical clearance from the KZN Department of Education is attached in Appendix C. I also asked for permission to carry out the study from the parents, learners, principals, and heads of departments and subject teachers of the participants. The copy of the consent letter for the school governing bodies and school principals is attached in Appendix E, the consent letter for teachers is attached in Appendix F while that for learners is attached in Appendix G. All the concerned parties were asked to sign the consent forms, which informed the participants about their rights and also described the steps to be followed and their purpose. The consent form also stipulated time frames that were to be expected for the completion of each instrument. Participants were informed that withdrawal from the study could be undertaken at any stage of the research without any prejudice to the them. Informed consent letters were given to each participant, where they and their parents were required to give consent for their participation. They were also re-assured that pseudonyms would be used so their identity would not be revealed in any way by the study. Participants were told of their right not to participate, their right to withdraw from the study at any given point and that they would not be coerced in any way to give any information against their will. Ten schools were initially targeted for the research but four schools did not give consent. Eight hundred and fifteen consent forms were issued to learners; 604 learners gave their consent, while 211 declined to participate.

3.11 Limitations

Interviews with the learners were supposed to be carried out to investigate why they performed the way they did, however because of time constraints the interviews were done with the teachers. This was mainly because by the time the first phase of the research was finished, learners had written their final examinations and it was impossible to set up meetings with them. A further limitation was that not all teachers who initially agreed to complete the questionnaire, did so and returned it to the researcher. A further limitation that affected some of the analysis was that the examination scripts were not re-marked. Hence if teachers did not follow the marking memorandum diligently, this may have affected some of the conclusions that were made.

CHAPTER 4 INITIAL ANALYSIS OF PERFORMANCE IN THE MATHEMATICS PAPER 2 TRIAL EXAMINATION

4.1 Introduction

This chapter provides a quantitative analysis of the overall performance of the Grade 12 learners across the six schools for their Mathematics Trial Examination Paper 2 scripts. It focuses on: overall results across sections; overall results per section for each school; results per question for each school.

4.2 Overall Results Across the Schools

The overall results for the schools on the Mathematics Paper 2 trial examination showed that the highest average score was on Question 3, which was based on analytical geometry, the average mark was 46.4%. The second highest average score was on Question 7, based on Euclidean geometry, where the average mark was 37.4%, followed by Question 9 (Euclidean geometry) and Question 6 (trigonometry), where the overall average marks were 35.8% and 30.4% respectively.

The overall average marks for all of the learners were below the pass mark of 30% for six of the questions (1, 2, 4, 5, 8 and 10). The lowest overall average percentage was obtained for Question 10 (Euclidean geometry), with an average mark of 14.1%, followed by Question 1 (statistics) with an average mark of 20.6% and Question 2 with an average of 23.6%.

The overall average mark for all of the learners for the whole paper was 27.8%, which is below the pass mark of 30%. This is an indication that the paper was challenging for more than half of the number of learners involved in the study. To get an overall picture of how the schools fared on the paper, the average percentages per school will now be discussed.

Table 4.1

Average percentage marks per school per question

	Question										
School	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Total**
AD	20.3	31.9	63.2	29.1	33.5	39	43.7	42.7	49.3	20	**36**
SS	42.9	26.5	63.7	40	40.3	40.7	45.7	34.2	46.9	23	**39.8**
KS	23.9	45.9	58.7	33.7	31.6	40.7	49.9	33.2	45.5	15.1	**35.7**
SB	4.8	10.6	39.2	16.7	22.5	26.5	31.5	20.6	21.7	12.4	**20.8**
LM	11.4	4.6	22.3	11.7	23	9.6	27.8	8.1	20.2	6.7	**14**
FT	8	13.3	23.7	11.9	42.8	19.6	19.8	10.3	24.3	4.9	**18.1**

Table 4.1 shows that school SS had the highest overall average score of 39.8% followed by school AD, with an average of 36% and school KS with an average of 35.7%. These three schools had average marks above the minimum pass mark of 30%, while the three lowest performing schools (SB, LM and FT), had average marks below the pass mark of 30%. The three lowest performing schools did not do well on Questions 1, 2, 4 and 10.

Schools which obtained the highest overall average percentage marks did not always get the highest average percentage marks for individual questions. For Question 1, only school SS had an average mark above the minimum pass mark of 30%, all of the other schools had averages below 30%, with school SB having the lowest average of 4.8%. Schools FT and LM also had low average percentages, which were 8% and 11.4% respectively. Question 1 required learners to use grouped data provided in a frequency table to estimate the mean height of the data and to draw an ogive which they then had to use to estimate the median height and interquartile range. The schools did poorly in this question as compared to other questions.

For Question 2, school KS had the highest average mark of 45.9%, while school LM had the lowest average mark of 4.6%. The trend continued for the rest of the questions, where the top three schools (AD, SS and KS) would interchangeably get the highest average marks, while the three lowest performing schools (SB, LM and FT) would also interchangeably get the lowest average mark, except on Question 5. For Question 5, the highest average percentage mark was obtained by school FT, at 42.8%. School SS had the second highest average score at 40.3%. This was the only question in where a school in the bottom half of the group, in terms of total average mark, had the highest average mark.

4.3 Overall Results per Section

The data were used to compare the performance of learners from the six schools across the different sections. The Mathematics Trial Examination Paper 2 had four sections; statistics (20±3 marks), analytical geometry (40±3), trigonometry (40±3) and Euclidean geometry (50±3). A summary of the average marks per section per school are shown in the Table 4.2.

Table 4.2

Average percentage marks for each section per school

	All	AD	FT	KS	LM	SB	SS
Statistics (Q1+Q2)	21.9	25.8	10.2	33.1	8.6	7.3	36.0
Analytical Geometry (Q3+Q4)	33.8	43.0	16.7	43.8	16.0	25.9	49.6
Trigonometry (Q5+Q6)	29.5	36.3	31.0	36.2	16.1	24.5	40.5
Euclidean Geometry (Q7+Q8+Q9+Q10)	24.5	34.6	12.0	30.6	13.0	19.2	33.6
Total	27.8	36.0	18.1	35.7	14.0	20.8	39.8

Table 4.2 shows that the highest overall average mark of 33.8% was obtained in analytical geometry, followed by trigonometry at 29.5%, Euclidean geometry at 24.5%, with the lowest overall average percentage obtained in statistics. The table also shows that school AD, school KS and school SS had average marks which were above the overall average mark for all of the schools in the four sections. Schools FT, LM and SB had average marks below the overage average percentage for all of the schools. There was one exception for trigonometry, for school FT which had an average mark of 31.0%, which was above the overall average of 29.5%. Figure 4.1 below shows the comparison between sections for the schools.

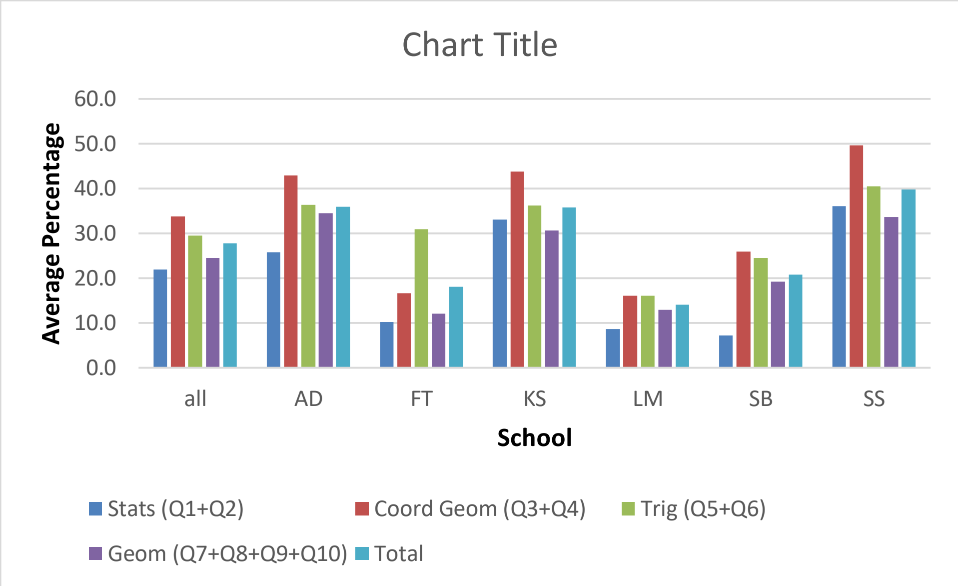


Figure 4.1. Learners' performance across sections for the six schools

Figure 4.1 shows that for the high performing schools (AD, KS and SS), the average mark for all sections was above 30%, except for one section in the case of school AD where the average mark was below 30% for statistics. School FT had relatively high average percentage for trigonometry, of above 30%, while for all other sections this school's average marks were less than 20%. School LM had average marks below 20% for all of the sections, while for school SB the average marks were more than 20% for two sections (coordinate geometry, trigonometry) and below 20% for statistics and Euclidean geometry.

4.3.1 Analysis of results for statistics

Questions 1 and 2 were based on statistics. The marks for each sub-question were recorded for each school, and analysed in Excel to try and understand how learners fared on these questions per school and overall. Table 4.3 shows the average percentages for each school for Questions 1 and 2.

Table 4.3

Average percentages for Questions 1 and 2

	Question							
School	**1.1**	**1.2**	**1.3.1**	**1.3.2**	**2.1**	**2.2.1**	**2.2.2**	**2.2.3**
All	7.6	37.6	12.4	12.0	39.9	24.8	21.9	15.6
AD	17.8	36.2	15.1	7.9	50.0	39.5	31.6	11.8
FT	0.5	19.5	1.9	1.6	37.5	4.8	20.7	6.7
KS	9.2	38.2	17.6	18.7	57.6	46.9	47.6	36.8
LM	2.2	28.7	1.1	1.5	29.2	0.7	1.7	1.1
SB	0.0	11.1	2.5	1.3	21.5	18.1	4.4	0.0
SS	14.1	74.4	27.9	30.0	37.4	31.3	16.8	23.7

Table 4.3 shows that for the sub-questions in Question 1 and 2 which covered statistics, three schools (AD, KS and SS) continuously had average percentage marks above the overall average percentage for all six schools. The other three schools (FT, LM and SB) had average scores below the overall average percentage mark for all of the sub-questions. The learners did not do well on Question 1.1, where the average mark for the schools was below 20%. The question required learners to determine the estimated mean from the grouped data that were provided. Learners in school SB scored an average of 0% and in school FT the average percentage was 0.5 for Question 1.1, which showed that the learners lacked knowledge of how to calculate the estimated mean from the grouped data. There was better performance on Question 1.2 which required learners to draw an ogive curve to represent given data. Learners needed to find the cumulative frequencies for the data and then draw the ogive. The average marks were below 30% for Questions 1.3.1 and 1.3.2. Question 1.3.1 required learners to use the ogive that they drew in Question 1.2 to find the median height of the palm trees, while Question 1.3.2 required learners to use the same ogive to estimate the interquartile range, which proved to be a difficult task for them. Question 1 is shown in Figure 4.2.

QUESTION 1

The table below shows the heights of palm trees in a park.

HEIGHT IN CM	NUMBER OF PALM TREES
$120 < x \leq 135$	1
$135 < x \leq 150$	15
$150 < x \leq 165$	45
$165 < x \leq 180$	28
$180 < x \leq 195$	1

1.1 Determine the estimated mean height of the palm trees in the park. (2)

1.2 Draw an ogive to represent this data. (4)

1.3 Use your ogive curve to estimate the:

1.3.1 median height of the palm tree. (2)

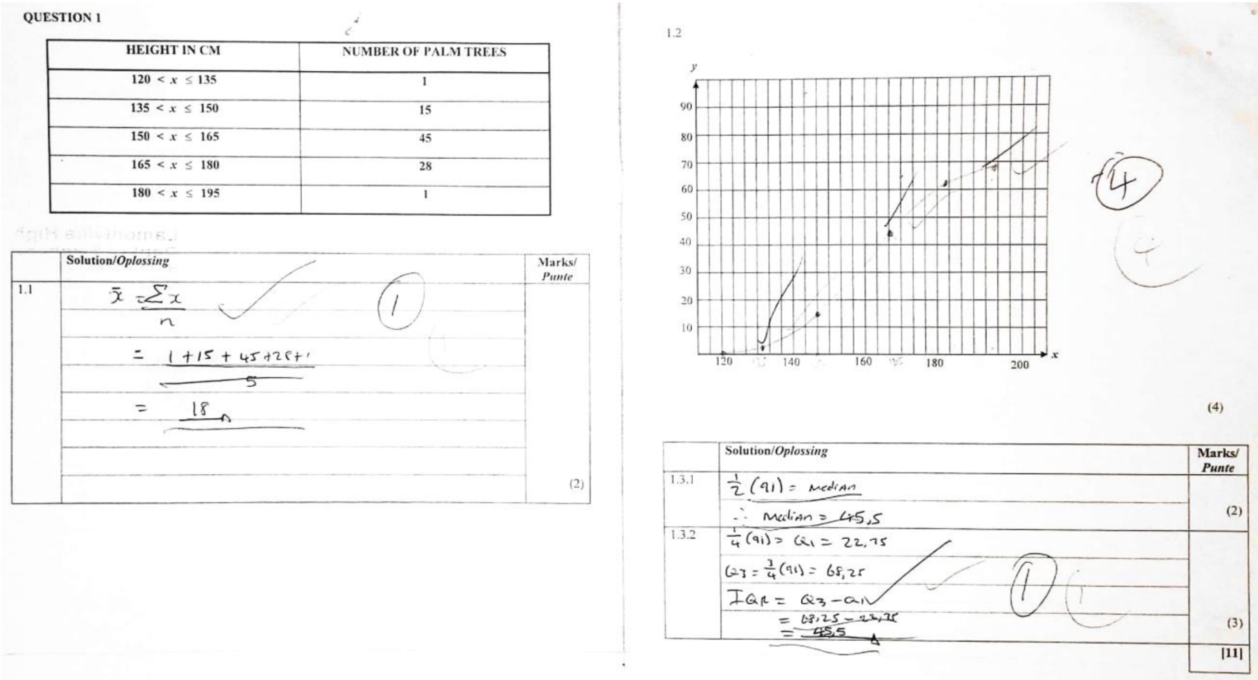
1.3.2 interquartile range (IQR). (3)
[11]

Figure 4.2. Question 1

Question 1.3 in Figure 4.2 requires that learners needed to use their ogive curve to answer Questions 1.3.1 and 1.3.2. The learners who were not successful in answering Question 1.2 were not able to respond correctly to Questions 1.3.1 and 1.3.2. Some learners who managed to draw the ogive correctly were not able to use their graphs to find the median and the interquartile range. Samples of the learners' responses are shown in Figure 4.3.

The common mistake made by learners in answering Question 1.1 was that of summing up the frequencies (number of palm trees) and dividing the sum by 5 (which represented the number of class intervals). This misconception is shown in Figure 4.3 and suggested limited conceptual understanding. The two responses in Figure 4.3 show that the two learners did not understand the difference between grouped data and ungrouped data. Many learners used a similar approach and did not get the question correct.

a.



QUESTION 1

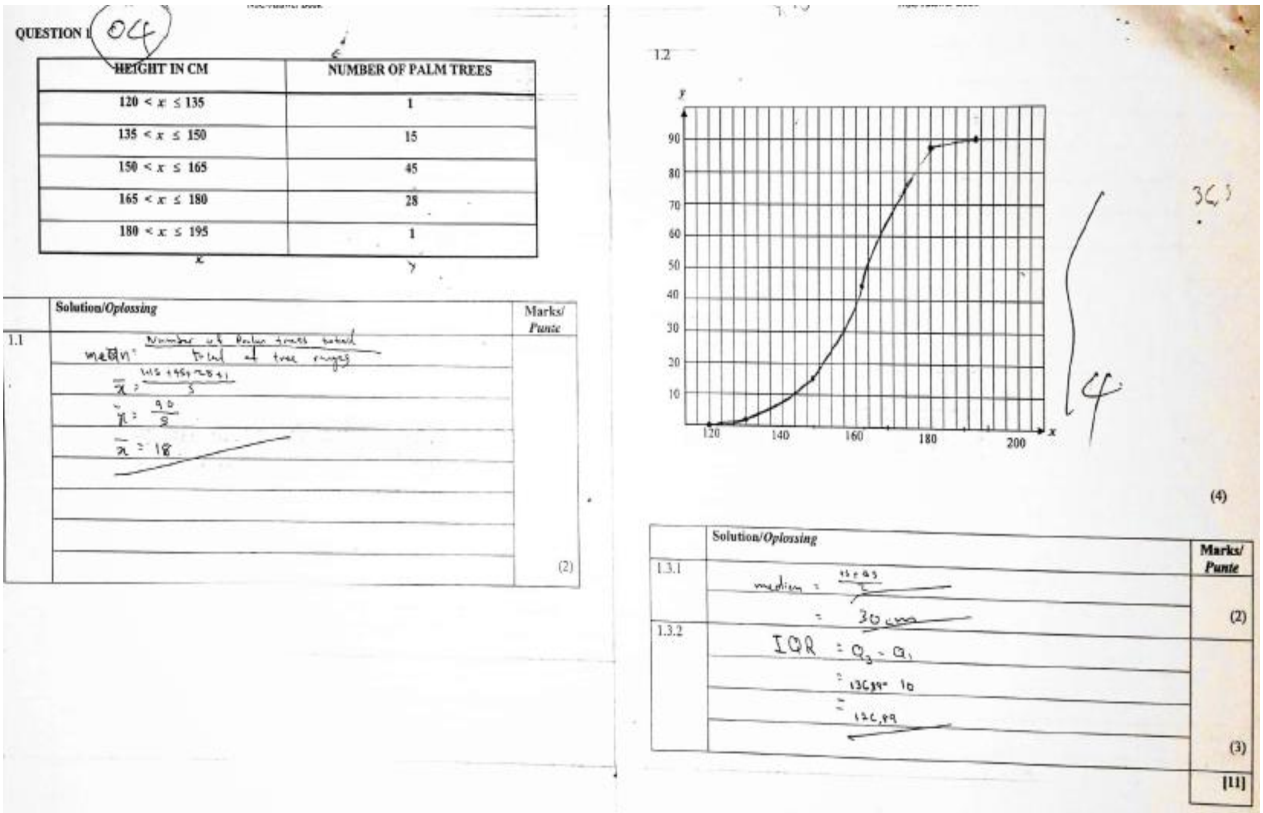
HEIGHT IN CM	NUMBER OF PALM TREES
$120 < x \le 135$	1
$135 < x \le 150$	15
$150 < x \le 165$	45
$165 < x \le 180$	28
$180 < x \le 195$	1

	Solution/*Oplossing*	Marks/ *Punte*
1.1	$\bar{x} = \frac{\sum x}{n}$ $= \frac{1+15+45+28+1}{5}$ $= 18$	(2)

1.2

	Solution/*Oplossing*	Marks/ *Punte*
1.3.1	$\frac{1}{2}(91) =$ median ∴ Median = 45,5	(2)
1.3.2	$\frac{1}{4}(91) = Q_1 = 22,75$ $Q_3 = \frac{3}{4}(91) = 68,25$ IQR = $Q_3 - Q_1$ = 68,25 − 22,75 = 45,5	(3)
		[11]

b



QUESTION 1

HEIGHT IN CM	NUMBER OF PALM TREES
$120 < x \le 135$	1
$135 < x \le 150$	15
$150 < x \le 165$	45
$165 < x \le 180$	28
$180 < x \le 195$	1

	Solution/*Oplossing*	Marks/ *Punte*
1.1	mean: Number of Palm trees total / total of tree ranges $\bar{x} = \frac{1+15+45+28+1}{5}$ $\bar{x} = \frac{90}{5}$ $\bar{x} = 18$	(2)

1.2

	Solution/*Oplossing*	Marks/ *Punte*
1.3.1	median = $\frac{15+45}{2}$ = 30 cm	(2)
1.3.2	IQR = $Q_3 - Q_1$ = 136,89 − 10 = 126,89	(3)
		[11]

Figure 4.3. Sample of learners' responses to Question 1

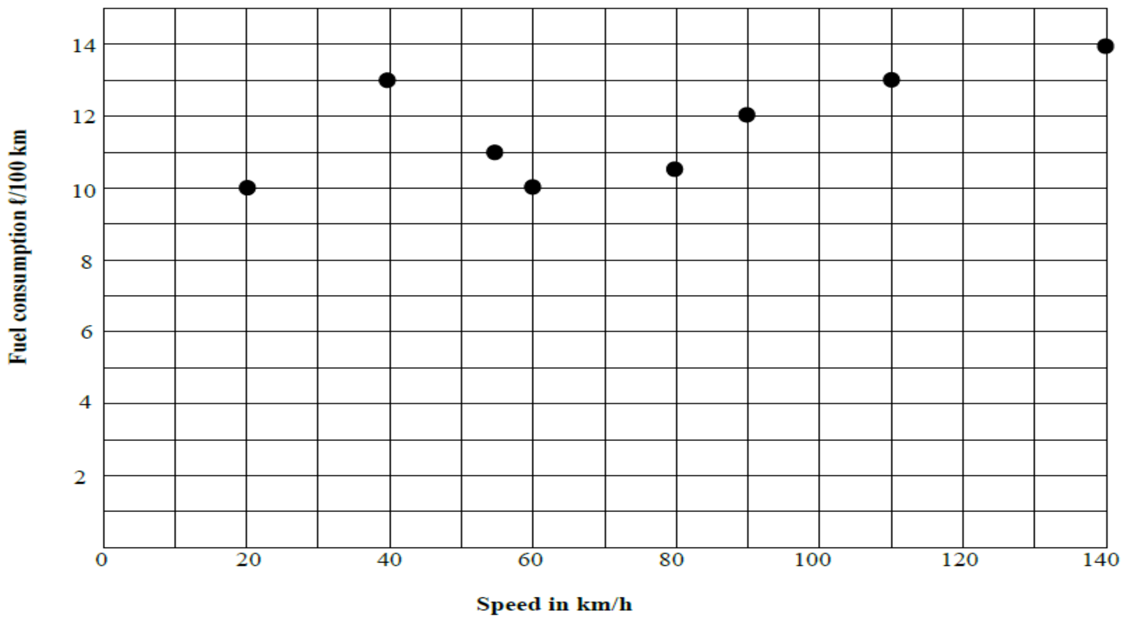
Students' performance on Question 1.2 was quite good with a number of learners getting full marks as shown in Figure 4.3. However some learners did not get full marks because they did not ground the graph to point (120,0). Another common mistake was learners sketching the upper limit of each class interval against the frequency of that interval instead of the cumulative frequency. Questions 1.3.1 and 1.3.2 were poorly answered as many learners who sketched the ogive graph correctly were not able to use the graph to determine the median height and the interquartile range, as shown by the responses in Figure 4.3. Many learners calculated the position of the median, and the position of the lower quartile and upper quartile and left those as the final answers shown, a misconception that was common among the responses.

For Question 2.1 learners were required to identify an outlier from a scatter plot and write down its coordinates. The learners' average mark was above 30% in four of the schools, for Question 2.1. Question 2.2.3 required learners to determine the average fuel consumption of the motor car from the data given on a scatter plot. As the data were given on a scatter plot, the fuel consumption was represented on the vertical axes. Learners were then supposed to add all the y-coordinates of the points and divide the sum by the number of points. Learners did not do well on this question, with school SB getting an average of 0%. Question 2 is shown in Figure 4.4.

Many learners were able to identify the outlier and wrote the coordinates at (40,13). How-ever some learners confused cases when one is given two sets of data with cases when one is given one set of data and hence wrote the outlier as the point on the extreme right of the scatter plot which was point (140,14).

QUESTION 2

The scatter plot below shows the fuel consumption versus the speed of a motor car.



2.1 Identify an outlier. Write down its co-ordinates. (1)

2.2 Determine:

2.2.1 the equation of the regression line excluding the outlier. (3)

2.2.2 the correlation coefficient excluding the outlier and explain the type of correlation. (2)

2.2.3 the average fuel consumption of the motor car. (2)
[8]

Figure 4.4. Question 2

The responses of the learners to sub-questions 2.2.1, 2.2.2 and 2.2.2 showed that very few learners were able to input all of the coordinates into the calculator to determine the equation of the least squares regression line, the correlation coefficient and the average fuel consumption. Learners obtaining the wrong values may be as a result of errors in reading off the coordinates of the scatter plot. Responses of two learners to Question 2 are shown in Figure 4.5.

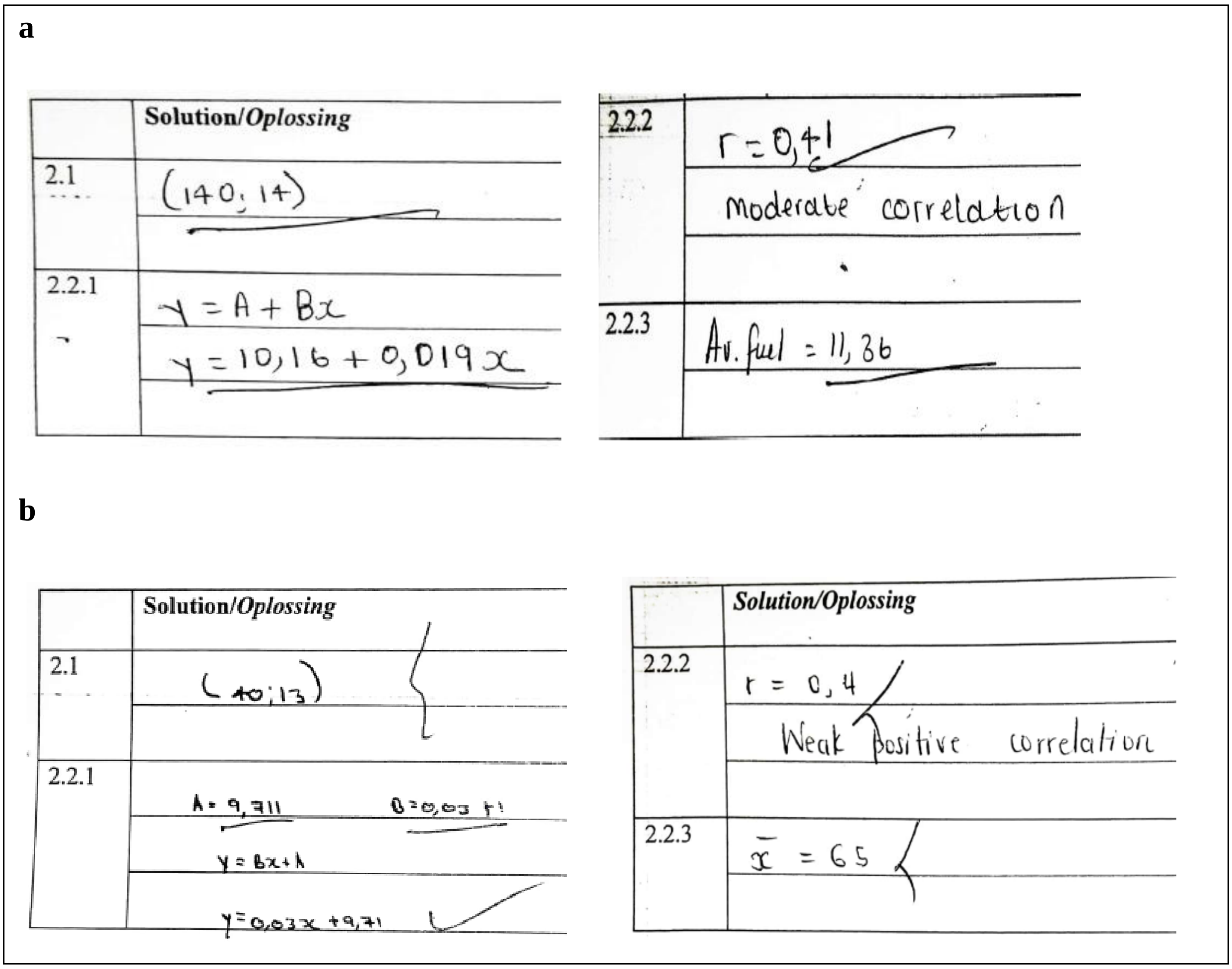


Figure 4.5. Responses by two learners to Question 2.

The responses by many learners to Questions 2.2.1, 2.2.2 and 2.2.3 show no clear pattern, as shown in Figure 4.5. The responses in the figure showed that the values were not derived from the data set and hence learners did not access the necessary statistical information to answer these questions successfully. The responses of the two learners to Question 2.2.1 showed that they understood the equation that was supposed to be determined, as they were able to write down the formula correctly, but the values of A and B were both incorrect.

Comparison between the schools on the sub-questions for Questions 1 and 2 showed that three schools (SS, KS and AD) performed above the overall average percentage mark, while the other three schools (LM, SB and FT) performed below it. In Questions 1.2, 1.3.1 and 1.3.2, school SS had an average percentage mark higher than those of all the other schools, while for Questions 2.1, 2.2.1, 2.2.2 and 2.2.3, school KS had the highest average percentage mark. The difference in performance in Questions 1 and 2 showed that learners in school SS understood grouped data better than those in the other schools, while learners in school KS had a better understanding of bivariate data on which Question 2 was based. No one school obtained the lowest average mark continuously for Questions 1 and 2, but the three lowest performing schools (LM, SB and LM) obtained the lowest mark interchangeably.

The average mark for Question 1 for all of the participating schools was 20.6%, with an average of 2.27 out of the possible 11 marks. Three of the schools had average marks which were above the overall average of 20.6%, school SS having an average of 42.9%, school KS an average of 23.9% and school AD an average of 23.3%. The other three schools had average percentage marks below the overall average for all of the schools. School LM had an average of 11.4% and school FT an average of 8%, while school SB obtained the lowest average of 4.8% on Question 1. The poor performance by the three lowest performing schools on Question 1 showed that the learners did not understand grouped data well enough and hence could not determine the estimated mean, estimated median and interquartile range and could not represent the data using an ogive curve. Question 1 received the second worst scores after Question 10.

Question 2 had an average score of 23.6% for all of the participating schools. Three schools had average percentage marks above the overall average for all six schools. School KS had the highest average percentage mark of 45.9%, followed by school AD at 31.9% and school SS at 26.5%. Learners from these three schools were able to respond successfully to questions involving bivariate data, by identifying the outlier, determining the least squares regression line, correlation coefficient and the average fuel consumption. The other three schools had average marks below the overall average percentage score for all of the schools, indicating a lack of understanding of the bivariate data. School FT had an average score of 13.3%, School SB had an average percentage score of 10.6%, while the lowest average percentage was from school LM with an average percentage score of 4.6%. Figure 4.6 compares the performance of the schools in statistics.

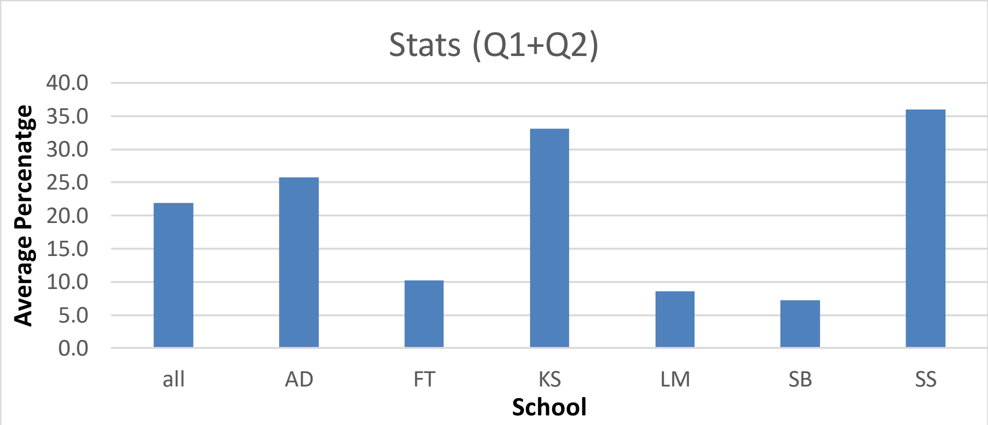


Figure 4.6. Overall average percentages for schools in statistics

Figure 4.6 shows that schools SS and KS performed better on the Statistics questions (Questions 1 and 2) with an average mark above the 30% pass mark, while the other four schools did not achieve an average above the pass mark. The lowest average marks were recorded at schools LM and SB, average marks of 8.6% and 7.3% respectively, showing that learners at these schools did not understand the concepts of grouped data and bivariate data.

4.3.2 Analysis of results for analytical geometry

Questions 3 and 4 were based on Analytical Geometry. The trend observed for Questions 1 and 2 was also observed for Questions 3 and 4 and their sub-questions. The three schools, AD, SS and School KS have average scores above the overall average scores. There was an exception in Question 3.1, where School KS had an average mark of 78%, which was below the overall average of 79.1%. The average percentages for Question 3 and 4 are shown in Table 4.4.

Table 4.4

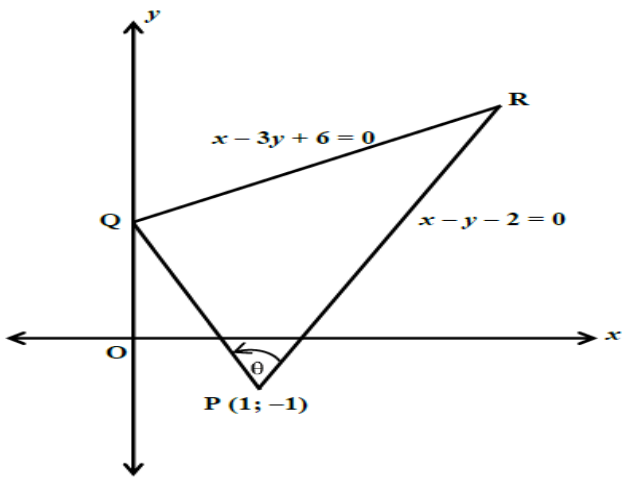
Average percentage marks for Questions 3 and 4

	Question											
School	**3.1**	**3.2**	**3.3**	**3.4**	**3.5**	**3.6**	**4.1.1**	**4.1.2**	**4.1.3**	**4.1.4**	**4.2.1**	**4.2.2**
All	79.1	47.0	40.8	46.3	32.9	45.9	31.4	41.1	34.9	20.9	17.5	3.1
AD	89.5	65.1	61.8	67.5	49.3	57.9	39.7	52.0	48.4	19.4	13.4	0.0
FT	64.9	19.7	15.9	23.4	11.3	19.2	21.2	15.9	11.1	11.3	6.7	0.5
KS	78.0	50.8	40.8	57.9	53.4	77.2	24.0	60.0	55.0	29.6	29.8	6.8
LM	65.7	23.0	23.6	22.8	4.2	12.4	31.2	9.6	7.0	9.0	3.6	0.0
SB	78.5	43.7	32.9	41.8	16.8	43.0	30.6	34.2	15.8	12.0	7.1	0.0
SS	95.0	72.9	64.9	59.8	50.4	54.6	42.1	62.6	57.1	34.7	32.5	7.3

Table 4.4 shows that learners did well in sub-questions 3.1, 3.2, 3.3, 3.4, 3.6 and 4.1.2. Question 3.1 required learners to show the coordinates of point Q, which was the y–intercept of line QR. The equation of the line QR was provided and learners were required to substitute into the equation the x-coordinate ($x = 0$) to get the y-coordinate. In Question 3.2, learners were required to write down the gradient of the line QR. Since the equation of line QR was provided, learners were supposed to make y the subject of the formula and then write down the value of the gradient. Learners also did well on Question 3.3 which required them to prove that an angle was a right angle, by multiplying the gradients of the two lines which met to form the angle. The performance of the learners was also better for Question 3.4, which required learners to calculate the coordinates of R, the point of intersection of two lines (line QR and PR) whose equations were given. Question 3 is shown in Figure 4.7.

QUESTION 3

In the figure below, PQR is a triangle with P(1 ; −1) . Q is a point on the y-axis. The equations of QR and PR are $x - 3y + 6 = 0$ and $x - y - 2 = 0$ respectively. Given $Q\hat{P}R = \theta$.



3.1	Show that the co-ordinates of Q are (0 ; 2).	(2)
3.2	Write down the gradient of QR.	(2)
3.3	Prove that $P\hat{Q}R = 90°$.	(2)
3.4	Calculate the co-ordinates of R.	(3)
3.5	Calculate the area of Δ PQR.	(4)
3.6	Calculate the length of PR. (leave your answer in the simplest surd form).	(2)
		[15]

Figure 4.7. Question 3

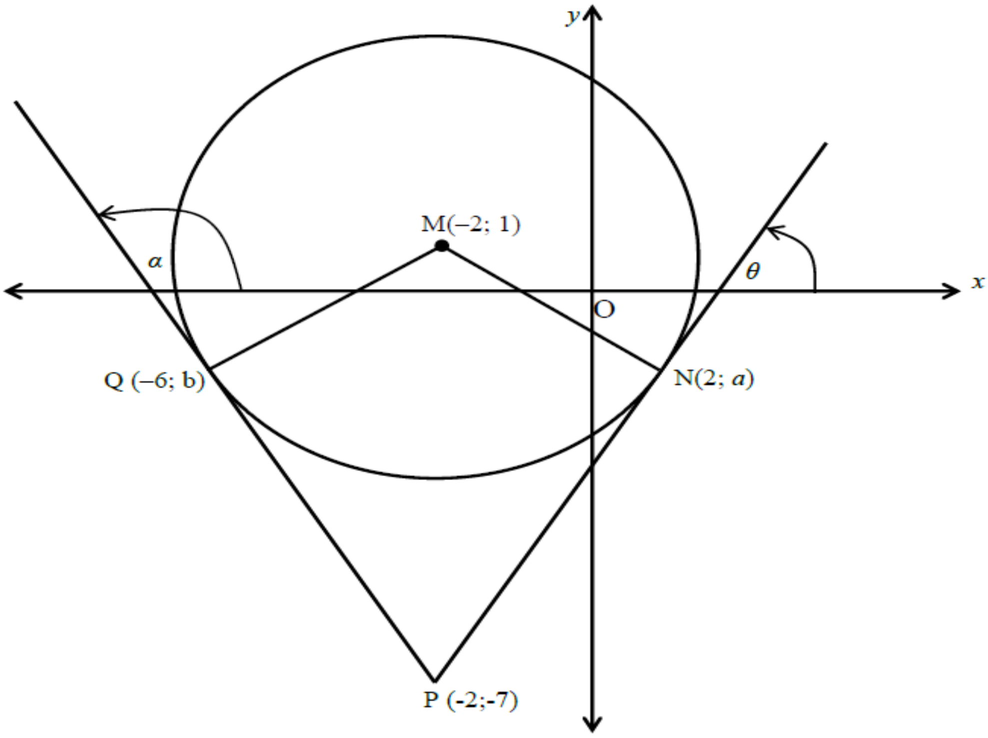
Most of the learners were able to determine the coordinates of Q, which is the y-intercept of line QR and line QP as shown in Figure 4.7. The high average percentage shown in Table 4.4 for Question 3.1, 79.1%, provide evidence that many learners understood the concept of the y-intercept.

The average mark for Question 4.1.2 was 41.1%. Question 4.1.2 required learners to determine the equation of the circle. The centre of the circle was provided, but for the radius learners were supposed to use the value that they were required to deduce from the previous question. Learners could use the value that they were supposed to deduce in Question 4.1.1 even if they failed to deduce it and hence the average percentage was better than that for Question 4.1.1.

Questions 4.1.4, 4.2.1 and 4.2.2 were poorly answered, and the averages for these questions were below the 30% pass mark. Question 4.1.4 required learners to show that $tan^2\alpha + tan^2\theta = 2$, where θ and α were the angles of inclination of the tangent lines PN and PQ respectively. The learners were supposed to use the gradients of the two tangent lines calculated from the previous question. Question 4.2.1 required learners to show that two circles whose equations were provided, touch internally. The question required learners to compare the distance between the centres of the two circles and the sum of their radii. Question 4.2.2 achieved the lowest score among the analytical geometry questions, with an average of 3.1%, this question required learners to determine the equation of the common tangent to the two circles. Question 4 is shown in Figure 4.8.

QUESTION 4

4.1 In the diagram below, MN is a radius of a circle with centre M(–2 ; 1). The co-ordinates of N are (2 ; a) and $a < 0$. The co-ordinates of P are (-2'-7). PQ and PN are tangents to the circle at Q and N respectively.
The coordinates of Q is (–6 ; b). PM is parallel to the y – axis.



4.1.1 Deduce that $a = -3$. Show all your workings. (5)

4.1.2 Determine the equation of the circle. (2)

4.1.3 Calculate the gradient of the tangents at Q and N. (4)

4.1.4 If the angle of inclination of lines PN and PQ are θ and α respectively, without using a calculator, show that $\tan^2 \alpha + \tan^2 \theta = 2$. (4)

4.2 The circle defined by $(x + 1)^2 + (y - 1)^2 = 16$ has centre C and circle defined by $x^2 + y^2 - 2y = 8$ has centre D.

4.2.1 Show that the two circles touch each other internally. (5)

4.2.2 Determine the equation of the common tangent to the circles. (2)
[22]

Figure 4.8. Question 4

Table 4.4 showed that three of the participating schools had average scores of 0% for Question 4.2.2 and one school had an average score of 0.5%. Many learners did not even attempt to answer this question, with the few who responded showing a lack of conceptual understanding. Some of the learners' responses to Question 4.2.2 are shown in Figure 4.9.

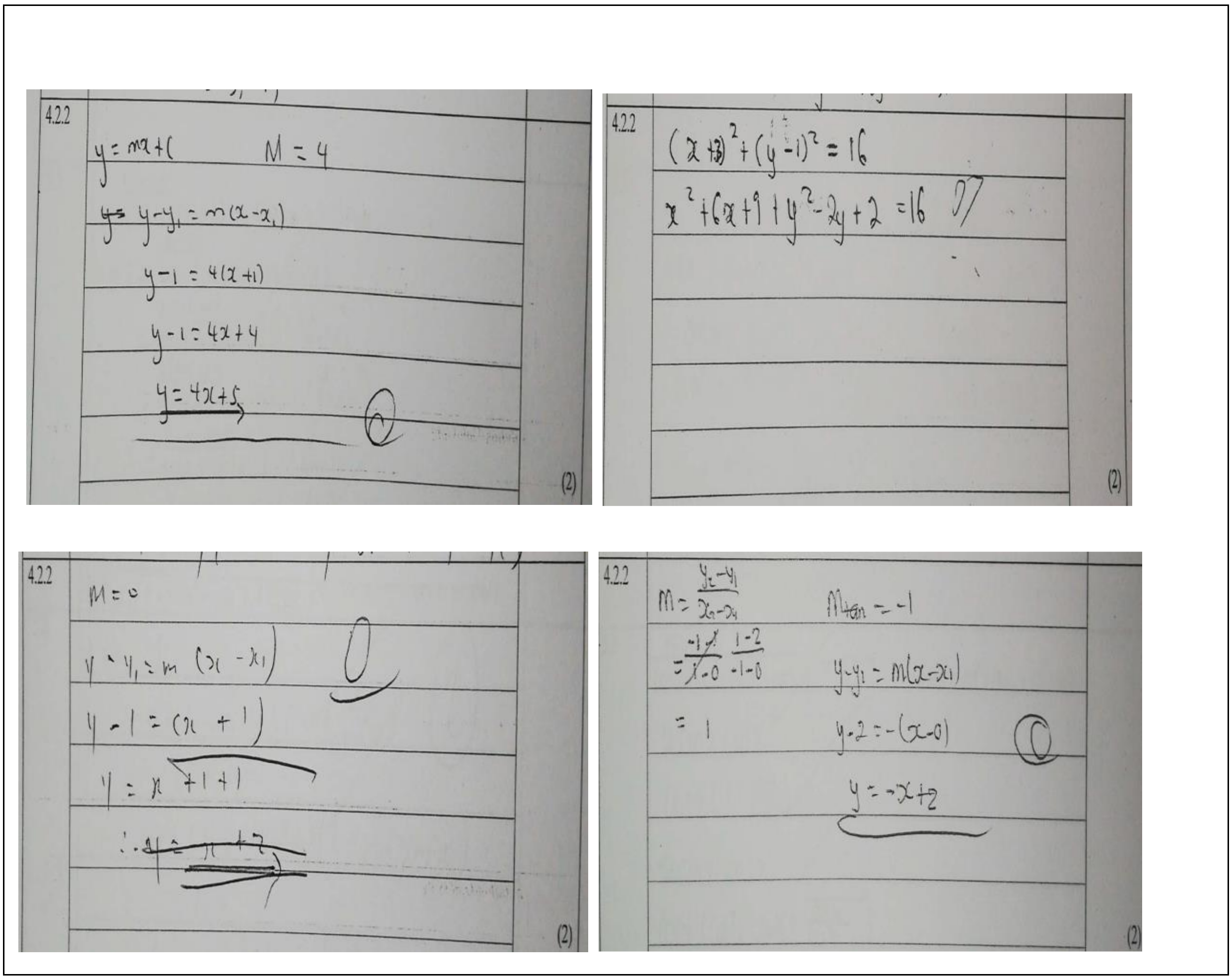


Figure 4.9. Responses by four learners to Question 4.2.2

Learners who responded to Question 4.2.2 showed poor understanding of the requirements of the questions. Instead of them proving that the two circles have one point of contact, many tried to find the equation of the common tangent as shown by the responses in Figure 4.9. Learners' work showed different methods of calculating the gradient of the common tangent as illustrated by three of the responses in Figure 4.9, where they used different values for the gradient as well as different points to calculate the value of the constant in the equation of the common tangent.

Comparison of the average percentage marks for the six schools showed that three of them constantly performed above the overall average, while the other three scored below it in all of the sub-questions except for a few cases. For Question 4.1.1 one of the highest performing schools in all the other sub-questions, School KS, performed below the overall average of 31.4%, with an average of 24%. The learners from this school were not able to use the fact that a tangent to a circle is always perpendicular to the radius at the point of contact, and thus the product of their gradients is equal to negative one. School KS had the second lowest average mark, the lowest was obtained by school FT, with an average of 21.2%. In all of the other sub-questions school KS had an average score that was above the overall average.

Question 3 was fairly well attempted in comparison to the other questions, with an overall average mark of 46.4%, possibly implying that the learners understood analytical geometry better than the other sections. Three schools did well on this question, managing to score averages above 50%. School SS had the highest average score of 63.7%, followed by school AD at 63.2% and school KS at 58.7%. The average marks for the three lowest performing schools improved, even though they were still below the overall average mark, school SB had an average of 39.2%, school FT an average of 23.7% and school LM an average of 23.3%. Question 3 was the one the learners performed best on since it had the highest overall average (46.4%).

The overall average mark for Question 4 was 25.3%. School SS had the highest average mark of 40.0%, followed by school KS at 33.7% and school AD at 29.1%. The other schools had average percentages below the overall average mark. School SB had an average of 16.7%, and school FT an average of 11.9%, with the lowest average mark from school LM at 11.7%.

Figure 4.10 shows the overall performance of the schools in the analytical geometry section after combining the scores for Questions 3 and 4.

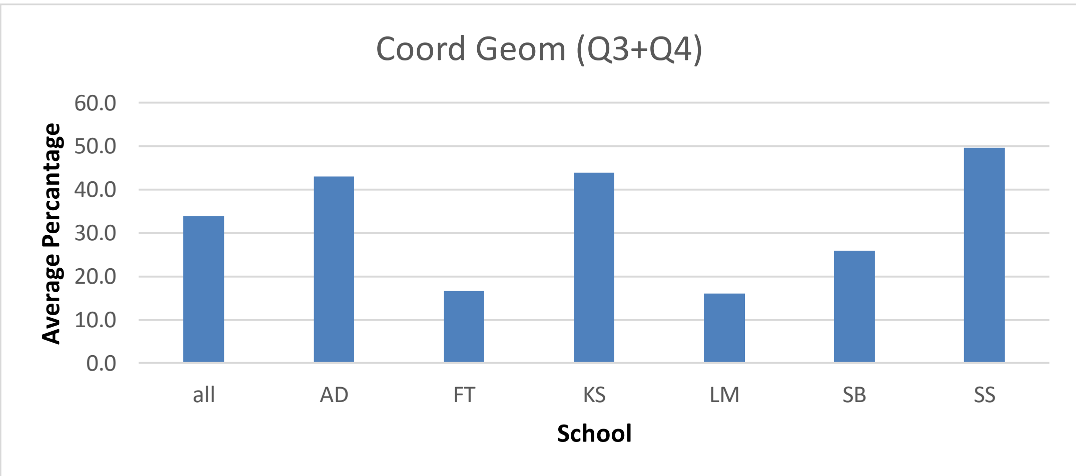


Figure 4.10. Average percentage marks for schools in analytical geometry

Figure 4.9 show school SS had the highest average mark at 49.6%, followed by schools KS and AD with average marks of 43.8% and 43.8% respectively. The lowest average marks were obtained by schools FT and LM at 16.7% and 16.0% respectively. Learners from schools SS, KS and AD showed better understanding of the analytical geometry concepts, with average marks above the pass mark of 30%.

4.3.3 Analysis of results for trigonometry

Questions 5 and 6 were based on trigonometry, and the general patterns that were observed in statistics and analytical geometry were also observed in this section of trigonometry. Three schools (AD, SS and KS) constantly had average percentage scores above the overall average percentage mark for all schools except for a few cases.

Table 4.5

Average percentage marks of schools for trigonometry

	Question							
School	**5.1**	**5.2**	**5.3**	**5.4**	**6.1.1**	**6.1.2**	**6.2**	**6.3**
All	53.1	15.1	33.8	10.7	43.5	13.5	54.6	17.1
AD	72.1	13.3	39.2	6.6	53.5	17.1	73.5	19.5
FT	40.0	5.8	14.4	3.5	22.8	4.7	47.1	5.8
KS	40.3	24.2	42.6	15.7	65.1	15.5	71.2	24.6
LM	52.8	9.5	21.6	7.1	16.5	8.0	8.4	9.0
SB	49.1	6.9	25.1	10.1	28.3	6.7	60.5	12.2
SS	67.2	23.4	51.1	16.8	61.3	24.4	61.6	26.0

Table 4.5 shows that learners did well on Questions 5.1, 5.3, 6.1.1 and 6.2 where the average were all above the 30% pass mark. Question 5.1 required learners to show without the use of a calculator the value of a trigonometry expression involving the reduction formula and special angles. The question was answered well with many learners able to reduce the angles and use the special angles. In Question 5.3, learners were required to prove an identity. For Question 6.1.1, learners were required to write down the expansion for $\sin(x + 30^{\circ})$ and to leave the answer in surd form. Learners were expected to use the compound angle formula and the special angles. Question 6.2 was also answered well, with learners sketching the graphs of $f(x) = 2cosx$ and g(x) $= \sin(x + 30^{\circ})$ for the interval $x\epsilon[-180\,;270]$. Question 6.2 was the best performed question in the trigonometry section, with an average score of 54.6%.

Table 4.5 also shows that Questions 5.2, 5.4, 6.1.2 and 6.3 were poorly answered with average marks all below 30%. In Question 5.2, learners were required to prove that $\frac{1+\sin\theta}{\cos\theta} = \frac{n+1}{n-1}$ if $sin\theta = \frac{2n}{n^2+1}$. Learners struggled to find the value of $cos\theta$ in terms of n, which was supposed to be done with the aid of a diagram. Some of the learners' responses are shown in Figure 4.11.

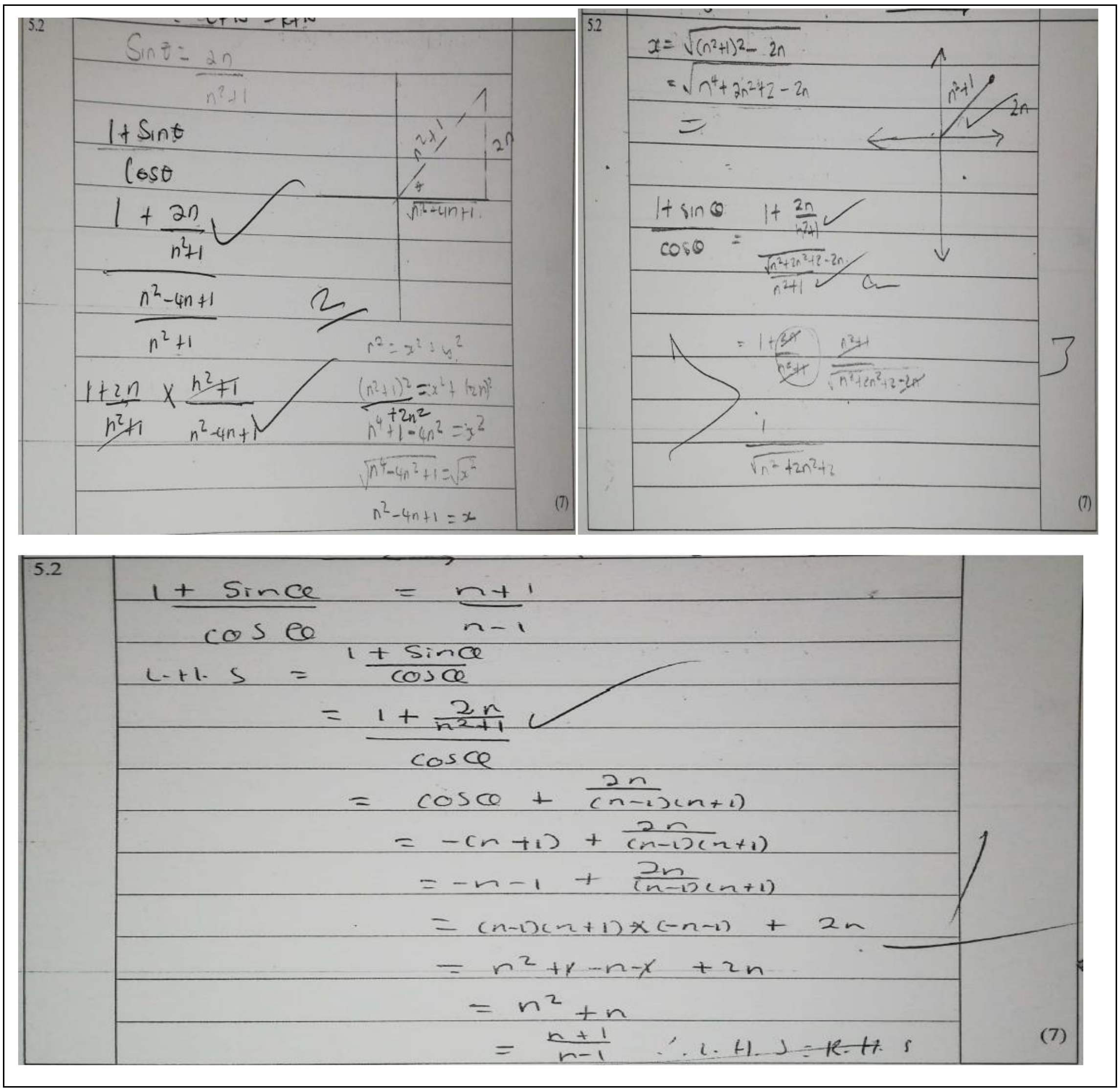


Figure 4.11. Some learners' responses to Question 5.2

The learners' responses for Question 5.2 showed that many of them were able to substitute for $Sin\theta$, but they could not write $Cos\theta$ in terms n, and hence were not able to get all the marks. The few learners who were able to write $Cos\theta$ in terms of n were not able to simplify algebraically and were therefore not able to provide the required proof.

Question 5.4 required learners to calculate the maximum area of a triangle that was inside a semi-circle, and present the answer in terms of x. The question required learners to use properties of triangles and circles and apply the area rule. Many learners did not respond to Question 5.4. The few learners who responded were able to write down the area rule and substituted side AO and side OB with x as these sides were the radii of the semi-circle. Learners were not able to deduce that for the area of the triangle to be maximum, then angle $A\hat{O}B$ must be equal to 90°.

In Question 6.1.2, learners were required to solve a trigonometric equation $2cosx = \sin(x + 30°)$ for the interval $x\epsilon[-180\ ; 270]$, which proved to be a challenge to many of them. Many learners were not able to expand $\sin(x + 30°)$ and use special angles. The common mistake was using the co-ratio for $cosx$ which was of little help because of the coefficient 2. Question 6.3 involved a two-dimensional shape and learners were required to use the sine rule and the trigonometric ratios. Many learners were unable to identify and solve for the common side between the vertical plane and the horizontal plane (see Appendix J).

In Question 5, based on trigonometry, four schools had average percentage scores above the overall average mark of 28.6%. The surprise inclusion in this group was school FT with an average mark of 42.8%, this is the only question where this school had an average above the overall average percentage for all schools. The other schools with average marks above the overall average percentage were school SS at 40.3%, School AD with an average of 33.5% and School KS with an average 31.6%. The two schools with average marks below the overall average percentage were school LM (23%) and school SB (22.5%).

For Question 6 the overall average was 30.4%, with school SS, school KS and school AD having average percentages above the overall average for all the schools. The school with the highest average marks were schools KS and SS at 40.7% each, followed by school AD at 39.0%. The lowest average mark was from school LM at 9.6%, followed by school FT with 19.6% and school SB with 26.5%. Figure 4.12 shows the performance of the schools in trigonometry.

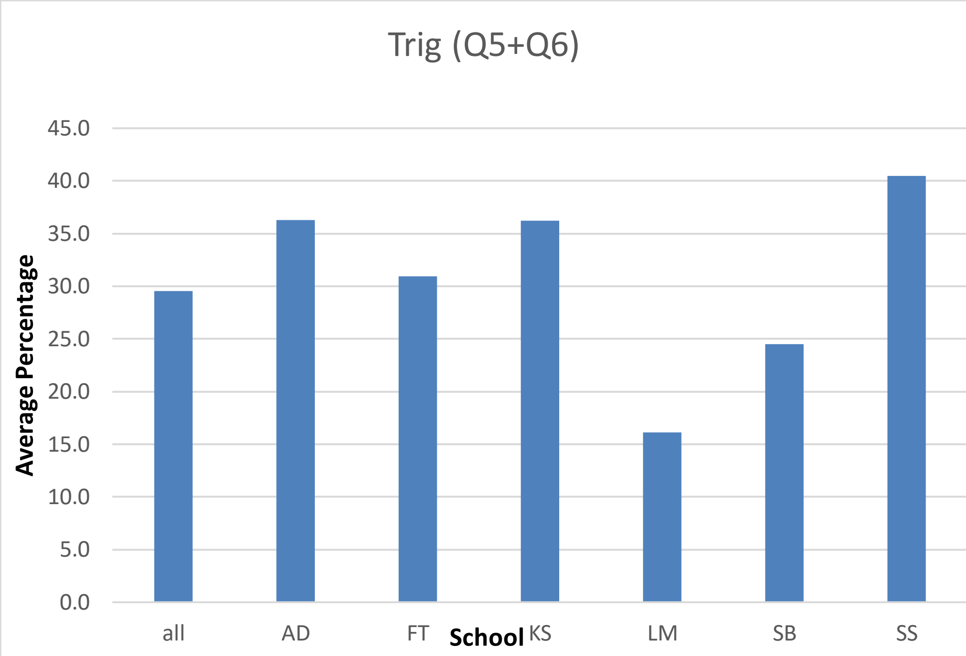


Figure 4.12. Performance of schools in trigonometry

Figure 4.12 above shows that four schools have average percentages which are above the pass mark of 30% in trigonometry. School SS had the highest average percentage in trigonometry at 40.5%, followed by schools KS and AD with average marks of 36.2% and 36.3% respectively. The lowest average mark was recorded by school LM at 16.1%.

4.3.4 Analysis of results for Euclidean geometry

The section on Euclidean Geometry had 12 sub-questions. Only five of the sub-questions (7, 8.1, 8.3, 9.1 and 10.2.1) had overall average marks above the pass mark of 30%. The highest overall average mark of 58.7% was obtained for Question 9.1, followed by 37.4% and 34.5% for Questions 7 and 8.3 respectively. Table 4.6 shows the full analysis per sub-question in the Euclidean geometry section.

Table 4.6

Average percentage marks of schools for Euclidean geometry

	Question											
School	7	8.1	8.2	8.3	8.4	9.1	9.2	10.1	10.2.1	10.2.2	10.2.3	10.2.4
All	37.4	32.7	16.2	34.5	15.0	58.7	17.5	10.4	37.9	17.2	8.0	3.8
AD	43.7	53.1	27.2	58.6	26.8	76.6	27.4	19.5	58.2	13.2	8.2	2.1
FT	19.8	18.9	4.2	14.7	1.9	47.4	5.8	3.4	19.7	2.4	0.2	0.0
KS	49.9	31.2	20.8	48.4	27.5	67.6	27.8	6.1	47.0	29.2	6.8	3.0
LM	27.8	18.4	7.9	5.1	2.2	32.0	10.8	10.8	6.2	16.9	2.0	1.3
SB	31.5	24.5	16.5	26.3	13.5	38.9	7.8	3.1	42.7	10.1	10.1	3.8
SS	45.7	47.8	20.4	47.9	16.0	78.6	21.5	18.8	50.4	24.4	18.1	10.2

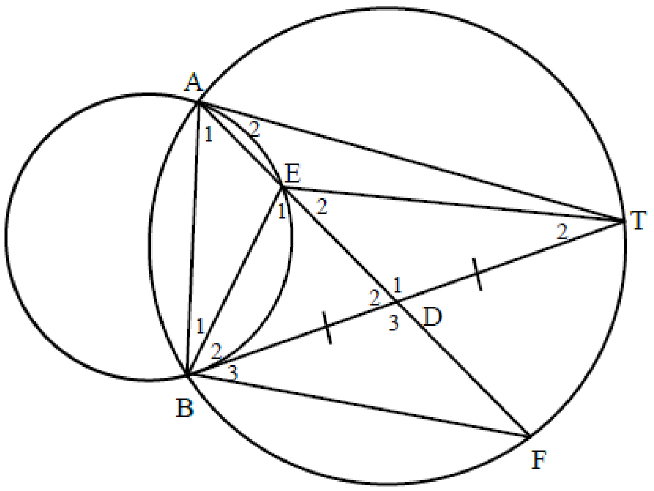
Table 4.6 shows that Question 9.1 was answered well. Learners were required to calculate the length of GC, which required the use of the proportionality theorem, which says a line drawn parallel to one side of a triangle divide the others two sides in the same proportion. The average mark for all of the schools was above the 30% pass mark for Questions 7, 8.1, 8.3 and 10.2.1. In Question 7, learners were given a diagram with a circle, a cyclic and a tangent to the circle. Learners were required to write down five other angles each equal to x and to give the reasons as to why this was so. These five other angles were supposed to be identified using properties and theorems such as alternate angles on parallel lines are equal, corresponding angles on parallel lines are equal, angles opposite equal sides on a triangle are equal, the exterior angle of a cyclic quadrilateral is equal to the opposite interior angle and the tangent-chord theorem.

Learners also did well in Question 8.1, where they were supposed to prove a cyclic quadrilateral. In this case the cyclic quadrilateral was supposed to be proved by showing that the opposite angles are supplementary or by showing that the exterior angle is equal to the opposite interior angle. Question 8.1 also required learners to have knowledge of the angle in a semi-circle or angle subtended by the diameter on the circumference of the circle. Learners also did well in Question 8.3, which required learners to prove that two triangle were similar. Similarity between the two triangles was supposed to be proved by showing that the corresponding angles of the two triangles were equal. Question 10.2.1 was also answered well and it also required learners to prove that two triangles were similar by showing that the corresponding angles were equal (see Appendix J).

Table 4.6 also shows that Questions 8.2, 8.4, 9.2, 10.1, 10.2.2, 10.2.3 and 10.2.4 were poorly answered, the averages for all these questions were below the 30% pass mark. Question 8.2 required learners to prove that a given triangle is an isosceles triangle. The question required learners to use the tangent-chord theorem and the theorem that states that the exterior angle of the cyclic quadrilateral is equal to the opposite interior angle, to prove that two of the angles in the triangle were equal. For Question 8.4, learners were required to calculate a missing side in a triangle. The question required learners to use the similar triangles proved in Question 8.3 to write down the sides in the same proportion, then substituting the known sides. For Question 9.2, learners were required to apply the proportionality theorem to find the length of a given side. They were supposed to write the sides in the same proportion and then substitute the known sides.

Question 10.1 required learners to prove the theorem that states that the sides of two equiangular triangles are in the same proportion. All of the sub-questions in Question 10.2 were poorly answered, with Questions 10.2.3 and 10.2.4 having the lowest average scores of 8% and 3.8% respectively. Question 10.2 is shown in Figure 4.13.

10.2 In the figure below, two circles intersect at A and B. TB is a tangent to the smaller circle at B. The line through D and A cuts the circles at E and F such that BD = DT. AB, BE and EA are joined.



10.2.1 Prove that ΔTDA ||| ΔFDB. (4)

10.2.2 Prove that $TB^2 = 4FD \cdot AD$. (2)

10.2.3 Prove that $BD^2 = DE.AD$. (4)

10.2.4 Deduce that ET = BF. (5)

[22]

Figure 4.13. Question 10.2

All the sub-questions in Question 10.2 were proof type questions as indicated in Figure 4.13. Question 10.2.2 was also poorly answered, learners were required to use the solution in Question 10.2.1 to write down the sides in the same proportion and then simplify the answer. Question 10.2.3 required learners to first identify two triangles using the sides given and prove that the two triangles are similar, then use the similar triangles to write down the sides in the same proportion. Question 10.2.4 was the one that learners performed worst on. Learners were supposed to prove that two sides were equal by using a number of theorems and deductions. The solution to Question 10.2.3 was required in the process of answering Question 10.2.4. Learners were also required to prove that two sides are equal b using congruency in Question 10.2.4. Figure 4.14 shows some of the learners' responses to Questions 10.2.3 and 10.2.4.

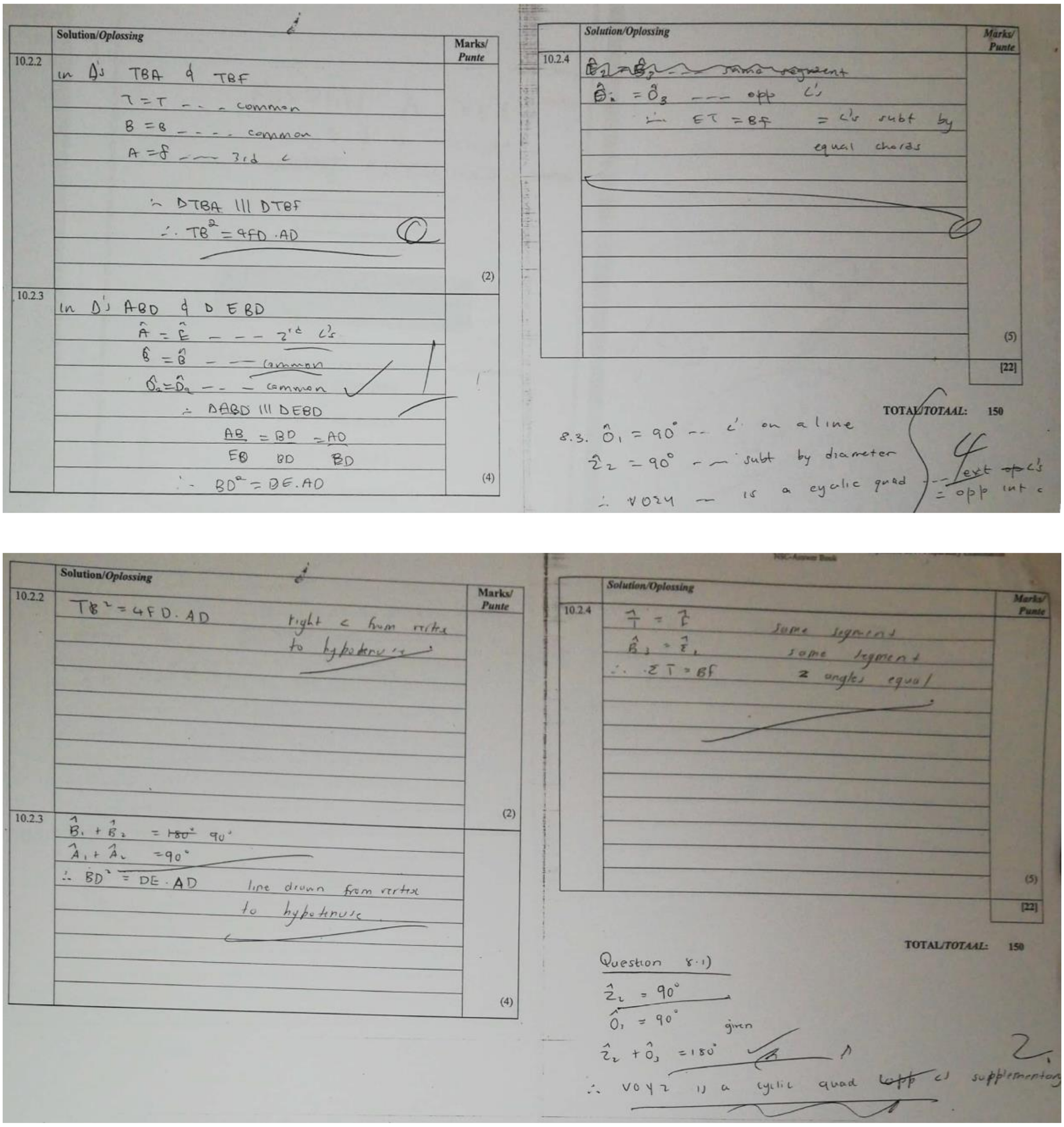


Figure 4.14. A sample of learners' responses to Questions 10.2.3 and 10.2.4.

The learners' scripts showed that many of them did not respond to Questions 10.2.3 and 10.2.4, and left blank spaces. The few learners who responded to these two questions showed a lack of conceptual understanding, as shown in Figure 4.14. In Question 4.2.3 some learners tried to provide the proof by first proving similarity, but a number of them could not link the correct corresponding angles and in some cases corresponding angles were identified correctly, reasons were not provided as to why those angles were equal. Learners who responded to

Question 10.2.4 showed no evidence of understanding the question. The majority of learners who responded to Question 10.2.4 were not able to achieve marks.

The overall average mark for Question 7, which was based on Euclidean geometry, was 37.4%, and three schools had average marks above this. School KS had an average of 49.9%, school SS had an average of 45.7% and school AD an average of 43.7%. The lowest average mark for Question 7 was obtained by school FT, at 19.8%, followed by school LM at 27.8% and school SB at 31.5%.

The trend of the same three schools having average marks above the overall average, and three schools having average marks below the overall average was also observed for Question 8. Schools AD (42.7%), SS (34.2%) and KS (33.2%) had average marks above the overall average for all of the schools, which was 25.3%. The three schools with average marks below the overall average were schools SB (20.6%), FT (10.3%) and LM (8.1%).

For Question 9 the highest average mark was 49.3%, from school AD, followed by school SS with 46.9% and school KS with 45.5%. The overall average mark for all of the schools was 35.8%. Schools LM and SB had the lowest average marks at 20.2% and 21.7% respectively.

For Question 10 the overall average for all of the schools was 14.1%, with three schools, SS (23.0%), AD (20.0%) and KS (15.1%) having averages above this. The lowest average mark was obtained by school FT with an average of 4.9%, followed by school LM at 6.7% and school SB at 12.4%. Figure 4.15 summarises the performance of the schools in Euclidean geometry.

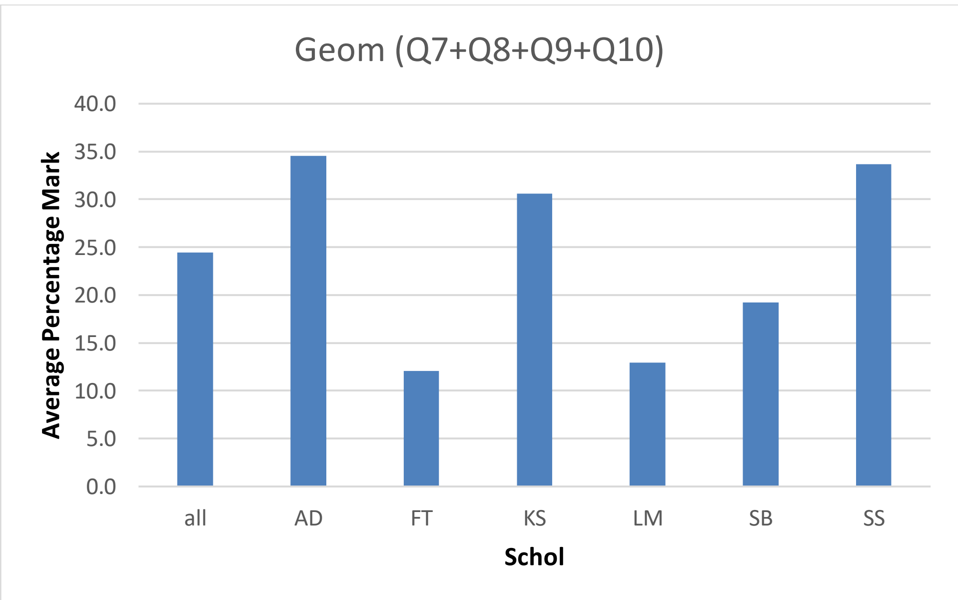


Figure 4.15. Performance of schools in Euclidean geometry

Figure 4.15 shows that school AD had the highest average mark of 34.6%, followed by schools SS and KS with 33.6% and 30.6% respectively. The lowest performing schools in Euclidean geometry were school LM with an average of 13% and school FT with an average of 12.0%. The trend observed in other sections, where three of the schools always performed above average was also observed in the section on Euclidean geometry.

For all 10 questions, three schools had average marks above the overall average. These were schools AD, KS and SS. Schools SB, FT and LM had average marks below the overall average percentages for all of the questions except for Question 5, where school FT had the highest average mark for all the schools. The overall average mark for all of the schools was 27.8%, which was below the pass mark of 30%. The highest overall average mark was 39.8% obtained by school SS, followed by school AD with 36.0% and school KS with 35.7 %. School LM had the lowest overall average of 14.0%, followed by school FT with 18.1% and school SB with 20.8%.

4.4 Analysis of zeroes and blank spaces

The number of questions which were not attempted and the number of zeroes obtained in some schools was quite worrying. Blank spaces refer to where no attempt was made to answer, while zeroes mean they attempted to answer but got no marks, while non-zero means that they got some marks for the question. A few questions will now be looked at to try and observe any trends.

Table 4.7

Number of blank spaces and zeroes per question per school for the first four questions

School AD	Q1.1	Q1.2	Q1.3.1	Q1.3.2	Q2.1	Q2.2.1	Q2.2.2	Q2.2.3	Q3.1	Q3.2	Q3.3	Q3.4	Q3.5	Q3.6	Q4.1.1	Q4.1.2	Q4.1.3	Q4.1.4	Q4.2.1	Q4.2.2
No of Learners	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76
Blank Spaces	7	7	19	19	4	13	13	20	8	7	12	9	16	11	20	6	14	39	29	49
Zeroes (0)	54	54	44	48	34	17	31	45	0	19	17	14	19	21	21	20	12	18	26	27
School KS	Q1.1	Q1.2	Q1.3.1	Q1.3.2	Q2.1	Q2.2.1	Q2.2.2	Q2.2.3	Q3.1	Q3.2	Q3.3	Q3.4	Q3.5	Q3.6	Q4.1.1	Q4.1.2	Q4.1.3	Q4.1.4	Q4.2.1	Q4.2.2
No of Learners	125	125	125	125	125	125	125	125	125	125	125	125	125	125	125	125	125	125	125	125
Blank Spaces	6	3	4	11	0	8	7	4	4	5	19	12	18	15	27	4	6	26	25	53
Zeroes (0)	105	48	99	90	53	42	49	70	23	55	45	34	28	8	59	25	22	53	41	61
School SS	Q1.1	Q1.2	Q1.3.1	Q1.3.2	Q2.1	Q2.2.1	Q2.2.2	Q2.2.3	Q3.1	Q3.2	Q3.3	Q3.4	Q3.5	Q3.6	Q4.1.1	Q4.1.2	Q4.1.3	Q4.1.4	Q4.2.1	Q4.2.2
No of Learners	131	131	131	131	131	131	131	131	131	131	131	131	131	131	131	131	131	131	131	131
Blank Spaces	10	4	13	17	3	10	24	39	3	4	19	14	24	22	21	4	12	35	16	61
Zeroes (0)	101	18	81	61	79	61	81	60	3	29	24	13	30	38	44	30	31	42	51	56
School SB	Q1.1	Q1.2	Q1.3.1	Q1.3.2	Q2.1	Q2.2.1	Q2.2.2	Q2.2.3	Q3.1	Q3.2	Q3.3	Q3.4	Q3.5	Q3.6	Q4.1.1	Q4.1.2	Q4.1.3	Q4.1.4	Q4.2.1	Q4.2.2
No of Learners	79	79	79	79	79	79	79	79	79	79	79	79	79	79	79	79	79	79	79	79
Blank Spaces	11	23	15	31	14	29	42	47	9	9	27	24	31	23	26	16	14	49	50	59
Zeroes (0)	68	37	62	46	48	13	33	32	7	35	23	22	29	20	23	28	44	20	23	20
School LM	Q1.1	Q1.2	Q1.3.1	Q1.3.2	Q2.1	Q2.2.1	Q2.2.2	Q2.2.3	Q3.1	Q3.2	Q3.3	Q3.4	Q3.5	Q3.6	Q4.1.1	Q4.1.2	Q4.1.3	Q4.1.4	Q4.2.1	Q4.2.2
No of Learners	89	89	89	89	89	89	89	89	89	89	89	89	89	89	89	89	89	89	89	89
Blank Spaces	18	17	22	35	21	43	56	50	15	12	27	18	35	24	38	27	28	42	41	50
Zeroes (0)	67	38	65	50	42	44	31	38	14	53	38	47	45	52	16	51	50	32	39	39
School FT	Q1.1	Q1.2	Q1.3.1	Q1.3.2	Q2.1	Q2.2.1	Q2.2.2	Q2.2.3	Q3.1	Q3.2	Q3.3	Q3.4	Q3.5	Q3.6	Q4.1.1	Q4.1.2	Q4.1.3	Q4.1.4	Q4.2.1	Q4.2.2
No of Learners	104	104	104	104	104	104	104	104	104	104	104	104	104	104	104	104	104	104	104	104
Blank Spaces	14	16	20	23	7	30	38	47	17	20	37	25	45	31	43	22	33	58	43	69
Zeroes (0)	89	64	82	78	58	69	33	50	18	56	48	50	44	52	39	65	51	32	50	34

The number of zeroes and blank spaces per question for the last two schools in the table (schools LM and FT) was quite worrying. Analysis of the first four questions showed that most learners from those two schools either did not answer the questions (left blank spaces) or attempted the questions and got zeroes. A good example is Question 1.1 where at school LM, out of 89 learners, 18 learners left some blank spaces and 67 got zeroes. That means that 96% of the learners at school LM either left blank spaces or got zeroes. A similar trend is also observed at school FT, where for Question 1.1 there were 14 learners who left blank spaces and 89 learners who got zeroes. This shows that 99% of the learners did not get any marks for Question 1.1. Question 1.1 required learners to determine the estimated mean of grouped data from a frequency table. To get the estimated mean, learners were supposed to first calculate the midpoint of each class interval and multiply it by the frequency of that interval. The sum of the products of the midpoint of each interval and the frequency was supposed to be divided by the total frequency to get the estimated mean. All of the schools performed poorly on Question 1, with more than half the number of learners in each school getting either a zero or leaving blank spaces.

For Schools AD, SS, KS and SB there were high number of zeroes and blank spaces for some of the questions (Question 2.2.2, 2.2.3 and 4.2.2); for other questions many learners were able to score some marks except for those from schools FT and LM. In school FT, across all the items in the test, more than half of the learners consistently received zeroes or had blank spaces except for Questions 3.1, 5.1, 6.2 and 7. At school LM, more than half of the learners received zeroes or blanks spaces for all of the questions except Questions 3.1, 5.1 and 7. The percentages of learners at schools FT and LM who received zeros and left blank spaces are presented in Figures 4.16 and 4.17 to illustrates this extreme situation. The two figures show the questions for which the percentage of zeroes and blank spaces were too high.

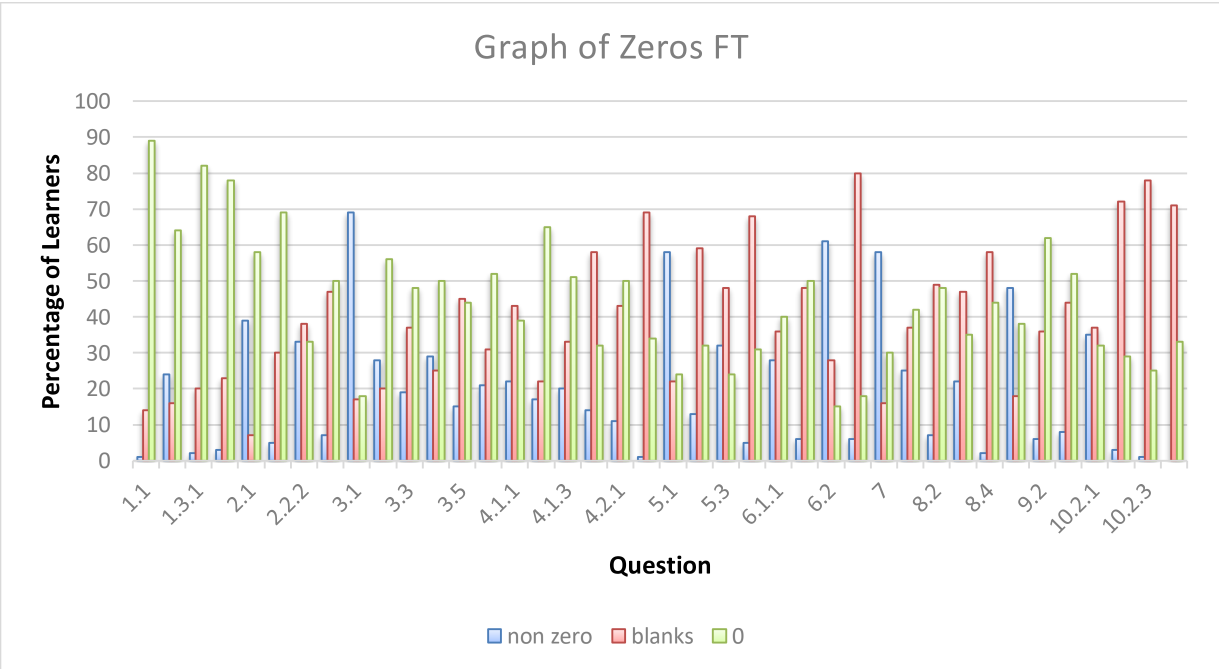


Figure 4.16. Graph of zeroes for school FT

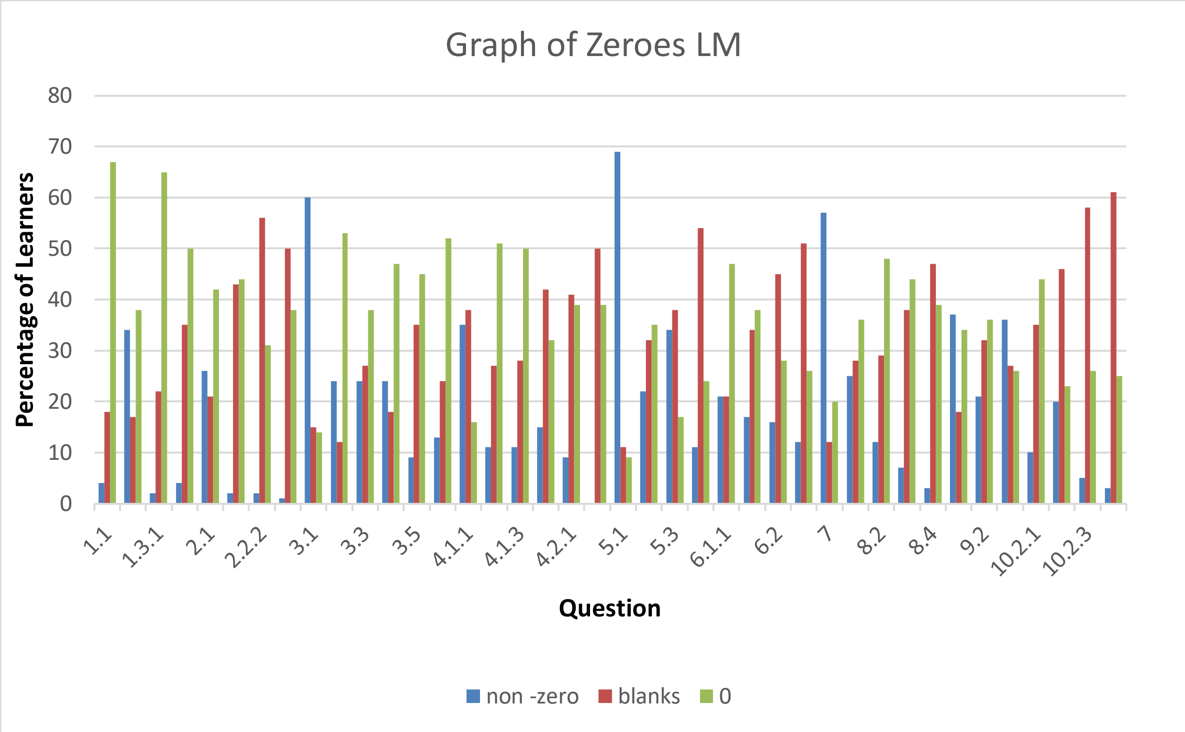


Figure 4.17. Graph of zeroes for school LM

Figures 4.16 and 4.17 clearly illustrate the high number of blank spaces and zeroes observed for learners from schools FT and LM. This shows that many of the learners' responses could be considered as being random, rather than actually demonstrating an understanding of the mathematics content that was being assessed. Hence any analysis using this test that tried to identify different levels of proficiency in mathematics among learners at these two schools would not be reliable. The responses were also expected to reduce the fit of the model to the data, because of the increase of 'noise' in the data. For the purposes of identifying ways in which the assessment instrument could be improved using insights from the Rasch analysis, it was therefore decided that the responses from those schools would be disregarded.

In the next chapter I first discuss the summary statistics and ordering of the item locations using the full data set. Thereafter the data from school LM and school FT were deleted from the sample for the subsequent analyses. The details appear in the next three chapters.

The zeroes and blank spaces were also high for some questions for the other four schools (AD, SS, KS and SB), but for most of the questions the numbers were low. On a question such as Question 1.1, many learners from all six schools left blank spaces or attempted to answer but their solutions were incorrect. But for many of the other questions, learners from the other schools performed better, except for those from schools LM and FT.

CHAPTER 5 ITEM FUNCTIONING FROM THE INITIAL RASCH ANALYSIS

5.1 Introduction

In this chapter I present an overview of the results of the initial Rasch analysis of all of the schools, to get a broad idea about the fit of the data and the person- item distribution. I looked at the empirical ordering of the items according to the analysis, and this is discussed in terms of the cognitive levels of the items. Finally, I looked at the issue of item fit and how misfit can be detected using the fit statistics, item characteristic curves (ICCs) as well as the category probability curves (CPCs). This allowed me to identify the mis-fitting items, after deleting some of the records as discussed at the end of Chapter 4.

5.2 Summary of statistics from the initial Rasch analysis

RUMM2030 was used in this study to report fit statistics in terms of item and person fit residual statistics, which provide an indication of the differences between the actual responses and those predicted by the model. The item trait interaction chi-square is also reported by RUMM2030, which reflect the property of invariance across the trait. The summary of statistics is shown in Table 5.1.

Table 5.1

Summary of statistics

	ITEMS [n=40]		PERSONS [n=604]	
	Location	Fit residual	Location	Fit residual
Mean	0.0000	-0.1405	-0.4746	-0.0747
SD	0.6333	2.5288	0.6628	0.8075
	Item Trait Interaction: Total Item Chi-Square = 801.8917 Total Degrees of Freedom = 360.00 Total Chi-Square Probability = 0.0000		Person Separation Index = 0.88714	

Table 5.1 shows the item mean as 0 (as set by the model) and the person mean as -0.4746, which is slightly below zero, showing that generally the students found the test difficult. The standard deviation (SD) for the item location is 0.6333, which is just below the ideal value of 1, while the SD of the person location is 0.6628, which is less than 1, suggesting that the distribution of the person locations is clustered together.

The mean for the item fit residual is -0.1405, which is close to zero. The SD of the item fit residual is 2.5288, which is larger than 1, which means that the fit residual varies more than expected. The mean of the person fit residual is -0.0747 and the SD of the person fit residual is 0.8075, which is slightly smaller than 1, showing that the distribution of the person fit residuals is slightly more clustered than the ideal situation. In terms of reliability, in RUMM2030 an estimate of the internal consistency reliability of the scale is the person separation index. The summary of statistics in Table 5.1 shows a person separation index of 0.88714, which is very good and is higher than the minimum of 0.85 advised by Tennant and Conaghan (2007). This shows that the estimation of the person's ability is consistent across the model. The figure of 0.88714 in this case indicates that the persons were separated well by the test. The item trait interaction figures have a chi-square value of 801.8917 with a probability value of 0.0000, which means the items are deemed to misfit the model expectations.

In Rasch analysis items are placed into graded levels of difficulty, from the greater at the top of a scale to a lesser at the bottom of the scale. Learners' proficiency and the item difficulty can be presented on one scale. The person-item location distribution provides the user with a comparison of learners and items, to better understand how appropriately the test measured learner proficiencies. Theoretically if persons and items are on the same location on the scale, the difficulty of the item and the ability of the person are comparable, so that the person has a 50% probability of answering the item correctly (Wright, 1996). Tennant and Conagnan (2007) explain that a comparison of the mean location score obtained for the persons with that of the value of zero set for the items provides an indication of how well targeted the items are for people in the sample. A well-constructed instrument should match the width of the target population ability distribution with the width of the distribution of the test (Wright, 1996)

The person-item threshold distribution is similar to the person-item location distribution, except that the difficulty location of each of the thresholds in each of the items is used when working out the item distribution, so for polytomous items, there are more than one threshold location. Hence the threshold distribution has a larger span than the item location distribution.

The person-item location distribution is represented in Figure 5.1. The item locations range from -1.848 logits to 1.637 logits. The person locations were estimated at between -3.082 logits and 1.862 logits (with a mean of -0.475), with 4 learners from schools FT and LM obtaining extreme scores, while 13 learners from these schools were located at below - 2 logits. The fact that the mean person location is smaller than the mean of the item location (set at 0), suggests that this trial examination was slightly difficult for the learners who took part in the study.

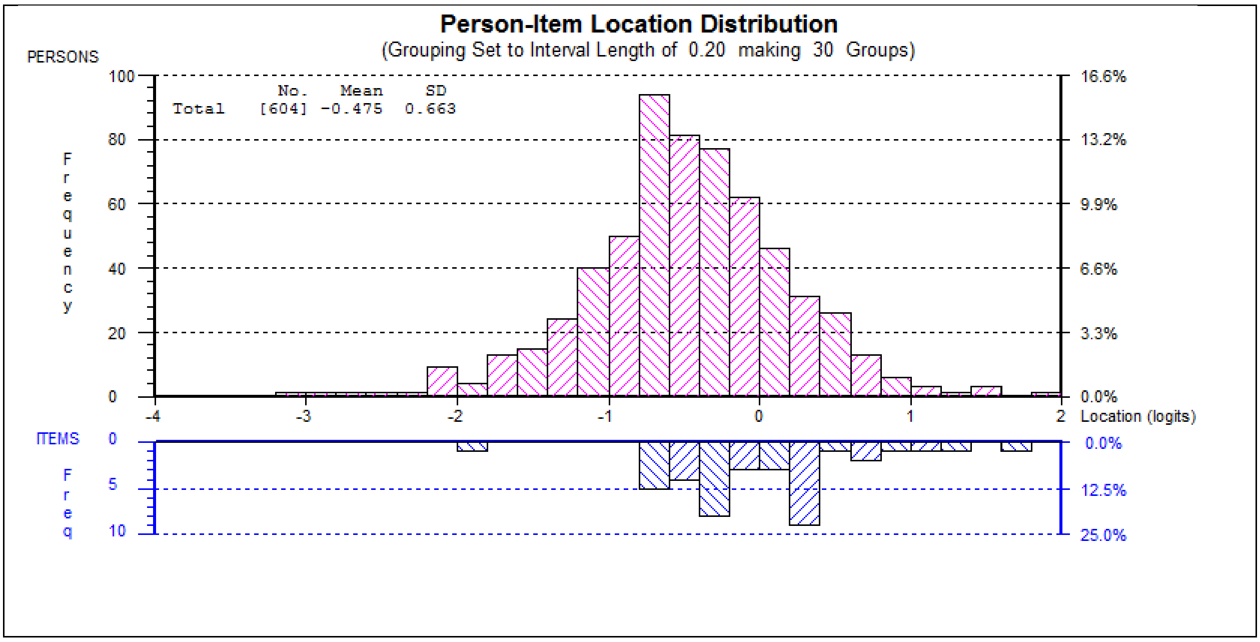


Figure 5.1. Person-Item Location Distribution

The learners whose location was below -2 logits found the trial examination extremely challenging, and these learners had a high chance of not getting all of the items correct. The easiest item was Question 3.1 where the learners were asked to show that the given answer was correct, and this was located at -1.848 logits in this analysis, while the second easiest item (Question 9.1) was located at -0.728 logits. Hence learners located at proficiency levels below -2 logits are predicted by the model to have less than a 50% probability of getting all of the items except item 3.1 correct. One learner's location was estimated at 1.862 logits, which was higher than any of the item locations, showing that this learner found the trial examination easy, and was likely to get all of the items correct.

The trial examination had polytomous items, and hence a person-item threshold distribution was useful to better understand the spread of the item thresholds. Figure 5.2 shows the person-item threshold distribution.

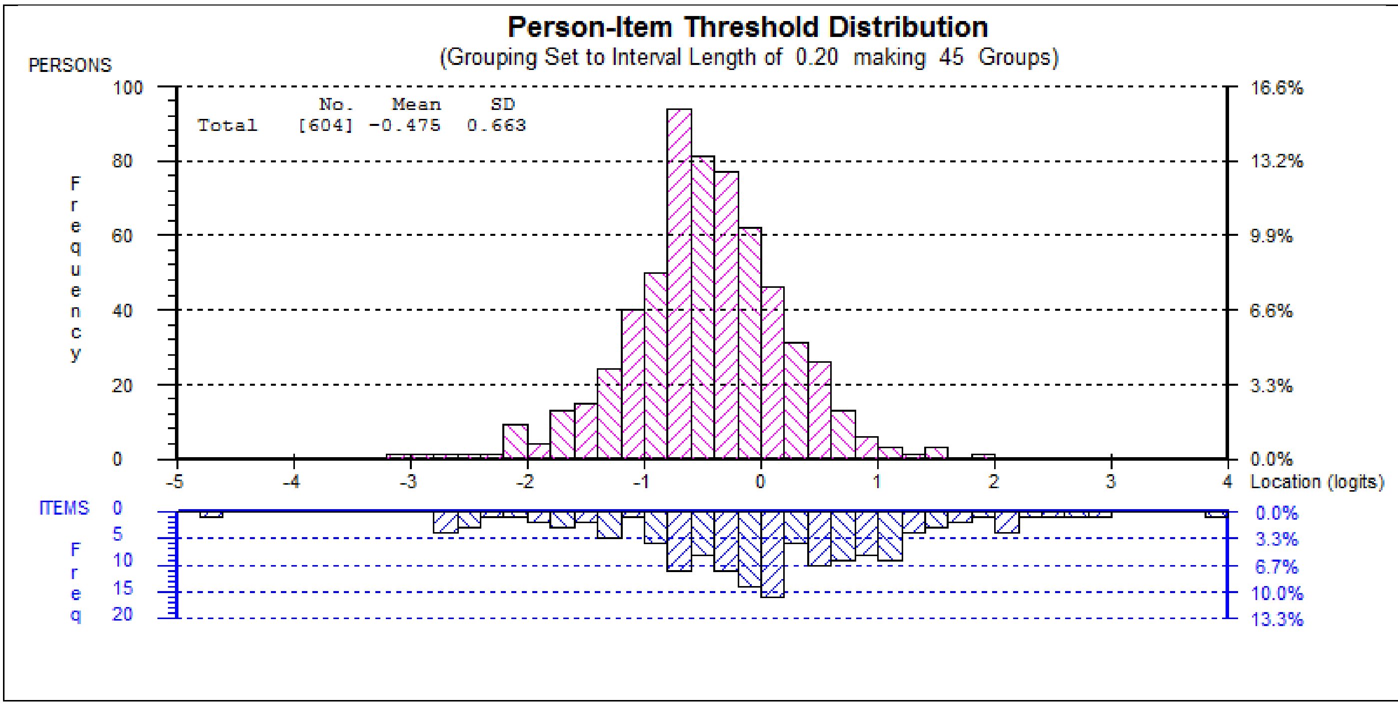


Figure 5.2. Person-item threshold distribution

Figure 5.2 shows that the thresholds ranged from below -4.8 to 4 logits, showing a much wider distribution than the items themselves.

5.3 Ordering of items according to the initial Rasch analysis

The Rasch model allows for the placement of items and persons on a proficiency level as indicated on the person-item and person-threshold maps. The map is basically a rough histogram with items on the right, ranging from relatively easy at the bottom to relatively difficult at the top. The same applies to person proficiency, which is arranged from relatively low proficiency at the bottom to higher proficiency at the top. When a learner is at an ability level that is equal to the difficulty level of an item (learner is at the same location as an item), then the Rasch model assumes that the learner has a 50% chance of achieving a correct response on that dichotomous item (Dunne, Long, Craig & Venter, 2012). If an item difficulty is above the ability location of any learner, then the learner has a less than 50% chance of achieving a correct response on that item. If the item is located lower on the scale than the person location, the learner would have a greater than 50% chance of achieving a correct response. The placement of the items according to difficulty level will later be compared to the taxonomy used for classification of items in assessments as recommended by the DoBE (2011).

In Figure 5.3, Item 4.2.2 is considered one of the most challenging questions for this group of learners, being placed at a difficulty level that is above the proficiency level of all of the learners. Items 1.1 and 10.2.4 were also relatively difficult for this cohort, as only 5 learners were placed at a proficiency level above the difficulty level of Item 1.1, and 5 learners were at the same proficiency level as the difficulty level of Item 10.2.4. Items 1.3.2, 10.2.3, 1.3.1 and 5.4 can also be referred to as difficult as very few learners were at a proficiency level that matched the difficulty level of the items.

```
RUMM2030        Project: PHDPROJECT        Analysis: PHDP2018
Title:  PHD PROJECT 2018
Display: ITEM MAP

-----------------------------------------------------------------------------------------------
LOCATION          PERSONS     ITEMS [locations]
-----------------------------------------------------------------------------------------------
  2.0                        |
                             |
                             | 4.2.2
                           × |
                             |   I.1
  1.0                      × | 10.24
                           × |   5.4
                         ××× | 10.23  1.3.1
                       ××××× | 1.3.2
                      ×××××× | 10.22  2.2.2  4.2.1    8.2  6.1.2    8.4    5.2   10.1  2.2.3
  0.0              ××××××××× |   6.3  2.2.1    9.2
                ×××××××××××× |   2.1    8.1  4.1.4
             ××××××××××××××× | 10.21    5.3    8.3    3.5  4.1.1  4.1.3    1.2      7
            ×××××××××××××××× |   6.2    3.3  6.1.1  4.1.2
         ××××××××××××××××××× |   9.1    3.6    3.2    5.1    3.4
 -1.0             ×××××××××× |
                    ×××××××× |
                       ××××× |
                         ××× |
                         ××× |
 -2.0                      × |   3.1
                          ×× |
                             |
                             |
                             |
 -3.0                        |
                             |
                             |
                             |
                             |
 -4.0                        |
-----------------------------------------------------------------------------------------------
            × = 5 Persons
```

Figure 5.3. Person-Item map indicating item difficulty locations

Only Item 3.1 was placed at a difficulty level of below -0.1 logits. Of the learners 110 were at a proficiency level below –1.0 logits, meaning these learners were highly likely to only get Item 3.1 correct, this corresponds to 18.2% of the cohort being most likely to manage to respond correctly to only one item.

The item difficulty as set by the Rasch model was then compared to cognitive levels of the items as set in the CAPS document for FET mathematics.

5.3.1 Cognitive levels of the items

According to the CAPS document for FET mathematics, there are four cognitive levels used to guide all assessment tasks and these are based on those suggested in the TIMMS study of 1999 (DoBE, 2011). The descriptors for the levels and the approximate percentages for tasks and examinations for each level, are summarised in Table 5.2.

Table 5.2

Descriptors for the cognitive levels

Cognitive levels	Description of skills to be demonstrated
Knowledge 20%	Questions require straight recall, identification of correct formula on information sheet. Questions will require use of mathematical facts and appropriate use of mathematical vocabulary
Routine procedures 35%	Questions in this category will require estimation and appropriate rounding of numbers. Proofs of prescribed theorems and derivation of formulae, identification and direct use of correct formula on information sheet [no changing of subject] and performing known procedures are some of the skills required in this category. Learners will also need to perform well known procedures, simple applications and calculations which might involve few steps, derivation from given information may be involved, identification and use [after changing the subject] of correct formula and solve problems generally similar to those encountered in class
Complex procedures 30%	In this category questions involve complex calculations and/or higher order reasoning There is often a not so obvious route to the solution and questions may not be based on real world contexts. Questions may involve making significant connections between representations and require conceptual understanding.
Problem solving 15 %	Questions in this category include non-routine problems and higher order reasoning and processing. Learners might be required to break down the question into its constituent parts.

Although the descriptors for each of the levels seem to be quite precise, there are sometimes differences in interpretation. In working out the predicted DoBE cognitive levels, I approached teachers for assistance, since they work with their learners and interpret these levels as part of their daily work. Seven educators who have experience in teaching Grade 12 were asked to independently classify the items. Where the teachers did not all place an item into the same cognitive level, the level selected by most teachers was considered as the cognitive level of the item. Items where teachers placed them into different cognitive levels included Questions 1.3.1, 1.3.2, 2.2.1, 2.2.2 and 2.2.3. The results of the classification are presented in Table 5.3.

Table 5.3
Cognitive levels grid for Mathematics Trial Paper 2, September 2017

Sub-Question Number	**Number of Marks on a Specific Level**			
	Level 1	**Level 2**	**Level 3**	**Level 4**
1.1		2		
1.2	4			
1.3.1		2		
1.3.2		3		
2.1	1			
2.2.1	3			
2.2.2	2			
2.2.3	2			
3.1		2		
3.2		2		
3.3		2		
3.4		3		
3.5			4	
3.6		2		
4.1.1			5	
4.1.2	2			
4.1.3		4		
4.1.4			4	
4.2.1				5
4.2.2				2
5.1		5		
5.2			7	
5.3		5		
5.4				3
6.1.1	3			
6.1.2		7		
6.2	6			
6.3			5	
7	9			
8.1		3		
8.2			3	
8.3		4		
8.4			3	
9.1		4		
9.2			5	
10.1		7		
10.2.1		4		
10.2.2				2
10.2.3			4	
10.2.4				5
Total Marks	**32**	**61**	**40**	**17**
%of Marks	**21%**	**41%**	**27%**	**11%**

From the breakdown in Table 5.3, it can be seen that the trial examination paper covered all of the cognitive levels, with the percentage for each level being slightly different from the requirements of the DoBE. The cognitive level grid indicates that Items 4.2.1, 4.2.2, 5.4, 10.2.2 and 10.2.4 were rated as the most difficult questions in this assessment tool.

5.3.2 Comparison between DoBE cognitive levels and difficulty levels according to the Rasch model

When comparing the DoBE cognitive levels with the Rasch difficulty levels it is important to note that the cognitive difficulty levels are theoretically derived, while the Rasch difficulty levels are empirically derived. When items are placed into different levels according to the DoBE taxonomy, this is based on theoretical justifications using the DoBE indicators or criteria. It is not always true that the learners will experience those items as easy or difficult as is expected by the cognitive levels. There may be a number of reasons for the differences, notably educational experience, or curriculum coverage. It is possible that an item placed at level 4 on the DoBE taxonomy is experienced as easy by the learners, because a similar item was discussed in class or a previous assessment. Figure 5.4 shows the item map with items ordered according to the difficulty levels identified by the Rasch model, and the items are highlighted with different colours to indicate their levels according to the DoBE taxonomy.

```
 RUMM2030        Project: PHDPROJECT        Analysis: PHDP2018
Title:  PHD PROJECT 2018
Display: ITEM MAP

-----------------------------------------------------------------------------------------------
LOCATION             PERSONS     ITEMS [locations]
-----------------------------------------------------------------------------------------------
  2.0                                 |
                                      |
                                      | 4.2.2
                                    × |
                                      |   1.1
  1.0                               × | 10.24
                                    × |   5.4
                                  ××× | 10.23  1.3.1
                                ××××× | 1.3.2
                               ×××××× | 10.22  2.2.2  4.2.1    8.2  6.1.2    8.4    5.2   10.1  2.2.3
  0.0                       ××××××××× |   6.3  2.2.1    9.2
                         ×××××××××××× |   2.1    8.1  4.1.4
                      ××××××××××××××× | 10.21    5.3    8.3    3.5  4.1.1  4.1.3    1.2      7
                     ×××××××××××××××× |   6.2    3.3  6.1.1  4.1.2
                 ×××××××××××××××××××× |   9.1    3.6    3.2    5.1    3.4
 -1.0                      ×××××××××× |
                             ×××××××× |
                                ××××× |
                                  ××× |
                                  ××× |
 -2.0                               × |   3.1
                                   ×× |
                                      |
                                      |
                                      |
 -3.0                                 |
                                      |
                                      |
                                      |
                                      |
 -4.0                                 |
-----------------------------------------------------------------------------------------------
```

× = 5 Persons   1.2= Cognitive Level 1.   1.1 = Cognitive Level 2   3.3= Cognitive Level 3   4.1.4 = Cognitive Level 4

Figure 5.4. Person-item map indicating item difficulty locations, with DoBE taxonomy levels indicated by different colours

On the person-item map in Figure 5.4 the items highlighted in yellow are those classified according to the DoBE taxonomy as level 1, while those highlighted in grey are classified as level 2, those highlighted in pink are classified as level 3, and those highlighted red are classified as level 4. Highlighting of the items was done to make for easy comparison between the classification according to the DoBE taxonomy and how the learners experienced the questions according to the Rasch analysis.

A more effective way of placing items at different difficulty levels in Rasch analysis is to use the individual item-fit, which arranges the items in location order to three decimal places. Table 5.4 shows the individual item-fit in location order from items experienced as easy by the learners to those items experienced as the most difficult according to this cohort.

Table 5.4
Individual item fit-location Order

RUMM2030 Project: PHDPROJECT Analysis: PHDP2018
Title: PHD PROJECT 2018
Display: INDIVIDUAL ITEM-FIT - Location Order

Seq	Item	Type	Location	SE	Residual	DF	ChiSq
9	3.1	Poly	-1.848	0.080	-0.311	523.48	16.978
34	9.1	Poly	-0.728	0.033	0.700	494.72	6.819
14	3.6	Poly	-0.691	0.055	-2.201	456.37	18.685
10	3.2	Poly	-0.619	0.051	-2.817	522.52	64.429
21	5.1	Poly	-0.617	0.030	4.386	520.60	26.807
12	3.4	Poly	-0.611	0.040	-2.484	477.46	26.866
27	6.2	Poly	-0.594	0.026	3.739	466.91	73.473
11	3.3	Poly	-0.551	0.057	-4.084	441.99	45.754
25	6.1.1	Poly	-0.534	0.041	-2.657	488.01	25.056
16	4.1.2	Poly	-0.418	0.056	-4.822	501.43	51.885
37	10.2.1	Poly	-0.391	0.037	-1.575	436.23	6.709
23	5.3	Poly	-0.312	0.038	-0.011	427.61	5.383
32	8.3	Poly	-0.293	0.037	-3.576	433.36	26.439
13	3.5	Poly	-0.271	0.036	-2.228	414.18	14.779
15	4.1.1	Poly	-0.253	0.030	6.117	410.35	130.202
17	4.1.3	Poly	-0.252	0.035	-1.751	472.67	14.580
2	1.2	Poly	-0.252	0.033	4.087	510.06	32.448
29	7	Poly	-0.239	0.024	2.621	526.36	7.261
5	2.1	Poly	-0.158	0.090	1.202	528.27	2.537
30	8.1	Poly	-0.155	0.044	-2.357	446.78	15.252
18	4.1.4	Poly	-0.041	0.042	3.565	338.44	2.345
28	6.3	Poly	0.005	0.041	-0.445	265.58	1.001
6	2.2.1	Poly	0.095	0.046	0.575	450.62	3.260
35	9.2	Poly	0.140	0.037	2.117	381.58	46.127
38	10.2.2	Poly	0.220	0.083	-0.920	271.33	8.707
7	2.2.2	Poly	0.253	0.064	2.962	406.51	8.116
19	4.2.1	Poly	0.265	0.038	-1.547	381.58	7.437
31	8.2	Poly	0.326	0.052	-2.811	394.05	16.471
26	6.1.2	Poly	0.355	0.031	-0.189	436.23	1.683
33	8.4	Poly	0.361	0.054	-1.092	346.11	4.085
22	5.2	Poly	0.362	0.035	2.431	398.84	5.335
36	10.1	Poly	0.389	0.034	2.202	402.68	24.089
8	2.2.3	Poly	0.399	0.067	1.025	380.63	5.444
4	1.3.2	Poly	0.568	0.056	-0.862	446.78	3.491
39	10.2.3	Poly	0.627	0.062	-0.243	236.81	6.705
3	1.3.1	Poly	0.741	0.071	-1.927	488.97	17.628
24	5.4	Poly	0.823	0.072	0.722	313.51	6.401
40	10.2.4	Poly	1.010	0.074	-1.179	225.31	3.298
1	I.1	Poly	1.217	0.091	-1.191	514.85	7.181
20	4.2.2	Poly	1.673	0.153	-0.791	250.24	10.745

Some items were categorised as level 1 according to the DoBE taxonomy, but on the Rasch ordering, they did not appear amongst the easiest items. For example, Item 2.2.3 (Determine the average fuel consumption of the motor car) appeared as the eighth most difficult item with a difficulty level of 0.399 logits, although it was classified as at level 1 in terms of the DoBE taxonomy. Other such cases were Item 2.2.2 (Determine the correlation coefficient excluding the outlier and explain the type of correlation) which was the fifteenth most difficult, Item 2.2.1 (Determine the equation of the regression line excluding the outlier) which was the eighteenth most difficult and Item 2.1 (Identify an outlier. Write down its co-ordinates) which was the twenty-second most difficult item for this cohort. Although the other items grouped as level 1 under the DoBE taxonomy were relatively easy, none of those items were among the six easiest items. Item 6.2 (Sketching two trigonometric graphs on the same set of axes) was the seventh easiest item, with a difficulty level of -0.594.

The items categorised as level 2 according to the DoBE taxonomy dominated the top 10 easiest items for this cohort according to the Rasch ordering, with 8 of the level 2 items falling within the top 10 easiest items. Item 3.1 (Show that the co-ordinates of Q are (0;2)) was the easiest item with a difficulty level of -1.848, followed Item 9.1 (Calculating (stating reasons) the length of GC) with a difficulty level of -0.728. All of the top 6 easiest items according to Rasch analysis were level 2 questions on the DoBE taxonomy. However, some of the most difficult items for this particular cohort were level 2 items. Item 1.1 (Determine the estimated mean height of the palm trees in the park) was the second most difficult item with a difficulty level of 1.217 logits, while Items 1.3.1 (Use your ogive to estimate the median height of the palm trees) and 1.3.2 (Use your ogive to estimate the interquartile range (IQR)) were the fifth and seventh most difficult questions respectively, for this cohort despite the fact that they were placed at level 2 on the DoBE taxonomy.

Some of the items categorised as level 3 according the DoBE taxonomy were placed among the easy items according to the Rasch ordering. One such case was Item 3.3 (Prove that $PQR = 90°$), which had a difficulty level of -0.551 logits, and was the eighth easiest item. Amongst the level 3 items which the Rasch model ranked as easy were Items 3.5 (Calculate the area of ΔPQR) and 4.1.1 (Deduce that $ET = BF.$) which occupied position 14 and 15 respectively among the easiest items. However, the other 6 items placed under level 3 according to the DoBE taxonomy, fell within the 20 most difficult items according to the Rasch ordering, with Item 10.2.3 (Prove that $BD^2 = DE.AD.$) being the sixth most difficult, with a difficulty level of 0.627 logits.

All of the items categorised as level 4 according to the DoBE taxonomy, appeared among the 20 most difficult items when ordered using the Rasch model. For example, the Rasch ordering considered Item 4.2.2 (Determine the equation of the common tangent to the circles) as the most difficult item among the 40 items, with a difficulty level of 1.673 logits. Item 10.2.4 (Deduce that $ET = BF.$) was also categorised as level 4, and according to the Rasch ordering, was the third most difficult item with a difficulty level of 1.011 logits.

It is important to note that when comparing the DoBE cognitive levels with the Rasch difficulty levels, the cognitive levels are theoretically derived, while the Rasch difficulty levels are empirically derived. There are many reasons for the differences, which include educational experience and curriculum coverage.

The person-item map and the individual item-fit location order did not always put items in the same order as the DoBE taxonomy, although in other instances the ordering seemed to match. The differences are explored in later sections.

5.4 Item fit

The detailed analysis by school in Chapter 4 (Figures 4.16 and 4.17) as well as the discussion of the initial Rasch analysis in section 5.2 above raised concerns about the very low results from the two schools FT and LM. It was found that all of the extreme scores were from these two schools. Furthermore, the mean person location for FT was -0.901, while for LM it was -0,931. It was noted from the initial analysis that all items except one were located higher than the means of these two schools. In terms of trying to identify ways in which the assessment instrument could be improved, it was decided that the responses from these two schools should not be considered in the next few stages of the analysis. Since there seemed to be inconsistency in the responses from these two schools, in going further to focus on the actual instrument, we (the researcher and supervisor) felt that we could get more relevant information about the functioning by limiting the responses to the 411 learners from the other four schools.

In this section I looked at the issue of item fit and how misfit can be noticed using the fit statistics, ICCs and the CPCs. These three analytic tools were used to identify items that were not functioning as well as expected by the model, so that they could be followed up by trying to understand the source of the misfit and to see if post- hoc rescoring could help resolve the misfit.

5.4.1 Using fit residuals to check for item misfit

According to Douglas (1982), the data fitting the model means the data set meets the model requirements, and there is thus a close relationship between the collected data and the Rasch model. Fit residual shows the extent to which a person’s response patterns match the expectations of the model (Smith & Plackner, 2009).

Initial analysis of the data using RUMM2030 identified eight items with fit residuals that were not in the recommended range of from -2.5 to 2.5 (Smith, 2002). These items are given in Table 5.5.

Table 5.5

Items with fit residuals outside the acceptable range of -2.5 to 2.5

Item/Question	Fit residual
3.3	-3.467
8.1	-3.340
8.3	-3.076
4.1.2	-2.955
4.1.1	2.579
6.2	2.841
9.2	2.887
1.2	4.488

According to Smith and Plackner (2009), items which discriminate more than other items have fit residuals that are large in magnitude and negative. Items 3.3, 8.1, 8.3 and 4.1.2 have fit residuals that are negative and large in magnitude, and hence are discriminating more than the other items. Fit residuals that are positive and larger than the upper boundary of the acceptable range (-2.5; 2.5), characterise items which discriminates less than the summary discrimination of the rest of the items (Smith & Plackner, 2009). In this regard, Items 1.2, 9.2, 6.2 and 4.1.1 are discriminating less than all of the other questions, since they have fit residuals which are positive and large in magnitude.

5.4.2 Using the item characteristic curves to check for item misfit

The Item Characteristic Curves (ICC) of items are used to investigate if they meet the model requirements or not. In order to investigate if the collected data meet the requirements of the model, the Rasch model put the learners into class intervals of nearly equal sizes. In this study, five groups were used. The mean ability of the five groups becomes the horizontal coordinate of points in the ICCs. With the ICCs learners are represented on the horizontal axis from low ability to high ability and the expected response is represented on the vertical axis.

If the data fit the model, the predicted curve and the observed/actual proportion fit into each other. In a case where the curve for the actual means and that predicted by the Rasch analysis fit into each other, this is an indication that questions are functioning in a way that was predicted by the model. Where the theoretical curves and the observed proportions deviate to a great extent, this is an indication that data do not fit the model requirements. When ICCs are used to determine if data meet the model requirements, four categories usually emerge; good fit, under-discrimination, over-discrimination and haphazard misfit. I will now identify and briefly discuss items from the assessment instrument that fit into each of these categories.

5.4.2.1 Fairly-good fit items

Some items fell into the first category, which is that of items with a fairly good fit. Fairly good fit items are observed when the theoretical and observed curves match or are very close to each other. Figure 5.5(a) and (b) illustrate the ICCs for Items 3.5 and 4.1.3.

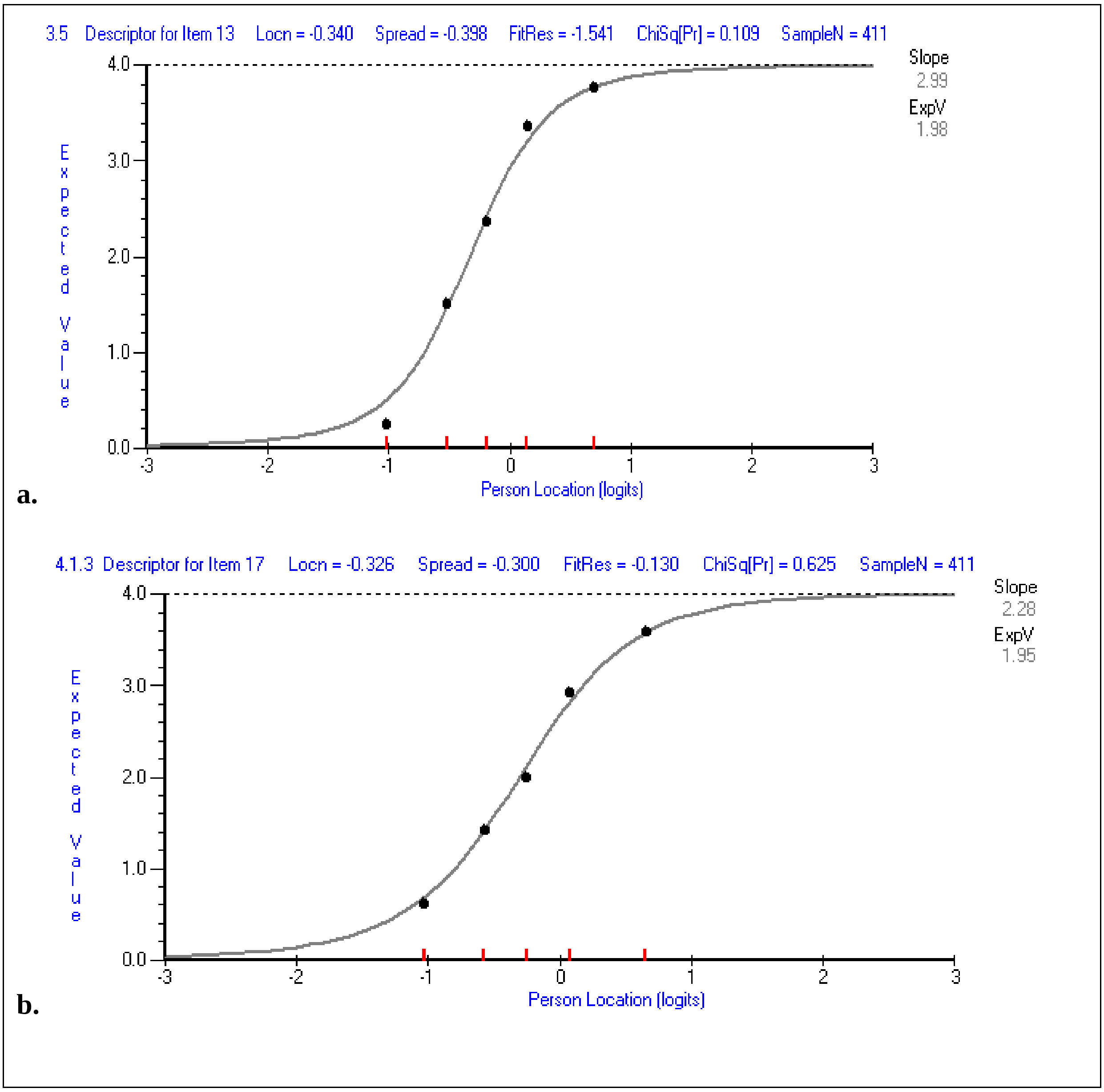


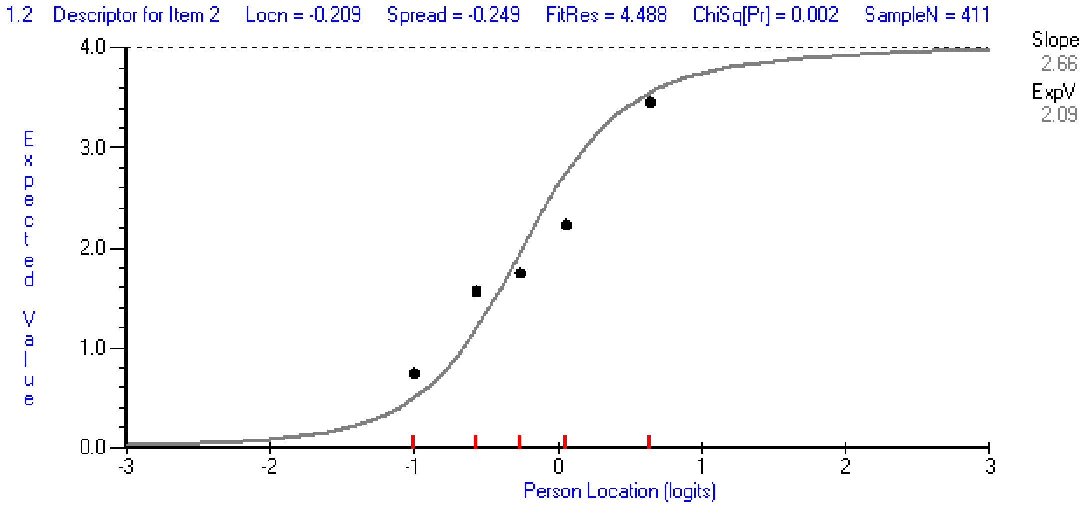
Figure 5.5. ICCs for (a) Item 13/Question3.5 and (b) Item 17/Question 4.1.3, indicating fairly good fit

In Figure 5.5 the actual proportions are in alignment with the theoretical curves, thus showing good fit. The fit residuals for Item 3.5 and Item 4.1.3 are -1.541 and -0.130 respectively both of which are relatively small and within the acceptable range of -2.5 to +2.5 (Smith, 2002). The two items in Figure 5.5 above also have chi square probabilities above the significant value of 0.05, indicating a fairly good fit to the model. A chi square probability value above the chosen criterion of 0.05, means that the difference between the observed proportions and the theoretical curve is not statistically significant, hence the observed differences might have occurred by chance (Andrich, 1988; Smith, 2002). Note that there were many other items besides those indicated above which showed relatively good fit.

5.4.2.2 Under-discrimination (under-fit) items

With under-discriminating items, the actual proportions are more gently sloping than the theoretical curve, meaning the items do not separate the leaners well enough. Under discriminating items indicate that low ability learners performed better than predicted by the model on the particular items. Consequently, because of the interactive nature of item difficulty and learner ability, the high proficiency learners are falsely estimated to respond to the items as if the items were much easier than they really were (Andrich, 1988). Items/Questions 1.2, 2.2.3, 4.1.1and 6.2 are some of those items that showed observed proportions which were flatter than the theoretical curves. The ICCs for Items 1.2 and 6.2 showing under discrimination are shown in Figure 5.6.

a. ICC for Question 1.2



b. ICC for Question 6.2

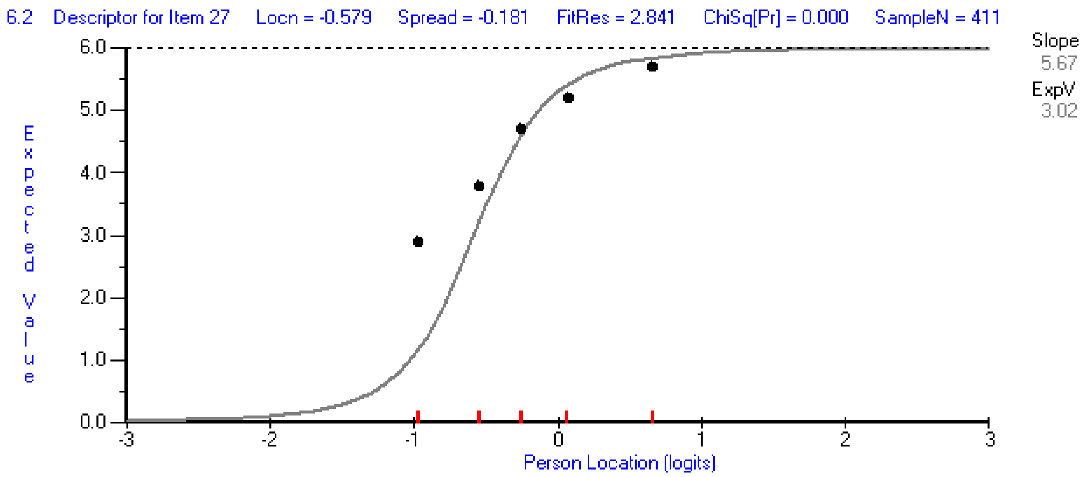


Figure 5.6. ICCs for Items 2/Question 1.2 and Item 27/Question 6.2 showing under-discrimination.

The fit residual for Item 2/Question 1.2 is 4.488 and that for Question 6.2 is 2.841, these are relatively higher than the recommended maximum of 2.5, indicating a big difference between the actual responses and those expected by the model.

For the lower proficiency groups, the actual proportions are above the theoretical curve, meaning they are performing above the expectations or what was predicted by the model. For example, for Question 1.2, a learner with a proficiency level of -1 logits, has an expected value of 0.5, while the observed value is 0.8. However, for the high proficiency groups, the observed proportions are below the theoretical curve, meaning that these are performing below the expectations of the model. For Question 1.2 a learner with a proficiency level of 0.1 logits has an expected value of 2.9 while the observed value is 2.2. The chi square probabilities for Items 1.2 and 6.2 are 0.002 and 0.000 respectively, which are below the significance value of 0.05 indicating that they are not fitting the model requirements (Smith, 2002). The other items with ICCs displaying under-discrimination are Questions 1.2, 2.2.3, 4.1.1, 6.2 and 9.2.

5.4.2.3 Over-discrimination items

Over-discrimination occurs when the curves for the observed proportions and the actual means do not match, and the theoretical curve is gentler than the observed proportions. According to the traditional test theory, high discriminating items are preferred, but with RMT, items which discriminate more raise concerns that there is response dependence in one form or another (Long et al., 2014). Examples of items showing over discrimination are shown in Figure 5.7.

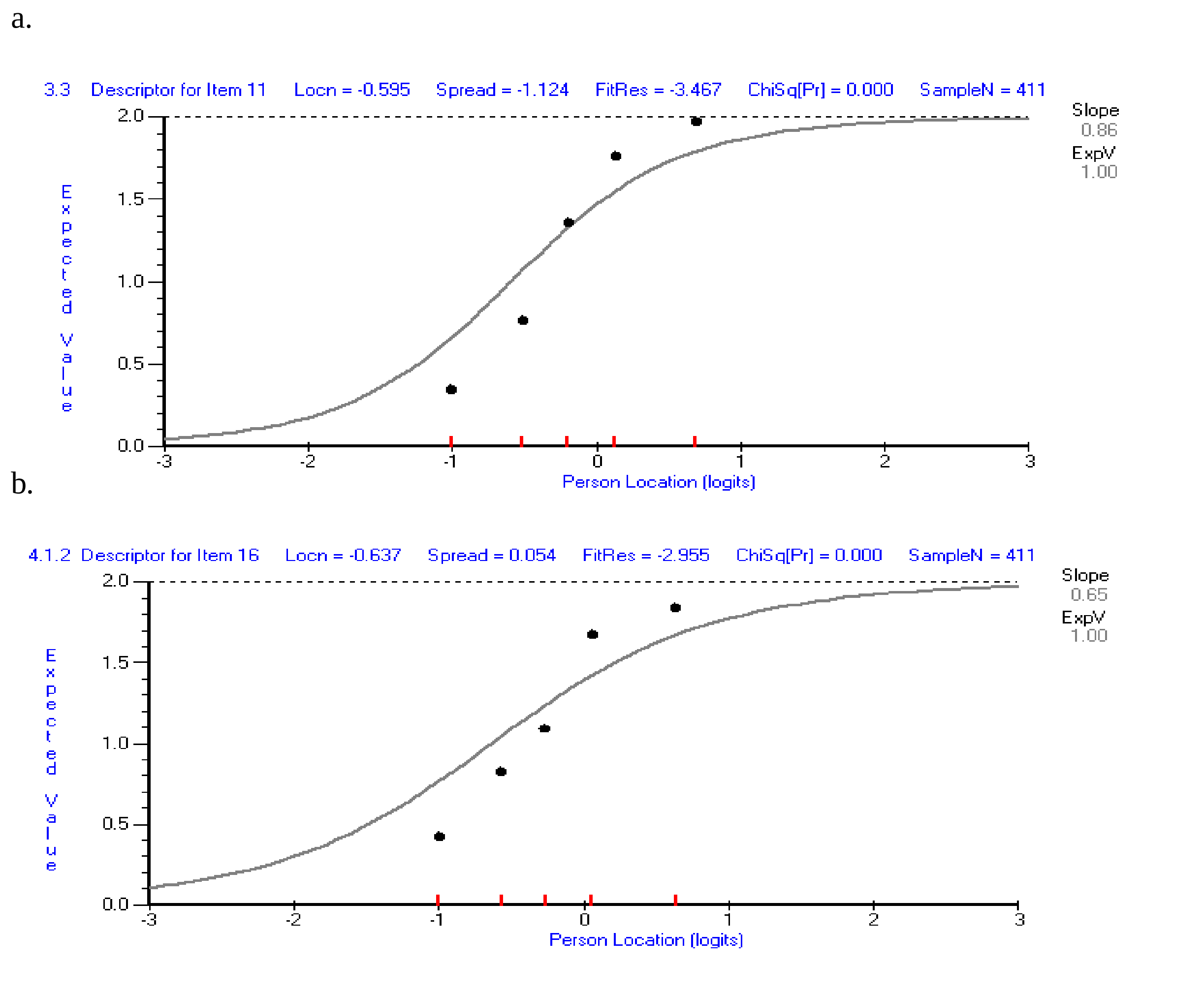


Figure 5.7. ICCs for Questions 3.3 and 4.1.2 showing over-discrimination

The items in Figure 5.7 have negative fit residuals (Questions 3.3 with a fit-residual of -3.467 and Questions 4.1.2 with a fit-residual of -2.995) which are not in the acceptable interval of -2.5 to 2.5. The other items with ICCs displaying over-discrimination are 3.2, 3.3, 3.4 4.1.2, 6.1.1, 8.1, 8.3 and 10.2.2. Poor fit to the model is signalled by both high negative fit residuals and high positive fit residuals.

Note that no items in the data set were observed with ICC's where the observed proportions were haphazardly and substantially different from the theoretical curves. All of the items which showed misfit to the model are analysed in later chapters.

5.4.3 Using category probability curves to check for the functioning of the scoring rubric

The items with fit residuals not in the interval (-2.5 to 2.5) were investigated and their CPC were considered, in order to gain more understanding of how the scoring rubric was functioning. The CPCs for other items were also investigated to check where the scoring rubric was working as expected by the model. According to Van Wyke and Andrich (2006), ordered thresholds are an indication that the hierarchy of responses identified in the scoring rubric reflects the underlying order of the proficiency scale. Disordered thresholds indicate that the scoring rubric for the item does not reflect the underlying proficiency continuum. Disorder in the thresholds is a sign that the item has failed to function as required by the model and hence the item needs to be rescored to try and reflect the order of responses. An item where the CPC showed disordered thresholds is shown in Figure 5.8.

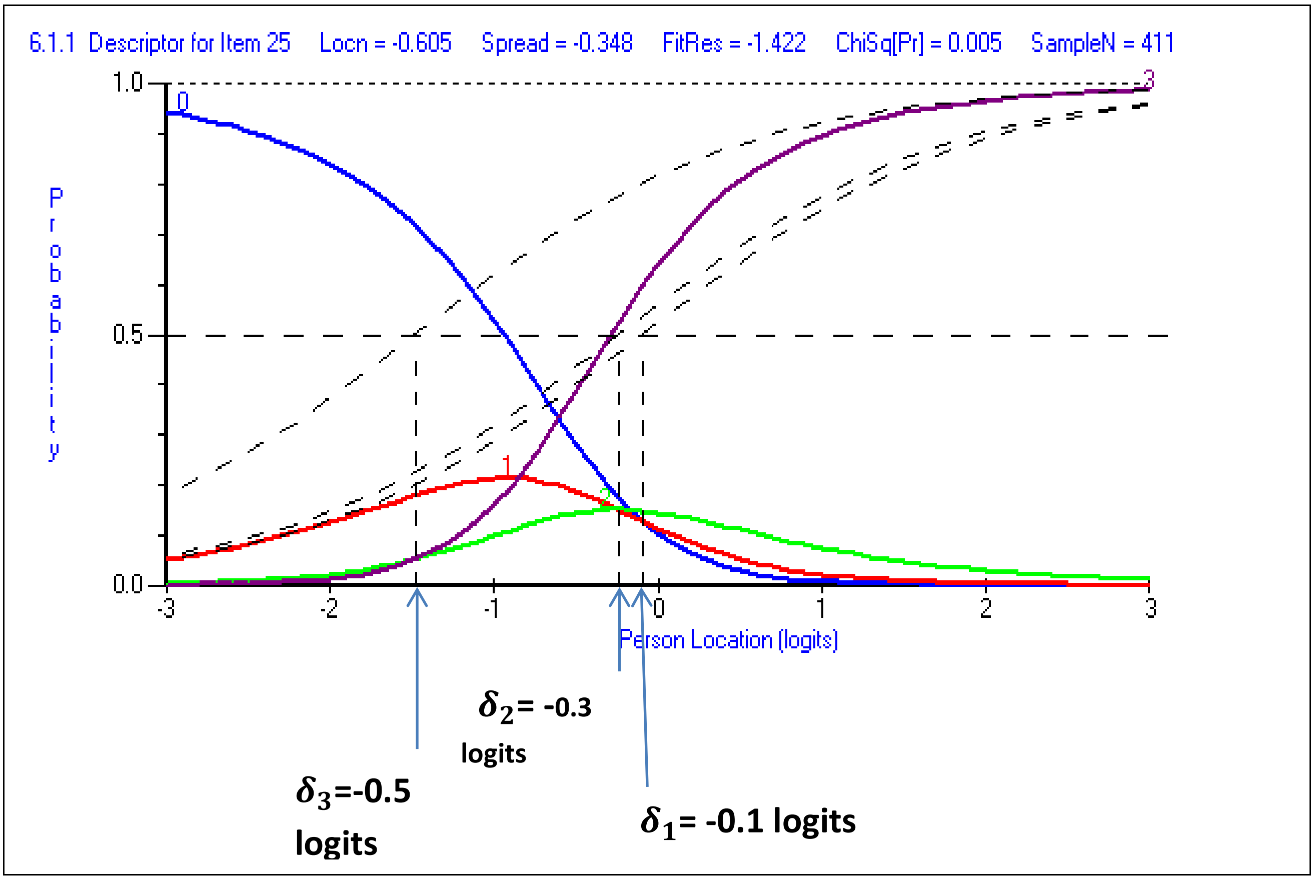


Figure 5.8. CPCs for Item 25/Question 6.1.1

Figure 5.8 shows the CPCs of Item 25 which is a typical polytomous item with four categories. This shows that the probability of a learner with a low proficiency (close to -3 logits) scoring a 0 is very high (0.95), while the probability of maximum marks (3 marks) at the same proficiency level of -3 logits is 0. The probability of scoring 1 mark at the ability level of -3 logits is 0.1. As the ability of the learner increases, the probability of them scoring 0 decreases. In addition, as ability increases the probability of achieving the maximum score of 3 increases.

At ability level 2 logits the probability of scoring a maximum of 3 marks is 0.95 while the probability of scoring a 0 is zero. The CPCs for the item shows disordered thresholds. The Figure 5.8 shows that the location of the first threshold ($\delta_1 = -0.1\ logits$) (the intersection of the curves for the score of 0 and 1) is greater than the location of the second threshold ($\delta_2 = -0.3\ logits$) and third threshold ($\delta_3 = -0.5\ logits$). Reversed thresholds are a result of the middle scores not working as expected. Very few learners got the middle scores, such that there is no point on the horizontal axis where the scores of 1 and 2 are most likely to be scored. Persons with low ability relative to the item's difficulty are still most likely to respond incorrectly and score 0, and persons with high ability relative to the item difficulty are still likely to respond correctly and score 3 marks. Learners with moderate ability, where the marks of 1 and 2 should be most likely, are still more likely to score either 0 or 3.

The disordered thresholds witnessed in Item 25 were also witnessed in many other items, where the middle scores did function well. The items with disordered thresholds provided me with the basis for rescoring, to try and correct the categories which were not functioning as expected. The rescoring process is covered in the next chapter.

CHAPTER 6 RESCORING OF ITEMS WITH FIT RESIDUALS OUTSIDE THE RECOMMENDED RANGE

6.1 Introduction

The items with fit residuals not within the recommended interval of -2.5 to 2.5 and where the CPCs showed disordered thresholds, were the first to be considered for rescoring to try and reflect the hierarchy of responses. The use of the CPCs helps to further explore anomalies in a data set besides the use of the fit statistics. The fit residuals helped to pinpoint the items that were problematic but a look at the CPCs showed that many items have thresholds that were disordered. When the scoring rubric did not reflect the underlying proficiency continuum, disordered thresholds appeared in the CPCs.

In trying to refine the instrument, I first identified items which did not meet the model requirements using the residual statistics and chi square probability. I also investigated the item characteristic curves (ICCs), to check whether the observed proportions were in line with the theoretical proportion. For items where there were anomalies, I looked closely at the questions and the marking guideline to check for any complications and deviations from a mathematics education perspective and an assessment perspective. In cases where the qualitative analysis confirmed the anomaly, rescoring was done. After every rescoring process, the item statistics were investigated and re-analysed using the new or revised scores. Where qualitative analysis could not find any theoretical reason to support the rescoring process for an item, then rescoring did not take place for that particular item.

6.2 Items with Fit Residuals Outside the Recommended Range

The items with fit residuals not in the recommended range are discussed and rescored in this section to try and improve the general fit to the model. Eight questions had fit residuals that were not in the acceptable interval, and these are given in Table 6.1.

Table 6.1

Items with fit residuals outside recommended range

Question	1.2	3.3	4.1.1	4.1.2	6.2	8.1	8.3	9.2
Fit residual	4.488	-3.467	2.579	-2.955	2.841	-3.340	-3.076	2.887

These items are discussed in detail in the next section, as well as the rescoring that was done to try and improve the fit to the model.

6.2.1 Analysis and rescoring of Question 1.2/Item 2

Question 1.2 had a Fit residual of 4.488, which is outside the recommended range. The ICC in Figure 6.1 for Item 2 (Question 1.2) shows extreme misfit. The chi square probability is 0.000 less than the significant result of 0.05. The item needs further probing to check if the misfit is a result of the scoring structure and whether rescoring may solve the misfit.

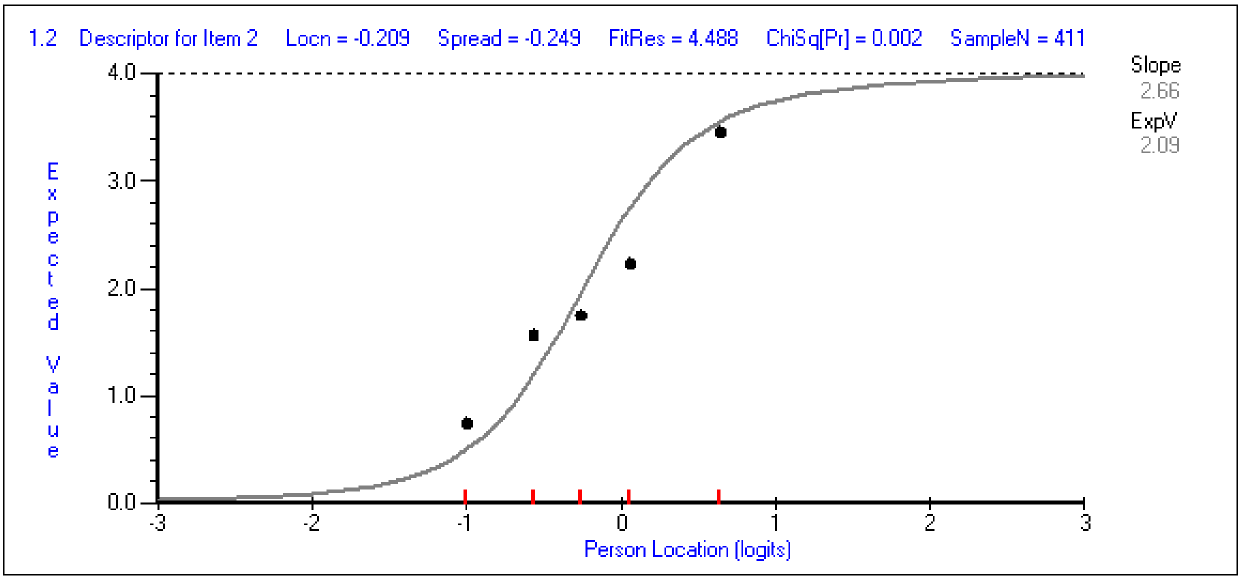


Figure 6.1. ICC for Item 2/Question 1.2

The item required learners to draw an ogive curve or a cumulative frequency graph. The item carried four marks or had four categories. The first mark was accorded for drawing the correct shape for an ogive curve even though the points/coordinates may be wrong, having an S-shape would guarantee a learner their first mark. Another mark was granted for grounding the ogive curve, and plotting the first coordinate (120; 0) correctly. The third mark was allocated to the middle point/coordinate (165; 61), noting that there was an error in the marking memorandum, as it was written as (60; 165) instead of (165; 61). The fourth mark was allocated to any other correct plotted points, to make up four marks in total. The mark allocation is clearly shown in Figure 6.2.

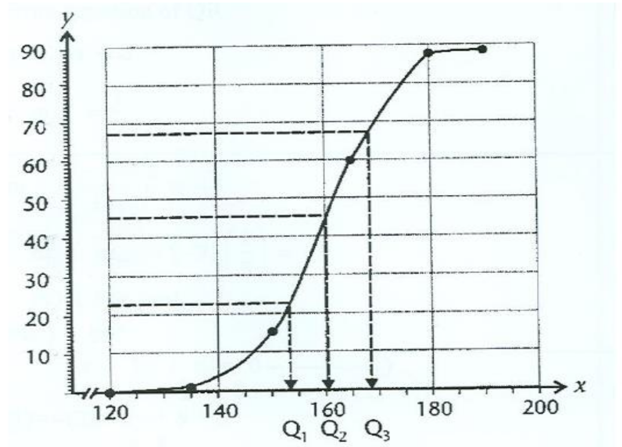
(a) QUESTION 1

The table below shows the heights of palm trees in a park.

1.1	Determine the estimated mean height of the palm trees in the park.	(2)
1.2	Draw an ogive to represent this data.	(4)

(b) **Marking memorandum for Item 1.2**

1.2



✓ A shape
✓ A (120;0)
✓ A (60;165)
✓ A for any other correct plotted points

(4)

Figure 6.2. (a) Question 1.2/Item 2 and (b) the marking guideline

All of the marks allocated for Question 1.2 were accuracy marks (represented by letter A on the marking memorandum) and any deviation from the required answer would result in the learner losing the mark. The distribution of categories of marks is shown in Table 6.2.

Table 6.2

*Category Response Frequencies for Question 1.2 (Item 2***)**

Category (mark)	0	1	2	3	4
Frequency (learners)	137	32	36	54	115

Table 6.2 shows that the frequencies for the scores of 1, 2 and 3 were very low compared to the scores of 0 and 4, an indication that the scoring might not be functioning well. The CPCs for the item are given in Figure 6.3.

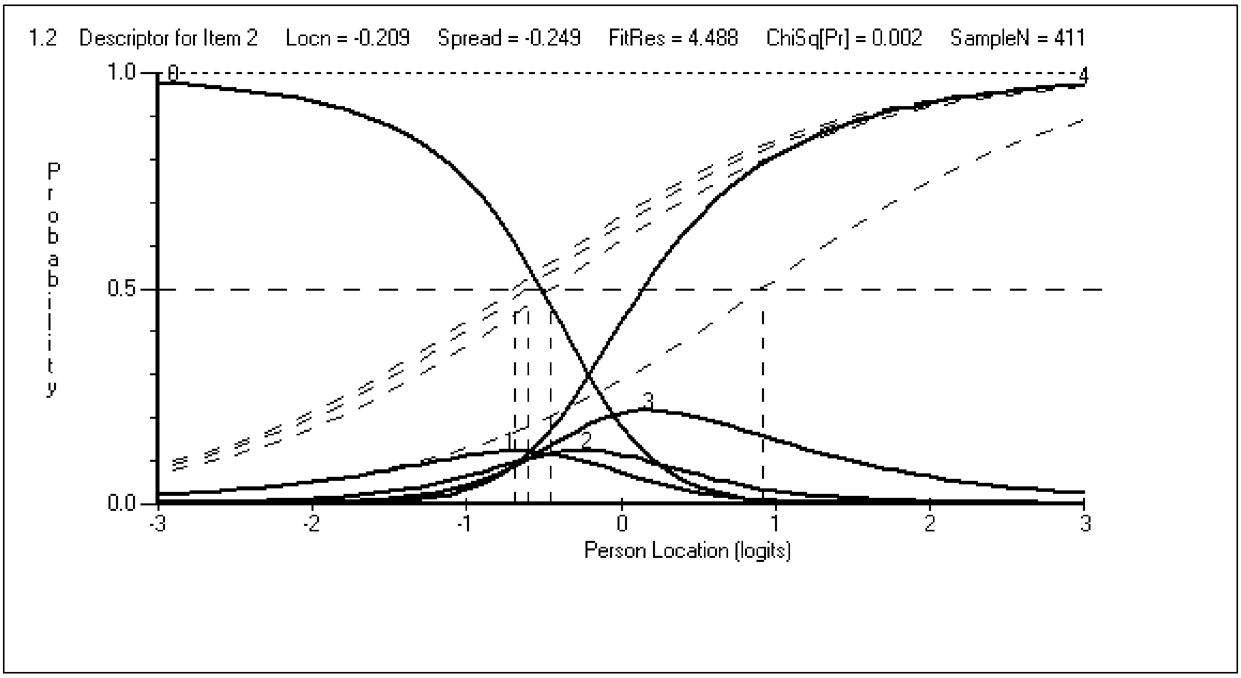


Figure 6.3. Initial CPCs for Question 1.2/Item 2.

Figure 6.3 shows that the scoring for categories 1, 2 and 3 were not working as intended as indicated in the initial CPCs. There are disordered thresholds, an indication that the item has failed to function as required by the model. The first threshold, which represents the point at which a score of 0 and 1 is equally likely and after which a score of 1 becomes more likely than a score of 0, has a location of 0.913392 logits. The second threshold is the point after which a score of 2 becomes more likely than a score of 1, and is approximately -0.451533 logits. The third and fourth thresholds are -0.605962 logits and -0.693363 logits respectively. The first threshold is at a location greater than that of the second, third and fourth thresholds. Along the horizontal axis, there is no point where the score of 1 is most likely, neither is there an interval or point where the scores of 2 and 3 are most likely. Hence the scores of 1, 2 and 3 are not functioning as intended. The most likely scores are 0 for low proficiency learners (below -0.2 logits) and 4 for high proficiency learners (above -0.2 logits). Table 6.2 shows that the categories most likely to be scored were 0 and 4, with category 0 having a frequency of 137 learners and category of 4 having a frequency of 115 learners

A look at the marking memorandum in Figure 6.2 above reflects why this might be so. A learner who is able to plot one or two coordinates on the graph is highly likely to be able to plot the other points, and hence to join the points to form an ogive curve. The third and fourth are considered as redundant marks as these are separate marks awarded for the same skill of sketching points on a graph. This disadvantaged the low proficiency learners as they were losing all the marks, and was of advantage to the high proficiency learners who were highly likely to get all marks. The learners (54) who scored 3 marks did not ground the ogive curve by not plotting the first point (120; 0) on the graph. Some of the learners' errors which resulted in them losing marks are shown in Figure 6.4.

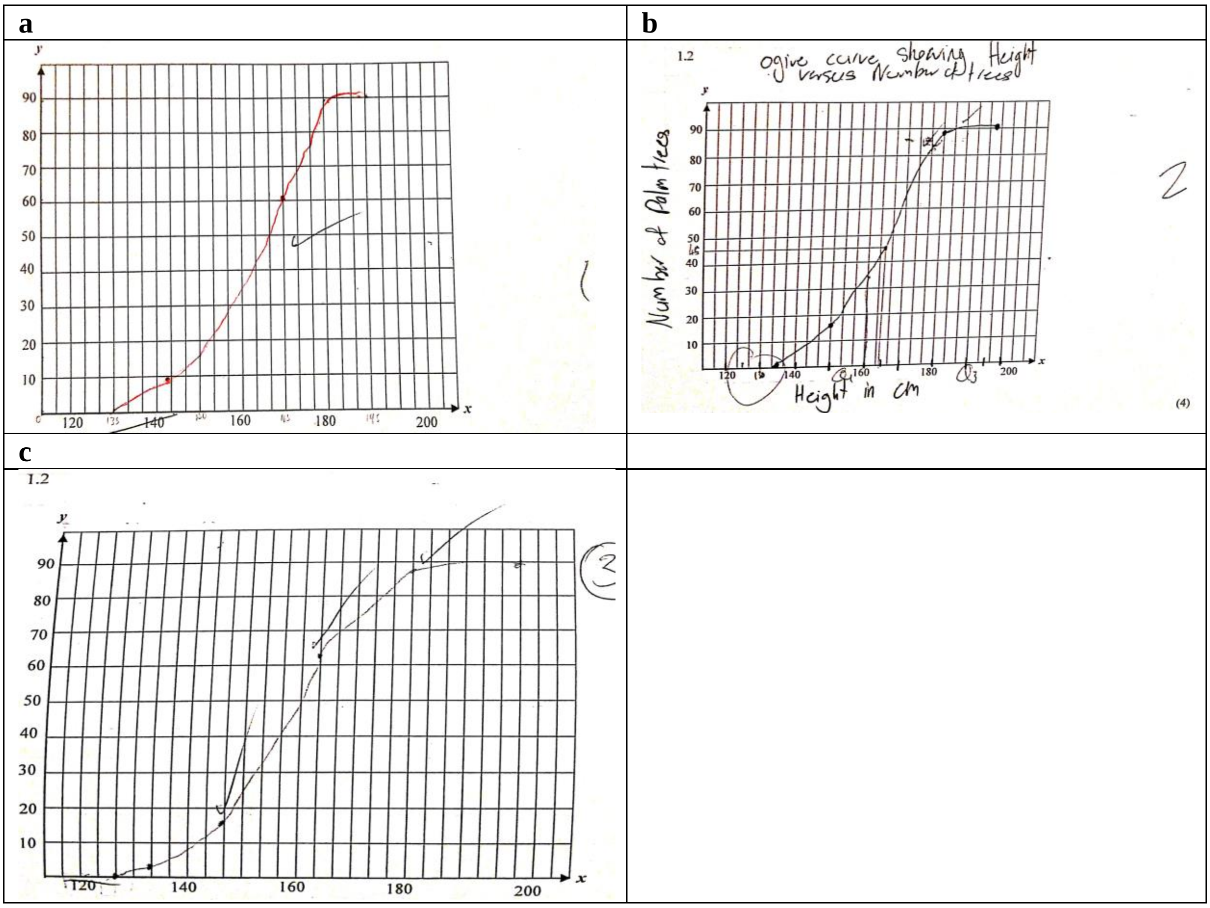


Figure 6.4. Examples of a graph of (a) a learner who scored 1 mark, (b) a learner who scored 2 marks, and (c) a learner who scored 3 marks on Question 1.2.

Most of the learners who got 1 mark on Question 1.2 were awarded it for the correct shape (S-shape) of the ogive curve, although the points were wrong. Thirty-two of the learners who responded to this question got a mark of 1, and in all cases the shape was the only correct aspect.

Those who were awarded 2 marks (Figure 6.4(b)) managed to draw the correct shape and at least one other point that was correct on the graph, while those who got 3 marks managed to draw the correct shape and all the other points besides the first point/coordinate as shown in Figure 6.4(c). In many cases the only problem was grounding the graph by drawing the first coordinates (120; 0).

Rescoring was done to try and improve the scoring rubric. Category 1 was collapsed to 0, categories 2 and 3 were collapsed into 1 and category 4 was rescored as 2. The category curves after the rescoring show that the categories are still not working. The initial CPCs and the CPCs after rescoring are presented in Figure 6.5.

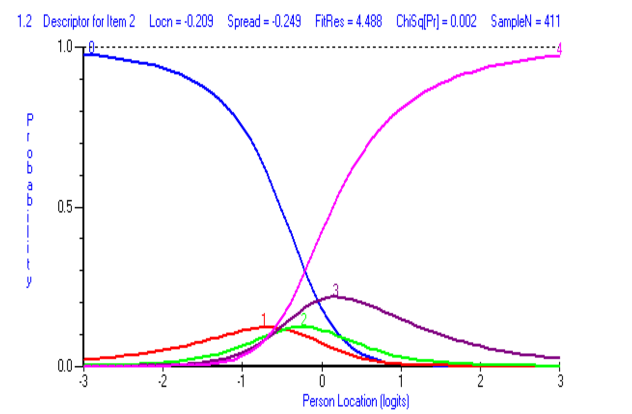
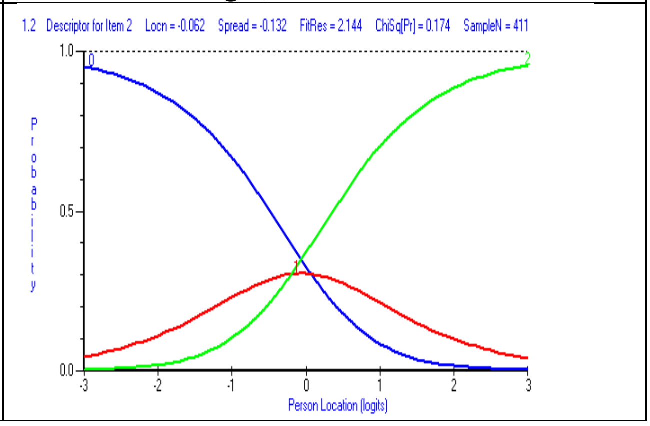
(a) Initial Category Probabilty Curves for Question 1.2	**(b) Category probability curves After Rescoring Question 1.2**
	

Figure 6.5. Initial CPCs and the CPCs after rescoring for Question 1.2

The CPCs for the rescored Item 2 (Question 1.2) showed a fit residual of 2.144 that is within the recommended range. The score of 1 is still not working well. The chi square probability is now 0.174, a value that is greater than the significant value of 0.05, indicating a better fit than before rescoring. The rescored item does not have redundant marks anymore, where separate marks are allocated for the same skill.

6.2.2 Analysis and rescoring of Question 3.3

The ICCs for Question 3.3 showed observed proportions which were not aligned to the theoretical curve. The actual proportions are much steeper than the theoretical curve. Learners with low proficiency (below -0.2 logits) performed less well than predicted by the model, while the high proficiency learners performed better than expected by the model. The fit residual for Question 3.3 is -3.467, which is not in the acceptable range of -2.5 to 2.5. The ICC for the item is shown in Figure 6.6.

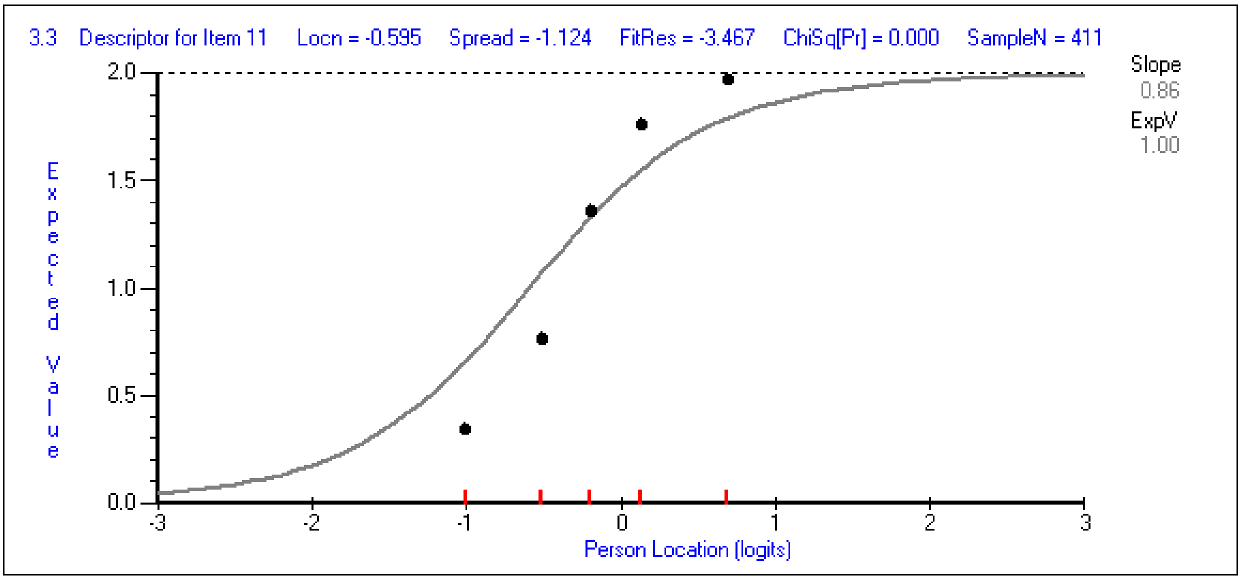


Figure 6.6. ICC for Question 3.3/Item 11

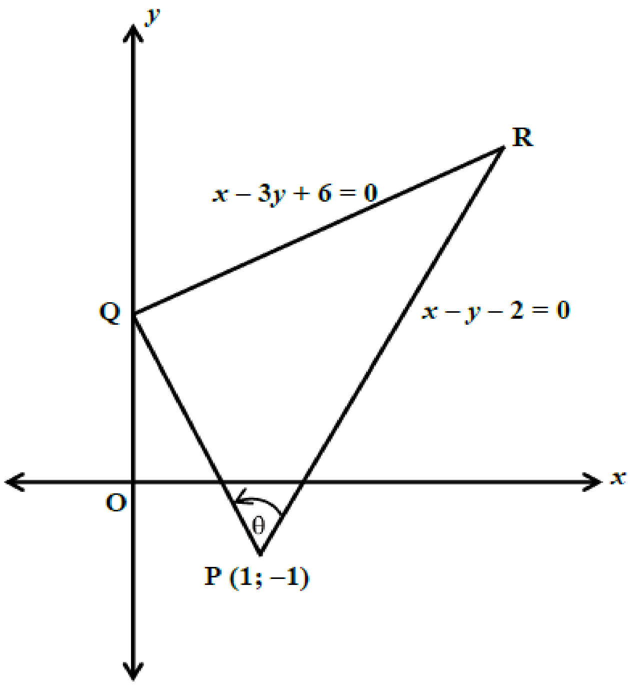
Figure 6.6 shows that the chi square probability for Question 3.3 is 0.000, which is below the set criterion of 0.05, and this shows that the differences between expected and observed outcomes was statistically significant. The item is showing over-discrimination.

Question 3.3 was based on Analytical Geometry, and learners were required to prove that angle $P\hat{Q}R$ is equal to 90˚. The question and the accompanying marking memorandum are given in Figure 6.7.

a.

QUESTION 3

In the figure below, PQR is a triangle with P(1 ; −1) . Q is a point on the y-axis. The equations of QR and PR are $x - 3y + 6 = 0$ and $x - y - 2 = 0$ respectively. Given $Q\hat{P}R = \theta$.



3.1	Show that the co-ordinates of Q are (0 ; 2).	(2)
3.2	Write down the gradient of QR.	(2)
3.3	Prove that $P\hat{Q}R = 90°$.	(2)

b.

3.3	$m_{PQ} = \frac{-1-2}{1-0} = -3$ $\therefore m_{PQ} \times m_{QR} = (-3)\left(\frac{1}{3}\right) = -1$ $\therefore PQ \perp QR$ Thus $P\hat{Q}R = 90°$	✓ A gradient of PQ ✓ A products of gradients (2)

Figure 6.7. (a) Question 3.3 and (b) the marking guideline

For the learners to answer this question successfully, they were required to determine the gradients of the lines PQ and QR and multiply to get a product of negative 1. The first mark was allocated for determining the gradient of line PQ. The gradient of line QR was already calculated in Question 3.2. The second mark was allocated for getting the correct product of the gradients. The 2 marks allocated for this question were all accuracy marks (A). The category frequencies from RUMM showed that 110 learners got 0, 35 learners got 1 mark and 189 learners got 2 marks. Very few learners scored 1 mark compared to the other categories. The CPCs for the item are provided in Figure 6.8.

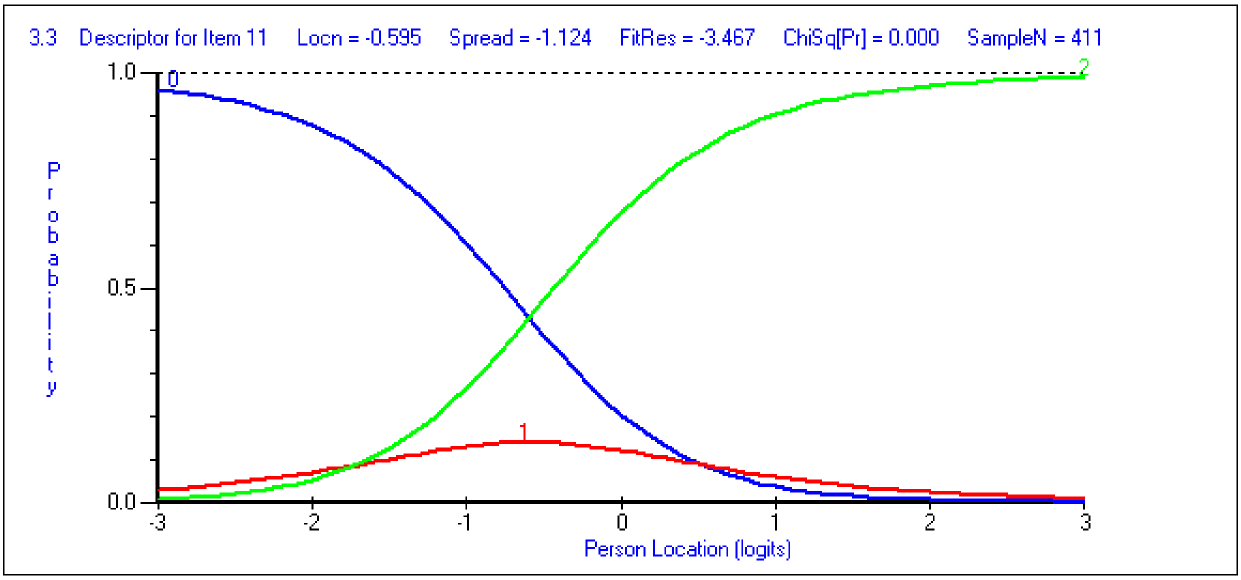


Figure 6.8. CPCs for Question 3.3

The second category (that for scoring 1 mark) was not working as intended. At no point along the horizontal axes is the score of 1 the most likely. The location where it is equally likely to get a score of 0 and 1 is at a higher location than the location where it is equally likely to get a score of 1 and 2.

From the learners' responses, many who managed to calculate the gradient of line PQ, were also able to multiply the gradients and get a product of negative 1. Very few learners managed to calculate the gradient and were not able to get the product of the gradients.

Rescoring was done to try and improve the scoring rubric by collapsing scores of 1 and 2 to 1 mark. The fit residual after rescoring was -2.606 which was still showing misfit as it was still outside the recommended range. The ICCs for the rescored item showed observed proportions which were still steeper than the theoretical curve and showed over-discrimination. The chi square probability after rescoring was still 0.000, showing that the difference between the observed and expected outcomes was still statistically significant. Since the rescoring could not resolve or improve the fit of this item, it was decided to retain the original scores and item dependence was later checked.

6.2.3 Analysis and rescoring of Question 4.1.1

Question 4.1.1 has a fit residual of 2.579, which is outside the recommended range of -2.5 to 2.5. The ICC also shows item misfit, as the observed proportions are flatter than the theoretical curve, showing that the item is under-discriminating. Learners with low proficiency were performing better than predicted by the model and those with high proficiency were performing less than expected by the model. The ICC for the item is shown in Figure 6.9.

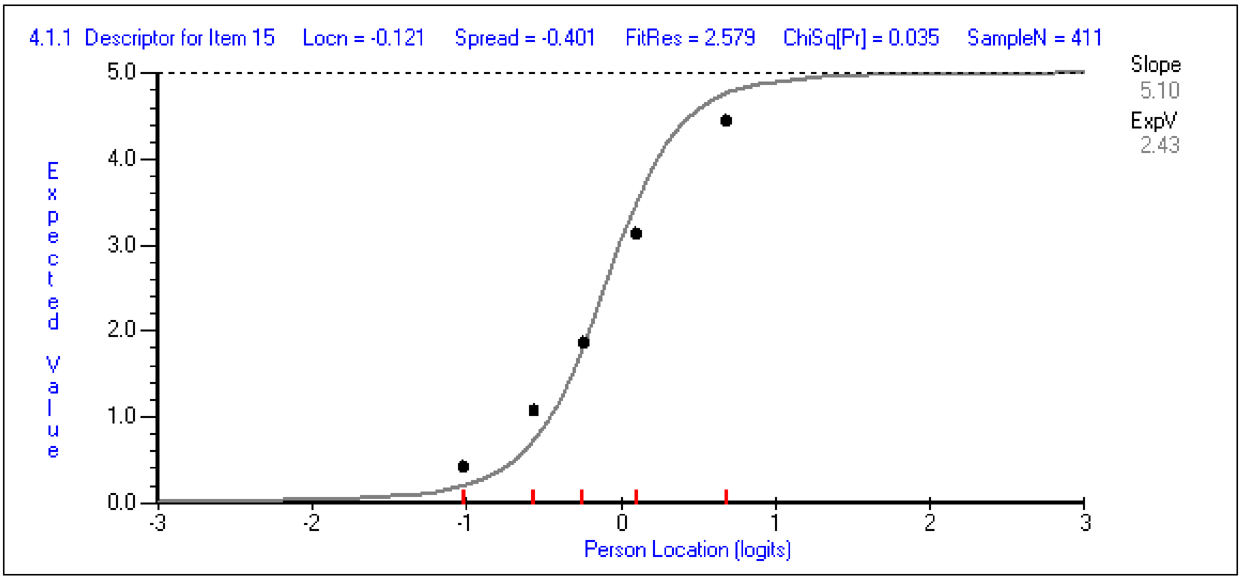


Figure 6.9. ICCs for Question 4.1.1

The chi square probability for Question 4.1.1 is p = 0.035, which is below the set criterion of 0.05, showing that there is a statistically significant difference between the observed and expected outcomes.

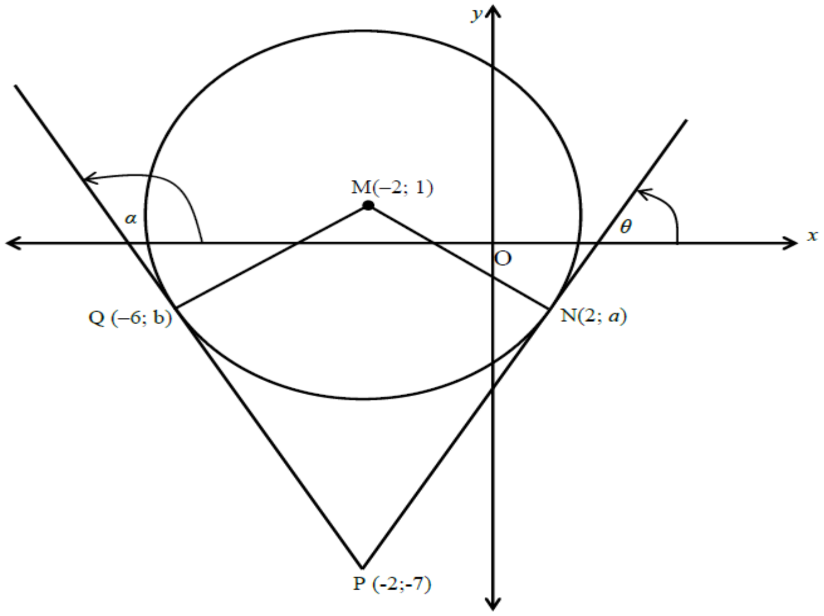
Question 4.1.1 was based on Analytical Geometry and required learners to show/deduce that $a = -3$. The question had 5 marks corresponding to 6 categories. All the marks were accuracy (A) marks. The question and the corresponding marking memorandum are shown in Figure 6.10.

The first mark was allocated for using the theorem that the radius of a circle is perpendicular to the tangent of the circle, hence multiplying the gradients of the two lines will give a product of negative 1. The second mark was allocated for substituting the gradients followed by simplification, writing in standard form and then factorisation to get full marks. In the case of accuracy marks, any deviation from the given answers will not be accepted.

a.

QUESTION 4

4.1 In the diagram below, MN is a radius of a circle with centre $M(-2\ ;\ 1)$. The co-ordinates of N are $(2\ ;\ a)$ and $a < 0$. The co-ordinates of P are (-2’-7). PQ and PN are tangents to the circle at Q and N respectively.
The coordinates of Q is $(-6\ ;\ b)$. PM is parallel to the y – axis.



4.1.1 Deduce that $a = -3$. Show all your workings. (5)

4.1.2 Determine the equation of the circle. (2)

b.

4.1.1	$m_{MN} \times m_{NP} = -1$ Radius ⊥ Tangent $\left(\frac{a-1}{2+2}\right)\times\left(\frac{a+7}{2+2}\right) = -1$ $\left(\frac{a-1}{4}\right)\left(\frac{a+7}{4}\right) = -1$ $\frac{a^2 - a + 7a - 7}{16} = -1$ $a^2 + 6a - 7 = -16$ $a^2 - 6a + 9 = 0$ $(a+3)(a+3) = 0$ $\therefore a = -3$	✓ A $m_{MN} \times m_{NP} = -1$ ✓ A substitution ✓A simplification ✓A standard form ✓A factorization (5)
4.1.2	$MN^2 = r^2 = (-2-2)^2 + (1+3)^2 = 32$ $(x+2)^2 + (y-1)^2 = 32$	✓ A value of radius ✓CA equation of circle (2)

Figure 6.10. (a) Question 4.1.1 and 4.1.2, and (b) the marking guideline

The CPCs for the item were investigated and Figure 6.11 shows that some of the categories were not working as expected.

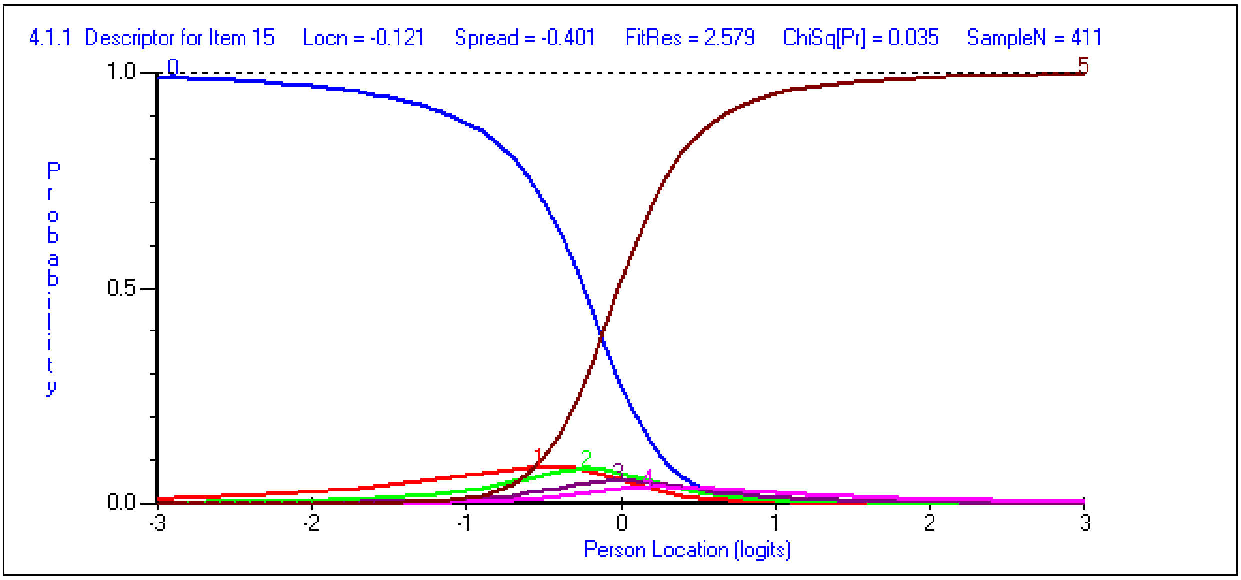


Figure 6.11. CPCs for Question 4.1.1

Figure 6.11 showed that the categories 1, 2, 3 and 4 were not working as intended. The thresholds are disordered and there is no point along the horizontal axes where the scores of 1, 2, 3 and 4 are the most likely to be scored. For the low proficiency learners (below -0.1 logits) the most likely score is 0, and for the high proficiency learners (above -0.1 logits) the most likely score is 5.

The scores of 1 and 2 were collapsed to 0, the score 3 rescored as 1, score of 4 rescored as 2 and the score of 5 rescored as 3. Rescoring improved the general fit of the item with the rescored item now having a fit residual of 1.78, which is within the recommended range.

6.2.4 Analysis and rescoring of Question 4.1.2

Question 4.1.2 showed misfit to the model with a fit residual of -2.955, which is outside the recommended range. The ICCs for this question showed observed proportions which are steeper than the predicted curve. The lower proficiency group performed less well than predicted by the model and the higher proficiency group performed better than predicted by the model. The item showed over-discrimination.

The question was based on analytical geometry and required learners to determine the equation of the circle. Two marks were allocated for the question, 1 accuracy (A) mark for calculating the correct value of the radius and 1 continuous accuracy (CA) mark for the equation of the circle. The question and the corresponding marking rubric are shown in Figure 6.10.

Rescoring was done by collapsing the scores 1 and 2 to 1. The fit residual after rescoring improved to -1.199 which is with the recommended range.

6.2.5 Analysis and rescoring of Question 6.2/Item 27

The initial analysis showed that Item 27/Question 6.2 needed further investigation, with the ICCs showing a fit residual of 2.841, outside the recommended range of -2.5 to 2.5. The chi square probability is 0.000, less than the significant value of 0.05, showing that the difference between observed and expected outcomes is statistically significant. The ICCs for Question 6.2 shows observed proportions that are flatter than the expected theoretical curve. The item is under-discriminating. The ICCs for Question 6.2 is shown in Figure 6.12. The lower proficiency group (less than 0.3 logits) performed better than predicted by the model, and the higher proficiency group (more than 0.3 logits) performed less well than predicted by the model.

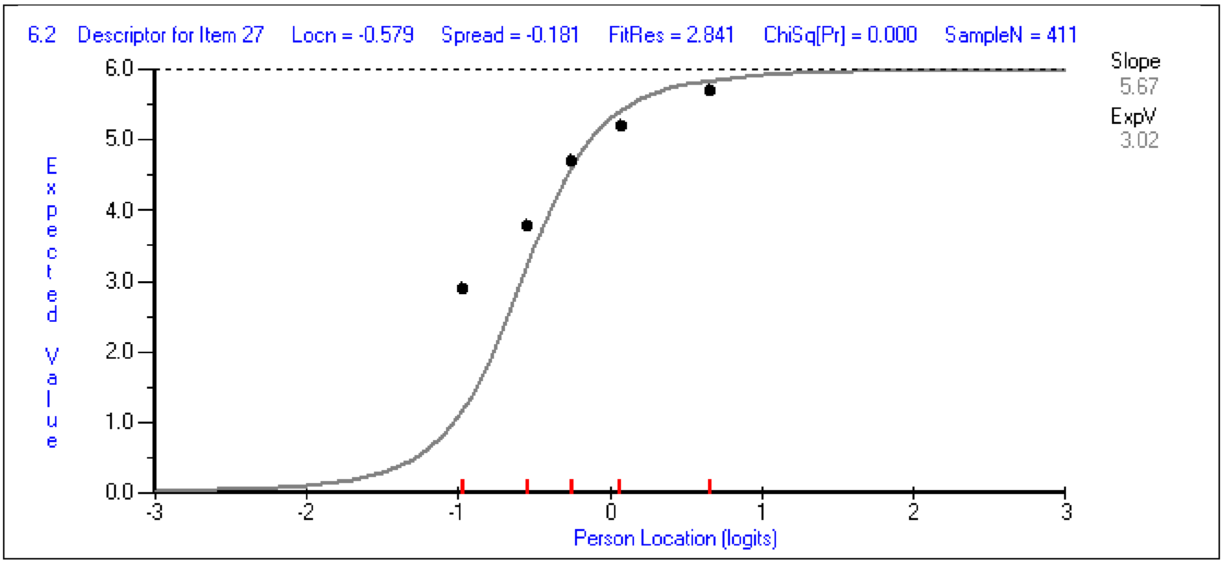


Figure 6.12. ICC for Item 27/Question 6.2

The ICCs also shows the observed proportions were not aligned to the theoretical curve, prompting the researcher to carry out further investigations.

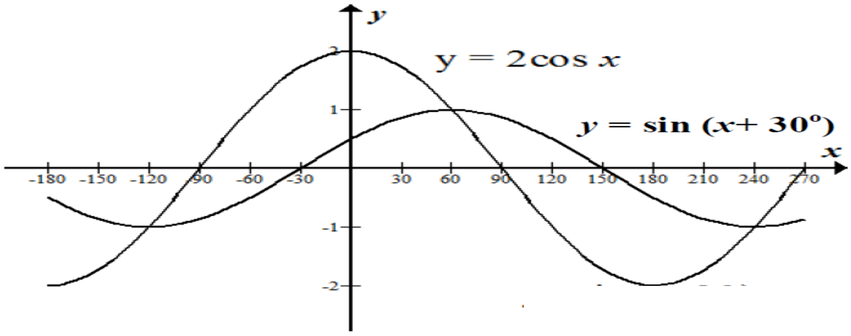
Question 6.2 is based on trigonometric graphs. The question required learners to sketch the graphs of $f(x) = 2\cos x$ and $g(x) = \sin(x + 30^{\circ})$ on the same set of axes which was provided. Calculator use was useful in coming up with the correct coordinates. By changing the calculator into table mode and inputting the trigonometric function with the interval, the learner would get all of the coordinates to plot. The question and the marking guideline are provided in Figure 6.13.

(a)

6.2 On the axes provided, in your answer book sketch the graphs of $f(x) = 2\cos x$ and $g(x) = \sin(x + 30^\circ)$ for the interval $x \in [-180^\circ ; 270^\circ]$. (6)

(b)

6.2



✓A✓A x - intercepts of EACH graph
✓A✓A - maximum and minimum values of EACH graph
✓A✓A - shape of EACH graph

(6)

Figure 6.13. (a) Question 6.2/Item 27, and (b) the marking guideline

The 6 marks allocated for this question were all accuracy marks. The initial CPCs for Question 6.2 are given in Figure 6.14.

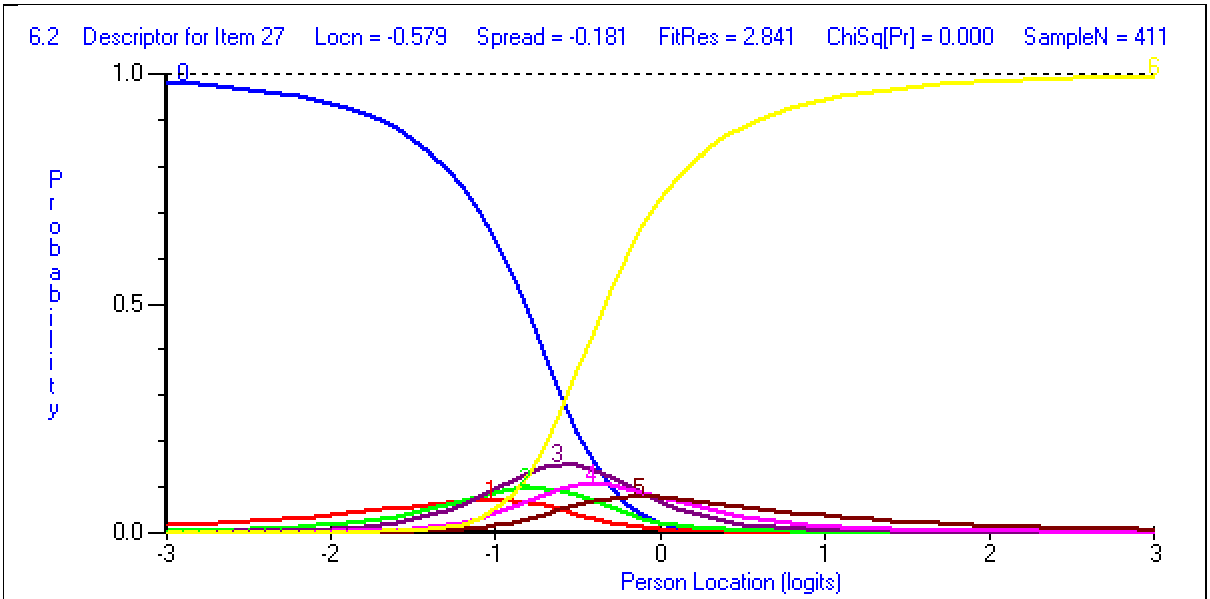


Figure 6.14. Initial CPCs for Question 6.2/Item 27

The initial CPCs in Figure 6.14 show disordered thresholds. The first threshold had a location of 1.205153 logits, the second a location of -1.205533, the third a location of -1.103001 logits, the fourth a location of -0.141312 logits, the fifth a location of 0.25475 logits and the sixth threshold a location of -2.256699 logits. These disordered thresholds are a result of the middle categories (corresponding to the scores 1, 2, 3, 4 and 5) not functioning as intended. The thresholds indicate that achieving a score of 6 does not require higher ability than achieving a score of 1,2, 3, 4 or 5. A closer look at the marking rubric in Figure 6.13(b) shows why this might be so. The question required learners to sketch the graphs of $f(x) = 2\cos x$ and $g(x) = \sin(x + 30^o)$. The category frequencies are given in Table 6.3.

Table 6.3

Category frequencies for Question 6.2/Item 27

Category (Mark)	0	1	2	3	4	5	6
Frequency	49	5	17	38	21	26	212

The category frequencies show that the middle categories are not working as expected, a finding which is supported by the initial CPCs in Figure 6.14.

Learners who used their calculators to find the table of contents or the coordinates required for plotting the graphs were able to draw both graphs and scored the maximum marks. Two marks were allocated for the x-intercepts of each graph, another 2 marks for the maximum and minimum values of each graph and the last 2 marks for the shape for each graph. The majority of the learners got 0 or the maximum number of marks, with 49 learners getting 0 and 212 getting all six marks. Out of the 411 learners who were involved in the study, 342 attempted to answer this question. Some cases where learners did not get all of the possible marks are shown in Figure 6.15.

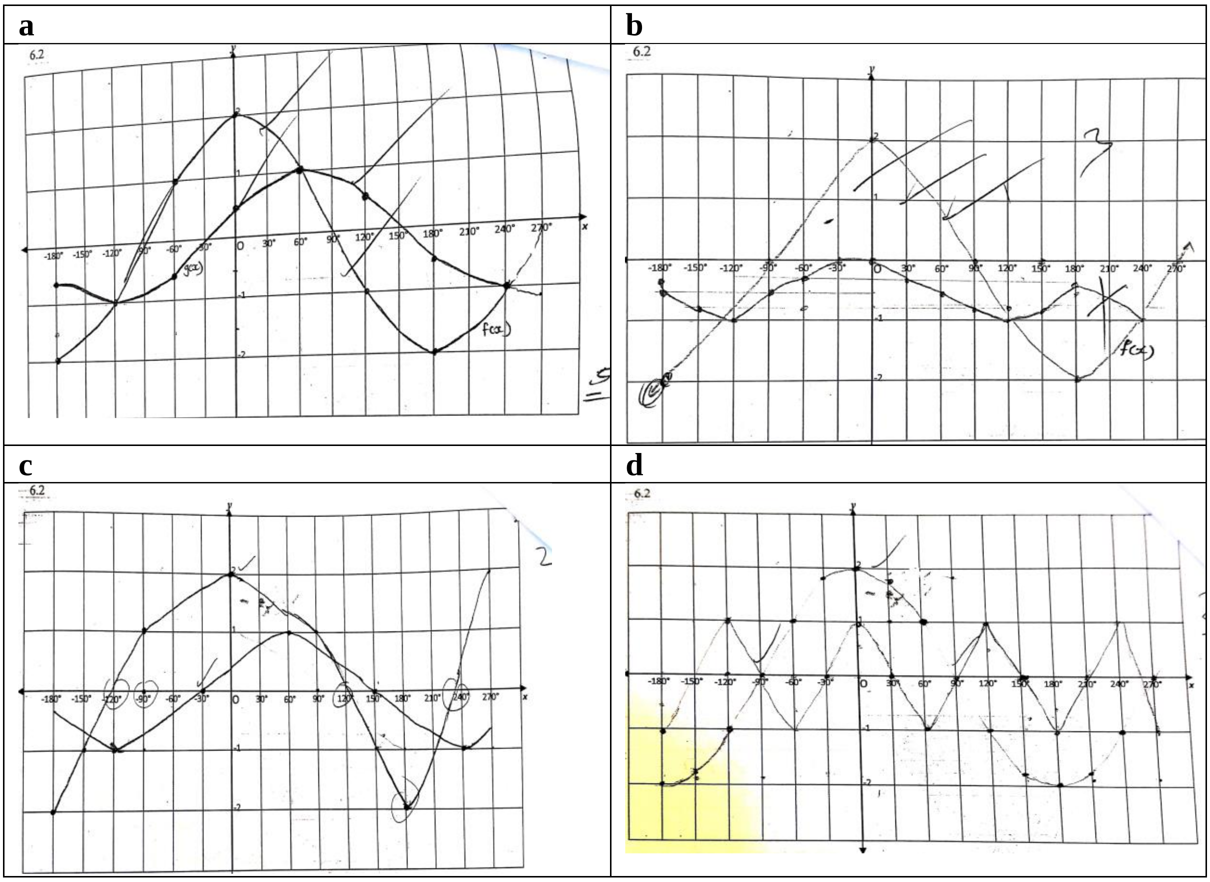


Figure 6.15. Sample of learners' responses to Question 6.2/Item 27

In (a) the learner got 5 marks out of the possible 6 because they did not plot all the required coordinates, from 240˚ to 270˚ for f(x). Because of the missing coordinates, the graphs were considered incomplete. In (b) the graph of f(x) is correct but the graphs of g(x) are incorrect, hence the learner got 3 marks out of the possible 6. In (c), few coordinates were correct and the learner was awarded only 2 marks. In (d) one graph was correct while the other was incorrect.

Rescoring was done to try and improve the scoring of the item. The scores of 0 and 1 were rescored to 0, and the scores of 2 and 3 were collapsed to 1, while scores of 4 and 5 were collapsed to 2 and the score of 6 was rescored to 3. Hence a learner who was able to sketch one graph and not the other would get 1 mark, and a learner who was able to sketch both would get 3 marks. Those who were not able to sketch both graphs would lose only 3 marks instead of all 6 marks. The fit residual changed after rescoring from 2.841 logits which is outside the recommended range, to 1.833 logits, which is within the acceptable range. The overall fit of the item improved.

6.2.6 Analysis and rescoring of Question 8.1

Question 8.1 has a fit residual of -3.340, which is outside the recommended range. The ICCs in Figure 6.16 for Question 8.1 shows misfit. The chi square probability is 0.000, which is less than the significant result of 0.05 and shows that the difference between the observed and theoretical outcomes is statistically significant. The item needed more probing to check where the problem lay.

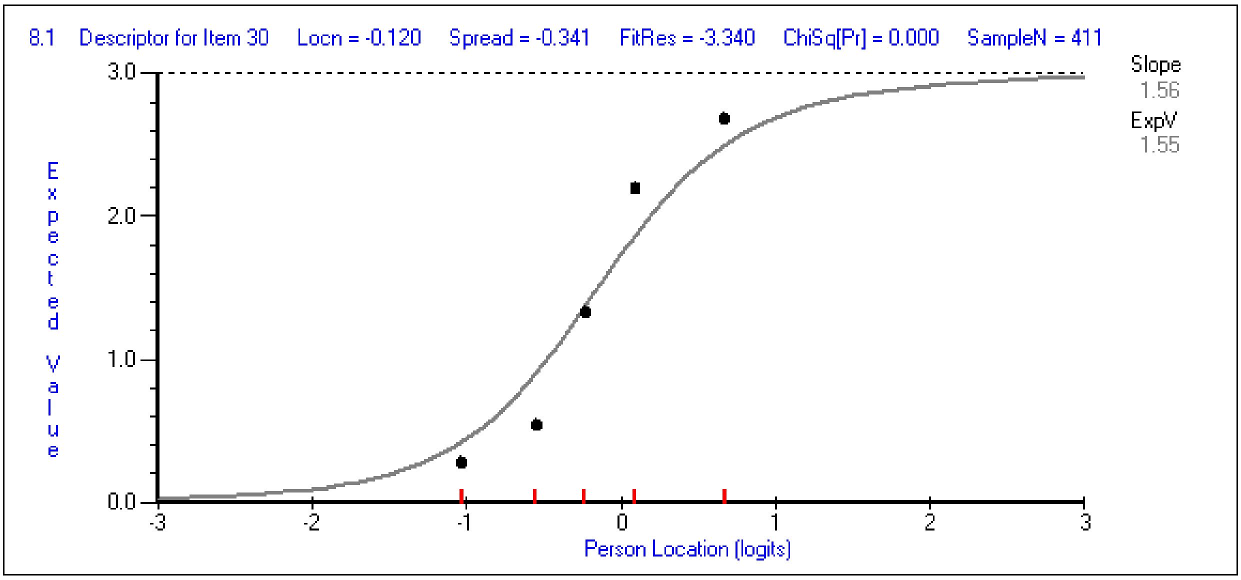


Figure 6.16. ICC for Question 8.1

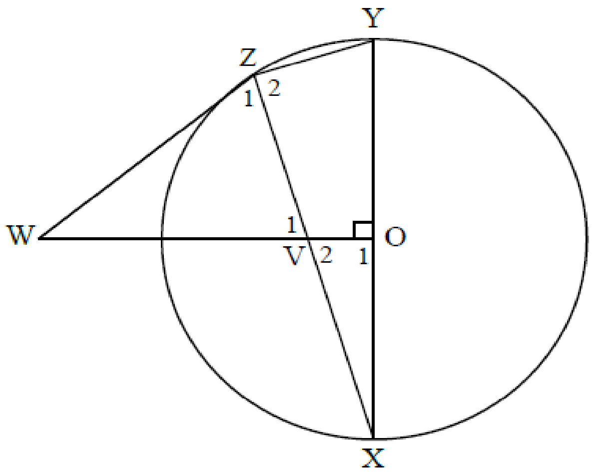
The ICC in Figure 6.16 showed observed proportions which are steeper than the theoretical curve, where the low proficiency group performed below what the model predicted and the higher proficiency group performed better than what the model predicted. The item was over-discriminating.

Question 8.1 (Item 30) was based on Euclidean Geometry and required learners to prove that given figure is a cyclic quadrilateral. The question carried 3 marks. The question and accompanying marking rubric are given in Figure 6.17.

a

QUESTION 8

In the figure below, O is the centre of the circle ZYX. WO intersects XZ at V and WZ is a tangent to the circle at Z. WO ⊥ XY.



8.1 Prove that VOYZ is a cyclic quadrilateral. (3)

8.2 Prove that ΔWVZ is isosceles. (3)

8.3 Prove that ΔXOV ||| ΔXZY. (4)

b.

8.1	$\hat{Z}_2 = 90°$ $\hat{Z}_2 = \hat{O}_1$ ∴ VOYZ is a cyclic quadrilateral ... (Converse opp angles of a cyclic quad)	✓S ✓R angle in semi circle each = 90° ✓Reason (3)
8.2	$\hat{Z}_1 = \hat{Y}$ (tan. chord theorem) $\hat{V}_1 = \hat{Y}$ (ext cyclic quad) $\hat{Z}_1 = \hat{V}_1$ ∴ Δ WVZ is isosceles (two equal angles)	✓S/R ✓ S/R ✓R (3)
8.3	In Δ XOV and ΔXZY $\hat{X}$ is common $\hat{O}_1 = \hat{Z}_2 = 90°$ (ext cyclic quad) ∴ $\hat{V}_2 = \hat{Y}$ (remaining angles) ∴Δ XOV ||| ΔXZY ∠∠∠	✓ S ✓S /R ✓ S/R ✓ R (∠∠∠) (4)

Figure 6.17. (a) Question 8, and (b) the marking guideline

The 3 marks allocated for Question 8.1 were given for the correct statement, the reason, and the final conclusion. The category frequencies for Question 8.1 are given in Table 6.4.

Table 6.4

Category frequencies for Question 8.1

Category (marks)	0	1	2	3
Frequency (learners)	139	36	54	112

These category frequencies showed that the majority of the learners either scored 0 or full marks for the item. The CPCs showed a disordered threshold, where the first threshold was at a higher location than the second threshold.

Rescoring was done by collapsing the scores of 2 and 3 to 2. The general fit of the question improved with a fit residual of -2.258, which is within the recommended range.

6.2.7 Analysis and rescoring of Question 8.3

Question 8.3 showed misfit, with a fit residual of -3.076, which is outside the recommended range. The ICCs showed observed proportions with a bigger gradient than the theoretical curve, a sign of an item that is over- discriminating. The chi square probability was 0.00 which showed that there is a statistically significant difference between the observed and predicted outcomes.

The CPCs for Question 8.3 showed disordered thresholds, where the second threshold was at a higher location than the third, another reason why rescoring was required. The category frequencies are shown Table 6.5.

Table 6.5

Category frequencies for Question 8.3

Category (Marks)	0	1	2	3	4
Frequency	69	73	42	47	114

The scores of 2 and 3 were not working as intended as there was no point along the horizontal axis where the scores of 2 and 3 were the most likely to be scored.

Question 8.3 and the marking rubric are shown in Figure 6.17. The question required learners to prove that two triangles are similar by proving that the corresponding angles are equal. Marks were allocated for each statement and a reason. No clarity was given regarding accuracy and CA marking. The middle scores were not working as intended and together with a fit residual outside the recommended range. The question was re-scored by collapsing the scores of 1,2 and 3 to 1 and 4 to 2. The fit of the item improved after re-scoring and the fit residual changed from -3.076 to -2.397, which is within the recommended range.

6.2.8 Analysis and rescoring of Question 9.2

Question 9.2/Item 35 had a fit residual of 2.887, which is outside the recommended range of -2.5 to 2.5. The chi square probability was 0.000 meaning the differences between the observed and predicted outcomes was statistically significant. The ICCs for Question 9.2 are given in Figure 6.18.

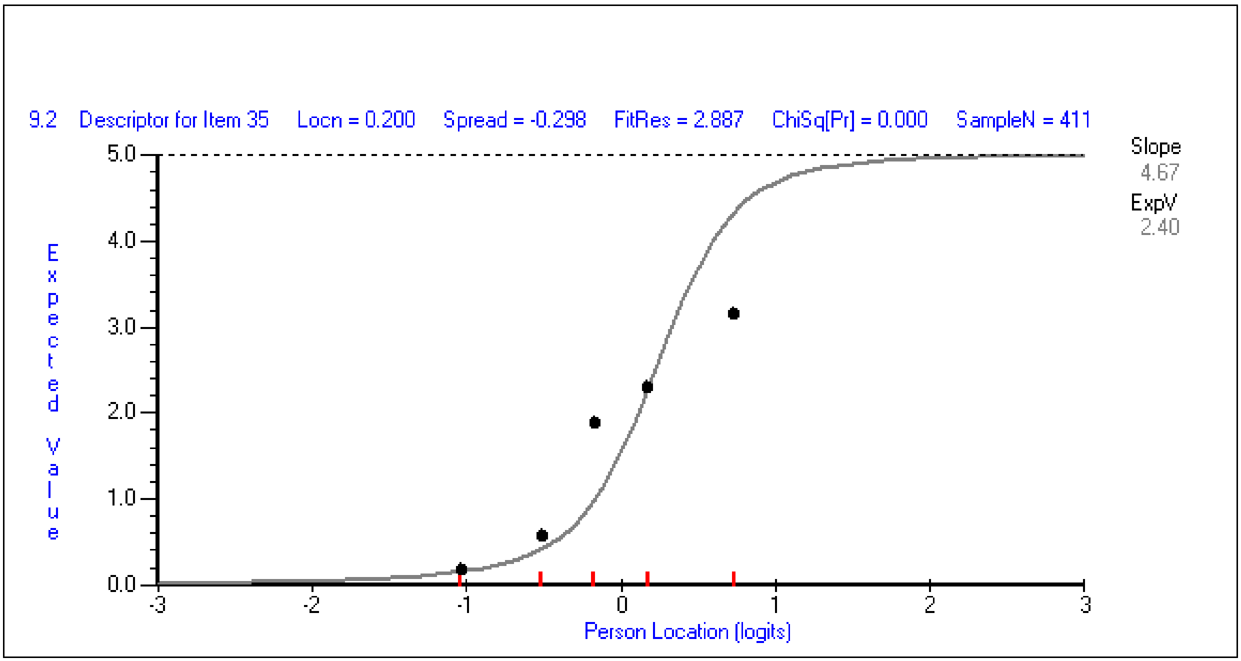


Figure 6.18. ICCs for Question 9.2.

The ICC shows under-discrimination, where the lower proficiency group is performing better than predicted, and the higher proficiency group is performing lower than expected. Question 9.2 was based on Euclidean geometry where learners were expected to calculate the length of line segment on a triangle. The question and the marking rubric are given in the Figure 6.19.

The marking rubric shows that CA were applied in the question, which needed use of the proportionality theorem. Learners who answered Question 9.1 incorrectly but continued to use those answers correctly on Question 9.2 were not supposed to be penalised because of the CA classification of marks.

a.

QUESTION 9

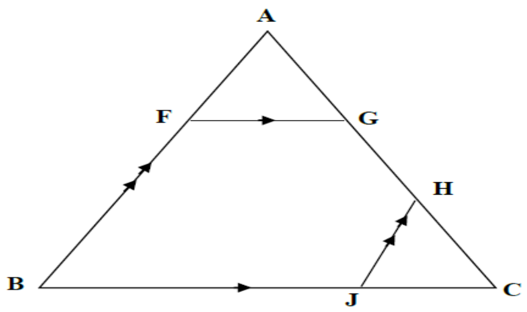
In the diagram below FG || BC, HJ || AB.
FA = 3 units
FB = 9 units
AG = 2 units

CJ: JB = 1:3

Calculate (stating reasons) the lengths of:

9.1 GC (4)

9.2 GH (5)



[9]

b.

9.1	$\frac{FA}{FB} = \frac{AG}{GC}$ (*Prop intercept theorem, FG* ∥ *BC*) $\frac{3}{9} = \frac{2}{GC}$ $3GC = 18$ $\therefore GC = 6$	✓S/✓R ✓Substitution ✓ Answer (4)
9.2	Let GH = x ∴ HC = $6 - x$ $\frac{6-x}{x+2} = \frac{1}{3}$... (HJ // AB prop theorem) $18 - 3x = x + 2$ Now $-4x = -16$ $\therefore x = 4$ ∴ GH = 4 units	(CA applies in the question) ✓M S✓R✓ ✓Simplification ✓ Answer (5)
		[9]

Figure 6.19. (a) Question 9.2, and (b) the marking guideline

The category frequencies from RUMM showed that for Question 9.2, 145 learners got 0, 30 learners got I mark, 19 learners got 2 marks, 9 learners got 3 marks 10 learners got 4 marks and 63 learners got 5 marks. This shows that the majority of the learners either did not get any mark or they got full marks for the question. The low frequencies for the middle categories show that the middle scores were not working as expected. Figure 6.20 shows the CPCs for Question 9.2.

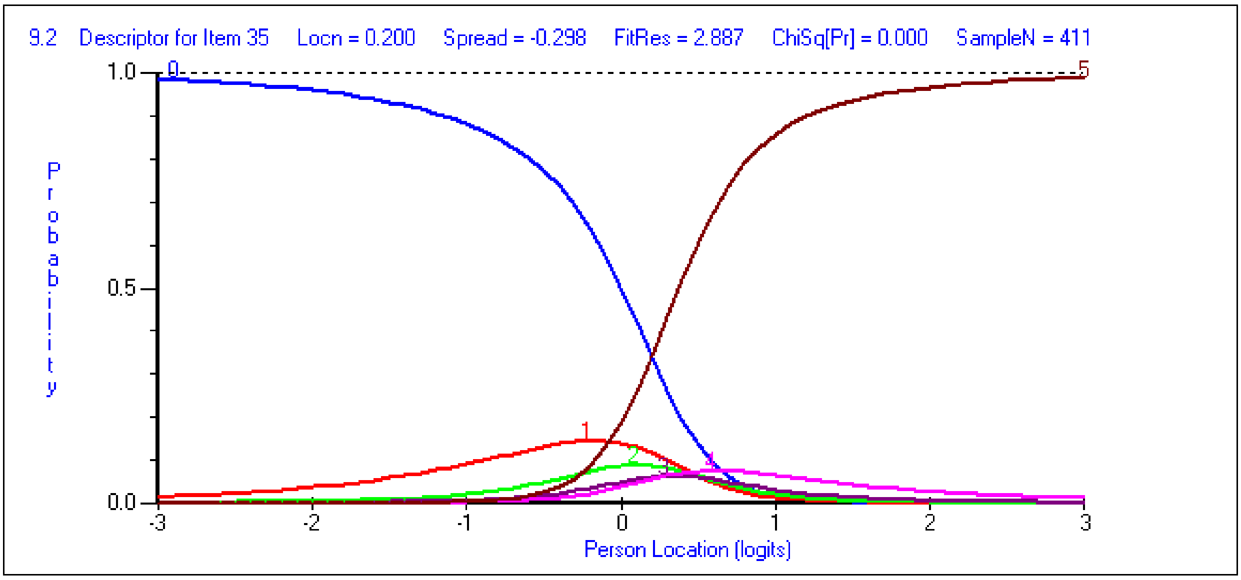


Figure 6.20. CPCs for Question 9.2

The CPCs for Question 9.2 show that the middle scores are not working as expected. The figure showed disordered thresholds, where the first threshold is at a higher location than the second and third thresholds. At no point along the horizontal axis are the scores of 1, 2, 3 and 4 most likely. Even though learners with low ability are still most likely to respond incorrectly and score 0, and high ability learners are still most likely to respond correctly and score 5, learners with average ability who are supposed to get scores between 1 and 4 (inclusive) are still more likely to get 0 or 5.

The item was rescored to try and improve the fit. The scores of 2 and 3 were collapsed to 2 and the scores of 4 and 5 were collapsed to 3. The general fit of the item improved with the fit residual 1.703 which was within the recommended range.

6.3. Items with Disordered Thresholds

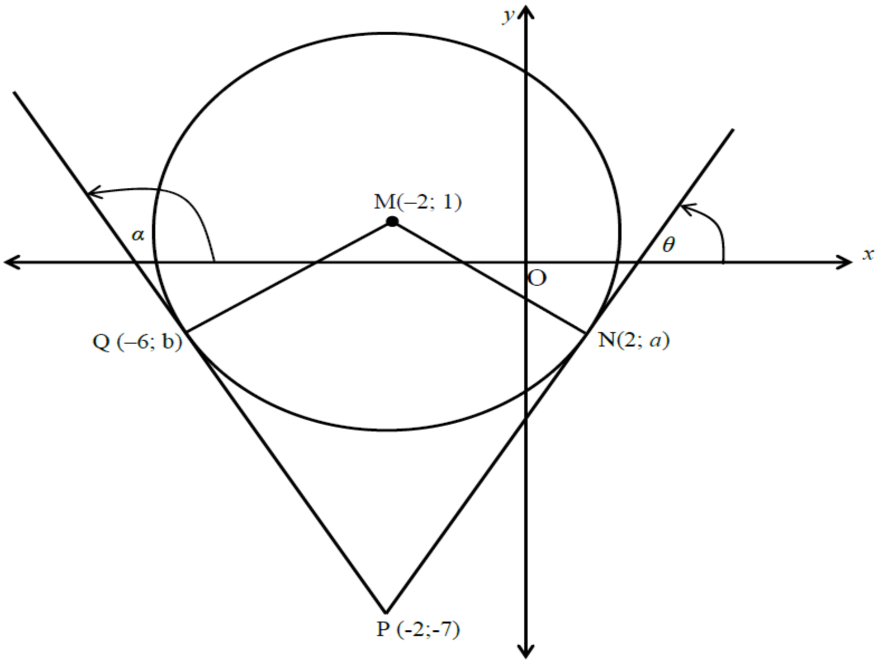
All of the other items with thresholds that were not ordered, were also considered for rescoring although their fit residuals were within the recommended range of -2.5 to 2.5. Rescoring was done according to the recommendations of Van Wyke and Andrich (2006). Questions 4.1.1, 5.1 and 7 are discussed in this section.

6.3.1 Analysis and rescoring of Question 4.1.4

Question 4.1.4 was on Analytical Geometry and required knowledge of the angles of inclination. For learners to calculate the angles of inclination, they first needed to calculate the gradients of the lines in question. The question had 5 categories, corresponding to 4 marks. The question is shown in Figure 6.21 along with the marking guideline or memorandum.

QUESTION 4

4.1 In the diagram below, MN is a radius of a circle with centre M(–2 ; 1). The co-ordinates of N are (2 ; a) and $a < 0$. The co-ordinates of P are (-2'-7). PQ and PN are tangents to the circle at Q and N respectively.
The coordinates of Q is (–6 ; b). PM is parallel to the y – axis.



4.1.1 Deduce that $a = -3$. Show all your workings. (5)

4.1.2 Determine the equation of the circle. (2)

4.1.3 Calculate the gradient of the tangents at Q and N. (4)

4.1.4 If the angle of inclination of lines PN and PQ are θ and α respectively, without using a calculator, show that $\tan^2 \alpha + \tan^2 \theta = 2$. (4)

(a)

(b)

4.1.4	$\tan\theta = 1$	✓A $\tan\theta = 1$
	$\tan\alpha = -1$	✓A $\tan\alpha = -1$
	$\tan^2\alpha + \tan^2\theta$	✓A Substitution (1)
	$=(1)^2 + (-1)^2$	✓A Substitution (-1)
	$= 2$	(4)

Figure 6.21. (a) Question 4.1.4 and (b) the marking guideline

The marking guideline showed that all 4 marks allocated for this question were accuracy (A) marks, meaning any deviation from the required answers would lead to learners losing marks. If the student found the gradients of the lines PN and PQ and equated them to tan α and tan θ respectively, the learner scored 2 marks. The other 2 marks came from substituting the values of tan α and tan θ. The initial CPCs for the question (Item 18) are presented in Figure 6.22.

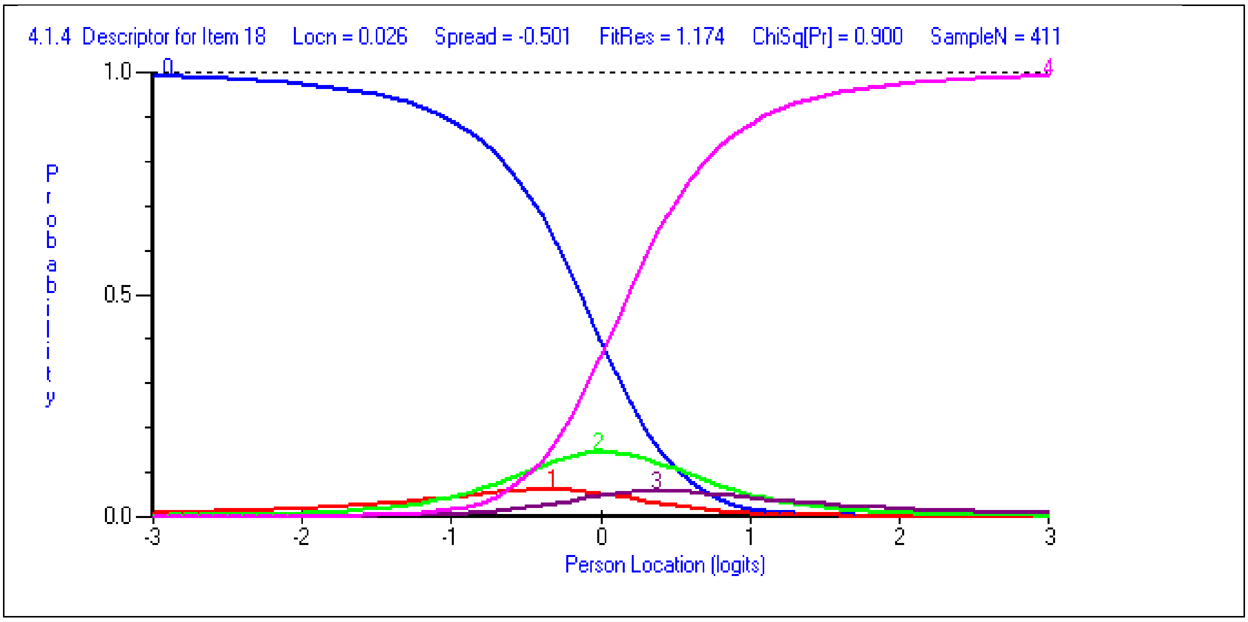


Figure 6.22. Initial CPCs for Question 4.1.4/Item 18

The initial category curves in Figure 6.22 for Question 4.1.4, indicate that the first threshold (for the scores of 0 and 1) is at a higher location (2.013650 logits) than the second threshold (at around -1.008868 logits). Also the third threshold for scores 2 and 3 (location 1.150219 logits) is at a higher location than the threshold for scores 3 and 4 (-2.0492448 logits). This is a result of the failure of the scores 1, 2 and 3 to function as per model requirements. At no point or interval along the x-axis are the scores of 1, 2 and 3 the most likely marks to be obtained.

The category frequencies generated from Rasch analysis are presented in Table 6.6. However, of the 411 learners involved in the study (excluding schools LM and FT), only 262 students attempted Question 4.1.4

Table 6.6

Category frequencies for Question 4.1.4

Category (marks)	0	1	2	3	4
Frequency (learners)	133	11	24	8	86

The category frequency table helps to explain why categories 1, 2 and 3 were not functioning as intended. Very few learners scored these marks, as compared to 0 and 4. The learners' scripts showed that many did not get any mark for Question 4.1.4 or got all marks. A learner who was able to identify the gradient of line PN was also able to determine the gradient of line PQ, and also substitute into the given expression hence they were able to get all 4 marks. Those who were not able to substitute the gradients received 0 and no part marks. The supporting evidence from the learners' scripts is presented in Figure 6.23.

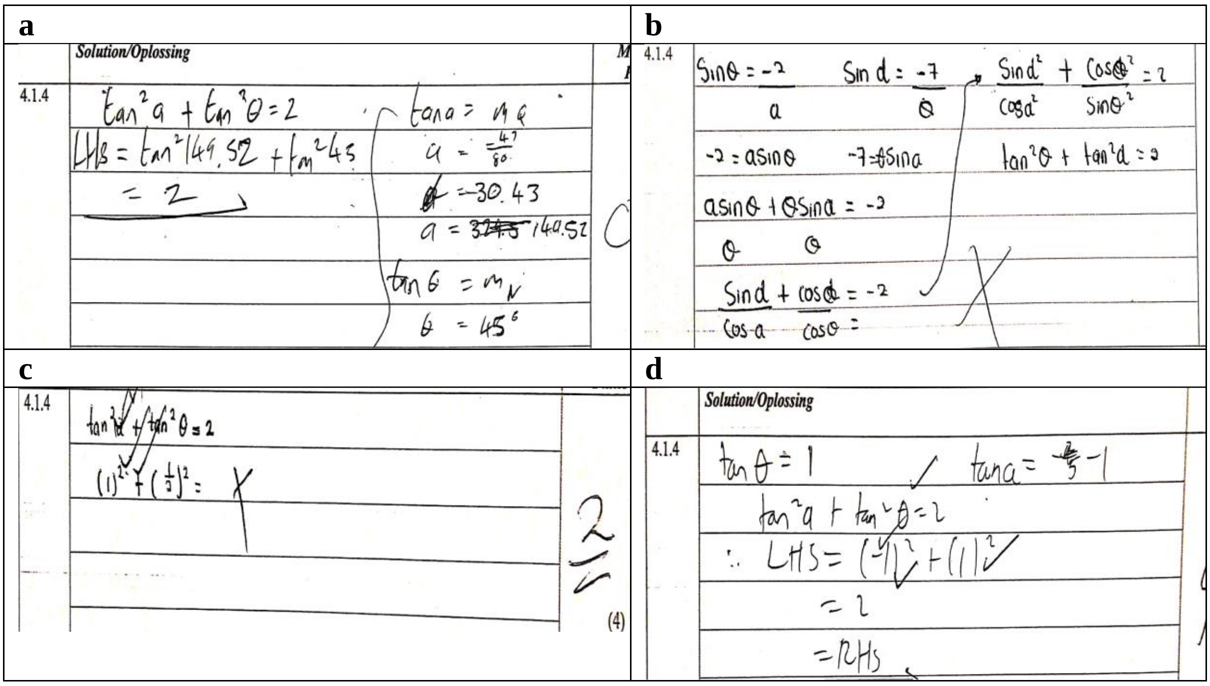


Figure 6.23. Samples of learners' responses to Question 4.1.4

The category frequencies indicate that many of the learners were ether getting zeroes or all the marks. Those who were getting part marks as shown in Figure 6.23(c) were able to find the gradient of PN and substitute but failed to do the same for line PQ resulting in them getting 2 marks.

The item was rescored, with scores of 1,2 and 3 being collapsed into 1 and the score of 4 rescored as 2. The resulting CPCs show that the score of 1 was still not functioning as intended. The first threshold is at a location higher than the second threshold. However, there was a great improvement in the fit residual from initial 1.174 logits to -0.358 logits which is within the acceptable range. The initial probability curves and the CPCs after rescoring are shown in Figure 6.24.

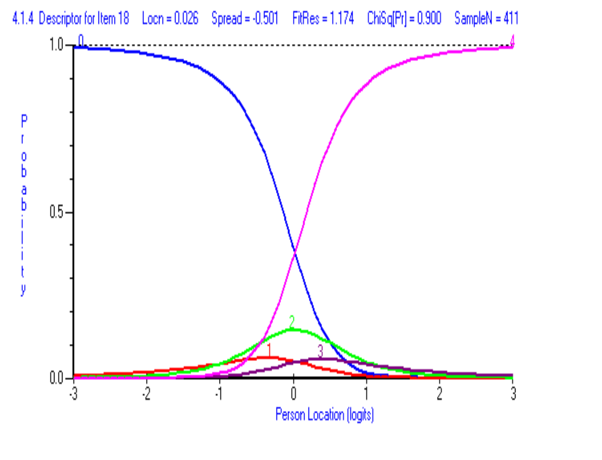
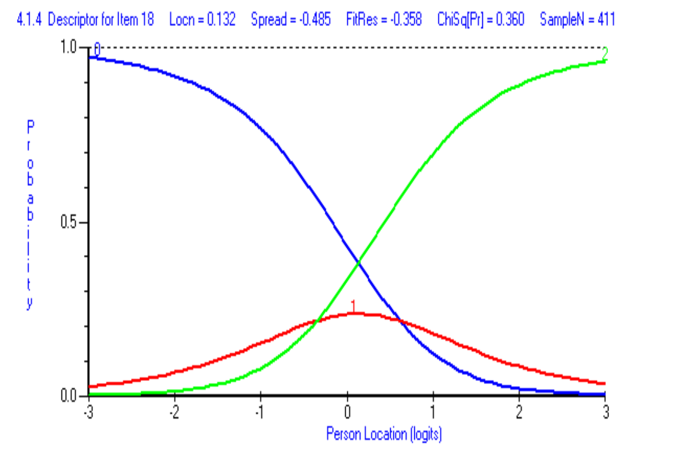
Initial probability curve for Question 4.1.4	Category Probability Curves After Rescoring Question 4.1.4
	

Figure 6.24. Initial CPCs and rescored CPCs for Question 4.1.4

The chi square probability for the rescored Question 4.1.4 is 0.360, a value that is greater that the set criterion of 0.05. This means that the difference between the observed outcomes and the predicted outcomes is not statistically significant.

6.3.2 Analysis and rescoring of Question 5.1/Item 21

The ICCs for Item 21/Question 5.1 showed a fit residual of 1.754 and a chi square probability of 0.462 which is higher than the value of 0.05. The ICCs showed observed proportions which were almost aligned to the theoretical curve as shown in Figure 6.25. However, the CPCs showed disordered thresholds.

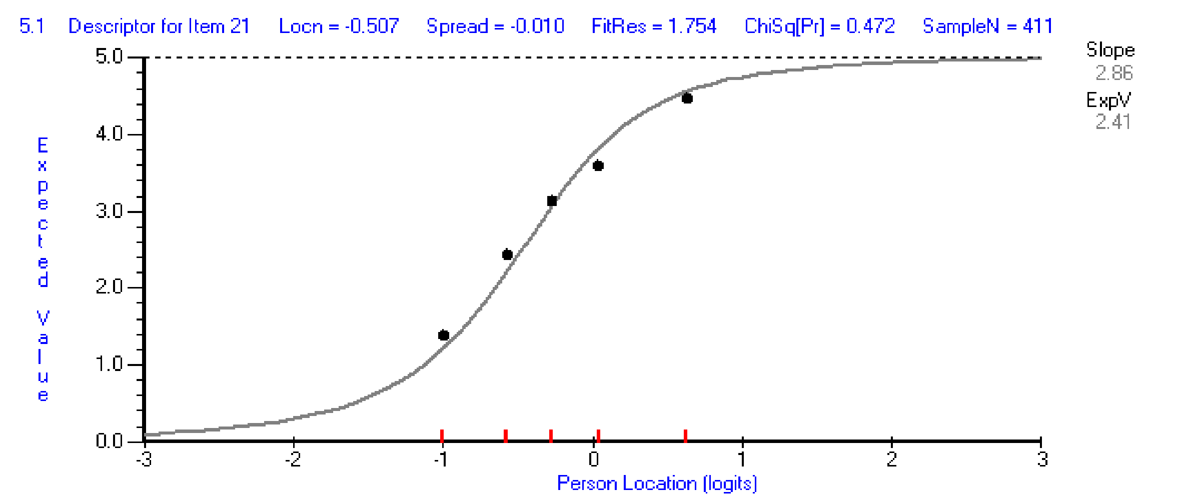


Figure 6.25. ICC for Question 5.1 /Item 21

Item 21/Question 5.1 was based on trigonometry and involved special angles and the reduction formula. The reduction formula was required to reduce Cos 210˚ and Tan840˚. This was basic trigonometry and the question carried 5 marks all allocated for accuracy. Knowledge of negative angles was also required to achieve a successful response to the question. The question and marking guide line is presented in Figure 6.26.

(a) 5.1 Show, without using a calculator, that

$$\sqrt{2}\cos\left(-45^{o}\right)+\cos 210^{o}-\tan 840^{o}=\frac{2+\sqrt{3}}{2}. \qquad (5)$$

(b)

5.1	$\sqrt{2}\cos\left(-45^{o}\right)+\cos 210^{o}-\tan 840^{o}$ $=\sqrt{2}\cos 45^{\circ}+(-\cos 30^{\circ})-(-\tan 60^{\circ})$ $=\sqrt{2}\cdot\frac{\sqrt{2}}{2}-\cos 30^{o}+\tan 60^{o}$ $=1-\frac{\sqrt{3}}{2}+\frac{\sqrt{3}}{1}$ $=\frac{2-\sqrt{3}+2\sqrt{3}}{2}$ $=\frac{2+\sqrt{3}}{2}$	✓A $\cos 45^{\circ}$ ✓ A $-\cos 30^{\circ}$ ✓ A $-\tan 60^{\circ}$ ✓A substitution of special angle values ✓ A simplification (5)

Figure 6.26. (a) Question 5.1 and (b) the marking guideline

The marking memorandum allocated 3 marks for reducing cos(-45˚), cos210˚ and tan840˚, then 1 mark for substituting the special angle values and the last mark for simplifying. The initial CPCs for Question 5.1 are shown in Figure 6.27.

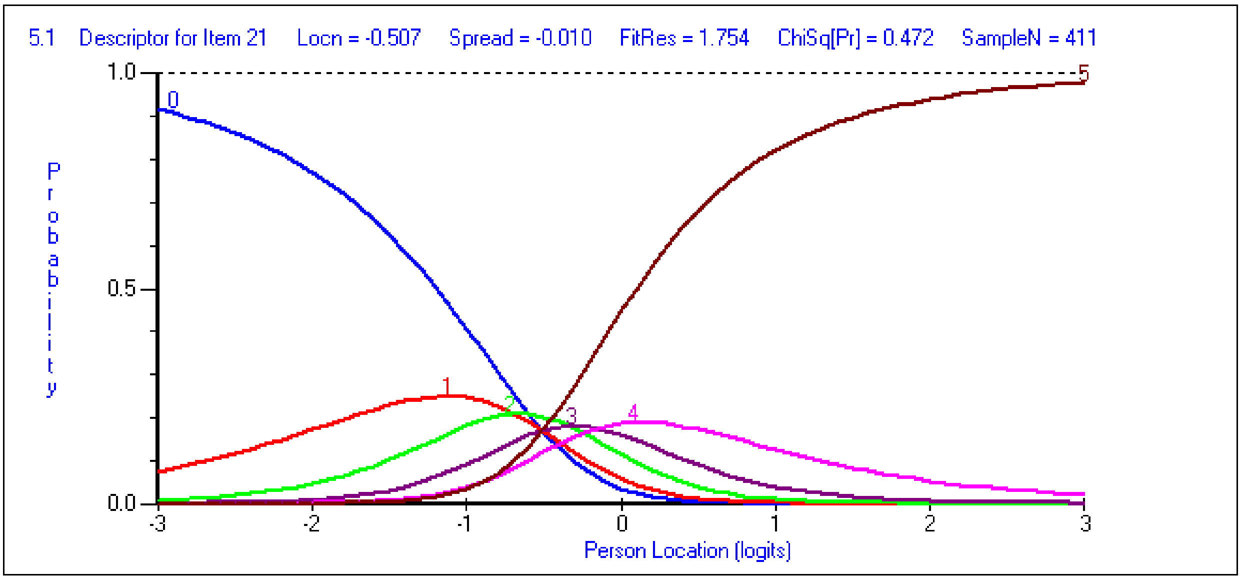


Figure 6.27. Initial CPCs for Question 5.1

Figure 6.27 shows that the first threshold (location of -0.5) is at a higher location than the second threshold with a location of -0.8. The third threshold (location of -0.4) and fourth threshold (location of 0.3 logits) are at a higher location than the fifth threshold with a location of around -1 logits. These disordered thresholds are a result of the scores of 1, 2, 3 and 4 not working or operating as intended, with no point or interval on the horizontal axis where these scores were most likely to occur. Although the categories 1, 2, 3 and 4 were not functioning as intended, the category response frequencies from Rasch analysis did show a big difference between scores on this item. The category frequencies are presented in Table 6.7.

Table 6.7

Category frequencies for Question 5.1

Category (Mark)	0	1	2	3	4	5
Frequency	55	47	46	60	49	129

Some of the learners' answers to Question 5.1 are shown in Figure 6.28.

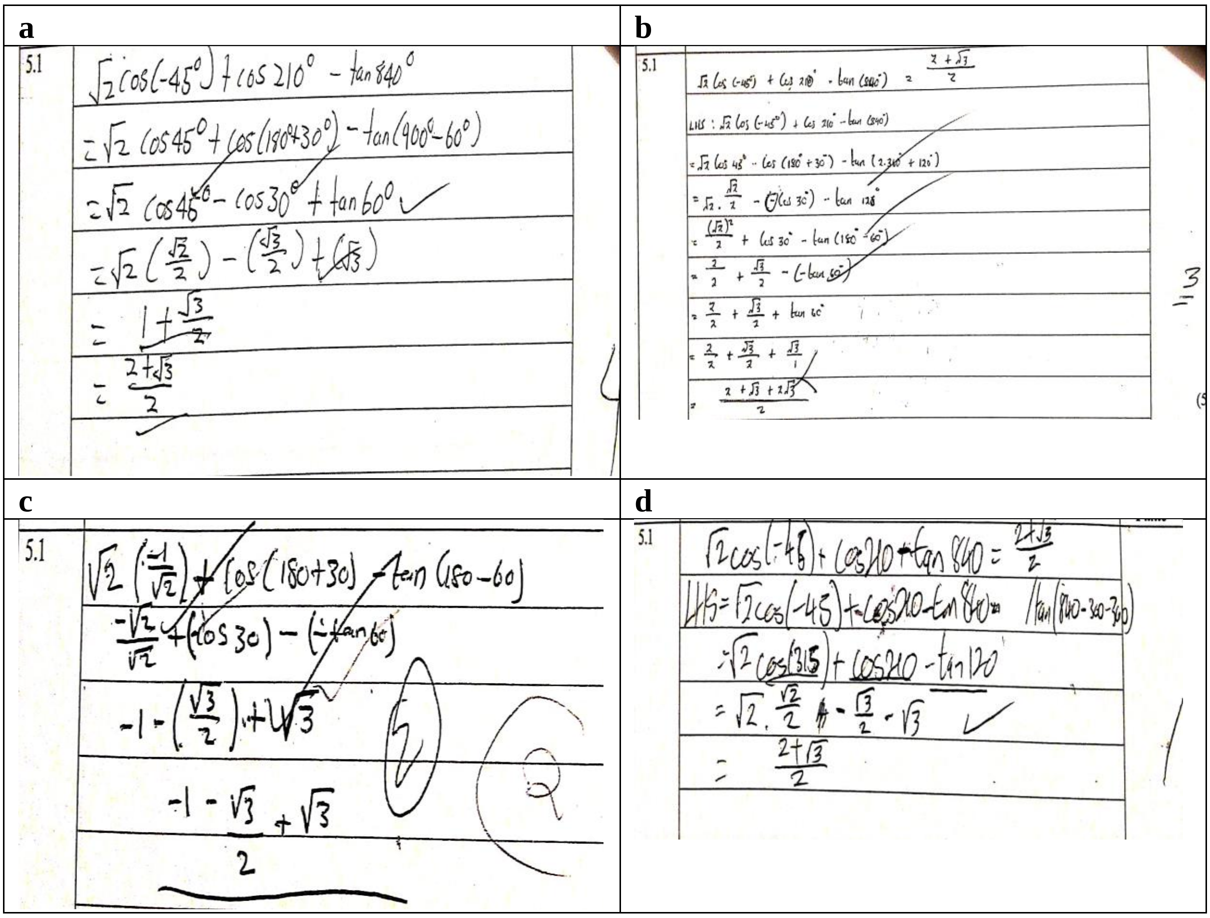


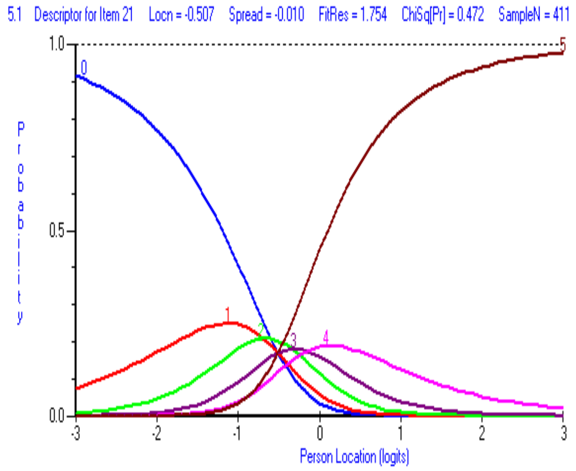
Figure 6.28. A sample of learners' responses to Question 5.1

Figure 6.28 showed that learners who scored 4 out of the possible 5 marks for Question 5.1 did not reduce one of the angles although they correctly substituted for the special angle or they reduced all the angles using reduction formula and were able to substitute the special angles but did not simplify as shown in Figure 6.28(a). Learners who scored 3 marks made many mistakes ranging from wrong substitution and wrong simplification to an incorrect reduction of given angles into acute angles.

Learners who got 2 marks and 1 mark made a range of mistakes ranging from wrong reduction to wrong substitution, and incorrect simplification.

The item was then rescored by collapsing the scores of 1, 2 and 3 into 1. The scores of 4 and 5 were collapsed into 2. Figure 6.29 shows the results of the rescoring procedure.

(a) Initial CPCs for Question 5.1



(a) Category Probability Curves After Rescoring Question 5.1

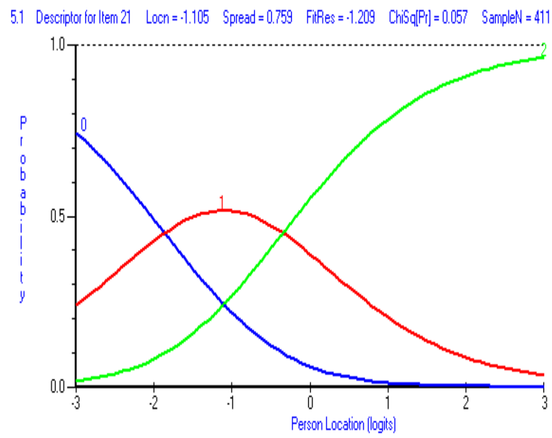


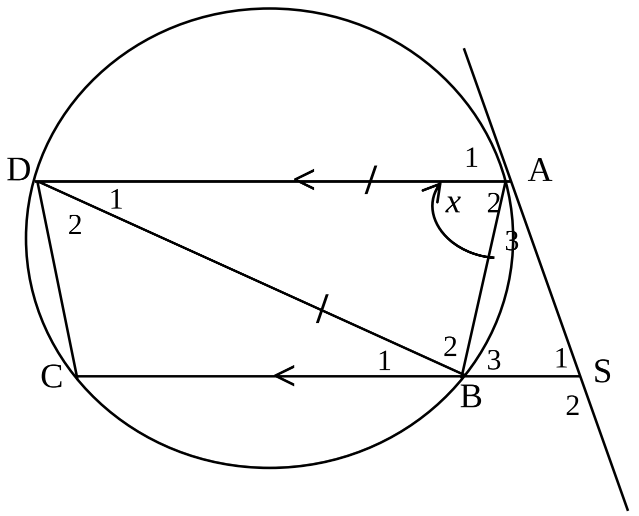
Figure 6.29. (a) Initial CPCs and (b) the CPCs after rescoring Question 5.1

Figure 6.29 shows that the CPCs of the rescored item showed an improved fit residual of -1.209, down from an initial fit residual of 1.754. The resulting CPCs showed categories which were now operating well, with all of the scores having an interval along the horizontal axis where they were most likely to occur.

6.3.3 Analysis and rescoring of Question 7/Item 29

The ICC for Item 29/Question 7 showed that the observed proportions were better aligned to the predicted curve and the chi square probability is 0.095 showing that there is no significant difference between the observed outcome and the predicted outcome. The fit residual was 2.216, which is within the recommended range. However, the CPCs showed that categories 1, 7 and 8 were not working well and this prompted further investigation and rescoring. Question 7 was based on Euclidean Geometry and learners were supposed to write five other angles equal to x and provide reasons for their answers. The question required knowledge of geometry theorems on cyclic quadrilaterals, parallel lines and tangents to circles. The question and the marking guideline are provided in Figure 6.30.

(a) 7. Refer to the figure below. ABCD is a cyclic quadrilateral. AS is a tangent to the circle at A. CB is produced to S. AD || SBC; AD = BD; $\hat{A}_2 = x$.



Write down, with reasons, FIVE other angles each equal to *x*. (9)

(b)

7	1. $\hat{B}_2 = x = \hat{A}_2$ ($\angle$s opp = sides)	S✓/R✓
	2. $\hat{A}_1 = x = \hat{B}_2$ ....(tan chord theorem)	S✓/R✓
	3. $\hat{A}_1 = \hat{S}_1 = x$ ....(corres $\angle$s, DA || CS)	S/R✓
	4. $\hat{B}_3 = x$ (alternate angles, DA || CS)	S✓/R✓
	5. $C\hat{D}A = x$ ……..(Exterior angle of a cyclic quad)	S✓ R✓
		(All Accuracy) (Penalize once for not stating parallel lines)
		(9)

Figure 6.30 (a) Question7/Item 29 and (b) the marking guideline

All the marks for this question were accuracy marks. This question carried the highest numbers of marks (9 marks) among all the sub-questions. The categories and frequencies are shown in Table 6.8.

Table 6.8

Category frequencies for Question 7/Item 29

Category/Mark	0	1	2	3	4	5	6	7	8	9
Frequency	78	41	88	61	103	37	55	22	26	38

The initial CPCs in Figure 6.31 show that the scores of 6, 7 and 8 are not functioning as intended. The score of 1 is at the boundary, on the threshold for the scores of 0, 1 and 2. The other remaining scores seem to functioning well. The first, second, third, fourth and fifth thresholds are ordered, the disorder starts from the sixth threshold to the ninth threshold, as a result of the scores 6, 7 and 8 not functioning as intended.

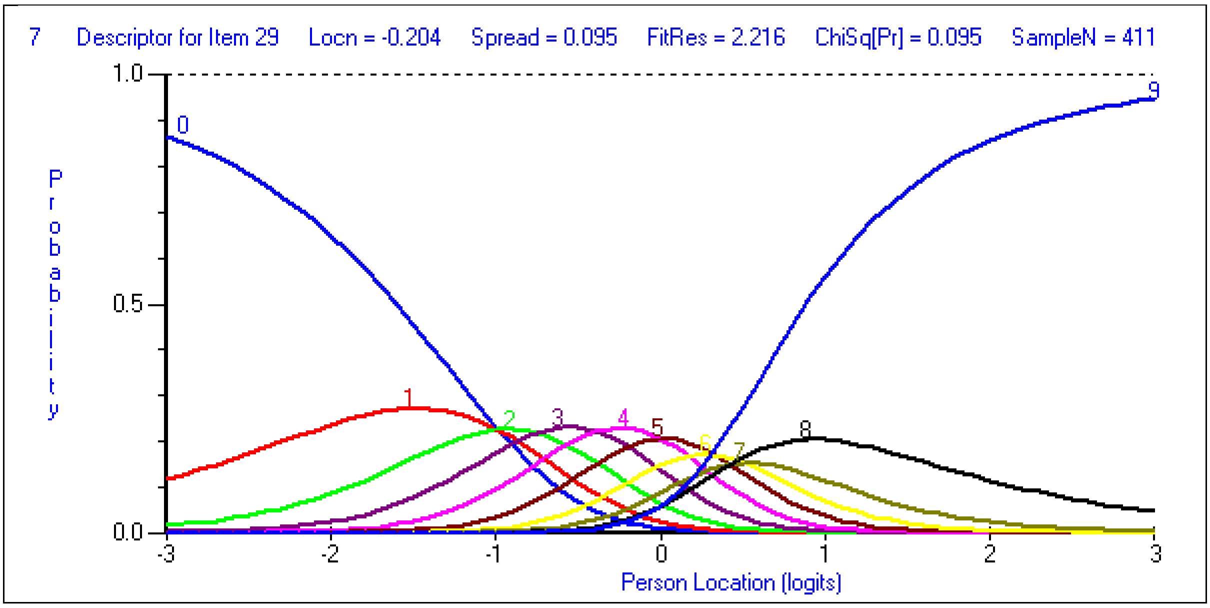


Figure 6.31. Initial CPCs for Question 7/Item 29

Question 7 required learners to write down with reasons, five other angles equal to x. The marking rubric shows that every answer and reason carried 2 marks, except for one case where the answer and reason was allocated 1 mark. The scores of 6, 7 and 8 were not functioning well as learners were not stating the parallel lines when mentioning corresponding angles or alternating angles as a reason or they were failing to give all of the valid reasons for their answers. Some responses from the learners' scripts are presented in Figure 6.32.

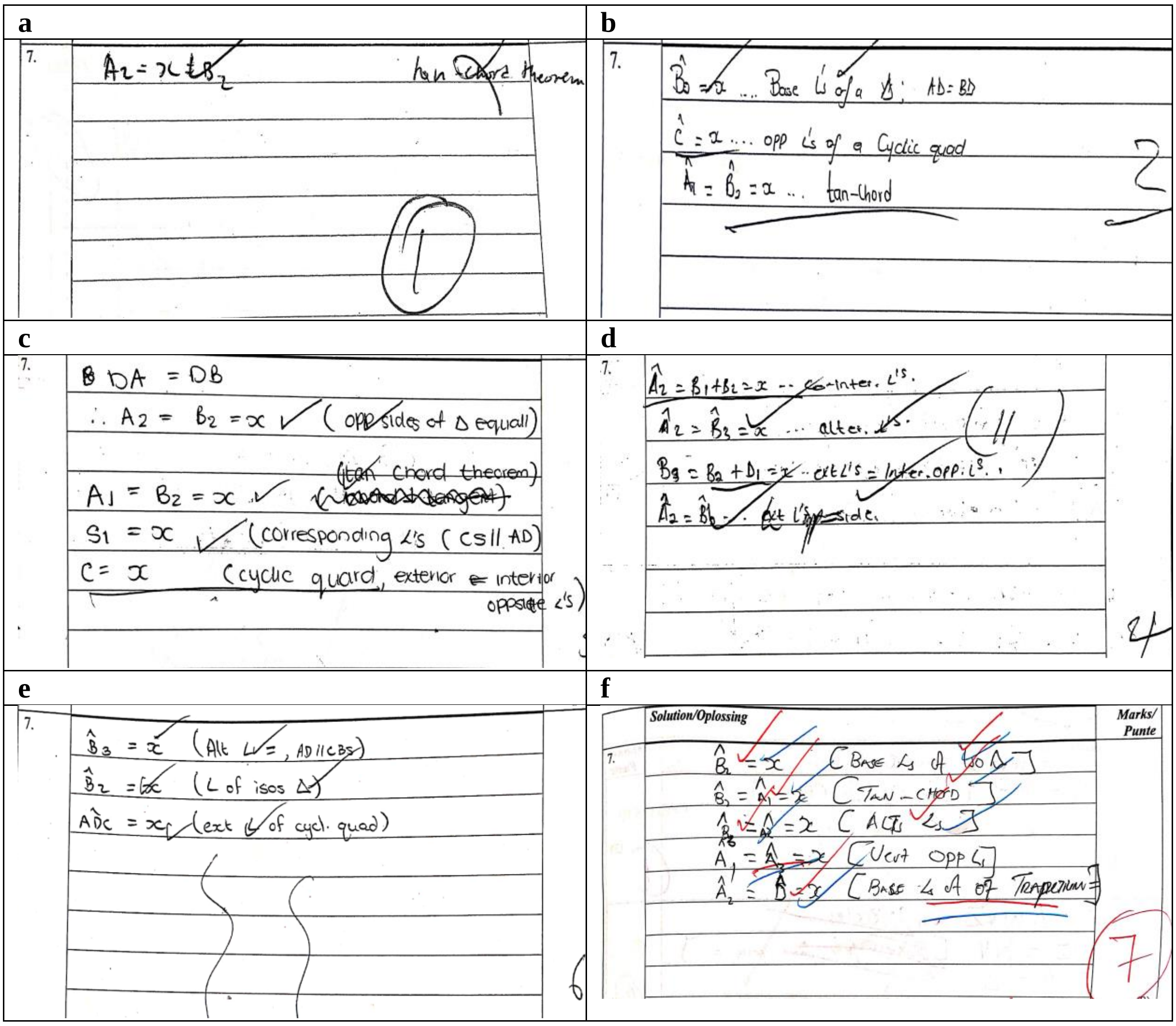


Figure 6.32. A sample of learners' responses to Question 7/Item 29

Figure 6.32 shows that learners got marks for identifying the correct angles and giving reasons for their answers. Those who got 1 mark were able to identify one angle but did not give a correct reason for their answer.

Rescoring was done by collapsing the scores 1 and 2 to 1, score of 3 and 4 to 2, scores of 5 and 6 to 3, scores of 7, 8 and 9 to 4. The rescored probability curves are shown in Figure 6.33.

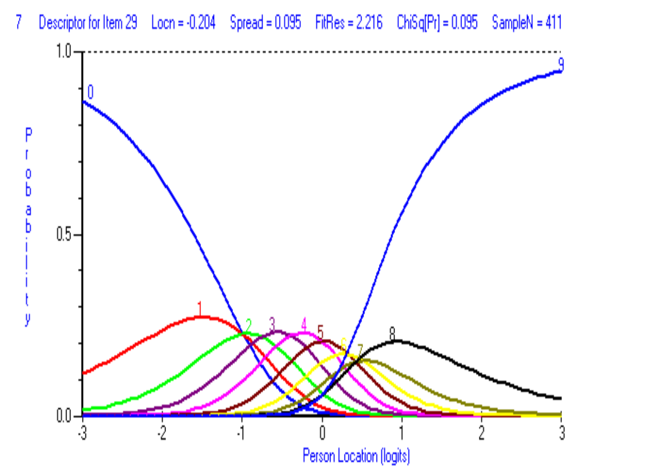
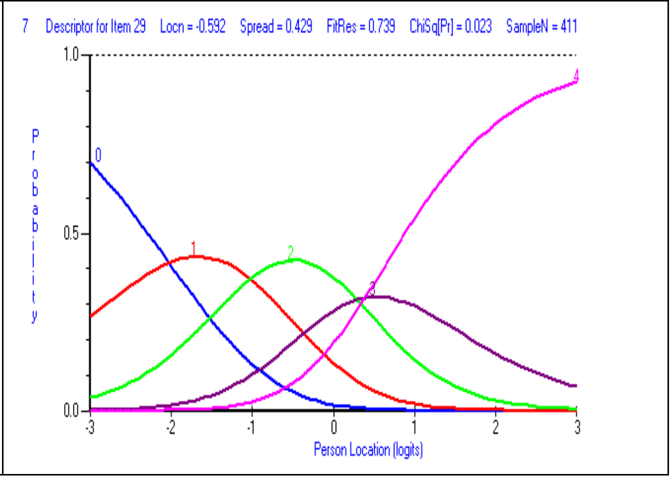
(a) Initial probability curve for Question 7	(b) Category probability curves after rescoring Question 7
	

Figure 6.33. (a) Initial CPCs and (b) CPCs after rescoring for Question 7/Item 29

The resulting CPCs are ordered, with all of the scores having a point or interval on the horizontal axis where it is the most likely score to be obtained. The fit residual also changed from 2.216 to 0.739 which is within the recommended range.

A summary of the rescoring and the rescored items appears in Appendix H of this thesis.

CHAPTER 7 DIFFERENTIAL ITEM FUNCTIONING, MULTIDIMENSIONALITY AND ITEM DEPENDENCY

7.1 Introduction

In this chapter I explore the aspects of differential item functioning (DIF), multidimensionality and item dependence. Item-splitting was used to try and resolve items displaying DIF and response dependency.

7.2 Exploring Differential Item Functioning

DIF is a result of questions not working in a uniform manner among different groups of people, who have the same ability (Andrich & Hagquist, 2015). DIF can be a result of various factors, some of which are discussed in later section and identified graphically or statistically.

Graphically, DIF was identified through the ICCs generated in RUMM2030, when the ICCs had locations which were different for different groups of people. Different locations of the curves for different groups of people with the curves having similar slopes or gradients, indicated uniform DIF, but where the curves had different slopes then there was none-uniform DIF (Tennant & Pallant, 2007).

While the graphical displays of ICC, provide the visual representation of the data, the DIF needs to be confirmed statistically by use of ANOVA which is used to check whether there is a meaningful difference between the means of two or more groups of people (Andrich & Hagquist, 2015). In ANOVA the F-ratios are constructed, which is a ratio of the estimated variance of residuals among groups and the estimated variance of residuals within groups. With the assumption that the means come from a single random set of residuals from within the groups, then this ratio should be 1.0. An F-ratio bigger than 1.0 indicate real difference between the group means, while an F-ratio smaller than 1.0 will lead one to conclude that the difference is not statistically significant (Andrich & Hagquist, 2015). Since we are working with estimates of variances in ANOVA, one cannot say with 100% certainty that an observed difference is real, since it may be a difference by chance or a peculiarity of the particular sample of residuals (Andrich & Hagquist, 2015). The statistical significance is usually determined using the criterion of 0.01 or 0.05. For this study the statistical significance of the DIF was set at the 0.05 level.

I checked for the DIF for language and gender using the Bonferroni adjustment. According to Bland and Altman (1995), there is a concern that with many tests of fit, some will be significant by chance (type 1 error). A Bonferroni correction is considered if it is imperative to avoid a type 1 error and a large number of tests are carried out without a pre-planned hypothesis (Bland & Altman, 1995). The Bonferroni correction was carried out by dividing the chosen probability value of significance (0.05 in this case) by the number of tests of fit (99 in this case). The Bonferroni adjustment = $\frac{Chosen\ Criterion\ (\propto)}{Number\ of\ tests\ of\ fit} = \frac{0.05}{99} = 0.000505$. No items were observed displaying DIF for gender. Five items were observed as displaying DIF for language after the Bonferroni adjustment and I discuss these in the next section.

For the language factor, learners were grouped according to whether they speak English as the first home language or as their second or third language (others). Details of the number of English first language speakers and English second language speakers appear in Table 3.2 of the methodology section.

Table 7.1 shows the items that showed DIF for language after the Bonferroni adjustment arranged starting with the item with the highest mean squares.

Table 7.1

*Items showing DIF for the person factor of language***.**

Item	**DIF by language**		
	F-ratio	**Probability**	**Mean squares**
2.2.3	41.42808	0.000007	38.11397
2.2.2	19.28979	0.000012	22.19647
2.1	20.11805	0.000003	20.59690
5.1	22.66829	0.000009	18.35958
3.6	15.47693	0.000100	11.78099

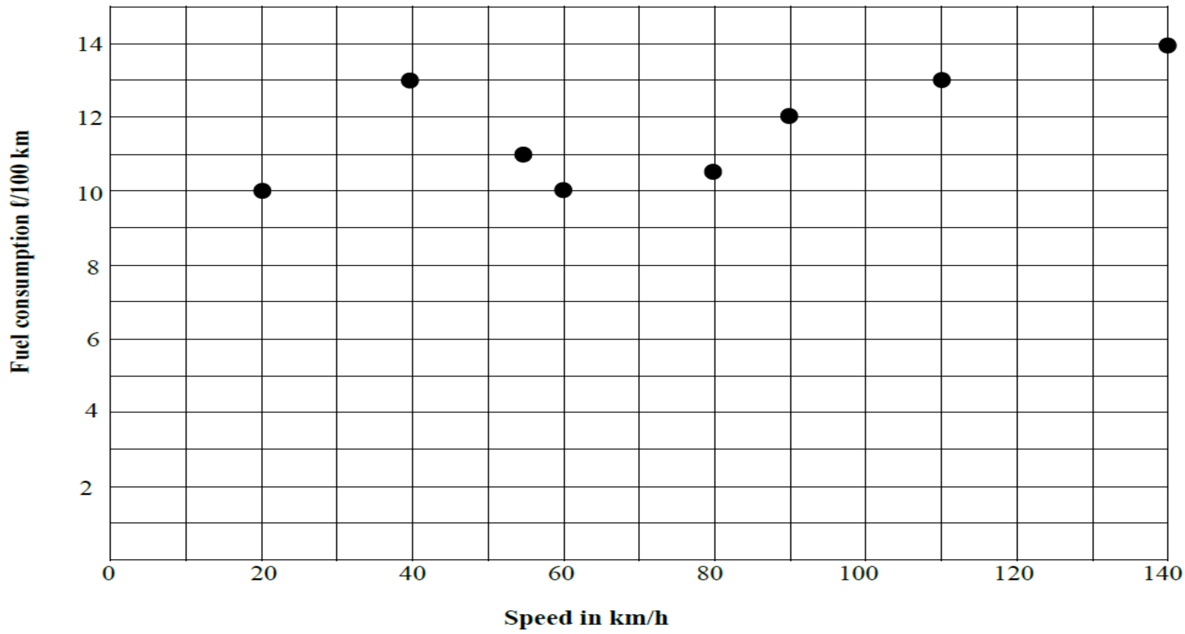
The items in Table 7.1 showed DIF for the person factor of language as the F-ratios were all greater than 1 and the associated probabilities were below the chosen criterion of 0.05, showing that the differences between the means were statistically significant. The five items displayed DIF for the person factor of language after the Bonferroni adjustment. The items are discussed in detail in the later sections. Note that items were considered in the order of the highest mean squares. Andrich and Hagquist (2015) explained that real DIF in one item favouring one group inevitably induces artificial DIF favouring the other group in all other items, a condition known as artificial DIF. Andrich and Hagquist (2015) showed that to distinguish between real and artificial DIF in items, a sequential procedure for resolving DIF in items is implied. If there is more than one item that shows DIF, then the one to deal with first is the one that has the highest Mean Square. The more general way to resolve the item that shows most DIF is by creating two new items from the item (item splitting), one that has one group responding and the other that has the second group responding (Andrich & Hagquist, 2015). This resolution of an item creates a data matrix with data missing from some cells. If it is the only item with DIF, then when the item is resolved and a new analysis is run, the artificial DIF effect will no longer be present. The resolved items only contain either the first group responses or second group responses, and no DIF can be shown from the ANOVA (Andrich & Hagquist, 2015). None of the items displayed artificial DIF when it was checked in the items below.

7.2.1 DIF for language for Question 2.2.3/Item 8

For Item8/Question2.2.3, learners were required to determine the average fuel consumption of a motor car from the data given on a scatter plot. This could be obtained by putting the data into the calculator that was used for calculating the equation of the regression line in Question 2.2.1 and for calculating the correlation coefficient in Question 2.2.2, but this time the learner was to include the outlier. The question and marking memorandum are shown in Figure 7.1.

QUESTION 2

The scatter plot below shows the fuel consumption versus the speed of a motor car.



2.1 Identify an outlier. Write down its co-ordinates. (1)

2.2 Determine:

2.2.1 the equation of the regression line excluding the outlier. (3)

2.2.2 the correlation coefficient excluding the outlier and explain the type of correlation. (2)

2.2.3 the average fuel consumption of the motor car. (2)

[8]

2.1	(40; 13)	✓ A (40; 13) (1)
2.2.1	$y = 0,04x + 8,64$	✓CA 0,04x ✓ CA 8,64 ✓ CA equation (3)
2.2.2	$r = 0,91$ strong, positive correlation	✓ CA 0,91 ✓ CA justification (2)
2.2.3	The car will be using about 11,69 *l*/100 km.	A✓CA✓ average 11,69 *l*/100 km (2)
		[8]

Figure 7.1. Question 2 and the marking guideline

The ICCs for Item 8 showed that for the same person locations, the slopes for the English first language speakers and the others (English second language speakers) are different and non-parallel. Figure 7.2 shows the ICCs for Question 2.2.3.

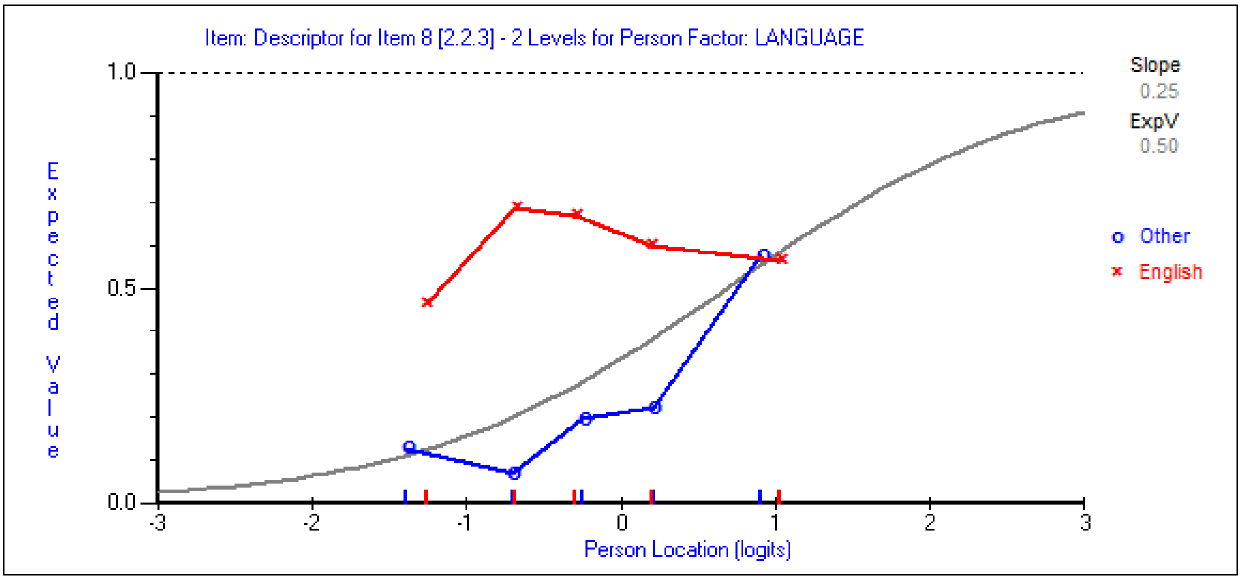


Figure 7.2. Graphical comparison between means of English first language speakers and others (English second language speakers) for Question 2.2.3 showing DIF effect for language.

For the proficiency level below 0.9 logits, the English first language group have a higher probability of responding positively to Question 2.2.3 than the English second language speakers (others). The observed means for the English first language speakers are higher than those of English second language speakers for proficiency levels below 0.9 logits. For this proficiency level group (below 0.9 logits), the English first language group have higher observed means than even predicted by the model, while the English second language group have lower observed means than predicted by the model. For the person ability levels above 0.9 logits, the others (English second language speakers) have higher observed means than the English first language speakers. ANOVA for Question 2.2.3 showed an F-ratio of 41.42808 and a probability value of 0.000007, implying that the difference between the group means is statistically significant.

Item 8/Question 2.2.3 was the first item to be split as it had the highest mean square value of 38.11397. From Item 8 (Question 2.2.3), I created two new items, one that just has the English first language group responding and one that has just the other language learners responding. I then run a new analysis in RUMM2030. The new analysis showed that there was still DIF in the other items, an indication that there was no artificial DIF. After splitting the location of the English first language speakers only was -0.471 and the location for the others (English second language speakers) was 1.171. This means that the English first language speakers experienced the question as easier than the English second language speakers did.

7.2.2 DIF for language for Question 2.2.2/Item 7

When Item 8 (Question 2.2.3) was split, the items in Table 7.1 still showed DIF. The item with the next highest mean square was Question 2.2.2 which is discussed in detail in this section.

For Item 7/Question 2.2.2 shown in Figure 7.1, a learner was required to determine the correlation coefficient for the data excluding the outlier and explain the type of correlation. A learner who had a problem with understanding what an outlier was, was likely to have problem with this item. Although CA was supposed to be applied in this item, a look at the learners' scripts showed that few teachers considered CA during marking. ANOVA showed an F-ratio of 19.28979, implying that there might be a real difference between the group means and the probability value of 0.000012 mean that the difference between the means was statistically significant. The DIF for language in Item 2.2.2 is also shown through the ICC shown in Figure 7.3.

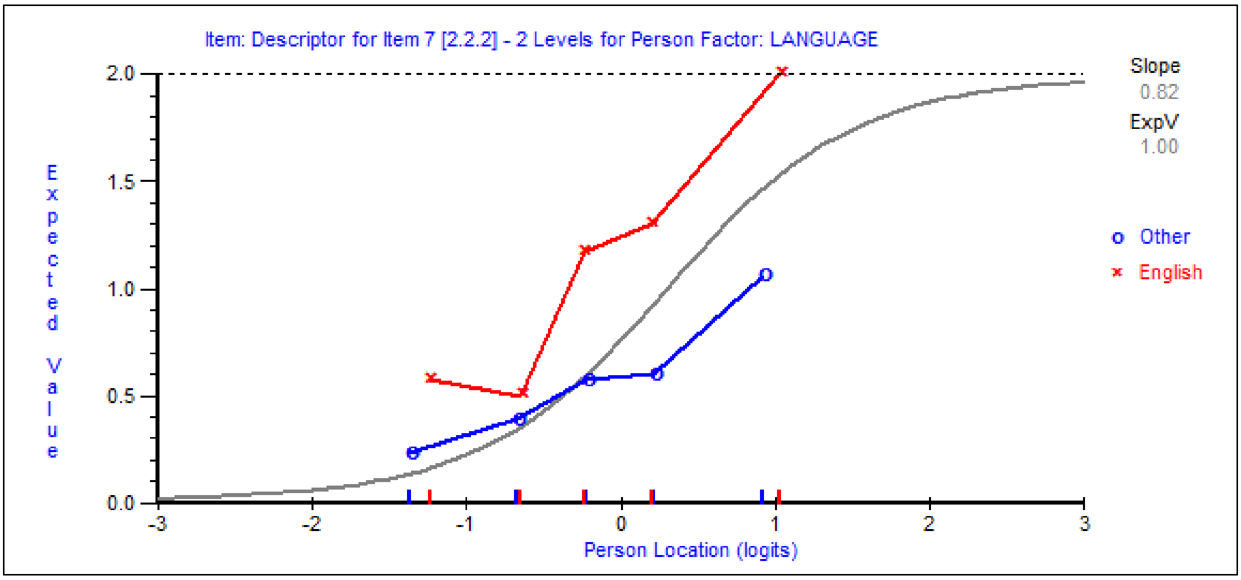


Figure 7.3. Graphical comparisons between the group means of English first language speakers and others for Item 7 (Question 2.2.2)

Figure 7.3 shows that English first language speakers have a higher probability of responding positively than the English second language speakers. At all levels of learner proficiency, the observed means for the English first language speakers is above the observed means for the English second language speakers (others), and is also above the theoretical curve predicted by the model. The observed proportions for the English second language speakers are close to the theoretical curve. Item 2.2.2 was the second to be split, as it had the second highest mean squares value, which was 22.19647.

When the item was split, the item location for the English first language speaking learners only was -0.511 while the item location for the English second language speaking learners (others) only was 0.504. Hence the English first language speaking learners experienced this item as being easier than the English second language speaking learners experienced it.

7.2.3 DIF for language for Question 2.1/Item 5

The ANOVA showed that there is DIF for the person factor off language for Item 5/Question 2.1 and an F-ratio of 20.11805 and a probability of 0.000003. The F-ratio was larger than the recommended value of 1.0, indicating that there might be a real difference between the group means. The probability value was also less than the chosen criterion of 0.05, implying that the difference between the group means was statistically significant.

For Question 2.1 (see Figure 7.1), a learner needed to understand what an outlier was before identifying it and writing down the coordinates. A learner who did not understand what an outlier was, was highly likely to respond incorrectly on this item hence the English first language speakers have a slight advantage over their English second language counterparts. DIF for language for Question 2.1 was also confirmed through the ICCs and these are shown in Figure 7.4.

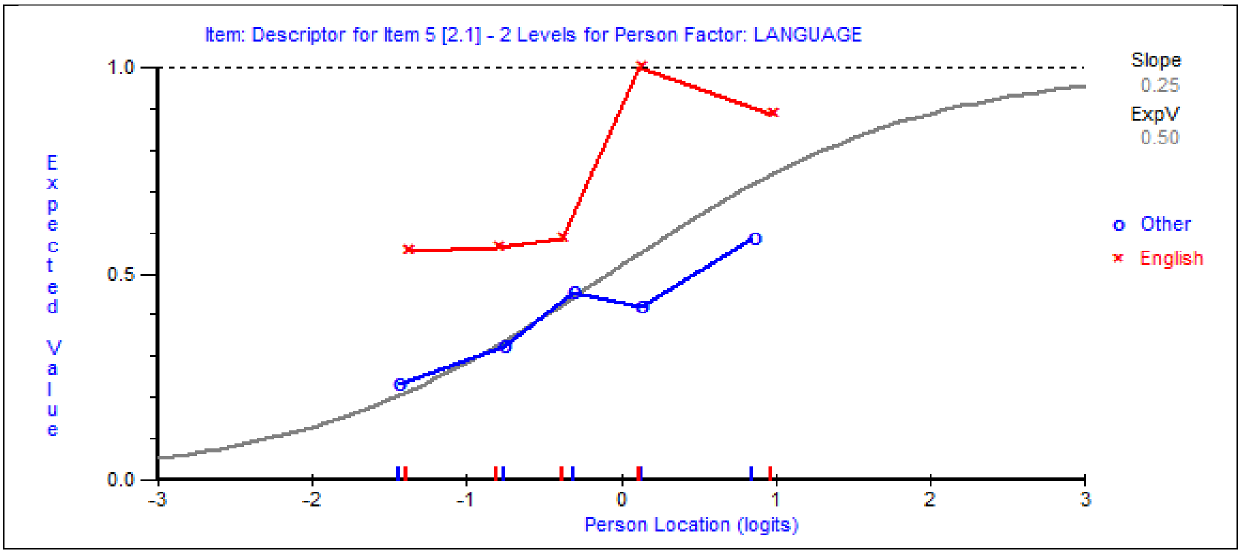


Figure 7.4. ICCs showing graphical comparisons between the group means of English first language speakers and others for Item 5/Question 2.1

Figure 7.4 showed that the English first language speakers had a higher probability of being successful in responding to Question 2.1 than the English second language speakers at all levels of learner proficiency. There was uniform DIF since at all the levels of proficiency the English first language speakers have a higher probability of responding positively than the English second language speakers. The English first language speakers' observed means are above the theoretical curve predicted by the model.

Question 2.1 was the third item to be split since it had the third highest mean squares (20.59690). When the item was split, the item location for the English first language speaking learners only was -1.234, while the item location for the others (English second language speaking learners) only was 0.195. The English first language speaking learners experienced this item as being easier than the others (English second language speaking learners) did.

7.2.4 DIF for language for Question 5.1/Item 21

The DIF analysis for Question 5.1 was somewhat different from that of the others, in that the English second language speakers had a higher probability of responding correctly (had a higher expected value) along the proficiency continuum than the English first language speakers. At first it was suspected that this item was a case of artificial DIF, that could have been induced by the DIF in the other items. As explained by Andrich and Hagquist (2015), the concept of artificial DIF shows when real DIF in one item favouring one group inevitably induces artificial DIF favouring the other group in other items. However, because the DIF remained after the items were progressively split in the order of highest to lowest mean squares, Andrich and Hagquist (2015) advise that this is not a case of artificial DIF. The Question 5.1/Item 21 is given in Figure 7.5.

QUESTION 5

5.1 Show, without using a calculator, that

$$\sqrt{2}\cos\left(-45^\circ\right)+\cos 210^\circ-\tan 840^\circ=\frac{2+\sqrt{3}}{2}.$$ (5)

Figure 7.5. Item 21/Question 5.1

Item 21 (Question 5.1) required learners to simplify, without using a calculator. This question did not require much understanding of the English language and learners who understood the reduction formula and special angles were able to come up with the solution, hence a strong English background and understanding was of no influence.

The ANOVA showed that for the person factor of language, the F-Ratio was 22.66829, indicating that there might be a real difference between the group means. The probability is 0.000009, which is below the set value of 0.05, meaning that the difference in the group means is statistically significant. This implied that there was DIF for the language factor for Item 5.1 which was also confirmed graphically by the ICCs.

The ICCs for the language group means are shown in Figure 7.6.

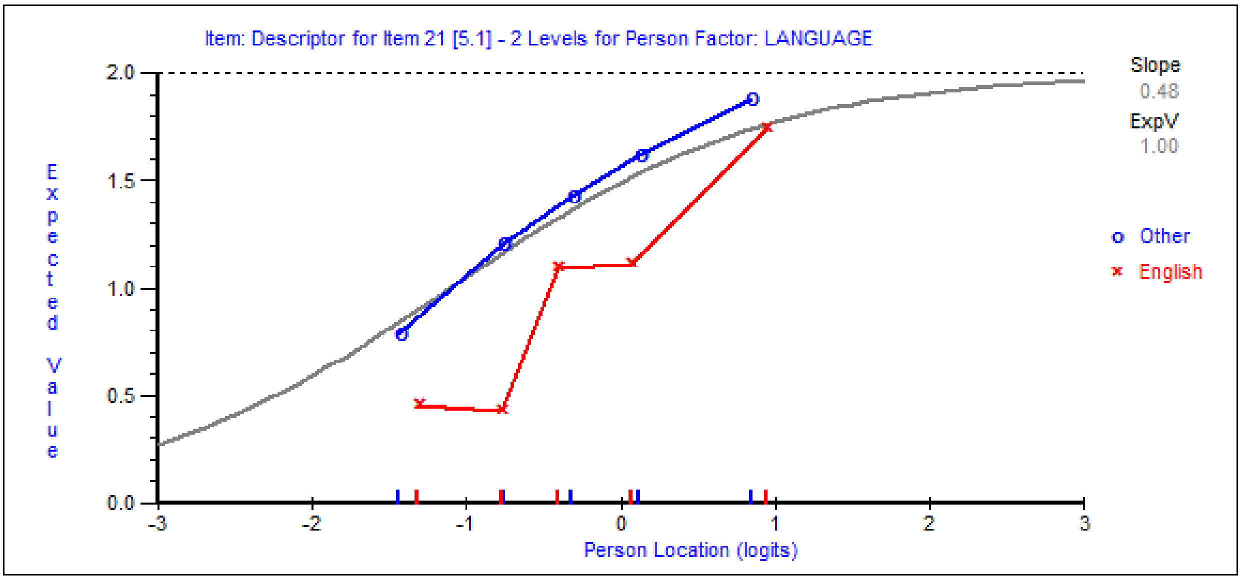


Figure 7.6. Graphical comparisons between the group means of English first language speakers and others for Item 21/Question 5.1

The ICCs in Figure 7.6 showed that at all proficiency levels, the English second language speakers had a higher probability of responding positively to Item 21 than the English first language speakers, as their observed means are all greater than those of the English first language speakers. The observed means for the English first language speakers are below the theoretical curve, meaning that they performed below what the Rasch model predicted. Item 5.1 was the fourth to be split as it had the fourth highest mean square value of 18.360.

When the item was split, the location for the English first language speakers only was -0.247 and the location for the others (English second language speakers) only was -1.228. The English second language speaking learners (others) experienced the item as much easier than the English first language speakers did.

7.2.5 DIF for language for Question 3.6/Item 14

For Item14/Question 3.6, learners were required to calculate the length of a side of a triangle and to leave the answer in simplest surd form. The ANOVA showed that the item had an F-ratio of 15.47693 implying that there might be a real difference between the group means. The probability value of 0.000100 meant that the differences between the group mean values were statistically significant. The DIF for language in Item 3.5 is also shown through the ICCs shown in Figure 7.7.

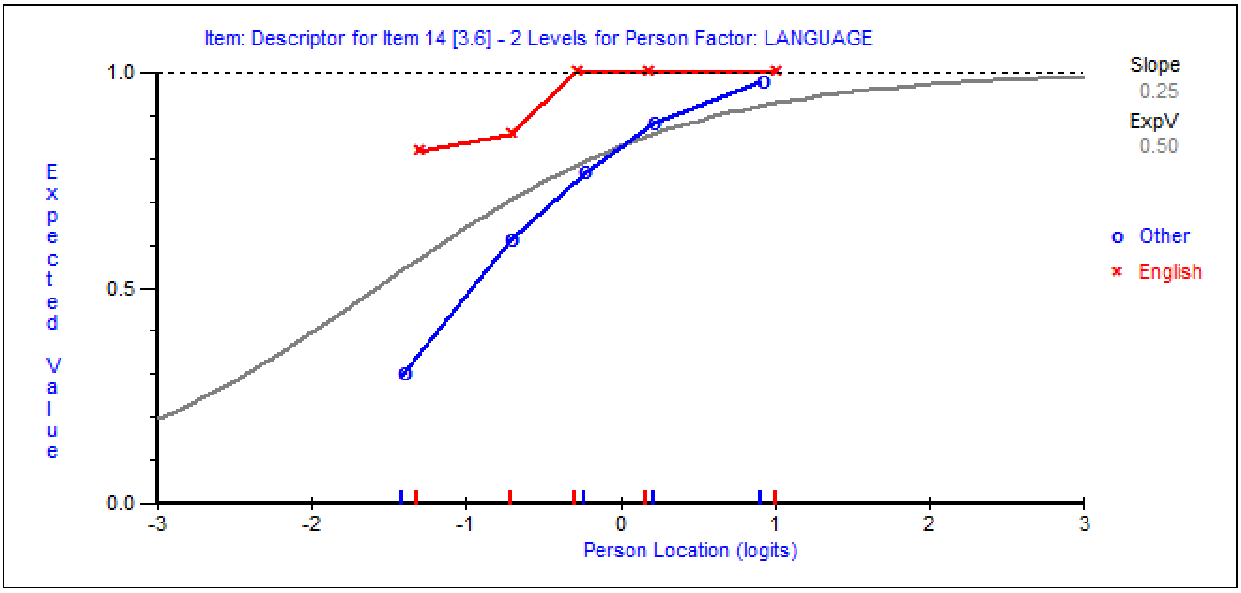


Figure 7.7. ICCs showing graphical comparisons between the group means of English first language speakers and others for Item14/Question 3.6

At all proficiency levels, the English first language speaking learners had a higher probability of responding positively to the item than the English second language speaking learners (others). The observed means for the English first language speakers were higher than the observed means for the English second language speakers at all proficiency levels. Item 3.6 was the fifth item to be split as its mean square value was the fifth highest at 11.788099.

When the item was split, the location for the English first language speakers only was -3.178, and the location for the others (English second language speakers) only was -1.226. The English first language speaking learners experienced the item as much easier than the English second language speaking learners did.

Below is a summary of the item splitting that was done for the items showing DIF for the person factor of language.

Table 7.2

Probability values before and after splitting the items for the person factor of language.

Item	Probability DIF language	Mean squares	Comments
2.2.3	P = 0.00007	3.11397	When split item location for English first language speakers only = -0.471, and for English second language speakers only = 1.171 logits
2.2.2	P =0.000012	22.19647	When split item location for English first language speakers only = -0.511, and for English second language speakers only = 0.504 logits
2.1	P =0.000003	20.59690	When split item location for English first language speakers only = -1.234 and for English second language speakers only = 0.970 logits
5.1	P = 0.000009	18.360	When split item location for English first language speakers only = -0.247, and for English second language speakers only= -1.228 logits
3.6	P = 0.000100	11.78099	When split item location for English first language speakers only = -3.178 and for English second language speakers only = -1.226 logits

All of the items in Table 7.2 showed genuine DIF for the language factor as the resolving of these items one at a time, starting with the item with the highest mean square, did not result in the disappearance of DIF in the other items. No item displayed artificial DIF.

7.3 Exploring Response Dependence

Local independence of responses is violated when the response to one question might depend on the response to a previous question, what is referred to as response dependence (Marais & Andrich, 2008). Response dependence is observed in items that over-discriminate. High correlations between standardised item residuals indicate a violation of the assumption of independence. I checked the residual correlation matrix to look for items with high correlations, and here I discuss the pairs of items with response dependence, and the ICCs for these items. I carried out content analysis to understand these items more clearly.

The correlations between the standardised item residuals were investigated after rescoring and item splitting for DIF for language. The residual correlation matrix is provided in Table7.3 and items with high correlations are highlighted.

Table 7.3

Residual Correlation Matrix showing correlations between the first few items

RUMM2030 Project: PHDPROJECT Analysis: SPLT36
Title: SPLITITEM3.6DIFLANGUAGE Date: 7 Nov 2020 07:47:04 PM
Display: RESIDUAL CORRELATION MATRIX

Item	I.1	1.2	1.3.1	1.3.2	2.2.1	3.1	3.2	3.3	3.4	3.5	4.1.1	4.1.2
I.1	1.000											
1.2	0.062	1.000										
1.3.1	0.015	0.233	1.000									
1.3.2	0.016	0.214	0.208	1.000								
2.2.1	-0.027	-0.108	-0.064	0.070	1.000							
3.1	-0.023	0.091	0.081	0.016	-0.064	1.000						
3.2	-0.030	-0.031	0.138	0.002	-0.138	0.166	1.000					
3.3	-0.034	0.026	0.180	-0.019	-0.178	-0.001	0.279	1.000				
3.4	-0.069	-0.145	-0.091	-0.013	0.047	-0.010	0.031	0.045	1.000			
3.5	-0.121	-0.015	-0.029	0.073	-0.051	-0.024	-0.080	-0.022	0.091	1.000		
4.1.1	0.033	0.001	-0.035	-0.058	-0.005	0.038	0.009	0.025	-0.008	-0.060	1.000	
4.1.2	0.010	0.025	0.040	0.122	0.058	-0.018	-0.018	-0.011	0.108	0.054	-0.057	1.000
4.1.3	0.044	0.000	0.029	0.111	0.015	-0.038	-0.060	-0.073	0.039	0.096	-0.054	0.108

In Table 7.3 the residual correlation between Item 1.3.1 and Item 1.3.2 was 0.208, and the residual correlation between Items 3.2 and 3.3 is 0.279. The residual correlation between Item 1.2 and 1.3.1 was 0.233. These residual correlations were considerably higher than the residual correlations of the other items. In this table only the residual correlations for the first few items are shown because of space restrictions. The high residual correlations show that the response to Item 1.3.2 depends on Item 1.3.1, the response to Item 3.3 depends on Item 3.2 and the response to Item 1.3.1 depends on Item 1.2. Christensen et al., (2017) suggested that any residual correlation greater than 0.2 above the average correlation would appear to indicate local dependence. The other pairs which had large residual correlations but which were not shown in Table 7.3 were Items 3.4 and 3.6 (residual correlation = 0.306), 4.1.1 and 4.1.4 (residual correlation = 0.248), 4.1.3 and 4.1.4 (residual correlation = 0.250), 8.2 and 8.4 (residual correlation = 0.263) and 10.2.3 and 10.2.4 (residual correlation = 0.280)

7.3.1 Response dependence: Questions 3.2 and 3.3 (Items 10 and 11)

Firstly, by looking at the ICC for Item 10 (Question 3.2) and Item 11 (Question 3.3), it seemed that there was greater empirical discrimination in Question 3.3 than in Question 3.2 as shown by the ICCs in Figure 7.8. Neither of the observed proportions fit the theoretical curves well. The observed proportions were much steeper than the theoretical curves which often indicates response dependence. However, the greater empirical discrimination in Question 3.3 confirmed that it was dependent on Question 3.2. Considering the item locations, the relative difficulties of Questions 3.2 and 3.3 were -0.696 and -0.556 respectively, showing that Question 3.2 was slightly easier than Question 3.3. The fit residuals for Questions 3.2 and 3.3 were -1.396 and -2.704 respectively, showing that Question 3.2 was better fitting, whereas Question 3.3 was poorly fitting.

Question 3.3 could not be resolved by rescoring and still showed misfit. The fit residual is -2.704, a value outside of the recommended range of -2.5 to 2.5. The chi square probability is 0.001, which is below the set criterion of 0.05 and hence indicating that it does not meet the model requirements. Question 3.3 will now be discussed in detail. Figure 7.8 shows the ICC for Question 33 after rescoring and item splitting for DIF.

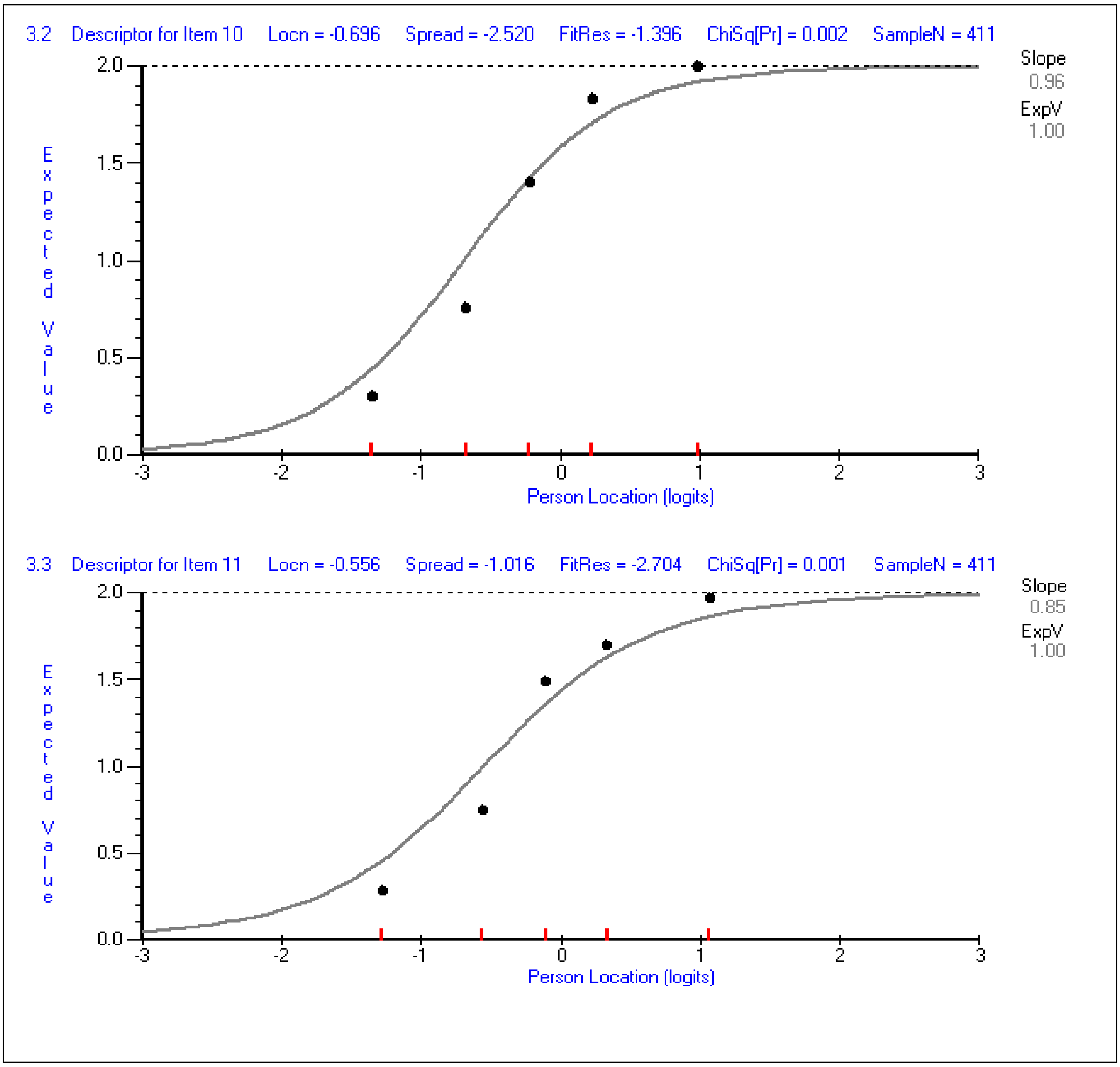


Figure 7.8. ICCs for Item 10 (Question3.2) and Item 11 (Question3.3)

In Figure7.8, the observed proportions were steeper than the theoretical curves for both items. For the low proficiency groups, the observed proportions were below the values predicted by the model, but for the high proficiency groups, the observed proportions were above the theoretical curve. In this case discrimination was greater than expected hence there was over-discrimination. Over-discrimination is a result of greater dependence among the responses hence there was response dependency. Content analysis will be carried out later in this section to determine some of the causes of response dependence in these items.

Item 11 (Question 3.3) was then split into 11S0, 11S1 and 11S2, corresponding to whether the learner got Item 10 incorrect, partially correct (1 mark) or completely correct (2 marks). All of the other items were anchored to their original estimates before the splitting. The individual item fit statistics for Items 11S0, 11S1 and 11S2 are given in Table 7.4.

Table 7.4

Individual item fit after splitting Item 11/Question 3.3

```
RUMM2030        Project: PHDPROJECT        Analysis: SPLT10R
Title:  SPLITITEM3.3RESPONSEDEPENDENCE                             08:37:46 PM
Display: INDIVIDUAL ITEM-FIT - Serial Order
```

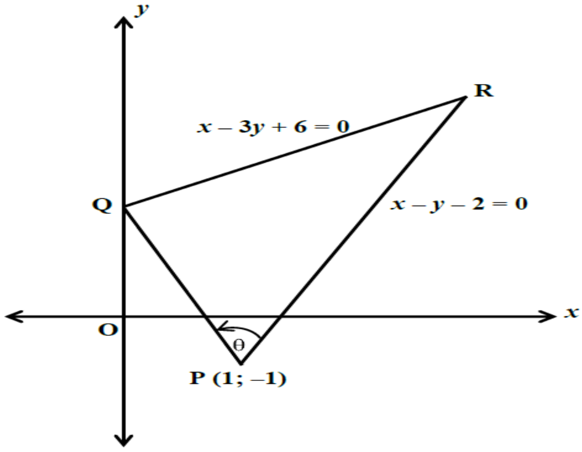
Seq	Item	Type	Location	SE	Residual	DF	ChiSq	DF	Prob
113	011S0	Poly	0.028	0.150	-0.648	93.42	1.539	4	0.819762
114	011S1	Poly	-0.379	0.395	1.513	8.58	4.234	4	0.375225
115	011S2	Poly	-1.145	0.110	-0.687	211.63	4.476	4	0.345392

Table7.4 indicates that for those learners who got Item 10 correct, getting a correct response on Item 11 had a difficulty estimate of -1.145, showing that it was by far the easiest task to do. For those learners who scored 1 mark for Item 10, getting a correct response in Item 11 was at a difficulty level of -0.379. This shows that it was relatively easy for them to do the task. For those learners who got Item 10 incorrect, getting a correct response for Item 11 was at a difficulty level of 0.028, showing that it was difficult for them to do the task.

The items and the accompanying marking rubric are shown in Figure 7.9. The two items were based on analytical geometry. A full description of the questions follows.

QUESTION 3

In the figure below, PQR is a triangle with P(1 ; –1) . Q is a point on the y-axis. The equations of QR and PR are $x - 3y + 6 = 0$ and $x - y - 2 = 0$ respectively. Given $Q\hat{P}R = \theta$.



3.1	Show that the co-ordinates of Q are (0 ; 2).	(2)
3.2	Write down the gradient of QR.	(2)
3.3	Prove that $P\hat{Q}R = 90°$.	(2)
3.4	Calculate the co-ordinates of R.	(3)
3.5	Calculate the area of Δ PQR.	(4)
3.6	Calculate the length of PR. (leave your answer in the simplest surd form).	(2)

3.2	$y = \frac{1}{3}x + 2$ $\therefore m_{QR} = \frac{1}{3}$	✓ A writing in standard form ✓ A answer (2)
3.3	$m_{PQ} = \frac{-1-2}{1-0} = -3$ $\therefore m_{PQ} \times m_{QR} = (-3)\left(\frac{1}{3}\right) = -1$ $\therefore PQ \perp QR$ Thus $P\hat{Q}R = 90°$	✓ A gradient of PQ ✓ A products of gradients (2)
3.4	$x - 3y + 6 = 0$........(1) $x - y - 2 = 0$..........(2) $(1) - (2): -2y + 8 = 0$ $y = 4$ subst $y = 4$ into (1) $x - 3(4) + 6 = 0$ $x - 12 + 6 = 0$ $x = 6$ R(6;4)	✓ M solving both equations simultaneously ✓ CA substituting $y = 4$ ✓ CA $x = 6$ (provided R is in first quadrant) (3)

Figure 7.9. Items 10 (Question 3.2) and Item 11 (Question 3.3) and the corresponding marking guideline

All of the marks in these two items were accuracy marks, meaning that any deviation from the correct solution would result in loss of marks. Item 11/Question 3.2 required learners to calculate the gradient of line QR, which meant learners were supposed to make y the subject of the formula in the given equation for the line QR. An alternative solution was to calculate the coordinates of point R by equating the equations of QR and PR, and use the points Q and R to calculate the gradient. The second method was not provided in the marking rubric. The solution for Question 3.2 was required in order for learners to respond successfully to Question 3.3. For the lines to be perpendicular, the product of the gradients is supposed to be equal to negative -1. Failure to respond to Question 3.2 meant that learners would not respond correctly to Question 3.3, hence there was response dependence. Also since the marks allocated in Question 3.3 were accuracy (A) marks, a learner with incorrect answers for Item 10 would then lose the marks in item 11.

7.3.2 Response dependence: Questions 1.3.1 and 1.3.2 (Items 3 and 4)

The ICCs for Item 3 (Question 1.3.1) and Item 4 (Question 1.3.2) showed that there was greater empirical discrimination in Question 1.3.2 than in Question 1.3.1, as shown in the ICCs in Figure 7.11. The observed proportions did not fit the theoretical curves well in both questions. The observed means were much steeper than the theoretical curve, which often indicate response dependence. Considering the item locations, the relative difficulties of Question 1.3.1 and Question 1.3.2 were 1.289 and 1.686 respectively, showing that Question 1.3.1 was less difficult than Question 1.3.2. The ICCs for the two items are presented in Figure 7.10.

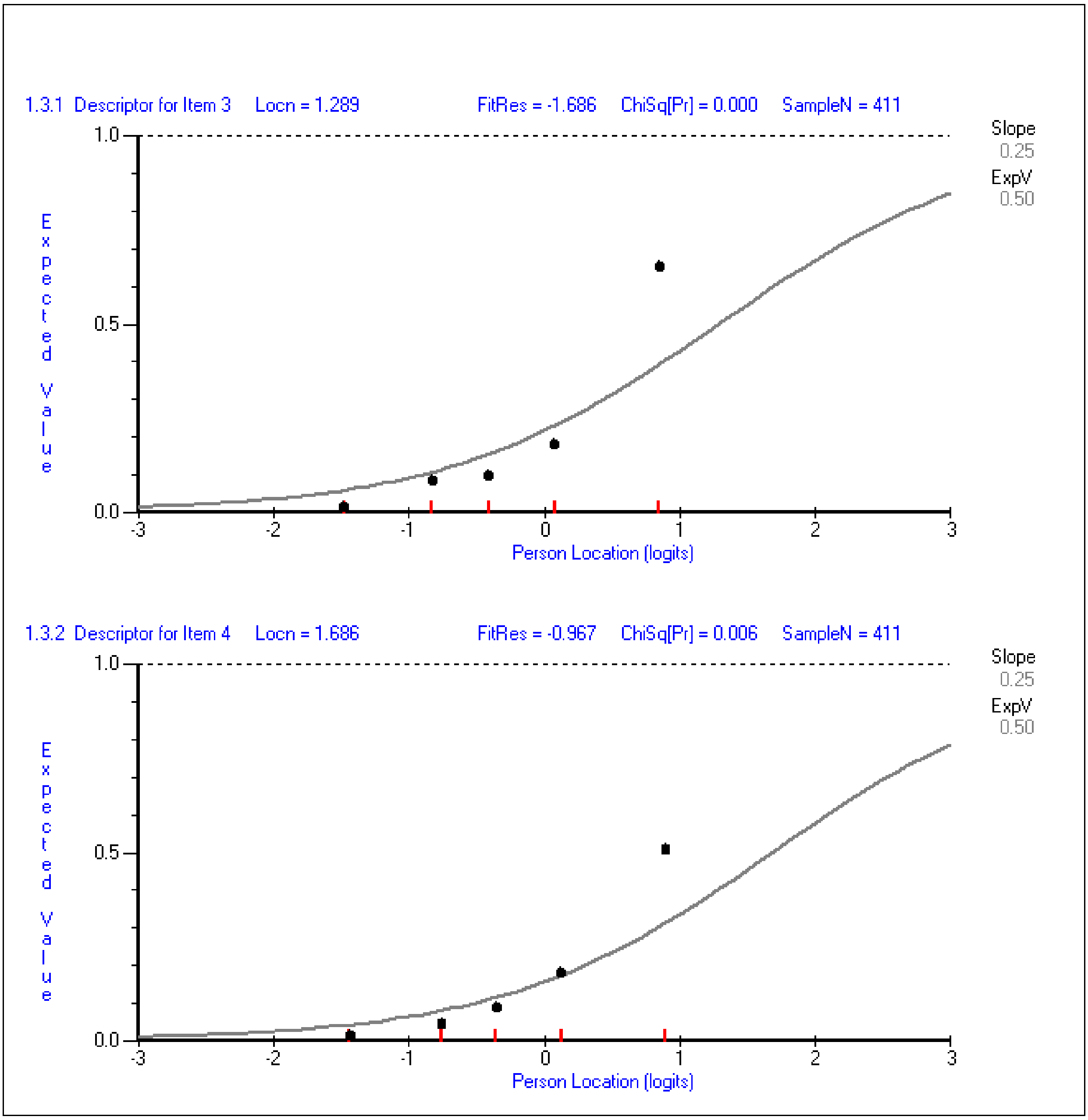


Figure 7.10. ICCs for Item 3 and Item 4

Item 4 (Question 1.3.2) was then split into 4S0 and 4S1 (since the item was rescored to two categories only), corresponding to whether the learner got Item 3 incorrect (0 marks) or correct (1 mark). All of the other items were anchored to their original estimates before the splitting. The individual item fit statistics for the Items 4S0 and 4S1 are given in Table 7.5.

Table 7.5

Individual item fit after splitting Item 4 (Question 1.3.2)

```
RUMM2030        Project: PHDPROJECT        Analysis: SPLT4R
Title:  SPLITITEM4FORDEP                                    08:36:49 PM
Display: INDIVIDUAL ITEM-FIT - Serial Order
```

Item	Type	Location	SE	Residual	DF	ChiSq	DF	Prob
4S0	Poly	2.072	0.236	-0.396	243.36	3.491	4	0.479221
4S1	Poly	0.805	0.258	-0.160	65.59	5.145	4	0.272758

For learners who got Item 3 incorrect, getting a correct response on Item 4 had a difficulty estimate of 2.072, showing that it was a very difficult task. For those learners who scored 1 mark in Item 3, getting a correct response in Item 4 was at a difficult level of 0.805, showing that it was difficult for them to do.

The items and the accompanying marking rubric are shown in Figure 7.11. The two items were based on statistics. A full description of the questions follows.

QUESTION 1

The table below shows the heights of palm trees in a park.

HEIGHT IN CM	NUMBER OF PALM TREES
$120 < x \leq 135$	1
$135 < x \leq 150$	15
$150 < x \leq 165$	45
$165 < x \leq 180$	28
$180 < x \leq 195$	1

1.1 Determine the estimated mean height of the palm trees in the park. (2)

1.2 Draw an ogive to represent this data. (4)

1.3 Use your ogive curve to estimate the:

1.3.1 median height of the palm tree. (2)

1.3.2 interquartile range (IQR). (3)

[11]

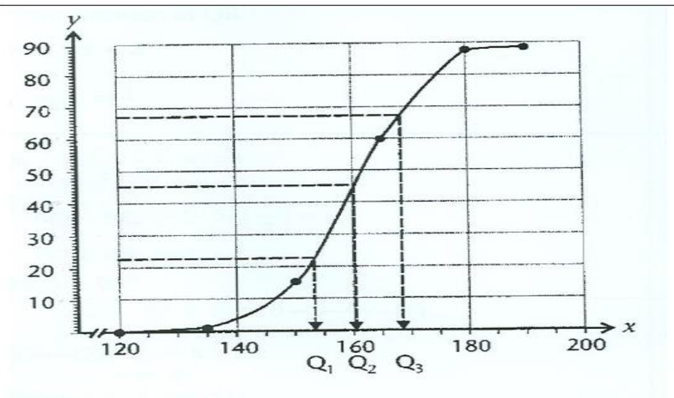
1.2		✓ A shape ✓ A (120;0) ✓ A (60;165) ✓ A for any other correct plotted points (4)
1.3		
1.3.1	median height = 161 cm	✓✓ A A Answer (Accept : 160 – 163) (2)
1.3.2	IQR = $Q_3 - Q_1$ = 169 – 152 = 17 cm	✓ CA Q_3 ✓ CA Q_1 ✓CA answer (Accept 1 mark deviation for Q_1 and Q_3) (3)

Figure 7.11 Item 3. (Question 1.3.1) and Item 4 (Question 1.3.2) and the corresponding marking rubric

Question 1.3.1 required learners to use their ogive curves to estimate the median height of the palm tree. The question required learners to determine the position of the median height and then use the graph to determine the height that corresponded to that position on the graph. The same concept of finding the position of the median in Question 1.3.1 is required in Question 1.3.2 in determining the position of the lower quartile and the position of the upper quartile. After determining the position of the lower and upper quartile, learners were then supposed to use the graph (ogive) to find the heights that corresponded to those positions. The upper quartile and lower quartile were then used to calculate the interquartile range (IQR = Upper quartile – Lower quartile).

The other pairs of questions with high residual correlations (eight pairs of questions) required the same approach of splitting to resolve response dependency, learners who were successful in the first question had a very high chances of responding to the next question correctly. A learner who was not successful in the first question had a very low chance of responding to the next question correctly as shown by the two pairs in Figure 7.11.

7.4 Multidimensionality

The principal component analysis (PCA) of the residual was run in RUMM2030 to detect any sign of multidimensionality. McGill (2009) explains that the PCA procedure groups items into sets that correlate with one another but are relatively independent of other internally consistent subsets of items. Figure 7.12 shows the principal component summary for the data set.

Principal Component Summary

Principal Component Summary | PC Normalised Vectors | PC Loadings | Varimax F

PC	Eigen	Percent	CPercent	StdErr
PC001	2,602	6,50%	6,50%	0,361
PC002	2,292	5,73%	12,24%	0,318
PC003	1,969	4,92%	17,16%	0,271
PC004	1,871	4,68%	21,84%	0,259
PC005	1,749	4,37%	26,21%	0,240
PC006	1,616	4,04%	30,25%	0,221
PC007	1,524	3,81%	34,06%	0,210
PC008	1,470	3,67%	37,73%	0,201
PC009	1,426	3,56%	41,29%	0,194
PC010	1,337	3,34%	44,64%	0,183
PC011	1,296	3,24%	47,88%	0,175
PC012	1,225	3,06%	50,94%	0,168
PC013	1,160	2,90%	53,84%	0,156
PC014	1,146	2,86%	56,70%	0,156
PC015	1,130	2,82%	59,53%	0,155
PC016	1,101	2,75%	62,28%	0,151
PC017	1,021	2,55%	64,83%	0,139
PC018	0,997	2,49%	67,32%	0,137
PC019	0,970	2,43%	69,75%	0,132
PC020	0,923	2,31%	72,06%	0,126
PC021	0,866	2,16%	74,22%	0,119
PC022	0,847	2,12%	76,34%	0,116
PC023	0,833	2,08%	78,42%	0,115
PC024	0,803	2,01%	80,43%	0,110
PC025	0,731	1,83%	82,25%	0,102

Figure7.12. Principal Component Summary

The residual PCA showed that the first principal component (PC1) had an eigenvalue of 2.602 which was considerably larger than those of the others. The principal component summary shows that 6.5% of the total variance is accounted for by PC1.

The principal component loadings are given in Figure 7.13, showing that 15 items load positively on PC1, while 25 items load negatively on PC1. It may be that the items that load positively or negatively on PC1 may share some characteristics with others in their subset, thus suggesting that they had a sub-dimension in common.

PC Loadings

Principal Component Summary | PC Normalised Vectors | PC Loadings | Varimax Rotation Loadings | Residual Correlation Matrix | Person-Item Residuals

Item	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
2.2.2	0,636	0,069	0,114	-0,386	-0,167	0,103	-0,126	0,012	0,251	0,163	-0,064	-0,036	-0,104	0,027	-0,002
2.1	0,452	-0,117	0,057	-0,318	0,108	0,073	0,249	0,006	-0,029	-0,102	-0,026	-0,147	-0,132	0,053	0,224
9.2	0,407	-0,208	-0,068	0,036	-0,234	-0,007	0,237	0,085	0,075	-0,102	-0,165	-0,033	-0,036	0,365	0,035
2.2.1	0,398	0,201	0,001	-0,312	-0,198	0,042	-0,158	-0,050	0,480	0,136	0,114	0,111	-0,216	-0,121	-0,140
2.2.3	0,263	0,147	-0,072	0,172	0,220	-0,405	-0,179	0,117	-0,095	-0,137	-0,203	-0,114	0,201	0,145	0,239
9.1	0,222	-0,355	0,099	0,081	0,014	-0,123	0,240	0,256	-0,073	0,112	0,157	0,364	0,206	0,232	-0,197
6.2	0,201	-0,254	-0,327	0,158	-0,048	0,066	0,008	0,033	-0,073	-0,201	-0,168	0,016	-0,318	-0,046	-0,173
6.1.2	0,129	-0,212	0,154	0,160	0,309	0,178	-0,211	-0,039	-0,129	-0,069	-0,092	0,214	-0,299	-0,071	0,128
3.5	0,124	0,322	0,218	0,419	0,092	0,038	0,141	-0,413	-0,087	-0,211	0,057	0,232	0,202	0,135	-0,010
5.3	0,121	-0,098	0,139	-0,001	0,581	-0,241	-0,014	-0,028	0,079	0,068	-0,066	0,220	0,130	-0,331	0,062
3.6	0,120	0,563	-0,127	0,412	0,195	0,185	-0,087	-0,033	0,181	0,093	-0,010	-0,086	0,045	-0,101	-0,019
5.2	0,091	0,354	0,008	0,215	-0,070	-0,067	0,026	-0,027	-0,394	-0,147	-0,181	0,115	-0,202	-0,149	-0,180
7	0,078	0,078	-0,328	-0,190	0,169	-0,316	-0,103	-0,330	0,085	-0,088	-0,092	-0,332	0,076	-0,055	-0,326
6.1.1	0,012	0,291	-0,015	0,022	0,292	0,149	-0,335	0,429	-0,125	0,208	-0,081	-0,206	-0,194	0,056	-0,058
10.1	0,007	-0,359	-0,091	0,078	-0,016	0,044	0,020	0,098	-0,168	0,110	0,556	-0,360	0,060	-0,201	0,315
I.1	-0,008	-0,045	0,142	-0,273	0,074	0,113	-0,174	0,270	-0,134	-0,042	-0,029	-0,055	0,540	-0,142	-0,204
4.2.1	-0,010	0,061	0,501	-0,081	-0,282	-0,289	0,024	-0,247	-0,005	0,370	-0,096	-0,187	0,107	-0,112	0,043
4.1.3	-0,026	0,363	0,145	0,087	-0,220	-0,243	-0,190	0,157	0,020	-0,362	0,143	-0,282	0,008	0,259	-0,061
8.3	-0,037	0,409	-0,360	-0,010	-0,018	-0,081	0,186	0,202	-0,017	0,061	0,249	0,008	-0,186	0,096	-0,013
8.4	-0,073	0,406	-0,271	-0,299	-0,012	0,061	0,147	0,024	-0,191	0,146	0,043	0,340	0,058	0,080	-0,032
8.2	-0,087	0,142	-0,435	-0,369	-0,121	0,041	-0,111	-0,356	-0,258	0,057	-0,013	0,044	0,116	0,109	0,375
10.22	-0,098	0,243	-0,106	-0,228	-0,175	0,025	0,280	0,295	-0,141	0,027	-0,343	0,132	0,056	-0,303	0,281
3.4	-0,104	0,271	-0,113	0,377	-0,127	0,282	0,069	-0,001	0,365	0,072	0,249	0,066	0,161	-0,109	0,150
10.21	-0,115	0,002	-0,376	-0,180	0,160	-0,073	0,133	0,353	0,223	-0,184	-0,022	-0,036	0,305	0,014	-0,199

Figure 7.13. Principal component loadings

Going further in the analysis, two item subsets were created from the items loading negatively and positively on the first residual factor in the PCA. Table 7.6 provides details of the items in each set.

Table 7.6.

Details of the items from each subset.

Item set 1		Item set 2	
		I.1	Ogive curve
		1.2	
		1.3.1	
		1.3.2	
2.1	Fuel consumption and regression line		
2.2.1			
2.2.2			
2.2.3			
		3.1	Analytical geometry
		3.2	Gradient
		3.3	Show <= 90
		3.4	Coordinates of R
3.5	Area of Δ		
3.6	Length PR		
4.1.1	Analytical Geometry, equation of circles		
4.1.2			
4.1.3			
4.1.4			
4.2.1			
4.2.2			
5.1	Show trig expression= given expression		
		5.2	Prove trig identity
		5.3	Prove identity
5.4	Maximum value of area		
		6.1.1	Expansion
		6.1.2	General solution
		6.2	Sketch 2 trig graphs
6.3	3D trig		
		7	State 7 <s with reasons
8.1	Prove cyclic quad		
8.2	Prove Δ isosceles		
8.3	Prove similar		
8.4	Similarity calculating length		
		9.1	Ratio & proportion calculations
		9.2	Ratio & proportion calculations
		10.1	Proving similarity theorem
10.2.1	Prove similarity		
10.2.2	Prove ratio expression		
10.2.3	Prove ratio expression		
10.2.4	Deduce from ratio expression		

Noting that the mean location of the items in Set 1 was -0.305, while the mean location for Set 2 was -0.3631, suggested that the items in Set 2 were experienced as less difficult than those in Set 1. Many proof type questions for Euclidean geometry (Questions, 8.1, 8.2, 8.3, 8.4, 10.2.1, 10.2.2, 10.2.3, 10.2.4) and questions which required applications (5.4, 6.3, 4.1.1, 4.1.4, 4.2.1) were all in Set 1 and literature has shown that these sections are challenging to learners (Luneta, 2015; Naidoo & Kapofu, 2020, Ngirishi & Bansilal, 2019). Set 2 consisted mostly of questions which require calculations and no justification of answers or steps was required.

I compared the person location estimates for each set using the t- test option in RUMM 2030. In this procedure, the person estimates derived from each set are compared using t-tests. The percentages of tests falling outside the 95% confidence interval are then evaluated. The results appear in Figure 7.14. It is assumed that any significant number of tests outside this interval would indicate the presence the multidimensionality (Andrich & Marais, 2019)

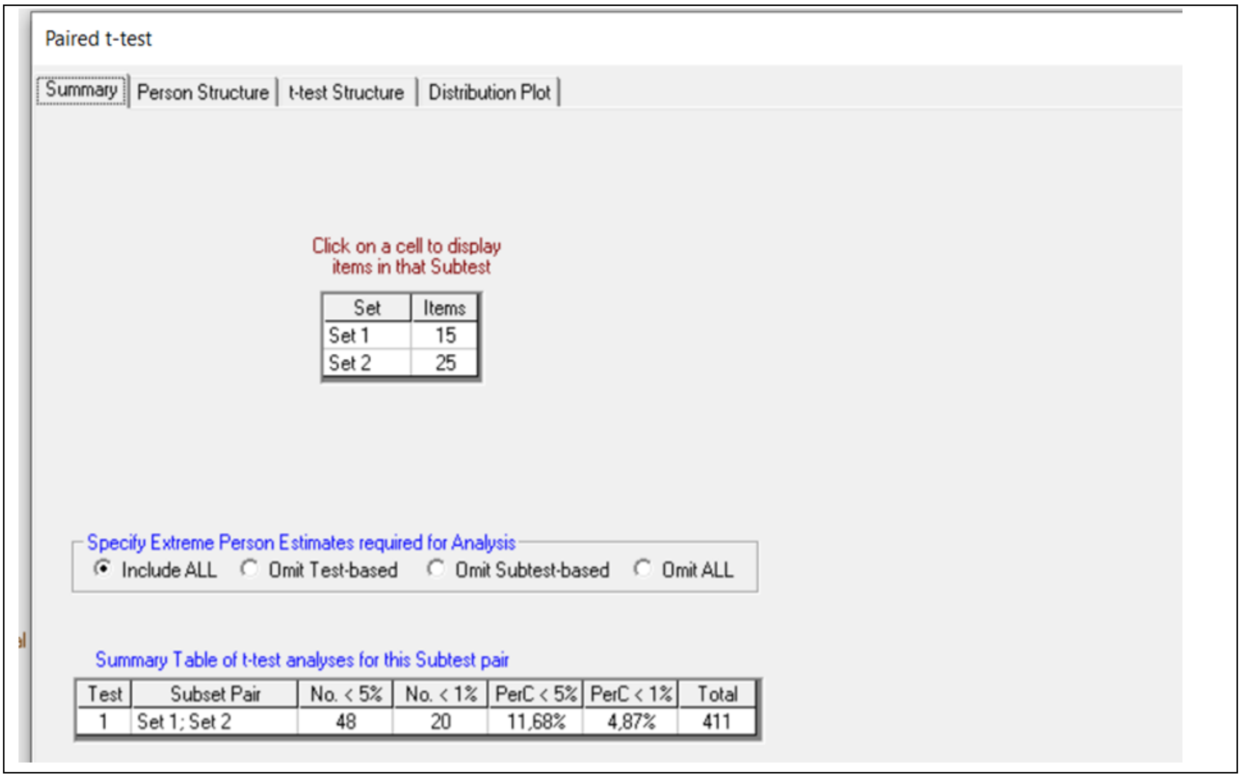


Figure 7.14. Summary of t-test analysis for the sub-set pair

In using the paired t- test option it was found that there were 48 persons (11.68%) where the differences in estimates between the two subscales exceeded the 5% level of significance, while there were just 20 persons (4,9%) where the differences in estimates exceed the 1% level of significance. Because the percentages were close to the value of 5%, we consider it to be a minor violation of independence that may be just a chance effect (Andrich & Marais, 2019, p.282). Andrich and Marais (2019) caution that because the items to be placed in the subtests were derived from a related analysis, and not an independent one, the reason for forming the subtests is a conservative way of checking for the presence of two dimensions. Hence we deduce that this is a minor violation of independence.

CHAPTER 8 QUALITATIVE ANALYSIS OF THE TEACHERS' RESPONSES

8.1 Introduction

This section provides an analysis of the seven teachers' responses to the questionnaires and their beliefs regarding the findings of the Rasch analysis. The teachers responded to most of the questions, but there were a few where not all teachers responded. Although the questions in the questionnaire were based on the findings of the Rasch analysis, which most of the teachers were not familiar with, the findings related to issues involving the teaching and learning of mathematics and the teachers could easily relate these to the contexts in their schools. The questions and the teachers' responses are discussed in the following sections.

8.2 Questions and teachers' responses

8.2.1 Teacher responses to Question 1

The first question that the teachers were asked about the paper was: From the Rasch analysis done, learners found this trial paper to be generally difficult. Can you please comment on this finding and give reasons why this is so?

Although the Rasch analysis revealed that learners found the mathematics trial examination paper to be difficult, two of the teachers who took part in the study felt the paper was well balanced in terms of the cognitive levels of the questions asked. Teacher 4, wrote:

> "The paper was generally fair, covered the outcomes and balanced in terms of short, medium and long responses as well as knowledge, application and synthesis."

The paper being well balanced meant that the average learner was supposed to pass the paper. Teacher 4 believed that the paper covered all the cognitive levels used to guide all assessment tasks. Teacher 1 mentioned the issue of the trial paper focusing on the most basic concepts of every section, and stated that teachers do not pay maximum attention or spend a lot of time on the basics of a section during revision time.

> "With an exception of Question 10.2, most of the question paper focused on the most basic concepts of every section."

The responses from these two teachers showed that they believed that the paper was fair enough, and had enough questions which covered knowledge, routine procedures, complex procedures and problem solving. However, although the Mathematics Trial Examination Paper 2 was perceived to be well balanced by the teachers, the Rasch analysis revealed that the learners found the paper difficult. The teachers, including Teachers 1 and 4 gave reasons as to why they believed the learners found the paper challenging. One of the reasons mentioned was the inclusion of many proof questions, which the learners found to be very difficult. Teacher 4 wrote:

> "There are a number of questions that require proof and many learners usually want/expect to find answers to problems rather than outlining proofs."

Teacher 3 also believed that the paper was challenging for the learners because of their failure to deal with proof-type questions. Teacher 3 wrote that:

> "Generally most learners are threatened by paper 2, especially topics that include circle geometry and especially geometry that requires proofs other than straight forward geometry questions that require giving a size of an angle and reasons. This question paper had lots of questions that require proving. They had to think a lot to get 7 marks in Question 10.1".

Questions 8.1, 8.2, 8.3, 10.1, 10.2.1, 10.2.2 and 10.2.3 were some of those where learners were supposed to give proofs.

Figure 8.1 shows the script of a learner who did not attempt to answer the proof questions 10.1, 10.2.1, 10.2.2 and 10.2.3. Proof Questions 8.1, 8.2, 8.3, 10.1, 10.2.1, 10.2.2 and 10.2.3 contributed 32 marks out of the possible 150 marks. Learners who struggled with the proof questions lost those 32 marks, meaning they did not attempt 21.3% of the paper.

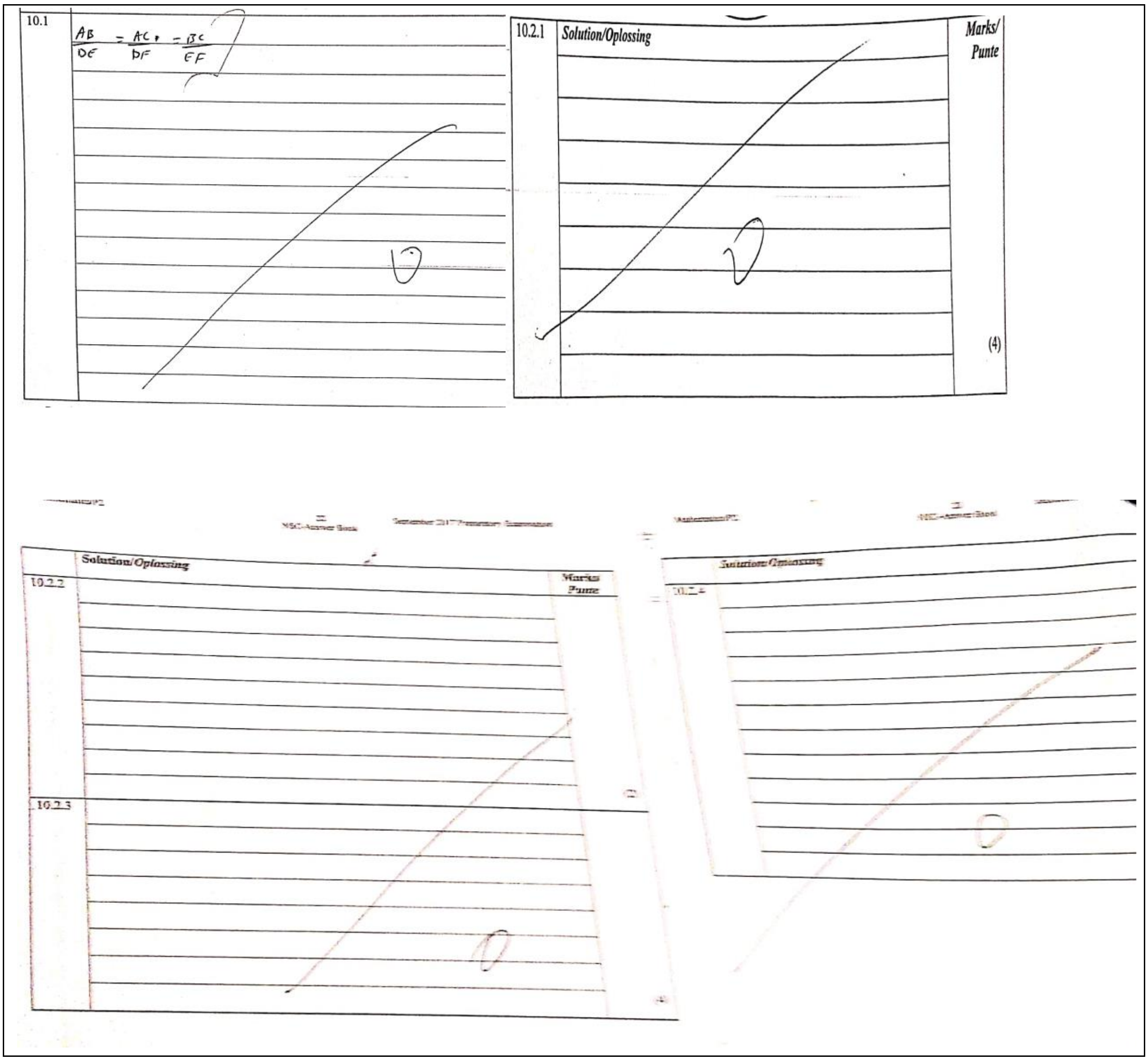


Figure 8.1. Learner SB493's script, showing blank spaces for Questions 10

Teacher 1 believed the learners found the paper challenging because the basics from the topics were not well covered during the teaching process nor during revision in preparation for examinations. Teacher 1 wrote:

> "During the teaching process and examination preparation time, teachers do not pay attention or spent more time on the basics. They just introduce the topic and quickly move on to the difficult questions. This action result in learners finding it very difficult to deal with the basics of the work covered in the syllabus or curriculum."

The concerns of Teacher 1 implied that if basics are not well taught, learners will end up lacking conceptual understanding. The learners will not understand the basics, and in most cases they do not even understand why certain steps are taken hence they will not be able to respond correctly to questions which require knowledge of the basics. After the introduction of new concepts, Teacher 1 said there is a need to make sure that all learners are comfortable working with the basics before attempting higher order questions. Failing to deal with the basic concepts is a very big disadvantage to learners, as knowledge of the basics is required for better understanding and for responding successfully to higher order questions.

Teachers 2, 1 and 4 mentioned the lack of adequate preparation as the reason why the learners found the paper difficult. Teacher 2 wrote the following:

> "It has been observed over the years that the trial examinations are generally harder and some learners might not have been adequately prepared for the examinations."

Teacher 4 was of the same view:

> "The other reason why learners found the paper to be difficult could be that the learners were not yet prepared to write a paper whose standard is the same as the final examination."

Adequate preparation has a strong impact on how learners will perform in an examination, regardless of whether the paper meets the standards set for examinations or not. Adequate preparation involves a number of steps to be taken by both learners and teachers, which includes the learners giving themselves enough time to study, using flow charts and diagrams, practising using old examination papers, explaining answers to others, organising study groups with friends, taking regular breaks, planning the day of the examination and even drinking water. According to Teachers 2 and 4, learners who do not usually prepare adequately are more likely to find any examination more challenging.

Teacher 1 believed learners need to give themselves enough time to study by drawing up a timetable that fits their study requirements and accommodates all of the learning areas and must avoid leaving things and important aspects for the last minute. There are exceptions where some learners only study during assessment time and succeed, but this way of partial studying is often not the best approach for examination preparation as the volume of work may be too great. Regarding last minute studying, Teacher 1 wrote:

> "For students who start studying when it is time for examinations, it is always difficult for them to write a mathematics paper and pass, because it is almost impossible to understand all the mathematics concepts in a short period of time. Last minute studying

might work for some subjects but not mathematics. Learners struggled in this examination because they did not give themselves enough time to study before the examination commenced."

Teachers 2, 6 and 7 believed the paper was difficult to the learners because questions demanded high cognitive levels and were of a high standard. Teacher 6 wrote:

"The level of questions began to be more difficult. Examiners went deeper into the CAPS document rather than generally testing the content. It was also to ensure that educators increase their standard of teaching and look much deeper into the content."

Also referring to the standard of the paper, Teacher 2 wrote

"It has been observed over the years that the trial examinations are generally harder and some learners might not have been adequately prepared for the examination."

Teacher 2's view seems to suggest that over the years the trial examinations have become harder and of a higher standard. Teacher 7 had a different view as to why the learners found the paper difficult. The teacher wrote:

"Learners have challenges when it comes to assessment, specifically examinations, let alone mathematics of which society regards it as most difficult and most challenging subject".

Teacher 7's comments seemed to suggest that it was expected that the paper would be difficult for the learners because of societal beliefs. This notion is supported by Brown et al. (2005), who mentioned that learners' performances in mathematics is often influenced by society's dislike for mathematics and the perception that the subject lacked relevance. Ercikan, McCreith and Lapointe (2005) claimed that the strongest predictors associated with mathematics achievement and participation are the student's attitude towards mathematics, the home environment and society, parents' expectations, teachers' expectations and mathematics anxiety.

Teacher 5 also believed that the paper was difficult because of how it was structured. To this end Teacher 5 wrote

"Most of the questions are trick in the sense that if a learner fails to get the first part of a question, they would lose most of the marks in the other parts of the same questions.

> They are not direct questions. Many mistakes can be made with many steps needed, for example, Question 4, using the product of gradients, completing the square. Also in this question paper, learners who could not understand the question had difficulties even if they had some clue. The questions required integration of different sections and topics."

This showed that the teacher did not understand the marking guideline, which allows for continuous accuracy (CA) marking. The teachers' views on CA marking will be explored further in sections to follow.

Teacher 5 also mentioned the inability of learners to understand questions as one of the reasons why some learners found the paper difficult. This will also be discussed in later sections.

8.2.2 Teacher responses to Question 2

The second question that the teachers were asked about the paper was: Despite the statistics questions (Questions 1.1, 1.3.1 and 1.3.2), being placed among the easy questions (levels 1 and 2 on the cognitive levels), the Rasch analysis and the results showed that these questions were among the top 7 most difficult questions. Why do you think these statistics questions were harder than expected?

Most of the teachers involved in this study (Teachers 2, 3, 4 and 5) blamed the learners' inability to respond correctly to statistics questions on complacency. According to the University of Cambridge Dictionary, complacency is a feeling of calm satisfaction with your own abilities or situation that prevents you from trying harder. Teacher 3 stated that learners underestimate statistics, and did not practice statistics questions in preparation for the exam. Teacher 3 wrote;

> "Most learners underestimate statistics, they did not think and apply themselves properly, they did not practice this section. They took it for granted and just answer and made unfounded assumptions and silly mistakes."

Teacher 2 echoed these sentiments when he wrote:

> "Statistics is normally done towards the end of the term and is taken to be easy by the learners and therefore not taken with same level of seriousness as other sections such

as trigonometry and geometry. Learners do not practice it with the enough attention because they think it is easy. They confuse the concepts because of lack of practice."

When learners do not practice a certain section of mathematics, silly mistakes become very common and concepts are usually mixed or misinterpreted as was witnessed in the case of the estimated mean for grouped data and the mean for ungrouped data.

Teacher 1 mentioned the technical aspect of the question as the cause of the poor performance by learners on statistics questions. The teacher noted that in most previous examinations, where learners are given grouped data in table form, the table will include columns for the midpoints of the class intervals and for the cumulative frequencies, which learners need to complete first. In Question 1, the columns for midpoints and cumulative frequencies were not provided, and hence the majority of learners struggled to respond successfully to these questions according to Teacher 1, who wrote:

"In most common cases (past exam papers and text books), where learners are given the grouped data in the form of a table, the table included midpoints and another column for cumulative frequencies to be completed first. In this question there was no such columns, therefore the majority of the learners might have forgotten that they will need these values in order to successfully answer the questions."

Question 1 is shown in the Figure 8.2.

QUESTION 1

The table below shows the heights of palm trees in a park.

HEIGHT IN CM	NUMBER OF PALM TREES
$120 < x \leq 135$	1
$135 < x \leq 150$	15
$150 < x \leq 165$	45
$165 < x \leq 180$	28
$180 < x \leq 195$	1

1.1 Determine the estimated mean height of the palm trees in the park. (2)

1.2 Draw an ogive to represent this data. (4)

1.3 Use your ogive curve to estimate the:

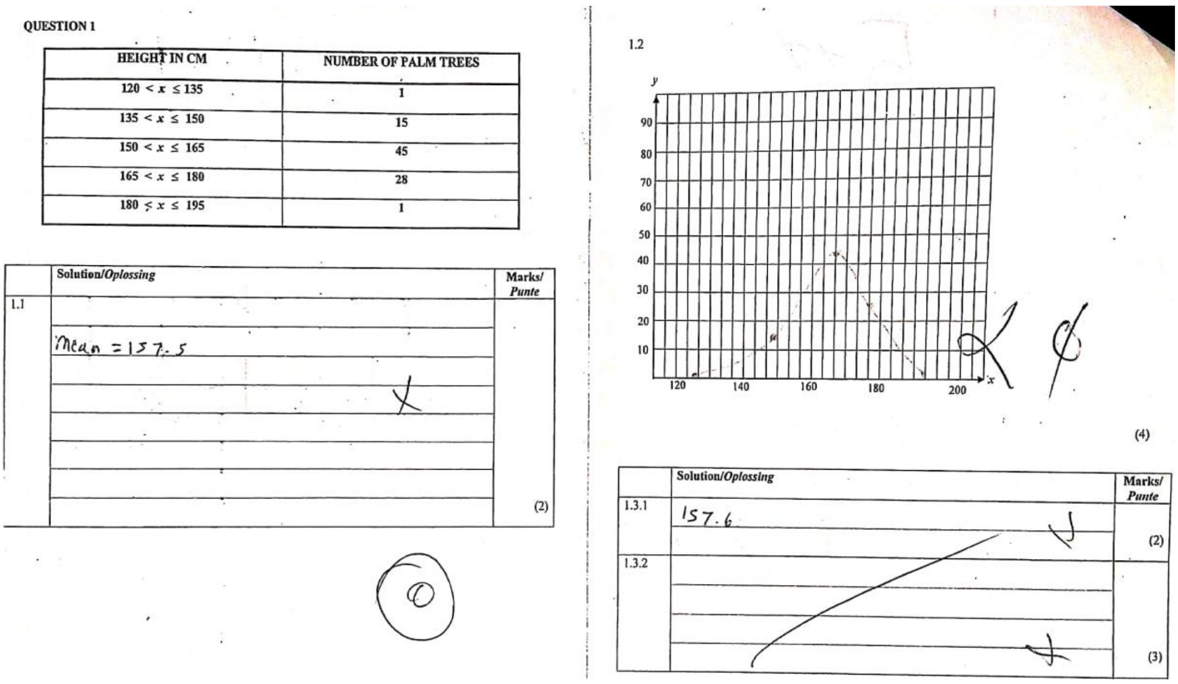
1.3.1 median height of the palm tree. (2)

1.3.2 interquartile range (IQR). (3)
[11]

Figure 8.2. Question 1

In Figure 8.2 the grouped data show the heights of palm trees in interval form and the frequencies for each interval. The columns for the cumulative frequencies and the points to plots were not provided. According to Teacher 1, the absence of these two columns made this question more challenging to the learners, as most of them did not answer Questions 1.1 and 1.2 correctly. Figure 8.3 shows the responses of two learners to Questions 1.1 and 1.2.

a. Learner FT 587's Response to question 1



b. Learner SB490's Response to Question 1.

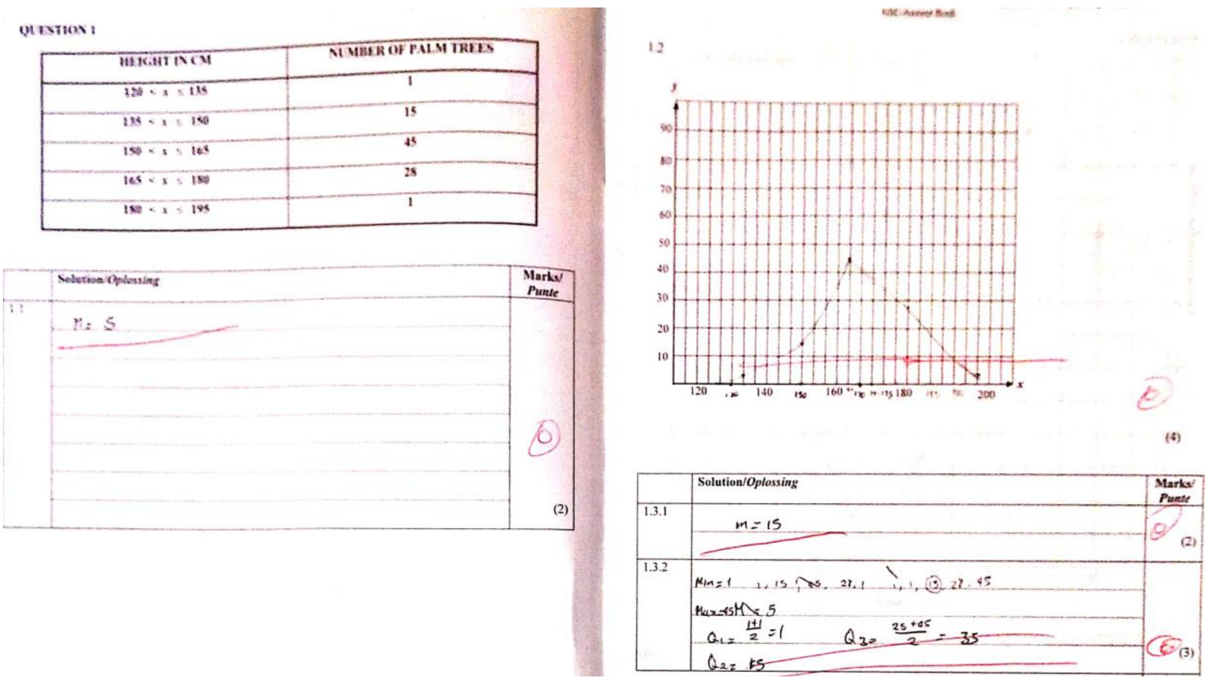
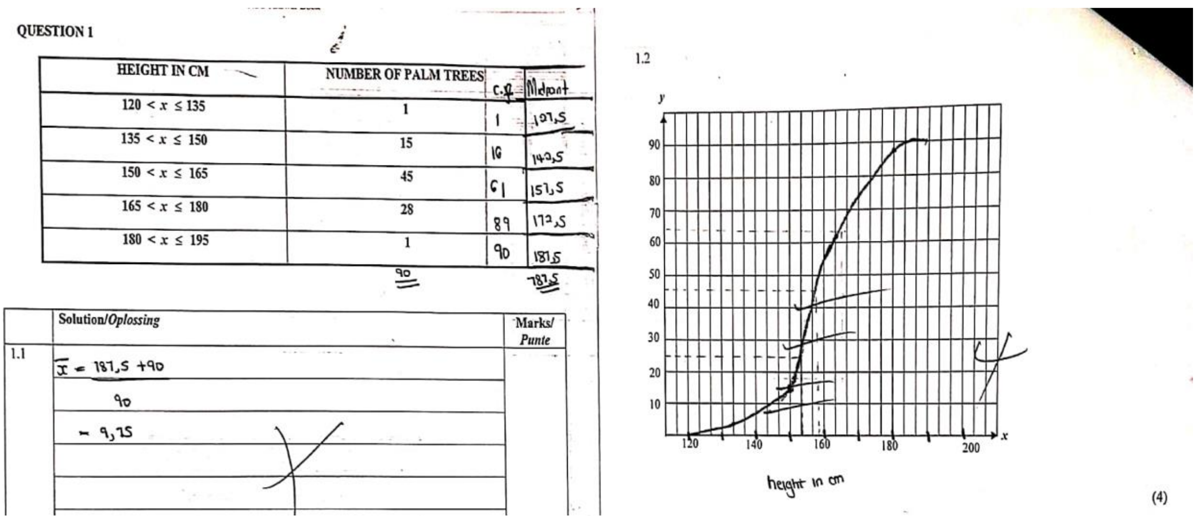


Figure 8.3. (a) Learner FT587's response to Question 1 and (b) learner SB490's response to Question 1.

Figure 8.3 shows that these two learners did not add the columns for the midpoints and the column for the cumulative frequencies. These two columns were important for learners to be able to respond to Question 1.2 successfully. As a result, these two learners did not score any marks. However, learners who realised that they needed the two columns for the midpoints and cumulative frequencies, and added these columns to their tables, were able to score some marks in Question 1.2 as evidenced in Figure 8.4.

a. Learner SS111's response to Question1.



b. Learner SS 200's response Question 1

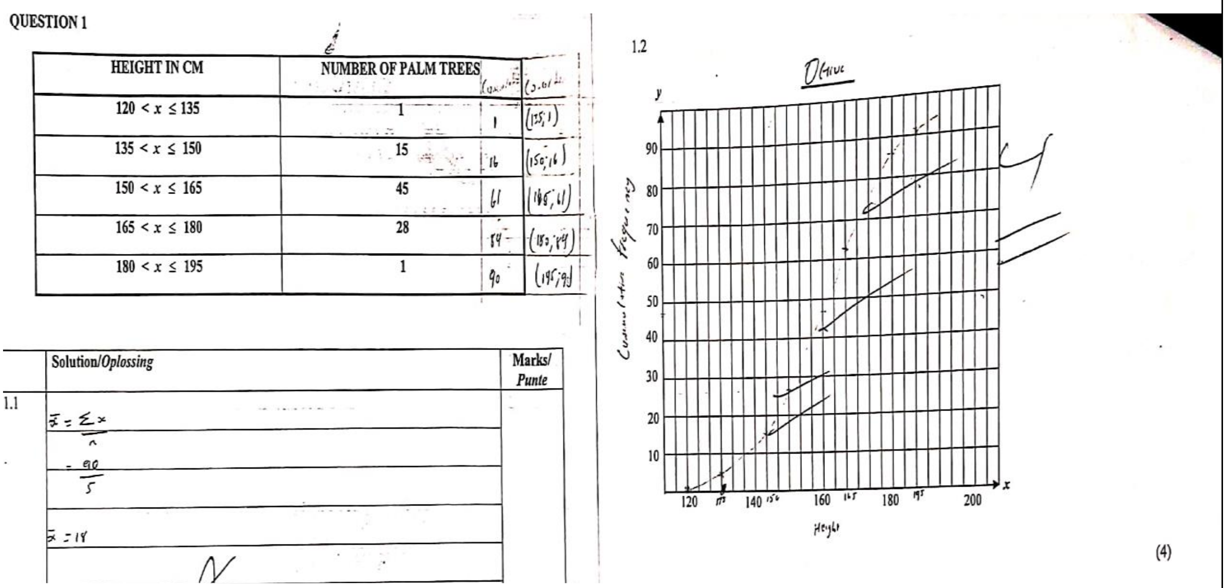


Figure 8.4. (a) Learner SS111's response to Question 1 and (b) learner SS200's response to Question 1.

In Figure 8.4, in (a) learner SS111 added the column for the cumulative frequencies and another column for the mid points. These two columns were very helpful when the learner was responding to Question 1.2 as the learners were supposed to plot the upper limit for the intervals against the cumulative frequencies. However, learner SS111 did not get Question 1.1 correct which required the learner to use the midpoints of the intervals and the frequencies.

In Figure 8.4(b), learner SS200 added the columns for the cumulative frequencies and another column for the coordinates to plot. These two columns enabled the learner to plot the correct graphs in Question 1.2. Learner SS200 did not respond correctly to Question 1.1.

The observation by Teacher 1 showed a lack of conceptual understanding on the part of learners. It shows that learners expected to see the questions asked in a given format or way and when the same question was given in a different format then they were not able to understand and answer it correctly.

Teacher 2 wrote that:

> "Statistics in 2017 was done towards the end of the year as was stated in the Annual Teaching Plan for the Department of Education KZN Province. At the same time learners and teachers treated statistics as easy, they don't take it with the same level of seriousness as they do on other sections of mathematics. Learners practice this section less and they end up confusing the concepts."

According to Teacher 2 the combination of taking the chapter for statistics for granted and putting in less practice resulted in learners struggling in this section. Lack of practice will normally result in misconceptions not being addressed by the teacher, since learners engage less with the concepts and hence consult less with the teacher.

Teacher 4 mentioned the issue of covering statistics when the trial examinations have already commenced. This implied that the teacher was still teaching this section when learners had already started writing trial examinations. Teacher 4 wrote:

> "Some schools do not cover the content according to work schedule, with some teachers covering statistics when trial examinations have already started."

Teachers 2 and 4's comments seem to be supported by the annual teaching plan from the KZN Department of Education. The annual teaching plan for the years 2016 and 2017, showing the dates when statistics is expected to be covered is shown in Figure 8.5.

a. KZN annual teaching plan for mathematics term 3, 2016

TERM 3			
TOPIC	CURRICULUM STATEMENT	SUGGESTED DURATION; START AND FINISH DATES	ACTUAL DURATION; START AND FINISH DATES
FINANCE, GROWTH AND DECAY	1. Solve problems involving present value and future value annuities. 2. Make use of logarithms to calculate the value of n, the time period, in the equations $A = P(1+i)^n$ or $A = P(1-i)^n$. 3. Critically analyse investment and loan options and make informed decisions as to best option(s), including pyramid schemes.	11 days 18 July – 1 Aug.	
COUNTING AND PROBABILITY	1. Revise 1.1 dependent and independent events; 1.2 the product rule for independent events: $P(A \text{ and } B) = P(A) \times P(B)$; 1.3 the sum rule for mutually exclusive events: $P(A \text{ or } B) = P(A) + P(B)$; 1.4 the identity: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$; 1.5 the complementary rule: $P(\text{not } A) = 1 - P(A)$. 1.6 solving of probability problems (where events are not necessarily independent) by using Venn-diagrams, tree diagrams, two-way contingency tables and other techniques. 2. Apply the fundamental counting principle to solve probability problems.	10 days 2 – 17 Aug.	
STATISTICS: REGRESSION AND CORRELATION	1. Revise symmetric and skewed data. 2. Use statistical summaries, scatterplots, regression (in particular the least squares regression line) and correlation to analyse and make meaningful comments on the context associated with given bivariate data, including interpolation, extrapolation and discussions on skewness.	5 days 18 – 24 Aug.	
TRIAL EXAMINATION to cover all the TOPICS dealt with in both Grades 11 and 12.			

b. KZN annual teaching plan for mathematics term 3, 2017

TERM 3							
DATES	TOPIC	CURRICULUM STATEMENT	ASSESSMENT	F/IF	DATE COM-PLETED	HOD: SIGNATURE and DATE	% COM-PLETED
24/07 – 25/07 (2 days)	FINANCE, GROWTH AND DECAY	1. Make use of logarithms to calculate the value of n, the time period, in the equations $A = P(1+i)^n$ or $A = P(1-i)^n$.					77%
26/07 – 07/08 (9 days)	FINANCE, GROWTH AND DECAY	2. Solve problems involving present value and future value annuities. 3. Critically analyse investment and loan options and make informed decisions as to best option(s), including pyramid schemes.					85%
08/08 – 17/08 (7 days)	COUNTING AND PROBABILITY	1. Apply the fundamental counting principle to solve probability problems.					92%
18/07 – 22/08 (3 days)	COUNTING AND PROBABILITY	2. Revise 2.1 dependent and independent events; 2.2 the product rule for independent events: $P(A \text{ and } B) = P(A) \times P(B)$; 2.3 the sum rule for mutually exclusive events: $P(A \text{ or } B) = P(A) + P(B)$; 2.4 the identity: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$; 2.5 the complementary rule: $P(\text{not } A) = 1 - P(A)$. 2.6 solving of probability problems (where events are not necessarily independent) by using Venn-diagrams, tree diagrams, two-way contingency tables and other techniques.	TEST SBA Weighting: 10	F			95%
23/08 – 28/08 (5 days)	STATISTICS: REGRESSION AND CORRELATION	1. Revise symmetric and skewed data. 2. Use statistical summaries, scatterplots, regression (in particular the least squares regression line) and correlation to analyse and make meaningful comments on the context associated with given bivariate data, including interpolation, extrapolation and discussions on skewness.					100%
29/08 – 29/09 (23 days)	REVISION and TRIAL EXAMINATIONS		TRIAL EXAM SBA Weighting: 25	F			
TRIAL EXAMINATION to cover all the TOPICS dealt with in both Grades 11 and 12.							

2017

Figure 8.5 The KZN Department of Education annual teaching plan for mathematics for the years 2016 and 2017 showing when the statistics chapter was covered.

The extracts from the annual teaching plans for mathematics Grade 12 in KZN shows that statistics was planned to be covered in late August for 2016 and 2017, when the trial examinations were due in September in both years. In some schools where teachers will be moving slowly due to a number of factors, the trial examinations will start when they are still covering this chapter of statistics. According to Teacher 4, learners will be focusing on the subjects that they will be writing while other teachers will still be calling them for lessons. Calling learners for lessons where they are taught new information can have negative effects on the learners and may disturb their preparation for trial examinations. Teachers 4 and 2 be believed that the divided attention might be the cause of poor performance in statistics.

Teachers 4, 2, 3, 5 and 6, held the same view that the statistics questions were straight forward and should have been easy for the learners if the work was covered and revised as expected. Without adequate preparation, misconceptions are common as evidenced in the learners' responses.

Teacher 2 wrote:

> "Learners do not practice it with enough attention because they think it is easy. They also confuse the concepts because of lack of practice."

Teacher 4 and Teacher 5 mentioned the issue of other teachers not applying continuous CA marking as disadvantaging learners during the marking of Question 1 which had statistics questions and this made the section look as if it was more challenging than it was supposed to be. The teachers' views on CA marking will be discussed in detail in section 8.2.7.

The Rasch model picked up the statistics questions as some of the most difficult sections as learners were failing to get marks, even though according to the cognitive levels these questions were among the easy ones.

8.2.3 Teacher responses to Question 3

The third question the teachers were asked about the paper was: The results showed the 6 easiest items//questions were dominated by analytical geometry questions (Questions 3.1, 3.2, 3.4 and 3.6) which were level 2 questions on the cognitive levels. Can you give reasons on what you might have done that made learners to perform better in these items?

All of the teachers in this study mentioned the drilling technique in the Analytical Geometry section as the approach they used in preparing their learners for the trial examinations. Drilling means listening to a model, provided by the teacher, or another student, and repeating what is heard. This is a repetition drill, a technique that is still used by many teachers when introducing new language items to their students and also when preparing their students for examinations. Teacher 1 wrote:

> "All these questions are based on the geometry of a line. In order to ensure that the learners perform better in these questions, I put a lot of focus on the straight lines. I drilled my learners a lot on how to handle a line joining points A and B, like:
>
> i. Finding the distance between them.
> ii. The gradient of a line joining them.
> iii. The equation of a line joining them.
> iv. Points of intersection of 2 or more lines.
> v. The values of x and y on both the y- axes and the x-axes.
> vi. The gradients of perpendicular lines."

Teacher 2 also mentioned that he drilled these concepts from Grade 10 with his learners, focusing on the formulas, substitution and simplification. Teacher 2 added the fact that the formulas that were used for the Analytical Geometry questions were provided in the formula sheet and that he emphasised and trained his learners earlier to make use of the formula sheet. Teacher 2 wrote:

> "The concepts asked in these questions have been drilled from Grade 10 and the formulae for solving them are provided in the formula sheet that is attached to the question paper."

The issue of the formula sheet was also mentioned by Teacher 3, who wrote that with the formulas provided, learners were supposed to identify the correct formula and do substitution and simplification to get maximum marks.

For Teacher 3, Questions 3.2 and 3.6 were at the level of Grade 10, with Question 3.2 requiring learners to write the equation of the straight line in standard form and then the gradient of the line. Teacher 3 also mentioned that in Question 3.6 learners were required to substitute into the distance formula after identifying points P and R. Teacher 3 wrote:

"Questions 3.1, 3.2 and 3.6 are easy questions of mere substitution. Question 3.1 is an easy question, where learners required to substitute and to prove that the coordinates (0;2), can be obtained if substituted on that straight line. Question 3.2 is one of the easiest as it is merely a grade 10 question of just writing an equation of a straight line in standard form and then see the gradient. Question 3.6 is also a grade 10 question of substituting in a distance formula by just identifying points. Question 3.4 can be intimidating to learners as most learners still do not like working with simultaneous equations."

The fact that Teacher 3 mentioned simultaneous equations as intimidating to learners and that learners do not like them, means the learners are not motivated enough regarding the aspect of simultaneous equations. The fact that the teacher identified that the learners do not like simultaneous equations is a step towards rectifying the problem. The educator was then supposed to design and arrange instruction that aroused interest and motivated the learners to do more (McDonough, 2007).

Teacher 3's views were supported by Teacher 4 who wrote that most of the Analytical Geometry questions follow from the Grade 10 and 11 work, but added that proper revision and examination preparation led to greater success in this section. Teacher 4 added that the proper use of the information sheet also played a role in the success of learners in analytical geometry as most of the formulas were provided in the formula sheet. Teacher 4 wrote:

"The questions on analytical geometry follow from grade 10 work and generally learners in Grade 12 will be revising the work. The questions are mostly the same with only coordinates, gradients and intercepts values changing. A drilling approach to teaching will yield the desired results for this section since the learners will just have to perform a series of steps. An understanding of the distance formula was worth 6 marks. A thorough knowledge of simultaneous equations was also needed as well as the equation of a straight line."

For Teacher 5, the use of past examination papers helped the learners a lot as the teacher mentioned that most of the subsections in Question 3 were familiar and similar to a number of questions that they were using as preparation from past examination papers. Teacher 5 wrote:

"The questions for analytical geometry were similar to quite a number of questions in the textbooks and past examination papers. The section is also easy for learners to understand. I did questions such as 3.1, where learners were required to calculate the coordinates of Q, where Q is a y-intercept. I have done many y-intercepts. The moment a learner sees it, he or she knows that the x-coordinate is 0. Then use the equation of the line that intercepts the y-axis. As for Question 3.2, I have done so many similar questions which require a learner to rewrite the question in the form $y = mx + c$, where m is the gradient. In Question 3.4, I have taught them that for the coordinates of a point of intersection, simply solve for the equations simultaneously and it was easy for learners to answer the question. For Question 3.6, teaching the learners to look at the formula sheet if one has forgotten the distance formula helped learners to respond correctly to this question. So these are questions which needed recalling of information".

According to Teacher 5, it meant adequate examination preparation that include drilling, lots of practice from past examination papers and the proper use of the information sheet led to the success of learners in responding to questions in this section. Spending more time on activities and past examination papers can help improve a learner's self-confidence.

8.2.4 Teacher responses to Question 4

The fourth question that teachers were asked about the paper was: In general, the analysis showed that many items/questions which were at the cognitive levels 1 and 2, were not always found to be easy by the learners. What are some of the possible causes of this?

Teacher 1, 2 and 3 attributed the challenges faced by learners in responding to level 1 and 2 questions, to the lack of adequate preparation for the examinations. As mentioned earlier, lack of adequate preparation leads to unnecessary mistakes during the examinations and learners displaying many misconceptions as described in section 8.2.

For Teacher 4, lower cognitive level questions are challenging to learners who lack the understanding of concepts (lack conceptual understanding). Teacher 4 further mentioned that some educators use examination-centred approaches to teach their learners, using question papers to teach, meaning teachers go straight into problem solving, without a proper introduction and explanation of concepts and definitions of terms. Teacher 4 also mentioned that learners struggle with level 1 and 2 questions because teachers put more emphasis on higher level questions thereby neglecting lower level questions. This observation means learners struggled with higher cognitive level questions and at the same time the learners will not respond correctly to the lower level questions. Teacher 4 believed that less time was spent on lower level questions. Teacher 4 wrote:

> "Questions with lower cognitive levels may be challenging to learners who lack the basic understanding. This may be due to an exam centred approach to teaching, for example using question papers to teach, where the teachers go straight to solving problems without a proper introduction of the defining terms. For example, a learner may use the gradient formula effectively without understanding gradient as a measure of slope or rate of change. Educators put more emphasis on higher level questions there by neglecting lower level questions."

How educators design and arrange instruction has a great deal to do not only with what is learned, but also with how the learners use what is learned. Exam-centred approaches do not lead to conceptual understanding. There are studies that indicate that deep approach learning is positively related to achievement (Amirali, Huon & Kevin, 2004; Roman, Cuestas & Fenollar, 2008). Several studies have also shown a relationship between approaches to learning and study success (Diseth, 2003; Watters & Watters, 2007). It is therefore without doubt that the approaches that educators use in teaching mathematical concepts have a greater impact on how learners learn new approaches and uses them in new scenarios. For Teacher 4, the exam-centred approach did not work in this case, as the learners struggled to answer level 1 and 2 questions.

For Teacher 5, learners struggled with some level 1 and 2 questions because they did not take time to check their answers properly and correct small mistakes. According to this teacher, learners should thoroughly check their answers, for example when dealing with the centre of a circle, and when completing the square, learners should check the signs. When learners make small mistakes and fail to score marks in level 1 and 2 questions, the Rasch model will pick that up and label these questions as difficult for the learners. Teacher 5 wrote:

> "The learners failed to check their answers properly for any small mistakes, for example, for the regression coefficient, using a calculator, do it twice or thrice, centre of a circle, checking the signs should be done, also completing the square. Where $r^2 = 9$, learners should remember that r=3 and not 9. Some learners would know how to do the calculations but would fail to understand what the question is asking for, for example proving that $P\hat{Q}R = 90°$, they would have to show that $m_1 \times m_2 = -1$, but

fail to understand that 90° means perpendicular lines. Other learners failed to analyse diagrams properly and identify things such as the triangles (Questions 6.3; 7 and 8)".

Teacher 5 also mentioned the case of learners who knew the steps for calculations but were unable to understand what the question was asking for, as in the case of Question 3.3. In Question 3.3, learners were supposed to prove that $P\hat{Q}R = 90°$, but they did not understand that if two lines meet at 90°, then the two lines are perpendicular to each other. This shows lack of conceptual understanding and that the basics were not properly covered by the teacher. Teacher 5's observation also showed the disadvantage of drilling and the exam centred approach to teaching as the learners will struggle to answer the basics. Other learners failed to analyse diagrams properly and could not identify shapes such as triangles, as in Questions 6.3, 7 and 8, a situation that shows lack of conceptual understanding.

8.2.5 Teacher responses to Question 5

The fifth question that the teachers were asked about the paper was: In general, the analysis showed that many items (questions) which were at the cognitive levels 3 and 4 were not found to be always difficult by the learners (for example Questions 3.3, 4.1.1, 4.1.4, 5.3, 6.3 and 9.2). What do you think are some of the reasons for this?

Teachers 1, 2, 3 and 4 believed that when learners spend more time practicing higher order questions, they eventually become easy to the learners. Teacher 1 wrote:

> "The questions were very authentic, they are found in most working materials like past examination papers, revision documents provided by the Department of Education, and the teachers and learners had maximum exposure to these materials."

Regarding the same idea of more exposure to similar questions, Teacher 2 wrote:

> "These questions were easy and familiar to learners who have worked on previous question papers. Most teachers focus on the more challenging (level 3 and 4) questions because they carry more marks and assume that the level 1 and 2 questions will be easy to the learners."

From these two teachers' findings, it is therefore clear that spending more time on the level 3 and 4 questions will help learners in clearing up some of their misconceptions, and give them a better understanding of the challenging concepts. The amount of time and effort that a learner can put into achieving goals is important in learning and has significant impacts on how students engage in learning (Reid, Duval & Evans, 2007). Self-doubts can easily creep in when learners hit obstacles, experience frustration and judge themselves as incompetent, so when teachers spent more time with their learners and practising the level 3 and 4 questions, these obstacles can be addressed easily and help learners gain more self-confidence.

To the same effect Teacher 4 wrote:

> "Educators put more emphasis on these type of questions and learners are more familiar with what is required of them. More time was spent on the challenging questions. Workshops were done to address the difficult sections hence the results showed good performance."

For Teacher 6, these are the types of questions that teachers drilled their learners on. Teacher 6 went on to say that level 3 and 4 questions that have been set or seen before, will no-longer be seen as level 3 and 4, but becomes more routine problems. Teacher 6 wrote:

> "These are the questions that the educators spent more time on and made these questions routine. More over level 3 and 4 questions that have been set or seen before no longer becomes level 3 and 4."

However, spending more time on the level 3 and 4 questions does not automatically make these questions easier for the learners, but must involve a lot of planning on the part of the teacher. To Romberg (1983), educators need to devise instructional strategies and methods that provide the learners with learning situations where they can develop and apply high-order operations which are critical for mathematical development.

Teacher 1 added that the correct use of the formula sheet as one of the reasons why the learners found these questions relatively easy. To this effect, Teacher 1 wrote:

> "All the learners needed to do were to get the right formula, substitute and then use basic algebra, making it easy to get the right answers."

These comment from Teacher 1 implies that teachers need to over-emphasise the correct use of the formula sheet as it is central to a positive response in most of the questions in mathematics.

8.2.6 Teacher responses to Question 6

The sixth question that the teachers were asked about the paper was: There were many instances where marking was not consistent with the marking guideline, for example, very few educators in their marking were using continuous accuracy (CA) marking, hence depriving learners from getting potential marks (Refer to Question 2.2.2 for example). Can you give your views on this issue and why teachers sometimes do not adhere to this aspect?

CA marking is a very important aspect in the assessment process as indicated by the teachers' responses, and should be part of every teacher's marking process for fairness and to avoid disadvantaging some learners. Emotions should not cloud teachers' judgements and influence how they mark learners' scripts. To this effect Teacher 1 wrote:

> "If you apply CA marking in Question 2 for example, the entire question is out of 8 marks, and if a learner gets 2.1 wrong, they are still more likely to get the remaining 7 marks because of continuous accuracy marking. Many teachers use emotions when marking, they assume that since they have taught the concepts in class and the questions are easy, the learners have no excuse in getting wrong solutions. They get annoyed that they spent so much time on the topic and then learners get the answers wrong, they deliberately penalize the learners".

For Teacher 1, teachers should not use emotions in the marking process; the marking guideline should be followed in applying CA marking. Teachers who do not apply CA marking disadvantage their own learners, as they deprive them of potential marks and in the end their learners will have low school-based assessment (SBA) marks. In some instances, a learner would have made a genuine mistake in a part of a question, but they will rectify that mistake in the questions which follow. A teacher who does not apply CA marking will then penalise the learner for all of the questions.

For Teacher 2, the failure by some teachers to use CA marking is due to their lack of knowledge of how to use it. Teacher 2 wrote:

> "The concept of continuous accuracy marking is not well understood by all the teachers, especially those that do not have experience of marking NSC examinations. Teachers need to be trained on how to apply continuous accuracy marking correctly."

Teacher 2's comment implies that there is large group of teachers who might not be using CA marking correctly. These could be teachers who might have been in the teaching field for many years but have never been appointed as NSC markers and teachers who have recently joined the teaching field. These groups of teachers need to be trained on how to apply CA marking, which at the moment is mainly trained or compulsorily applied at the NSC marking centres.

For Teacher 4, not using CA marking is not a good practice and may give a negative impression about learners' achievements. Teacher 4 wrote:

> "Not using CA marking is unfair to learners, and gives unreliable evidence of the learners' achievements. Some educators do not apply CA marking because of the huge number of scripts they have to mark in less time. In the end they look at the model answer and give marks or penalize the learners. It may also be not possible to award marks if some steps are missing.
>
> Wrong answers require the marker to make sense of the learners' responses in a short period of time. Some educators may also not give marks during tests and trial examinations as a way of encouraging heir learners to be more accurate in preparation for the final examinations."

Teacher 4 raised three issues which may influence teachers in their use of CA marking. Not using CA marking will lead to unreliable results in terms of the learners' achievements, as many learners will generally score very low marks in questions where mistakes and errors were common. A mistake in the first part of a question will result in the teacher overlooking the responses in the other parts of the question. It is therefore assumed that the learner is unable to solve all of the aspects in the question, which will be not the case, as the learner will only have challenges with the first aspect.

The second aspect raised by Teacher 4 regarding CA marking was the large number of scripts some teachers have to mark in a short period of time. When trial examinations are written, the provincial Department of Education normally sets dates for completion of marking after which teachers must submit the marks to the departmental offices. Some schools have a very large number of learners with few teachers. CA marking is time consuming as teachers are expected to check each and every response that different learners give, to see if the answers correspond to those given in the first part of the question. This will require more time than is usually allocated. In the end, teachers will ignore CA marking and only consider the model answers as they try to meet the targeted time.

The third issue mentioned by Teacher 4 is the idea of teaching learners to be accurate in preparation for the final examinations. Teacher 4 believes that some teachers will not give CA marks during tests and trial examinations as a way of teaching and encouraging their learners to be accurate when solving problems. The idea might be good but this disadvantage the learners as they will end up having low SBA marks as mentioned by Teacher 1.

For Teacher 5, the learners' line of thinking should be considered when marking. Teacher 5 wrote:

> "Learners should be given marks sometimes based on their line of thinking instead of CA marking. For example, in Question 2.2.2, if a learner might have included the outlier in calculating the correlation coefficient, I feel that the learner is supposed to be given part marks, even if r is wrong but the justification is correct, marks should be awarded to such learners as they are showing that they understand the meaning of the coefficient correlation. Also if the answer which is supposed to be used in the question that follows is wrong, but the learner used it and the steps are correct, learners should be awarded marks for understanding the steps."

Teacher 5's views are almost the same as those of Teacher 4, with the exception that Teacher 5 feels that learners deserve marks for having an idea about the concept in question (line of thinking). However, Teacher 5's suggestions require even more time than CA marking as they require more scrutiny from the teachers, which is very difficult when schools have large number of learners.

8.2.7 Teacher responses to Question 7

The seventh question the teachers were asked about the paper was: In case where an item carries 4 marks or has 5 categories or more, and all the marks are accuracy marks, for example in Question 4.1.4, results showed that the majority of learners will either get zero or get all marks. Very few learners got marks in between. Do you think this is a fair practice or unfair situation? Please explain.

All of the teachers who participated in the study indicated that allocating 4 accuracy marks or more on a question is an unfair practice to the learners, as they will either get all of the marks or get none of them. Allocating too many accuracy marks will result in the question not differentiating between those who partially know the answer and those who do not know how to solve the problem at all. Teacher 1 wrote:

> "The situation is unfair due to the fact that for the coordinates of Q, the value of b was not given, that means learners had to find it on their own. Chances are that some learners may make some mistakes and come up with different values of b. The different values of b will result in different values of $tan\theta$, which has a direct impact on proving that $tan^2 \propto + tan^2\theta = 1$. Therefore, CA marking was supposed to be considered in this question because a learner must be examined on their mathematical understanding not on whether they are able to get the examiners' answers right."

For Teacher 1, when a question involves a number of steps, then mark allocation should be on a CA basis. This will cover for mistakes and errors which may occur in different steps, and thus learners will get marks in different categories depending on proficiency. For teacher 2, accuracy marks for a question that carries many marks is an unfair practice especially for learners who are able to do only a few steps but fail to get the final answer correct, as these learners will get no mark instead of getting part marks. Teachers 1 and 2's comments are supported by Teacher 3, who wrote:

> "This is very frustrating as those who are not getting all marks, will get zero. There are no marks in between the full marks and zero."

Teacher 3 implies that a learner who is able to do a number of the correct steps but fails to get the final answer will not be rewarded for the effort, and those who make a mistake in the earlier stages, but continued correctly, will also not be rewarded for the effort. This will be frustrating for both the teacher and the learners.

Teacher 4 also believes that allocating many accuracy marks on a question is an unfair practice. To this effect Teacher 4 wrote:

> "It is an unfair practice. The learner may never get to be tested for a concept on a subsequent question because of failing the first part of the question."

Teacher 4's comments means that if a question has accuracy marks and has a number of concepts to be tested which comes one after the other, a leaner will then be penalised for a mistake in one of the concepts that comes earlier in the question, and the rest of the concepts will not be considered because of the mistake in the first one. This was also supported by Teacher 5, who wrote:

> "I think this is an unfair situation. Accuracy marks always deprive learners of crucial marks. Learners can easily make slight mistakes on parts of the question but know how to do the calculations for the rest of the question. Some parts of the question would require some kind of thinking, which might be worth some marks, but because of accuracy marking these marks will not be rewarded."

From these findings, it is clear that the allocation of marks should not be based on the final correct answer only. The steps carried out before the final answer should be considered as well as the calculations after the first mistake. Accuracy marking does not differentiate between learners of different abilities as learners either get all of the marks or no mark at all.

8.2.8 Teacher responses to Question 8

The eighth question the teachers were asked about the paper was: The Rasch analysis showed that in some questions, learners who had the same total scores, from either the same school or from different schools did not perform in the same way on most of the items. For example, a learner from School SS did better on a particular item than those from School AD, while for some items it was the other way round. What are some of the contributing factors for this observation?

A number of reasons are attributed by the teachers who took part in this study to the different performances by learners from different schools. These included different methodologies, teacher strength (teachers' PCK) and preferences and different learner abilities. Teacher 1 wrote:

> "Teachers teaching the subject probably do not meet and discuss topics in preparation for lesson delivery in their respective schools. They do not share information as teachers resulting in them teaching learners differently. Either the subject teachers themselves or with the help of subject advisors need to do topic discussions together to avoid all the misconceptions resulting in teaching the same concept differently."

Teacher 1's views can be linked to different performances both at school level and between different schools. When teachers in the same schools do not plan lessons together and decide on what to teach together, different methodologies will then be used in the same school for the same concept. The result is that learners in a certain class will perform better in a certain aspect but fail to do well in another aspect, in which a different class might be excelling. When ideas and methodologies are shared only the best methods can be adopted for the best results, even though with different learners it may mean different methodologies will work for the different groups.

For Teacher 2, different teachers have different areas of strength and this can be directly transferred to learner performances. Teacher 2 wrote:

> "Different people; teachers and learners, have different sections as their areas of strength and weaknesses. Different teachers are strong in certain sections or chapters than others and this has a direct impact on how they teach learners."

Teacher 2's observation implies that teachers will teach more efficiently on chapters covering aspects that they are good at and tend to spend more time on those than the chapters covering aspects that they are not so good at. This might have a direct influence on learner performance, as the learners will also do better in those chapters which were taught well, than in the chapters where the teacher was not so well equipped. Such challenges can be addressed by Teacher 1's suggestion that teachers should have regular meetings and workshops to discuss content and methodology, as a way of bridging the gaps in content and methodology. Teacher 2's suggestion is supported by Teacher 3 who wrote:

> "I think it is the strength of their teacher's content knowledge on some topics or which areas or topics were emphasised the most by the teachers, these are the areas where the learners will tend to do better."

For Teacher 2, the subject teacher may teach all of the topics but put more emphasis on certain topics. This may result in different performances between learners in different classes or schools. The same points were also raised by Teacher 4, who thinks some teachers may even avoid certain chapters on areas in which they are not strong. Teacher 4 wrote:

> "This is the input of particular teachers' teaching strategies, strength and weaknesses. Some educators are stronger in some sections and it shows with their learners' understanding. Some educators even avoid some topics which they are not good at. Due to time, some schools may have more engagements for a particular section than others. Some schools may have some intervention programmes while others do not."

Teacher 4 raised another important point of school intervention programmes. From my experience as a mathematics teacher, these intervention programmes may involve one school or different schools coming together and revising in one centre with their subject teachers or subject specialist or tutors who may come to assist. For Teacher 4, these engagements and intervention programmes may improve learners' performances in certain topics in certain schools.

For Teacher 5, different approaches by different teachers will result in different performances and to this effect the teacher wrote:

> "This mainly is a result of teachers' approach to these topics. A teacher at a particular school emphasises or drills learners on questions that they think will be easy for their learners to get marks especially if the teacher realises that the learners are struggling to grasp the concepts. A teacher might decide not to spend more time on such sections even if the teachers themselves are excellent on those concepts. Some teachers may face challenges with certain topics (lack content depth) and concentrate on what they know better. On the side of learners, if they feel that questions from certain topics are difficult, they avoid practising them much. Some teachers explain certain topics better than others, and these will influence directly the learners' performances."

Teacher 5 raised a number of issues mentioned by the other teachers above, but added the learner effect. The teacher mentioned that if the learners in their engagement with the chapters discover that they have more challenges in certain chapters, they tend to spent less time on those chapters and prefer to spend more time on chapters that they are more successful with.

This will show when it comes to examinations as they will perform better on those chapters that they engaged with more.

For Teacher 6, the different performance between schools is a result of the markers' different levels of experience and understanding of mathematics. This teacher's comments can be related back to the issue of some teachers applying CA marking while others did not. Schools with teachers who apply CA marking are likely to get higher marks than those schools where the teachers do not apply CA marking.

8.2.9 Teacher responses to Question 9

The ninth question that the teachers were asked about the paper was: Rasch analysis indicated items where learners who used English as the first home language performed better than learners who use English as a second language (Questions 2.1, 2.2.1, 2.2.3, 3.5 and 6.1.1). However, a different scenario was observed for Questions 4.1.1, 5.1 and 10.1 where the learners who use English as a second language performed better than the English first language speakers. What are some of the factors contributing to this finding?

The issue of language in education is a very broad subject which has been debated often in education circles. It was also touched on earlier, when differential item functioning (DIF) was considered for language. All of the teachers who took part in this study agreed that the language issue impacts negatively on learner performance. Teacher 1 wrote:

> "This could be based on the language barrier. The questions 2.1; 2.2.3; 3.5 and 6.1.1 consists of some critical words that determine the overall understanding of the questions, words like outlier, exclude and expansion. For English second language speakers, the challenge could be failing to interpret the question or understanding what is required of them. With questions 4.1.1; 5.1 and 10.1, the questions are proof type questions. The questions are procedural, book work and generally involve reproducing what the learner practised, with little or no need for interpretation or understanding of concepts."

For Teacher 1, where the questions required understanding, analysing and interpreting the question, English home language speakers performed much better than English second language speakers. Teacher 1 identified terms or words like outlier, exclude and expansion which are not commonly used which might have made it difficult for the English second language speakers to understand.

Teacher 2 agreed with Teacher 1 that the first group of questions was more about conception and interpretation, and in these questions the English first language speakers did better than the English second language speakers. In contrast, the second group of questions were more procedural, and success could be achieved by reproducing information from the textbooks.

Teacher 3 also wrote:

> "Even though maths has a language of its own, interpretation of questions is the key. Learners who better understand English or have teachers who mostly teach in English have an advantage. In our teaching, language is a barrier, learners who speak French in our school, also suffer understanding concepts and it's difficult to give clarity when a learner cannot hear or understand what you are saying or explaining."

For Teacher 5, when a question had terms which are commonly used in daily life, learners tended to understand better than when unfamiliar words or terms were used. Teacher 5 wrote:

> "In most cases where there is an English word that is not commonly used, for example 'excluded' in Question 2.2.1, learners who are not English first language speakers will not do or perform well. English is also a contributing factor to a certain extend. To understand a question is to understand it in English. Some of my learners cannot read properly questions in English and they do not perform well. But if the question is explained to them in their mother tongue, they quickly understand what they are supposed to do. Question 2.2.3 required conceptual understanding and can be difficult for some learners to understand words such as 'expand' in Question 6.1.1."

Teacher 5's comments were supported by Teacher 6, who also believed that for questions which required more understanding and interpretation, English first language speakers did better while English second language speakers did better on questions which were more procedural.

CHAPTER 9 SUMMARY, CONCLUSION, RECOMMENDATIONS AND LIMITATIONS

9.1 Introduction

This chapter presents a discussion of primary findings, conclusions, limitations and recommendations based on the results of the research. The chapter provides answers to the research questions, makes connections between the findings in this study and existing literature, and offers concluding remarks. The first research question is about the trends in performance as shown by the analysis of the overall results initially, as reported in detail in Chapter 4, before any Rasch analysis was conducted. This initial analysis was done to look at performance in the different sections of the Mathematics Trial Examination Paper 2 and to look at how the participating schools did, in order to identify broad areas of concern. The second Research question then delves further into the targeting and functioning of items, by using Rasch analysis methods as covered in Chapter 5. It also covers the results of the initial Rasch analysis and a discussion of what can be inferred about whether the test was well targeted or not. The second research question is also about identifying those items which were mis-fitting according to the fit statistics. The next stage (which relates to Research Question 3) was to carry out rescoring, DIF analysis, dimensionality analysis and item splitting to see if the item fit could be improved post hoc. Chapters 6 and 7 provided details of what was done in respect of these procedures; results of the final analysis after these procedures are covered in the answers to the three sub-questions of Research Question 3. Research Question 4 entailed taking these results to teachers to get their opinions about the diagnosis of the examination paper according to Rasch analysis, and how the functioning of the examination paper could be improved in future. This fourth Research Question was covered in Chapter 8.

To reiterate, the research questions of the study were as follows:

- What are some trends in performance of learners in the Grade 12 mathematics assessment instrument?
- What does a Rasch analysis reveal about the targeting and functioning of the assessment instrument as a whole?
- How can the Rasch analysis be used to improve the functioning of the mathematics assessment instrument?

- To what extent are the items functioning as expected?
  - How can the use of Rasch analysis contribute to improvement of the scoring rubric?
  - To what extent do the items display DIF, multidimensionality and item dependency?
- What are the teachers' views about some of the findings of the Rasch analysis?

Accordingly, this final chapter focuses on the following:

- A discussion of the primary research findings and the conclusions drawn from the findings, in terms of providing answers to the research questions.
- Recommendations which were derived from this study and for further research.
- Limitations that reflect the shortcomings of this study.
- A brief conclusion that summarises the study as a whole.

9.2 Research findings

In order to provide a logical sequence to this section, I have aligned the headings with the critical findings that emerged from the major research questions. This approach also provides a basis for the discussion of the conclusions drawn from the findings from both the literature review and the empirical study conducted.

9.2.1 What are some trends in performance of learners in the Grade 12 mathematics assessment instrument?

This first research question focused on the trends in performance of the learners in the Grade 12 mathematics assessment instrument.

9.2.1.1 The overall results for the examination were very poor

The results from Chapter 4 showed that overall the learners performed very poorly in this examination paper, with an average percentage for all of the learners of 27.8%, which is below the pass rate of 30%. The low average is an indication that the paper was challenging for most learners. There were three schools whose average was above 30%: school SS -39.8%, school KS - 35,7%), and school AD - 36%. These poor results are consistent with reports from other studies. The NSC diagnostic reports (DoBE, 2020, 2019) show that in the years 2016 - 2020, less than 36% of learners were able to score over 40% in the NSC mathematics examinations. Many studies have lamented these poor national results as well as the poor performance in international assessments, and there seems to be little improvement over the years (DoBE, 2012, 2019, 2020; Fleisch, 2008; Reddy, 2006; Bansilal et al., 2014)

9.2.1.2 The results varied across the schools

The results from Chapter 4 show that the schools did not all perform in the same way in the mathematics trial paper. Three of the schools (AD, SS and KS) performed better than the other three schools (SB, LM and FT) in all sections of Trial Paper 2. The top-performing schools had overall averages above the pass mark of 30% (school AD - 36%, school SS - 39.8 % and school KS - 35.7%), while the bottom three schools had overall average percentage marks far below the pass mark (school SB - 20.8%, school LM - 14% and school FT - 18.1%). Of the three top-performing schools, two were quintile 4 schools, one being a boarding school (school AD) and the other a rural school which was well resourced (school SS). The third top-performing school was a quintile 5 school and very well resourced (school KS). Of the bottom three top-performing schools, one was a quintile 4 school (school LM), located in a high density suburb and not well resourced, while another other was a quintile 2 school (school SB) located in a semi-urban area and poorly resourced, while the third (school FF) was an independent school located in the central business district and well resourced. The three top-performing schools were all well-resourced and as revealed in literature were expected to perform well (Reddy et al, 2016). However the same cannot be said about school FF, which was an independent school and well-resourced, but did not perform well, a finding that requires further research.

Various studies have shown that learners in different school environments perform differently which may be because of the different quality of education at the schools (Hoadley, 2012; Carter, 2010; Wood et al., 2011). It may be that different approaches and methods used by teachers lead to different results in different schools, while availability of resources and overcrowding also contribute to different performances (Mji & Makgato, 2006; O'Connor & Geiger, 2009). For example, a study by Mulkeen (2006) revealed that learners in rural schools receive low quality education because of there being few qualified educators teaching there and limited resources when compared to schools in town.

9.2.1.3 Learners in some schools may not have covered all the content

The poor results in schools LM and FT were especially apparent in the analysis of zeros and blank spaces presented in Table 4.7. In school FT, across all the items in the test, more than half of the learners consistently obtained zeroes or had blanks except for items 3.1, 5.1, 6.2 and 7. For school LM, more than half of the learners obtained zeroes or blanks for all items except 3.1, 5.1 and 7. The graphs showing this disturbing trend were presented in Figures 4.16 and 4.17. One example is in school LM, where 96% of the learners either left blank spaces or got zeroes for Questions 1.1 and 1.3.2 while 98% left blanks or got zeros for Question 1.3.1. Similarly, in school FT, 99% of the learners did not get any marks for Questions 1.1 and 1.3.1 and 98% did not get any marks for Question1.3.2. These low rates of participation suggest that the learners were not given sufficient opportunities to engage within the content which involved working with grouped data and the ogive curve to find the mean and interquartile range. Research focusing on the curriculum have revealed very low curriculum coverage rates at schools in South Africa, and studies point to the low content coverage as a reason for the persistently poor learner outcomes in mathematics (Mkhwanazi, Ndlovu, Ngema & Bansilal, 2018; Bansilal, Zondi & Shabalala, 2016; Stols, 2013; Taylor 2011).

9.2.1.4 Performance in Statistics

This study showed that learners performed worst in statistics, with an overall average percentage mark of 21.9% for this section. Only in two schools did learners get overall average marks above 30% (school KS-33.1% and school SS-36.0%) while the other schools had average scores below 30%. The NSC diagnostic reports show that learners' performance in Statistics has been poor ever since it was introduced to schools' mathematics (DoBE, 2012 2017, 2018, 2020), compared to traditional concepts such as algebra, or calculus which has been in the curricula for many decades. North, Gal and Zewotir (2014) noted that the scope of statistics taught at school level was limited before the introduction of Curriculum 2005. This means that mathematics teachers trained prior to this date, had little or no training in statistics (North et al., 2014).This may account for the poor performance since teachers are not as confident about teaching statistics compared to teaching other traditional mathematics concepts (Umugiraneza, Bansilal & North 2017).

Some concepts seem to have been more challenging than others. Questions based on the ogive curve were handled very poorly. One of the questions in statistics with the worst performance was Question 1.1, with an average percentage of 7.6%, where learners were required to determine the estimated mean from grouped data. The average percentages were also very low for questions where learners were required to calculate the interquartile range as well as the estimated median from the ogive curve. These questions require conceptual understanding of the ogive curve and it seems that learners struggle with interpreting the ogive curve. Some studies have found similar challenges with determining measures of central tendency and measures of dispersion when given different types of data representations such as ogives, box and whisker diagrams and histograms (Edwards, Ogun-Koca & Barr, 2017). Diagnostic reports for the NSC examinations indicate that learners often use frequencies instead of cumulative frequencies when plotting the ogive, and use the lower limit of the class interval instead of the upper limit, while some learners used the midpoint of the class intervals (DoBE, 2020).

The concept of the correlation coefficient, emerged as an area of difficulty where the average mark for Question 2.2.2 was 21.9%. Noting that the correlation coefficient is only taught in term 3 of the Grade 12 year, these poor results suggest that the learners have not had sufficient time to master the concept. The low average percentage observed for Question 2.2.3 (15.6%) also showed that the learners struggled to read off the y coordinates of the scatter plot, which were going to be used to determine the average fuel consumption. This contributes to the findings from other studies that learners struggled with reading off data given in different representations by identifying a specific difficulty associated with understanding a graphical representation (Edwards et al., 2017; Cooper & Shore, 2008; DoBE, 2018).

9.2.1.5 Performance in Analytical Geometry

The results from this study showed that learners did best in Analytical Geometry where the overall average percentage for all learners was 33.8%, a value above the pass mark of 30%. Although the learners did better in Analytical Geometry as a whole, they did not do well in the questions based on showing that two circles touch internally and finding the common tangent of the two circles (Questions 4.2.1 and 4.2.2) where the overall average percentage marks were 17.5% and 3.1% respectively. In fact, for Question 4.2.2, all of the learners from three schools

(SB, LM and FT) got zeros, showing that they had not encountered this type of problem previously. Question 4.1.4 required learners to show that $tan^2\alpha + tan^2\theta = 2$, which required substitution, but the average score for this question was 20.9%. Question 4.1.4 and 4.2.1 were more proof type questions as they required learners to show a given answer. The study by Naidoo and Kapofu (2020) further revealed that learners viewed Analytical Geometry as easy as it required the simple application of rules and formulae, but when the geometry activities required them to provide reasons and justifications for their solutions, it became a problem. This explains why the learners did not do well in Questions 4.1.4 and 4.2.1, where they were supposed apply some reasoning and justification.

9.2.1.6 Performance in Geometry

Learners obtained an overall average percentage mark of 24.5% for the Geometry section which is an indication that they found it hard to respond to Euclidean Geometry questions. Table 9.1 gives a summary of the poorly answered questions and the concepts they covered.

Table 9.1

Questions where learners performed poorly in Euclidean Geometry

Question	Concept	Average %
8.2	Prove that $\triangle WVZ$ is isosceles	16.2%
8.4	Calculate VO, if $X = 16$ units, $Y = 12$ units and the radius of the circle is 10 units.	15%
10.1	Prove that $ET = BF$	10.4%
10.2.3	Prove that $BD^2 = DE.AD.$	8%
10.2.4	Deduce that $ET = BF$	3.8%

It is clear from Table 9.1 that many of the questions where the learners scored very low marks were proof and deduction type questions. The study revealed a lack of conceptual understanding of the proofs, not showing any knowledge of what is required by the question, because there were many blank spaces left in learners' scripts. For those learners who understood what was required by the proof type questions, many of them were not able to give reasons to justify their answers. The study by Naidoo and Kapofu (2020) also found that learners found Geometry challenging especially when they had to remember, link and carry out proofs of different statements and theorems. In their study Ngirishi and Bansilal (2019), also found that Grades 10 and 11 learners struggled with proof type questions in Geometry.

Many studies have found that FET learners in South Africa find geometry extremely difficult and lack conceptual understanding, which is evident when they write examinations (Oberdorf & Taylor-Cox, 1999; Bowie, 2009; Roux, 2003; Van der Sandt, 2007). The literature review section discussed some of the challenges, errors and misconceptions when learners work with Euclidean geometry and these include challenges with determining angles and giving reasons to justify answers and proof type questions (Naidoo & Kapofu, 2020; Luneta, 2015; Ngirishi & Bansilal, 2019; DoBE, 2020, 2018). The literature review revealed that learners have challenges in proving cyclic quadrilaterals, parallel lines, confusing theorems and their converse theorems (DoBE, 2018, 2019, 2020). In many cases learners assume certain conclusions or deductions which have not been provided on diagrams and have not been proved, for example making a conclusion that a line is a diameter, or that lines are parallel (DoBE, 2020). Literature also showed that learners struggle with the ratio of sides of triangles, which they confuse with the actual length of the sides of the triangles (DoBE, 2018). The results from this study corroborate these earlier findings that learners' performance in geometry was poor. The analysis of results in Chapter 4, showed that learners did not do well in questions which required them to provide proofs and to make deductions (In Question 8.2, which required learners to prove that the given triangle was an isosceles triangle, in Question 10.1, which required learners to prove that the corresponding sides of two equiangular triangles were in the same proportion, in Question 10.2.4, which required learners to deduce that two sides are equal). This study showed that learners struggled to use similar triangles and proportionality theorem to determine unknown sides on triangles (Questions 8.4 and 9.2). Evidence of lack of conceptual understanding in geometry in this study was revealed in the analysis of Questions 10.2.3 and 10.2.4, where few learners who attempted to answer these questions could not link the correct corresponding angles, while in other cases wrong reasons were cited. The majority of the learners did not attempt to answer these two questions.

9.2.1.7 Performance in Trigonometry

Learners did not perform well in the trigonometry section, with an average score of 29.5%, which seems to support other studies which found that trigonometry is one of the topics in mathematics that learners find difficult (Atagana, et al, 2009; Siyepu, 2015; Demir, Sutton-Brown & Cermick, 2012). Many learners struggled with basic trigonometry in Question 5.2, where they failed to write cosθ in terms of n. Those who managed to write an expression for cosθ, struggled to simplify the resulting expression using algebra. Learners displayed lack of conceptual understanding when asked to calculate the maximum area of a triangle that was inside a semi-circle in Question 5.4 (average score was 10.7%). The learners were not able to apply the fact that for maximum area to be obtained then the size of one of the angles was supposed to be 90 degrees. An earlier study revealed that learners found it difficult to identify the right concepts and apply them with a high degree of accuracy (Yulandari, 2012), which is similar to the learners' struggles here.

The study also showed that learners had challenges when solving trigonometric equations which require the use of compound angles. Learners displayed a lack of conceptual understanding as they used the wrong concepts of changing co-ratios in Question 6.1.2, where they could have simply used compound angles and added like terms. Similar errors were reported in the diagnostic report (DoBE, 2017) where, when asked to solve trigonometric equations with compound angles, learners opted to use compound angle formulae which made the problem far more complicated than it was. The low average score (13.5%) for Question 6.1.2 was further evidence that the learners struggled with the trigonometric equations. Solving trigonometric equations seem to be a challenge for many learners according to various studies (Chigonga, 2016; Rohimah & Prabawanto, 2019). Learners struggle to solve trigonometry equations that require solutions in a given interval, because of misconceptions about periodicity and have challenges dealing with negative angles as they fail to identify the relevant quadrants and make valid inferences, while other learners struggle to factorise trigonometric expressions and applying the general solutions (Chigonga, 2016; Rohimah & Prabawanto, 2019).

The study showed that learners struggled with solving for triangles, Question 6.3, where they were required to use the sine rule, cosine rule and trigonometric ratios. Learners used the wrong formulae, both for the right angled triangles and non-right angled triangles, an indication of lack of conceptual understanding. Many learners were not able to use reduction formula in simplifying $sin(180^{\circ} - (x + Y))$. The findings of the study are in line with what was mentioned in the literature review that learners struggled with the reduction formula, especially with the signs of the reduced trigonometric ratios, while others have difficulties in seeing triangles in different three-dimensional shapes, where the answer requires learners to link two triangles in different planes (DoBE, 2016, 2017, 2018). Some reports have further revealed that learners were not able to identify the correct rule to use when working with non-right angled triangles, while others displayed poor algebraic manipulation skills (DoBE, 2017; DoBE, 2018).

9.2.2 What does Rasch analysis reveal about the targeting and functioning of the instrument as a whole?

9.2.2.1 Representing the person proficiency and item difficulty on a common scale

It was noted in section 3.8.5 that a key feature of the Rasch model is that the difficulty of items is located on the same scale as the ability of the persons attempting those items as. The focus of the model is on the interaction between a person and an item and is based on the probability that a person v with an ability βv will answer correctly, or partially correctly, an item i of difficulty δi. Recall that the equation that relates the ability of learners and the difficulty of items is given by the logistic function: $P\{X_{vi} = x\} = \frac{e^{x(\beta_v - \delta_i)}}{1+e^{x(\beta_v - \delta_i\ )}}$. Applying this equation, we can see that if a person v is placed at the same location on the scale as an item i, then βv = δi, that is, βv − δi = 0, and the probability is thus equal to 0.5. Thus from the model, any person has a 50% chance of achieving a correct response to an item whose difficulty level is at the same location as the person's ability level. Similarly, if an item's difficulty is above a person's ability location, then from the equation, the person has a less than 50% chance of obtaining a correct response on that item, while for an item with difficulty level below that of the person's ability the person would have a greater than 50% chance of producing the correct response. This equation allows us to compare the probabilities of persons of different ability being able to get an item of certain difficulty correct and also allows us to compare the probability of a person getting items of different difficulty, correct.

This feature of representing both person ability and item difficulty allows us to see how well the test is targeted as a whole, to the particular cohort. The person -item map (see Figure 5.3) can give us more nuanced information about particular students who are struggling and particular items which were too easy or too difficult. A test where the mean of the persons' location is close to 0, to which the mean of the items has been set, is one that is well targeted (Tenant & Conaghan, 2007)

9.2.2.2 Interpretation of the overall fit statistics

The Rasch analysis showed that the person mean location was -0.4746, which was below the item mean of zero, indicating that generally the learners found the Mathematics Trial Examination Paper 2 a bit difficult. In terms of the distribution of the person locations, the standard deviation of person location was 0.6628 which is below 1, which implies that the person locations were not as spread out as they could be. This implies that there was not much variation between the locations of the learners in terms of the total scores, showing that the items did not have a good diagnostic ability in making a fine-grained distinction between a wide range of ability (Modzuka, Long & Machaba, 2019).

Matters (2009) points out that assessment involves making inferences about student achievement based on the evidence that is provided by the instrument. If the test is not reliable, then the inferences that are made cannot be trusted. For example, if one sets a test made up of items that learners have seen before, and everybody gets all items correct, then there is not much that can be inferred. One cannot infer that all the learners know the content equally well because the test did not work well in distinguishing between the proficiencies of the learners in mathematics. In RMT we use the person separation index as a measure of the reliability of a test. It is similar to the Cronbach's alpha statistic and provides an indication of how well the test works in distinguishing between the proficiency of the learners. The greater the separation index, the greater the spread of persons relative to the standard errors (Andrich & Marais, 2019). The Rasch analysis showed a very good person separation index of 0.88714, which indicated that the estimation of the person's ability was consistent across the model, as the value was above 0.85 as advised by Tennant and Conaghan (2007). The learners were separated well by the test. The hierarchical ordering of the items varied across the trait as the item trait interaction figures had a chi-square value of 801.8917 and a probability of 0.0000. The chi-square probability value of 0. 0000 is less than 0.05 (or a Benferroni-adjusted value). With the Rasch model the item chi-square statistic relates to the ordering of the items, and the corresponding probability statistic needs to be greater than 0.05, so that the null hypothesis can be accepted (rather than rejected). Hence this value of p, suggests that the hierarchical ordering of the items may vary across the trait.

The analysis also revealed that there were 110 learners almost 20% of the cohort of 604 learners involved in this study, who were placed at proficiency levels below the difficulty level of 39 of the 40 items in the study. This shows why there were so many scores of zeroes and blank spaces during statistical analysis in Chapter 4, as 98% of the items were beyond the reach of 18.2% of the learners.

The item locations ranged from -1.848 logits to 1.637 logits while the person locations were estimated between -3.082 logits and 1.862 logits (with a mean of -0.475), with 4 learners from schools FT and LM obtaining extreme scores, while 13 learners from these schools were located at below - 2 logits. This may be explained by the phenomenon of non-participation of learners in many items as shown in Table 4.7 and Figure 4.16 and 4.17, where there were an overwhelming number of blanks and zeros, particularly from schools FT and LM.

9.2.2.3 Ordering of Items

The ordering of items according to the Rasch model depended on how the learners responded to the items during the assessment process, and items were arranged from the easiest (Item 3.1) to those of the most difficulty (Item 4.2.2), which represents the empirical order, as experienced by the learners. It is of interest that the theoretical ordering as guided by the DoBE assessment taxonomy (DoBE, 2011) differed from the empirical evidence provided by the Rasch analysis. It must be noted that the Rasch model ordering may differ if the test was taken by another cohort since the p-value of the chi-square statistic was less than 0.05. For this cohort, some items which were considered as level 1 according to DoBE taxonomy (Items 2.2.1, 2.2.2 and 1.3.1) were among the most difficult according to the Rasch model. These items were based on grouped data as well as the concept of the regression line taking into account of outliers. It has been noted that one of the reasons for the poor performance in items based on the regression line, could be related to the fact that these are only covered in Term 3 just before the trial examination so learners would not have had time to become familiar with the applications of these concepts. In contrast, some items classified as level 3/4 according to the DoBE taxonomy (Items 5.3, 3.3 and 4.1.1) were not experienced as so difficult as revealed by the Rasch model. It may be the case that learners were able to score part marks especially for Q4.1.1 where the expected answer was part of the instruction. Question 5.3 required learners to prove an identity and the marking memo allocated marks for simplification of the various ratios even if they did not lead to the final step.

9.2.3 How can the Rasch analysis be used to improve the functioning of the mathematics assessment instrument?

The detailed answer to this research question is provided in Chapters 6 and 7 which describes the various steps that were undertaken post-hoc.

This research question was further divided into three sub-questions:

To what extent are the items functioning as expected?

How can the use of the Rasch analysis contribute to the improvement of the scoring rubric?

To what extent do the items display DIF, multidimensionality and item dependency?

The findings regarding for these sub-questions are discussed below.

9.2.3.1 To what extent are the items functioning as expected?

The construct that is being assessed in this examination paper is mathematics proficiency in the sections of statistics, analytic geometry, Euclidean geometry and trigonometry. The RMT was used to investigate the extent of the reliability of the test in providing a measurement- like representation of proficiency in mathematics (Long et al., 2014). When carrying out a Rasch analysis, an assumption is that the data must fit the model. Before we can make inferences about assessment data, it is necessary to check the fit of the data to the model. When a test adheres to the requirements for measurement- like interpretations, then they allow for inferences related to comparisons of item and person proficiency locations. However, when the data does not fit the model, the identification of the anomalies contributes to a deeper understanding of the items assessment instrument as a whole, the specific items as well as the concepts that are being assessed. Much of the research using RMT in education is focused on this diagnostic function (Andrich & Marais, 2019; Modzuka et al., 2019; Bansilal, 2015; Long et al., 2014). With RMT, fit statistics are used to help detect discrepancies between the Rasch model prescriptions and the data that are collected in practice, and for this research question we look more closely at fit statistics for individual items to try to identify whether they are working as well as they are supposed to. If items work well then they contribute to a scale made up by the total score of the test. Then we have confidence that the total score is a useful tool to distinguish between mathematics proficiency of the cohort and it is meaningful to make comparisons of proficiency between learners who have higher or lower score. We apply RMT to investigate the validity and accuracy of the test in providing a measurement-like representation of mathematics proficiency in terms of person proficiency and item difficulty. A valid and reliable test would provide teachers with some indication of the levels of mastery of curricular elements and of developing proficiency in mathematics. It should also provide the KZN Department of Education with an overview of the learner cohort taking ML. An application of the Rasch model will help us to identify anomalies and inconsistencies among these assessment items and the accompanying scoring rubrics and working memoranda.

As part of the analysis, each item was checked in terms of the fit residuals and a summary is provided in Table 9.2. The fit residual is the difference between the person's response to an item and the response that is expected according to the model. The Rasch analysis showed eight items (questions) whose fit-residuals were outside the recommended range of -2.5 to 2.5 (Andrich & Marias, 2019). According to Van Wyke and Andrich (2006), a small fit residual implies that the difference between the observed response of each item and the expected response is small, while a larger fit residual implies the converse. Details of the mis-fitting items are presented in Table 9.2

Table 9.2

Mis-fitting items and summary of the analysis

Item	Concept and Skill	Summary of the Analysis
1.2	Draw an ogive	Level 1 according to DoBE cognitive levels. According to Rasch analysis, the item had a difficulty location of -0.252 logits. Average score was 37.6 %. Disordered thresholds. Item characteristic curve (ICC) showed under discrimination
3.3	Prove that $PQR = 90^{\circ}$	Level 2 according to DoBE cognitive levels. Location of -0.551 logits in Rasch analysis. Average score 40.8%. The item was easy for the learners. Disordered thresholds. Item characteristic curve (ICC) showed over discrimination
4.1.1	Deducing that $a = -3$. 'a' is the y-coordinate of the point of intersection of the circle, tangent and radius	Level 3 according to DoBE cognitive levels. Location of -0.253 logits on difficult scale, Rasch analysis. Average score was 31.4%. Disordered thresholds. Item characteristic curve (ICC) showed under discrimination
4.1.2	Determine the equation of the circle	Level 1 according to DoBE cognitive levels. Location of -0.418 logits, Rasch analysis. Average score was 41.1%. Disordered thresholds. Item characteristic curve (ICC) showed over discrimination
6.2	Sketching graphs of $F(x) = 2cos\ x$ and $g(x) = sin(x + 30)$	Level 1 according to DoBE cognitive levels. Location of -0.594 logits, Rasch analysis. Average score 54.6%. The item was easy for the learners. Disordered thresholds. Item characteristic curve (ICC) showed under discrimination
8.1	Prove that $VOYZ$ is a cyclic quadrilateral	Level 2 according to DoBE cognitive levels. Location of -0.155 logits, Rasch analysis. Average score 32.7%. Disordered thresholds. Item characteristic curve (ICC) showed over discrimination
8.3	Prove that $\triangle XOV /// \triangle XZY$	Level 2 according to DoBE cognitive levels. Location of -0.293 logits, Rasch analysis. Average score 34.5%. Disordered thresholds. Item characteristic curve (ICC) showed over discrimination
9.2	Calculating (stating reasons) the length of GH. (from proportionality theorem)	Level 3 according to DoBE cognitive levels. Location of 0.140 logits, Rasch analysis. Average score was 17.4. The item was difficult for the learners. Item characteristic curve (ICC) showed haphazard fit

Recall from section 3.8.7 that when the actual proportions are steeper than the predicted curves, the discrimination is more than predicted and the item is therefor considered as over-discriminating. Four of the mis-fitting items displayed over-discrimination: Items 3.3, 8.1, 8.3

and 4.1.2. An example of the ICC for Item 3.3 showing over-discrimination is given in Figure 9.1.

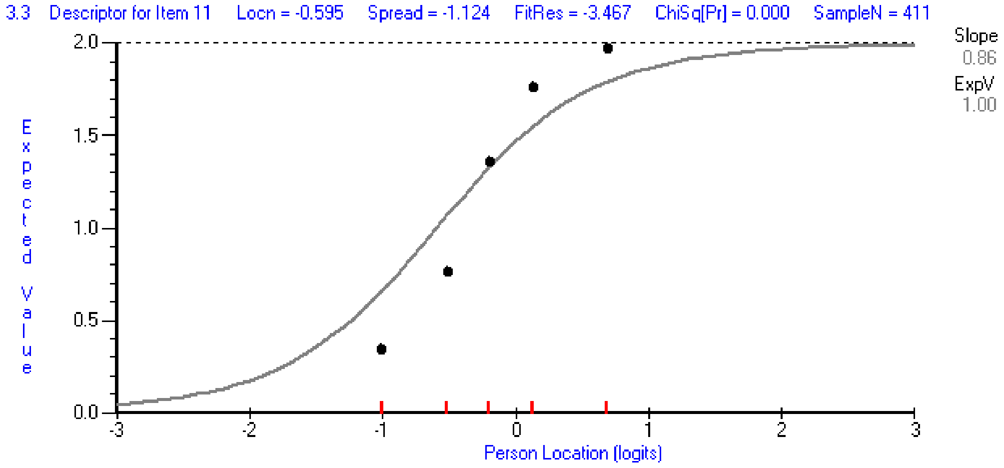


Figure 9.1. The item characteristic curve (ICC) for Item 3.3 showing over-discrimination

A seen in Figure 9.1, learners with higher proficiency performed better than predicted while learners of lower proficiency performed lower than predicted. Thus over-discrimination in the item may have unduly advantaged high proficiency learners, while disadvantaging learners of lower proficiency. While some educators may feel that the greater the item discrimination the better, the concern in RMT is that items which discriminate more are sometimes the result of response dependence (Long et al., 2014). It may be that the responses to Question 3.3 are dependent on the learners' responses in another question, so learners who got the other question correct were more advantaged than those who did not get it correct. It turned out that Item 3.3 is dependent on Item 3.2 as discussed in section 7.3 and highlighted again in section 9.2.3.3.

The phenomenon of under-discrimination is discussed in section 3.8.7 in detail. Three of the mis-fitting items in this study were under- discriminating. The fit residuals for Items 1.2, 6.2 and 4.1.1 were positive and large in magnitude, which is associated with a response pattern whose empirical discrimination tends to be less than that of the summary discrimination of the rest of the items (Smith & Plachner, 2009). We say that such items are under-discriminating as explained in Section 3.8. An example of the ICC for Question 6.2 showing under-discrimination is given in Figure 9.2.

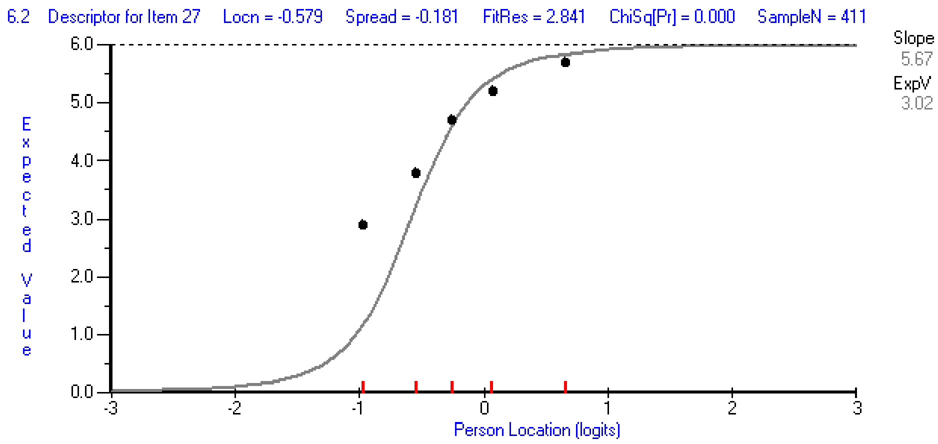


Figure 9.2. ICC for Question 6.2 showing under discrimination

Figure 9.2 shows a pattern of under discrimination for Question 6.2 in that learners of lower proficiency appear to perform better than expected and consequently because of the interactive nature of item difficulty and learner ability, the high proficiency learners are falsely estimated to respond to an item that appears easier than it really is (Bansilal, Long & Juan. 2019). This is a typical pattern that is observed when there is guessing for example in multiple choice items (Bansilal et al., 2019) and has also been observed in assessment items which ask for some reason for an answer and any given reason is allocated a mark, irrespective of whether the reason was relevant or not (Long et al., 2014), and hence the mark for the reason was redundant. The resulting fit is similar to the case of guessing where learners with low proficiency provide an answer that is given a mark and hence they do better than predicted by the theoretical curve. For Question 6.2 the analysis in Chapter 6 showed that the marking memorandum was based on an assumption that learners would substitute various points and draw the trigonometric graphs point by point. However most learners use calculators that can generate the table of values and hence learners were able to score marks easily and hence many of the marks in the rubric were redundant. Some of these issues can be addressed by post- hoc rescoring, and this is done to illustrate what needs to be taken into account when similar examination papers are set in future. Answers to the next research sub question in the next section, provides more details.

9.2.3.2 How can the use of the Rasch analysis contribute to the improvement of the scoring rubric?

As set out in the introduction to this chapter this part of the research question is concerned with a detailed look at the mis-fitting items and seeing whether a post-hoc rescoring would help improve fit statistics. Chapter 6 described how the items with disordered thresholds were identified as well as how the rescoring was determined and finally whether the rescoring resolved the disordered thresholds.

In Chapter 6 it was found that the mis-fitting items had disordered thresholds. A study of the category probability curves showed items with categories (scores) that were not functioning as expected. The rescoring resulted in acceptable fit residuals for all except Item 3.3 within the recommended range of (-2.5; 2.5) (Tennant & Conaghan, 2007). As explained in Chapters 3 and 6, the disordered thresholds indicated that the scoring rubric for the item did not reflect the underlying proficiency continuum (Long et al.,2014). According to Van Wyke and Andrich (2006), disorder in the thresholds signals that the items have failed to function as intended, which presents an opportunity to re-score them to try and reflect the hierarchy of responses.

As detailed in Chapter 6, questions (items) which showed disordered thresholds and had fit residuals outside the recommended range were considered for rescoring first. Some items showed disordered thresholds but were not rescored because the researcher could not find reasons to support the rescoring process as no reasonable explanation could be provided as the basis for rescoring. A summary of the rescoring process for the mis-fitting items which were successfully rescored is presented in Table 9.3. The rest of the rescoring process is provided in the Appendix section of this thesis.

Table 9.3

Summary of the rescoring process for the mis-fitting items

Item	Item with comments	Fit residual (FR)	Decision run 1	Results of rescoring 1 on FR	New score/ comment
2	1.2 (4 marks) ICC shows extreme misfit Disordered thresholds; categories 1, 2 and 3 not working. Chi square is 0.00, less than significant value $p<0.05$	4.087	Rescore 1, 2, 3-1, 4-2	FR = -0.126 ICC shows better fit Chi square = 0.428, indicating a better fit	2 marks
11	3.3 (2 marks) Disordered thresholds ICC shows misfit	-3.467	Rescoring did not work	FR = -2.709 Outside the recommended range	Original scores retained
15	4.1.1 (5 marks) Disordered thresholds ICC shows misfit	2.579	0, 1, 2-0 3-1 4-2 5-3	FR =1.78 ICC shows better fit	3 marks
16	4.1.2 (2 marks) Disordered thresholds Chi-square probability = 0.00 ICC shows misfit, observed proportions are steeper than theoretical curve	-2.955	1, 2-1	FR = -1.199	1 mark
27	6.2 (6 marks) Disordered thresholds Chi square 0.000 less than the significant value ICC shows misfit	2.841	0, 1-1 2,3-1 4,5-2 6-3	FR = 1.33 Chi square = 0.477 indicating a better fit ICC shows observed proportion more aligned to theoretical curve All categories now working as intended	3 marks
30	8.1 (3 marks) Disordered thresholds ICC shows misfit	-3.340	0-0 1-1 2,3-2	FR = -2.258 ICC shows better fit	2 marks
32	8.3 (4 marks) Disordered thresholds Over-discrimination	-3.076	1,2,3-1 4-2	FR = -2.397	2 marks
35	9.2 (3 marks) Disordered thresholds Under-discrimination	2.887	2,3-2 4.5-3	FR=1.703	3 marks

A comparison of the general fit of the summary statistics before and after rescoring will now be considered to determine if the rescoring improved the fit of the data to the model. The summary is shown in Table 9.4.

Table 9.4

Summary statistics before and after rescoring

	ITEMS [n=40]				PERSONS [n=604]			
	Location		**Fit Residual**		**Location**		**Fit Residual**	
	Before	**After**	**Before**	**After**	**Before**	**After**	**Before**	**After**
Mean	0.0000	0.000	-0.1405	-0.0505	-0.4746	-0.3368	-0.0747	-0.0806
SD	0.6333	0.9973	2.5288	1.5229	0.6628	0.8381	0.8075	0.8696
	Item Trait Interaction: Total Item Chi-Square Before = 801.8917 After = 488.385 Total Chi-Square Probability = 0.0000 (probability the same before and after rescoring)				Person Separation Index = 0,88714 (Before rescoring) Person Separation Index = 0.9101 (After rescoring)			

The rescoring improved the overall person mean location from -04746 logits to -0.3368 logits. The person-mean moved closer to the item mean of 0 logits. Since the total score achieved by a person is used in the calculation of the person's proficiency location, a higher mean person location is linked to a high average score. This shows that by rescoring the items, the average marks improved because redundant marks were removed and as a percentage of the new total marks, the average score of the group improved. The instrument with its revised scoring rubric was still difficult for the cohort of learners, but not as difficult as it was before rescoring.

After rescoring the SD for the item location was 0.9793, which was now closer to the ideal value of 1, while the SD of the person location is 0.8341, which was also close but less than 1, suggesting that the distribution of the person locations was no longer as clustered together as it was before rescoring. The person separation index changed from 0.88714 before rescoring to 0.9101. The separation of the persons' ability was excellent and consistent with the Rasch model. Tennant and Conaghan (2007) advised a minimum of 0.85 for good estimation of ability by the Rasch model. The estimation of the person's ability was consistent across the model. The chi-square value after rescoring was 488.385 with a probability value of 0.0000, which signals that the ordering may vary across the trait. In simple terms it means that if the test was administered to a different cohort of learners, the ordering of the items may differ. Whenever rescoring is being done it is important to keep in mind the principles of best test design (Van Wyke & Andrich, 2006; Wright & Stone, 1979). I also applied multiple criteria where both quantitative and qualitative factors were taken into consideration before revising the scoring in the cases where the categories were not working according to the model. When I was analysing the scoring structure using the CPC to decide on the rescoring process, some trends emerged. These included but were not limited to: redundant marking where more than 1 mark is allocated for the same step/skill (same approach needed), too many marks all allocated as accuracy marks, disregard of the use of continuous accuracy (CA) marks which were specified in the marking memorandum for items requiring use of answers from previous items, or acknowledgement of correct reasoning in a proof type question even though the first step may not have been correct. These trends were discovered in items which did not have a good fit to the model and where the CPC showed disordered thresholds.

Redundancy in marks refers to when more than one mark is allocated to the same step or skill and was observed in some of the items which showed disordered thresholds, such as Question 1.2 (further examples are covered in detail in Chapter 6). As discussed in section 6.2.1, for Question 1.2 most learners either got 0 or maximum marks and very few learners scored marks in the middle categories. In Question 1.2, except for the 2 marks for the shape and grounding the ogive curve, 2 other marks were allocated for plotting the median point and any other point meaning the technique of plotting the points on the graph was rewarded twice. Hence those who were able to plot the points, were rewarded with 2 marks while those who may have made a slip in the calculations earlier or had misconceptions, were disadvantaged twice for the same error or misconception. The CPC signalled this by showing some disordered thresholds. In their study of an ML trial examination paper Long et al., (2014) found similar instances of indiscriminate mark allocation which contravened assessment and measurement principles. Another case which led to redundancy in marks was geometry items where marks were given for statement and reasons such as Question 7, where learners where asked to identify five angles equal to a given angle and to give a reason. Learners who were able correctly identify an angle where likely to give the correct reason while a learner who did not identify a correct angle would not have a reason that could be acceptable. Hence the marks associated with the reasons were largely redundant.

A somewhat similar problem that showed up in the CPCs as disordered thresholds was questions for which accuracy marks alone were given and no part marks were allocated for method or consistent accuracy. For example, in Question 4.1.4, 4 marks were given, all allocated as accuracy marks. A learner who made a mistake in the first step of determining the Tan θ, would then lose all 4 marks even if the method used in the other steps was correct. Learners mostly got 0 (195) or all 4 marks (101). Very few learners got the middle scores and this showed on the CPC as disordered thresholds. The type of questions which required accuracy marks were those which asked learners to show that something was true, and hence only accurate responses were recognised (Item 3.1: Show that the coordinates of Q are (0;2)). In the case of Item 3.1, the reason for using this format would be so that the learners who do not get the full question correct would be able to apply the correct answer in the subsequent sub-questions. However there were many other questions for which accuracy responses were not necessary (Item 4.1.4, the last sub-question in Question 4.1). This paper had a large number of such questions, where the responses were only accepted if they were accurate, which disadvantaged the many learners who made slips or errors in the beginning.

Another problem that emerged was when the memorandum specified a CA mark but teachers did not follow a learner's initial error to see that the rest of the working was correct. Hence learners who made an initial error were not given the concession where their methods were followed so they could be given part marks. Mudaly (2006) observed that for questions that required consistent marking or CA marking, no other educator apart from the chief marker allocated marks in a way that was identical to the marking memorandum, noting that the majority of experienced teachers allocated marks for accuracy. To Mudaly (2006) not applying CA marking disadvantages learners who make manipulative errors in the early stages of presenting a solution. The results from this study suggest that teachers need training to become more efficient in applying CA marks.

Most proof type questions/items showed misfit to the Rasch model (8.1, 8.2 and 8.3, for example). The CPC showed that many learners either got 0 or full marks as the most likely scores to be attained along the ability continuum. Very few learners got the middle scores. With geometry proof problems, one suggestion (Cetin & Ilhan, 2017) would be to use holistic marking rubrics, where the whole argument is considered and then part marks given based on completeness. With an analytic or step by step rubric as is the case for most geometry assessments in school, if a learner does not start at the expected step, then they can get zero marks.

9.2.3.3 To what extent do the items display DIF, multidimensionality and item dependency?

Recall that DIF refers to the differential item functioning of an item and is present when persons from different groups have differing probabilities or likelihoods of success on an item, but are otherwise equally matched on the ability of interest. That is, people from two groups, who have the same overall ability ranking, will perform differentially on the item with DIF.

There were four items (2.2.3, 2.2.2, 2.1 and 3.6) where English first language speakers had higher probabilities of responding positively than English second language speakers who were at the same proficiency ranking. Table 9.5 provides a summary of these items.

Table 9.5

Items showing DIF for language

Item	Concept and Skill	Summary of the Analysis
2.1	Identify an outlier. Write down the coordinates.	Level 1 according to DoBE cognitive levels. According Rasch analysis, the item had a difficult level of -0.158 logits. Average score was 39.9%.
2.2.2	Determine the correlation coefficient excluding the outlier and explain the type of correlation.	Level 1 according to DoBE cognitive levels. Location of 0.253 logits, Rasch analysis. Average score was 21.9%.
2.2.3	Determine the average fuel consumption of the motor car.	Level 1 according to DoBE cognitive levels. Location of 0.399 logits on difficult scale, Rasch analysis. Average score was 15.6%.
3.6	Calculate the length of PR. (leave your answer in the simplest surd form).	Level 2 according to DoBE cognitive levels. Location of -0.691 logits, Rasch analysis. Average score was 45.9%.

The items or questions in Table 9.5 contained words or terms which may not have been part of everyday language, like scatter plot, average fuel consumption in the statement for Question 2, outlier in Question 2.1, correlation coefficient excluding the outlier in Question 2.2.2. These seem to have presented additional challenges to learners who used English as a second language. Bansilal (2015), reporting on the Rasch analysis of mathematics teachers' proficiency, showed that there was a DIF for the language factor in an item which had an overload of textual information, where English second language speakers on the same ability level as their English first language speakers experienced such items as more difficult.

In contrast Question 5.1 displayed a DIF where learners who spoke English as a second language had a higher probability of responding positively than learners who spoke English as their first language. Question 5.1 required learners to simplify a trigonometric expression in symbolic form to a numerical expression, which did not contain any extraneous textual information. For this question the wording of the instructions was short and straight forward, hence the learners did not struggle to interpret what was needed in order to answer the question. There is much research that shows that the language complexity or readability of the tests may contribute to lower than expected outcomes (Bansilal & Khan, 2013; Prins & Ulijn, 1998; Abedi & Lord, 2001). Authors Prins and Ulijn (1998) administered three English versions of nine mathematics tasks – original, adapted and non-verbal to 108 students in South Africa. The adapted versions were modified to make the tasks more readable, while the non-verbal versions did not have any references to context. The students were made up of three groups: those who spoke English as their first language (E1), first language Afrikaans speakers (E2) or those whose first language was an African language. These authors found that the average score on the adapted versions were statistically significantly higher than on the original versions for students from all the language groups (Prins & Ulijn,1998). However, second language English speakers benefited more from modifications to mathematics test items than first language English speakers. They also found that students whose first language was an African language (E3) had the lowest score on the original and adapted versions, but performed equally well as their counterparts on the non-verbal version of the task. The authors comment that these results show that the E3 group had the same algebraic computational skills as their E1 and E2 counterparts, but factors related to language and culture inhibited their performance on the non-verbal versions (Prins & Ulijn,1998). The finding from this study that second language English speakers did better than their English first language counterparts of a similar ability in Question 5.1 with its pure symbolic representation is very significant. This finding suggests that students who struggle with the language of instruction may be incorrectly assessed as having a lower proficiency in mathematics whereas it may be that it is the language factor in the instructions which poses an additional barrier for them.

There was no DIF for the person factor gender, as there was no significant difference in the group means of the males and females, after the Bonferroni correction. All the items displayed DIF for the school person factor, meaning learners from different schools responded in different ways to individual items despite having the same proficiency levels. This issue of different learning opportunities at different schools have received much attention recently in literature about school effectiveness in South Africa which point to large differences between schools in the country (van der Berg, 2007; Reddy et al., 2016; Zuze et al., 2016) Van der Berg (2007) highlighted the large variance in the results of some international studies, that is unexplained, but most likely has to do with school effectiveness, which in turn is linked to the school economic status. The intra-class correlation coefficient (ICC) can be used to indicate the proportion of the total variance in the outcome achievement variable that lies systematically between schools and within schools. The ICC can thus serve as a measure of the inequality between schools by measuring the proportion of variation in marks overall that can be attributed to differences between schools. Zuze et al., reported that for the TIMSS 2015, the between-school variance accounted for 61 per cent of the total variance for Grade 9 mathematics learners, while the within-school variance explained 39 per cent of the total variance in results. The authors (Zuze et al., 2016) note that as a comparison, the within-school variation for learners in Finland who took part in the Programme for International Student Assessment (PISA) in 2012 was 92.5 per cent, thus indicating that only eight per cent of variation in marks was between schools. The low between-school variance in Finland illustrates that the quality of schooling is virtually the same in any school in the country, i.e., the results did not vary a lot between schools. In contrast in South Africa, the ICC was 61% showing that most of the variance in marks are attributable to differences in quality of schooling at the various schools.

Response dependence was observed in 8 pairs of items which had residual correlation values above the 0.2 value recommended by Christensen et al., (2017). The items were 1.3.1 and 1.3.2, 3.2 and 3.3, 1.2 and 1.3.1, 3.4 and 3.6, 4.1.1 and 4.1.4, 8.2 and 8.4 and 10.2.3 and 10.2.4. There were two observable reasons why the items showed response dependence; the first was that teachers were not using CA marking appropriately regardless of the fact that the marking memorandum stated that marks were to be accorded for CA. As mentioned in the interviews, teachers felt it cumbersome to follow learners' errors to check that their method was correct and that the answers obtained by learners who made an initial error were consistent for their calculations.

The second case where response dependence was observed was when the memorandum was silent about CA marking, yet the response to one item needed the response from a previous item. The first case is Questions 3.4 and 3.6, where the marking guideline for 3.6 clearly indicated that all marks were CAs. Learners needed answers from Question 3.4 to be able to respond correctly to Question 3.6. If a learner did not answer Question 3.4 correctly, then if CA marking was applied that learner would be awarded marks in Question 3.6 if the answer was used correctly. However, as teachers were not applying CA marking these learners did not receive the marks in Question 3.6. The second scenario was observed in Questions 10.2.3 and 10.2.4, where the answer to Question 10.2.3 was required in order to respond correctly to Question 10.2.4. The marking memorandum was silent about CA marks and hence teachers were using accuracy marking meaning that learners who got wrong answers for Question 10.2.3 were disadvantaged. All of the items displaying response dependence fell into these two scenarios. According to Long et al., (2014), dependency on an earlier item for a subsequent item is regarded as an unfavourable test practice; instead it is the independence of items that offers greater precision. A study by Mudaly (2006) revealed that in questions which require consistent accuracy/CA marking, no teacher marked according to the marking guideline, and the experienced teachers allocated the majority of marks to pure accuracy. Learners who made manipulation errors in the first step or in the first question would not obtain any marks.

In terms of the investigation of whether the test was multi-dimensional, the results showed that this was not so. The paired t- test option was used and it revealed that there were 48 persons (11.68%) where differences in estimates between the two subscales exceeded the 5% level of significance and 20 (4.9%) persons where differences in estimates exceeded the 1% level of significance. The percentages were close to the value of 5% hence I considered it as a minor violation of independence and it may just be a chance effect, as suggested by Andrich and Marais (2019). Hence it was concluded that the test could be considered as essentially unidimensional with the items all contributing to one main construct of mathematical proficiency.

9.2.4 What are the teachers' views about the findings of the Rasch analysis?

The assessment tool was well balanced according to the teachers who were involved in the study. Rasch analysis indicated that learners found the paper difficult. The teachers attributed this to, among other things, the failure of educators to deal with the basics of most of the chapters, lack of adequate preparation on the part of learners, the inclusion of many proof type questions in the assessment tool, response dependence on some of the items/questions (accuracy and CA marking) and the language barrier. These will be discussed in detail in the next sections.

9.2.4.1 Lack of basic understanding

The teachers who participated in the study attributed the failure by educators to focus on the basics of every chapter as the reason why the learners found the paper difficult. Some questions which were at level 1 and 2 on the cognitive levels under the DoBE taxonomy were shown by the Rasch analysis to be among the most difficult questions. Lack of understanding of the basic concepts was attributed to failure by the teachers to focus on the basics before they move to more complicated aspects, an approach that can be classified as a poor teaching strategy. Mji and Makgato (2006) found that teaching strategies contribute to low performance in mathematics. The role of teachers in the learning of mathematics cannot be emphasised enough. Bansilal (2002) stresses that a key part of any educational process is a well-trained teacher. A well-trained teacher knows the correct teaching strategy to use, how to support his teachings, and making sure that basics are covered and understood before complicated sections are discussed. In support of the participating teachers' findings, there are strong beliefs within the field of mathematics education that differences in learners' mathematics abilities and attitudes towards mathematics may be a reflection of teacher's content knowledge, teaching methods and pedagogies (Ross, McDougall, Hogaboam-Grey & Lesage, 2003; Wilkins, 2008). Not focusing on the basics of every chapter is a sign of a teacher not being competent and in this regard Brumbaugh, Ashe and Rock (1997) argue that successful mathematics teachers must be effective and competent in mathematics, be confident about teaching the subject of mathematics and be able to investigate new mathematical knowledge and relevant strategies of effective teaching. The teachers who believe that mathematics learning is about solving high-level questions will thus neglect the basics and focus on challenging sections. Low quality of teaching and learning is largely a result of poor subject knowledge and ineffective instructional strategies on the part of teachers (Delgado, Hightower, Loyd, Wittenstein, Sellers & Swanson, 2011). Wong and Lai (2008) also believe that the mathematical knowledge of the teacher is a basic prerequisite for successful student achievement and positive attitudes towards the subject, a lack of which will result in negative impacts.

9.2.4.2 Lack of adequate preparation and motivation

According to the teachers who participated in the study, lack of adequate preparation on the part of learners contributed to learners facing challenges during the assessment process. According to the teachers the learners did not prepare well enough for the examination by making use of the previous question papers and engaging the teachers and other learners to clear away all misconceptions. Lack of adequate preparation is closely associated with a lack of motivation on the part of the learners. According to Deci (1972) learners who are intrinsically motivated engages in activities because they lead to unrelated outcomes and these learners have high autonomy and sense of control. For Ames (1992) a motivated learner is one who knows his/her goals and what is important which in turn determines whether or not they will engage in the task. Teachers therefore need to assist learners to improve their results in order to bring about a positive change and attitude towards mathematics (Ma & Xu, 2004). When learners always encounter negative experiences in mathematics, they develop negative attitudes and so it is up to the teachers to create positive experiences that can lead to positive attitudes (McLeod, 1992).

Closely related to the lack of motivation on learner performance is the attitude of the learners towards mathematics as a subject. The teachers who took part in the study mentioned that many learners had a negative attitude towards the learning of mathematics. According to the teachers some learners were not even bothered by the low marks they got in mathematics assessments and did not even practice when about to write tests or examinations. McLeod (1992) noticed that the accumulation of negative experiences (for example a learner who experience constant failure in algebra or geometry) may develop a negative attitude to mathematics. This is in line with Tobias' (1995) finding that a learner's state mind has a great influence on his or her perception of mathematics, where negative perceptions about mathematics and the presence of mathematics anxiety impact negatively on learners' attitudes. Coleman and Conrad (2007) also evaluated the negative perceptions of learners towards mathematics and discovered that it plays a role in poor performance. In a related study, Depado and McLaren (2006) examined attitude and performance in both business statistics and calculus and suggested that attitude plays an important role in mathematics performance and hence should be addressed.

9.2.4.3 Socio- economic factors

According to the teachers involved in the study, some learners came to the examination having already given up hope of passing because their societies had already labelled mathematics as the most difficult and challenging subject. Regarding societal influences, Driver, Asoko, Leach, Mortimer and Scott (1994) stated that learning can be influenced by learners' cognitive frameworks that are related to prior experiences and their cultural influences. According to the teachers these learners did not put any effort into practising the subject or seeking help and thus would struggle during the assessment process. Societal influences are closely related to lack of motivation and concentration. The teachers' findings are supported by those of Broussard and Garrison (2004), who examined the relationship between classroom motivation and academic achievement in elementary-school-aged children and found that for a higher level of mastery, motivation was related to higher mathematics grades. According to Saritas and Akdemir (2009), mathematics education requires highly motivated students because it requires reasoning, making interpretations, solving problems and mathematical issues and understanding concepts. Learners must be focused and motivated to progress in mathematics.

9.2.4.4 Proof type questions

According to the teachers who participated in the study, the inclusion of many proof type questions (there were thirteen proof or show type questions out of the 40 questions in the assessment tool), was one of the reasons why learners found the paper difficult. The teachers indicated that learners had challenges with proof type questions or questions which requires them to show proofs. This finding is supported by research which has shown that learners have challenges with solving proof type questions (Healy & Hoyles, 2000; Moore, 1994; Weber, 2004). According to Moore (1994), learners who lack knowledge of definitions are likely to face challenges with proof questions. Lack of conceptual understanding also leads to learners failing to solve proof type questions as do other factors such as inadequate concept images, inability to understand and use mathematical language and notations and not knowing how to show the proof. To Weber (2004), writing a proof in a certain domain requires one to understand the concepts in that particular domain. Learners experience problems in proof questions because proofs are mainly provided as finished products in textbooks and this does not challenge them to think deductively (De Villiers, 2004). A study by Naidoo and Kapofu (2020) revealed that learners struggle when the have to link different proofs and carry out proofs involving different statements and theorems especially in geometry. Ngirishi and Bansilal (2019) also revealed that learners struggled with proof type questions.

9.2.4.5 Allocation of accuracy marks

The teachers involved in the study felt that allocating all marks as accuracy marks for a question/item that carried more than 2 marks was a practice that was unfair to the learners, as most either got all of the marks or got nothing. Very few learners got marks in between zero and the maximum possible mark is such questions. Allocating all accuracy marks does not differentiate well between learners of different abilities, as knowing a section of the question was not rewarded. A study by Long et al., (2014), found that when an item had 3 marks which were all accuracy marks, learners either got all 3 marks or they got none at all, learners were not getting marks in between.

9.2.4.6 Continuous accuracy marking

Teachers did not apply CA marking during the marking process and as a result the response to some items was found to be dependent on the response to other items, what is referred to in Rasch analysis as response dependency. This happened in questions where marks were allocated for method and for consistent accuracy. Teachers in the study felt that it was an unfair practice, but gave reasons why many teachers (including the ones who completed the questionnaire), did not use CA marking. According to the teachers involved in the study, some teachers did not apply CA marking because of the huge number of learners in their schools while others did not apply it because they had not been trained on how to use CA marking and lack experience of marking at NSC level. The experienced teachers mentioned the need for more time for marking if CA marking was to be used yet department officials were demanding marks a few days or weeks after the assessment. According to Long et al., (2014), not applying consistent accuracy unduly disadvantaged those who did not address the question correctly. A study by Mudaly (2006) also revealed that teachers, including the most experienced teachers, were not applying CA marking but instead were allocating accuracy marks to questions requiring CA marking. Mudaly (2006) revealed that very few or no workshops were conducted by DoBE officials on CA marking.

9.2.4.7 Language barrier

The language barrier was cited by the participating teachers as contributing to the paper being difficult for the learners. Some learners did not fully understand the questions and hence were unable to respond correctly. For the teachers, interpretation of questions is central responding successfully to mathematics questions which are not provided in multiple representations. For these teachers, learners with a better command of the English language stood a better chance of responding correctly to mathematics problems than those who did not have a good command of it. The findings reported here are consistent with those identified in the literature (Sibaya & Sibaya, 1996; Mji & Makgato, 2006). The educators verified the language problem and most felt that it was sometimes challenging to explain concepts in the vernacular because it brought confusion and misinterpretation of ideas. In their study, Prins and Ulijn (1998) reported that mathematics assessments presented in textual form are necessarily linked to language fluency since solution of the problems requires the learner to move between the mathematical symbolic register and the natural language register. Understanding the natural language is necessary to first decode the instruction, then mathematical skills will follow in solving the problem (Prins & Ulijn, 1998). A study by Bansilal and Khan (2013) also found that the readability of the instructions in mathematics was a major hurdle for learners, some did not attempt to demonstrate competences in the mathematics procedures because of their failure to understand the instruction. Bansilal and Khan (2013) also revealed that instructions that had high lexical density were difficult to understand. A study by Henderson and Wellington (1998) also revealed that language was the greatest barrier to learning mathematics in many African countries, including South Africa. Setati and Adler (2012) argued that the teaching and learning of mathematics in a classroom where the language is not the learners' main language is complicated since mathematics has elements that are similar to learning a new language because of the presence of specific registers and set of discourses.

9.2.4.8 Inconsistency in learner performance in questions with different levels of difficulty

Some level 3 and 4 questions were not found to be that difficult by the learners. The teachers attributed this to more time being spent on practising higher-order questions they believed would eventually make those questions easy to attempt, as misconceptions were addressed well ahead of the examination. I am of the belief that success in mathematics involves the ability to understand one's current state of knowledge, build on it, and make changes or decisions in the face of conflict, an observation also noted by Saritas and Akdemir (2009). To achieve this Romberg (1983) suggested, it requires problem solving, abstracting, inventing and proving. Providing learners with more opportunities to grapple with challenging questions is in line with the findings of Wilson (1996), who observed that for students to accomplish learning, teachers should provide meaningful and authentic activities to enable them to construct their understanding and knowledge of the subject domain. Spending more time on level 3 and 4 questions should be associated with instructional strategies where students actively participate in their own learning as this is critical for success (Bloom, 1976).

9.2.4.9 Different performances in individual questions but same raw total

Learners with the same total scores, either from the same school or different schools did not perform in the same way on most of the items. This was attributed by the teachers involved in the study to different teachers using different methodologies, teachers having different strengths in terms of content knowledge and chapter preferences and different learner abilities in different schools and within the same school. According to the teachers involved in the study, individual teachers taught chapters that they were more comfortable with and proficient in more efficiently, and spent more time on those than the chapters they were not comfortable with. This finding is consistent with those of many studies which have reported that what teachers know and believe about mathematics is directly connected to their instructional choices and procedures, and what they will finally teach in the classroom (Brophy, 1990; Brown, 1985; Thompson, 1992; Wilson, 1990; Gellert, 1999).

Gellert (1999) reported that in mathematics education research, it seems undisputed that the teacher’s philosophy of mathematics and his/her general content knowledge strengths have a significant influence on the structure of mathematics classes. I agree with Bransford, Brown and Cocking (2000) who stated that teachers do not need knowledge of particular subject matter only but also need to have pedagogical knowledge and knowledge of their students, as research has shown that teacher competence in these areas is closely linked to student thinking, understanding and learning in mathematics education. I concur with the finding that student achievement in mathematics education required teachers to have a firm understanding of the subject domain and the epistemology that guides mathematics education (Bill, 1993, Grossman et al., 1989, Rosebery et al., 1992), as well as an equal understanding of different instructional activities that promote student achievement, in line with an understanding of the kind of learners one is dealing with. I am of the opinion that teachers who are competent provide a roadmap to lead students to an organised understanding of mathematical concepts, to reflective learning where feedback is constantly given, to critical thinking and ultimately to mathematical achievement.

9.3 Recommendations

The findings of this study can be useful to different stakeholders within the education system. In this section recommendations based on the findings which emerged are provided for teachers, subject advisors and other education officials.

9.3.1 Recommendations for teachers

The study has shown some advantages of using the Rasch analysis for formative purposes in the education sector. It is recommended that teachers should try using Rasch analysis as a tool to improve their assessment instruments. It can assist teachers to see the level at which their learners are operating, to observe items which are challenging for specific groups of learners, and to design interventions accordingly. Rasch analysis can help to identify items which are not fitting the model then teachers can investigate and find reasons why those items cause problems and find ways to fix those problems. The use of the Person-Item map to order items and ability of learners provides decision-makers and teachers with an extensive but quick diagnostic summary of which items are within and those which are beyond the reach of most learners.

Teachers are encouraged to study the responses of learners to take account of the different approaches the use. Teachers should take time and apply CA marking in order not to disadvantage learners. Management should ensure that teachers are given enough time to carefully mark the scripts and apply CA marking. Teachers with large classes cannot be expected to do justice to fair marking practices.

9.3.2 Recommendations for subject advisors

Teachers need training on the application of CA marking at school level especially for the marking of common tasks like the mid-year examinations, trial examinations and the end of year examinations. This study has shown that teachers find it cumbersome to carry out CA marking and opt not to do it with their large classes. It will help if teachers can be given intensive training on the importance of CA marking, as well as training on how to carry out CA marking more effectively. It is the task of subject advisors to ensure that teachers are supported in this task. Teachers need assistance in developing the skills for CA marking.

The findings from the study have suggested that some schools did not cover certain concepts sufficiently, especially for those concepts which formed part of the Grade 12 programme for the third term. It is therefore recommended that the subject advisors provide struggling schools with more support to help them plan and implement their work schedules effectively. Subject advisors need to visit schools regularly so that they are well informed about the curriculum progression of the schools.

9.3.3 Recommendations for Department of Basic Education officials

The assessment section needs to take maximum considerations when deciding on the allocation of marks for assessment tasks. Redundant mark allocation is an unfair practice in education as this study has shown, when redundant marks are part of the memorandum, the overall marks are affected. Questions which require the use of answers from previous questions should be allocated CA marks to avoid disadvantaging learners because of slips or errors they may have made. Examiners should take care not to overload examinations with questions that have three or more marks awarded only as accuracy marks as this disadvantages those who partially know the answer or solution. The study showed that when 4 accuracy marks are awarded to a question, learners either get full marks or they get zero, and few get marks in-between. This will give an unfair advantage to those who can provide the full solution and disadvantage those who partially know the solution or make a mistake in any of the steps. The assessments need to take into account the use of technology, since most learners use calculators to solve the problems. For example, the graph sketching in trigonometry was based on an assumption that the coordinates of each point was calculated by hand. Hence there were many redundant marks since a learner who used a calculator to generate a table of values would get all points correct or all incorrect, so it does not make sense to give marks for separate points but rather for the holistic picture.

Marking guidelines should state clearly in every question whether marks are awarded as accuracy or CA marks, especially for questions with some response dependency aspects. Teachers must be given enough time to mark common tasks before submission, since CA marking takes a lot of time, as teachers consider every response that a learner gives as either correct or incorrect.

In terms of different performances in schools, departmental officials are encouraged to organise content and methodology workshops regularly, where teachers can share good practices and give each other support in terms of pedagogical content knowledge. This would help to bridge the gap between different schools in terms of learner performance.

9.3.4 Themes for further research

This study looked at the use of the Rasch analysis in diagnosing areas in which the assessment instrument could be improved. A suggested area for further study would be the use of Rasch analysis for individual ability identification and designing of appropriate learning assistance. The analysis could focus on the extent to which the learning intervention worked as planned. In this study it was not possible to remark all the scripts but a future study could plan for this aspect by identifying a small group of expect teachers who could first remark the scripts before the Rasch analysis is conducted. This would help identify more clearly item dependency between items. When teaches do not apply CA marking consistently, then it may give a false impression that one item is dependent on the other. However if the scripts were marked strictly according to the marking memorandum, this problem would be reduced and the phenomenon of item dependence would be easier to identify.

More studies are required on the efficiency of marking rubrics in school level examinations. Studies about how the wording of items affect first and second language speakers would be very useful to that test. Designers can ensure that students are not unfairly disadvantaged in responding to mathematics problems because their English language proficiency may not be as high as others.

9.4 Limitations of the study

The study was carried out with three quintile 4 schools, one quintile 2 school, one quintile 5 school and one independent school, which is not an equal representation of all schools within the context of South Africa. The representation of the English first language speakers was also very small in this study (64 out of 604 learners) and cannot be referred to as representative of all South African communities.

Interviews were supposed to be carried out with learners to obtain information about why they performed in the way they did, but because of time constraints, no such interviews were done and only the teachers filled in some questionnaires. This was because by the time the first phase of the research was finished, learners had already written their final examinations and it was impossible to set up meetings with them.

Another limitation was that not all teachers who were given the questionnaires completed and returned them. A further limitation which effected the results was that the scripts were not re-marked by the researcher for quality assurance purposes.

9.5 Conclusion

The Rasch model has great potential if it is used for formative purposes in South Africa. Items are ordered according to how the learners experienced them from easiest to most difficult. Persons are grouped according to ability from low proficiency levels to high proficiency levels. The Person-Item map provides an approximation of person proficiency and item difficulty on a common scale. This makes it possible to identify the items which are challenging to specific learners and to plan remedial action accordingly.

If the Rasch model is used together with professional judgements it is possible to identify anomalies in the marking rubric. The ICCs and CPC can show items which have some anomalies and needs further investigation. Qualitative investigation can then be used to identify any problems from a mathematics education and assessment point of view then scoring rubric is adjusted.

The Rasch model helps to bring awareness to stakeholders that assessment is not all about reporting learners' progress in mathematics classrooms, but also takes into consideration the validity of that assessment tool. The validity of the marking rubric and marking process must also be considered. If it is used efficiently in a formative way, the Rasch model can also help to determine remedial action.

REFERENCES

Abedi, J., & Lord, C. (2001). The Language Factor in Mathematics Tests. *Applied Measurement in Education, 14*(3), 219-234. doi: 10.1207/S15325818 AMEP4032

Ainsworth, S. (2006). A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183-198.

Ainsworth, L. (2015). *Common formative assessments: How teacher teams intentionally align standards, instruction, and assessments*. Thousand Oaks, CA: Corwin.

Airasian, P. W. (2001). *Classroom Assessment Concepts and Applications* (4 ed.). McGraw-Hill: McGraw-Hill Higher Education.

Ames, C. (1992). Achievement goals and the classroom motivational climate. In Schunk, D.H., & Meece, J.L. (Ed.), *Students perceptions in the classroom* (pp. 327-348). Washington: Lawrence Erlbaum Associates.

Anderson, D., Lai, C., Alonzo, J., & Tindal, G. (2011). Examining a grade-level maths CBM designed for persistently low-performing students. *Education Assessment, 16*, 15-34. doi: 10.1037//0096-3445.1302.224

Andrich, D. (1985a). An elaboration of Guttman scaling with Rasch model for measurement. In N. Brandon-Tuma (Ed.), *Sociological methodology* (pp. 38-80). San Francisco: Jossey-Bass.

Andrich, D. (1988). *Quantitative applications in the social sciences. Rasch Models for Measurement*: Sage Publications.

Andrich, D. (2011). Rating scales and Rasch measurement. *Expert Rev Pharmacoecon Outcomes Res, 11*(5), 571-585. doi: 10.1586/erp.11.59

Andrich, D., & Hagquist, C. (2015). Real and Artificial Differential Item Functioning in Polytomous Items. *Education Psychology Measurement, 75*(2), 185-207. doi: 10.1177/0013164414534258

Andrich, D., & Marais, I. (2019). *A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences*. Singapore: Springer Nature Singapore Pte Ltd.

Archer, A. L. (2011). Effective and Efficient Teaching. In A. L. Archer, & Hughes, C. (Ed.), *Explicit Instruction: Effective and Efficient Teaching*. New York: Guilford Publications.

Ashcraft, M. H., & Kirk, E.P. (2001). The relationships among working memory, maths anxiety and performance. *Journal of Experimental Psychology. General., 130*(2), 224-237. doi: 10.1037//0096-3445.130.2.224

Atagana, H. I., Mogari, L.O., Kriek, J., Ochonogor, E.C., Ogbonnaya, U.I., & Makwakwa, E.G. (2009). An analysis of educators' and learners' perceived difficult topics in mathematics, physical sciences and life sciences at the Further Education (FET) band in Gauteng province. *A report of the ISTE 2009 Winter school*.

Ball, D. L., Thames, M.H., & Phelps, G.C. (2008). Content Knowledge for Teaching. What makes it Special? *Journal of Teacher Education, 59*(5). doi: 10.1177/0022487108324554

Bansilal, S. (2015). A Rasch analysis of a Grade 12 test written by mathematics teachers. *South African Journal of Science, 111*(5-6), 1-9.

Bansilal, S. (2009). The Use of real Life Contexts in the CTA: Some unintended Consequences. *Pythagoras, 69*, 17-27

Bansilal, S., & Wallace, J. (2008). National Performance Assessment in a South African Context: A case study of issues of classroom implementation and task design. *Africa Journal of Research in Mathematics and Technology Education, 12*(1), 77-91. doi: 10.1080/10288457.2008.10740643

Bansilal, S., Brijlall, D., & Mkhwanazi, T.W. (2014). An Exploration of the common content knowledge of high school mathematics teachers. *Perspective in Education, 32*(1), 34-50.

Bansilal, S., Zondi, L., & Shabalala, L. (2016). Investigating Opportunities to learn: An analysis of Grade 9 Mathematics learners' notebooks. Paper presented at the 22nd Annual National Congress of the Association for Mathematics Education of South Africa (AMESA), 29 June-3 July, Nelspruit.

Baviskar, S., Hartle, R.T., & Whitney, T. (2009). Essential criteria to characterise constructivist teaching: Derived from a review of literature and applied to five constructivist-teaching method articles. *International Journal of Science Education, 31*(4), 542-550.

Bennet, R. E. (2010). Cognitively based assessment of, and as learning: A preliminary theory of action for summative and formative assessment *Measurement: Interdisciplinary Research and Perspectives, 8*, 70-91.

Bennet, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Police and Practice, 18*(1). doi: 10.1080/09669594X.2010.51368

Bessant, K. C. (1995). Factors associated with types of mathematics anxiety in college students. *Journal for Research in Mathematics Education, 26*(4), 327-345.

Black, P. (1988). *Testing: Friend or Foe?* New York: Routledge.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Police and Practice, 5*(1), 7-74. doi: 10.1080/0969595980050102

Bland, J. M., & Altman, D.G. (1995). Multiple significance tests: The Bonferroni method. *BMJ*. doi: 10.1136/bmj.310.6973.170

Bloom, B. (1976). *Human Characteristics and School Learning.* New York: McGraw Hill.

Bond, T.G., & Fox, C.M. (2007). Applying the Rasch model: Fundamental measurement in the human sciences (2nd). Mahwah, NJ: LEA.

Bond, T. G., & Fox, C.M. (2012). *Applying the Rasch Model* (2nd ed.). New York: Routledge.

Bond, T. G., & Fox, C.M. (2015). *Applying the Rasch Model* (3rd ed.). New York: Routledge.

Boone, W. J., Staver, J.R., & Yale, M.S. (2014). *Rasch analysis in the human sciences*. London: Splinger.

Bowie, R. D. (2009). Syntegration or Disintegration? Models of Integrating the Arts Across the Primary Curriculum. *International Journal of Education and the Arts, 10*(88).

Bransford, J. D., Brown, A.L., & Cocking, R.R. (2000). *How people learn: Brain, mind, experience and school: Expanded Edition*. Washington DC: National Academy Press.

Branson, N., Lam, D., & Zuze, L. (2012). Education analysis of the NIDS Wave 1 and 2 Data sets. *SALDRU Working Papers*: University of Cape Town.

Brophy, J. E. (1990). Teaching social studies for understanding and higher-order applications. *Elementary School Journal, 90,* 351-417.

Broussard, S. C., & Garrison, M.E.B. (2004). The relationship between classroom, motivation and academic achievement in elementary school-aged children. *Family and Consumer Sciences Research Journal 33*, 106-120. doi: 10.1177/10777727X04269573

Brown, C. A. (1985). *A study of the socialization to teaching of a beginning secondary mathematics teacher.* Unpublished doctoral dissertation. University of Georgia. Gorgia.

Brown, M. E., Trevino, L.K., & Harrison, D.A. (2005). Ethical Leadership: A Social Learning Perspectives for Constructive Development and Testing. *Organisational Behaviour and Human Decision Processes, 97*(2), 117-134. doi: 10.1016/j.obhdp.2005.03.002

Brumbaugh, D. K., Ashe, J.L., & Rock, D. (1997). *Teaching Secondary Mathematics*: L. Erlbaum Associates.

Bush, T. (2007). Authenticity in research-reliability, validity and triangulation. In M. Briggs. A.R.J. & Coleman (Ed.), *Research Methods in Educational Leadership and Management* (2nd ed.). London: Sage Publications.

Callingham, R., & Bond, T. (2006). Research in Mathematics Education and Rasch Measurement. *Mathematics Education Research Journal, 18*, 1-10. http://dx.doi.org/10.1007/BF03217432.

Carter, P. L. (2010). Race and cultural flexibility among students in different multiracial schools. *Teachers College Record, 112*(6), 1529-1574.

Case, A., & Deaton, A. (1999). School Inputs and Education Outcomes in South Africa. *Quarterly Journal of Economics August, 114*(3), 1047-1084.

Cassim, I. (2006). *An exploration study into grade 12 learners' understanding of Euclidean geometry with special emphasis on cyclic quadrilateral and tangent theorems.* Unpublished masters' thesis. University of the Witwatersrand. Johannesburg.

Cavanagh, R. F., & Waugh, R.F. (2011). The utility of Rasch measurement for learning environments research. In R. F. W. Cavanagh, R.F. (Ed.), *Applications of Rasch Measurement in Learning Environments Research*. Rotterdam: Sense Publishers.

Cetin, B., & Ilhan, M. (2017). An analysis of rater severity and leniency in open-ended mathematics questions rated through standard rubrics based on the SOLO taxonomy. Egitim ve Bilim, 42(189), 217-247. http://dx.doi.org/10.15390/EB.2017.502

Charmaz, K. (2015). Teaching theory construction with initial grounded theory tools: A reflection on lessons and learning. *Qualitative Health Research, 25*.

Chigonga, B. (2016). Learners' errors when solving trigonometric equations ad suggested interventions from Grade 12 mathematics teachers *International Society for Technology in Education*, 163-176.

Christensen, K. B., Makransky, G., & Horton, M.C. (2017). Critical Values for Yen's Q3 Identification of Local Dependence in the Rasch model using Residual Correlations. *Applied Psychological Measurement, 41*(3), 178-194. doi: 10.1177/0146621616677520.

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

Coleman, C., & Conrad, C. (2007). Understanding the negative graduate student perceptions of required statistics and research methods courses: Implications of programs and faculty. *Journal of College Teaching & Learning, 4*(3), 11-20.

Combrinck, C. (2018). *The use of Rasch Measurement Theory to address measurement and analysis challenges in social science research.* PhD, University of Pretoria, South Africa.

Confrey, J. (1990). A review of the research on student conceptions in mathematics, science and programming. *Review of Research in Education, 16*(1), 3-56.

Cooper, D. R., & Schindler, P.S. (2001). *Business Research Methods*. London: McGraw-Hill Higher Education.

Cooper, L. L., & Shore, F.S. (2008). Students' misconceptions in interpreting centre and variability of data via histograms and stem and leaf plots. *Journal of Statistics Education, 16*(2).

Craig, T., & Campbell, A. (2013). The development of engineering students' algebra skills: findings from an algebra skills test refined by Rasch analysis. In Z. Davis & S. Jaffer (Eds.). *Proceedings of the 19th Annual Congress of the Association for Mathematics Education of South Africa, 1,* 59-69. Cape Town. AMESA.

Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches* (3rd ed.). London: Sage Publications.

Creswell, J. W. (2011). *Educational Research Plan: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*: Pearson Education.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods approaches* (3rd ed.). CA: Sage.

Cronbach, L. J. (1971). Test Validation. In R. Thorndike (Ed.), *Education Measurement* (2nd ed.). Washington DC: American Council on Education.

Dawit, T. T., An, V., & Jan, A. (2014). Effectiveness of Critical Thinking Instruction in Higher Education: A Systemic Review of Intervention Studies. *Higher Education Studies, 4*(1). doi: 10.5539/hes.v4n1p1

Debba, R. (2018, January 17-18). *An investigation of the scoring rubric used in mathematical literacy preparatory examination paper in KZN, South Africa.* Paper presented at the 7th International Conference of Probabilistic Model for Measurement. Perth, Western Australia.

De Villiers, M. (2004). Using dynamic geometry to expand mathematics teachers' understanding of proof. *International Journal of Mathematical Education in Science and Technology, 35*(5), 703-724.

De Villiers, M., & Jugmohan, J. (2012). Learners' conceptualisation of the sine function during an introductory activity using sketchpad at Grade 10 level. *Edu. Matem., Pesq., Sao Paulo, 14*(1).

Deci, E. L. (1972). Intrinsic motivation, extrinsic reinforcement and inequity. *Journal in Personality and Social Psychology, 22*(1), 113-120.

Deci, E. L., & Ryan, R.M. (1985). *Intrinsic motivation and self-determination in human behaviour*. New York: Plenum.

Delgado, C. R., Hightower, A.M., Lloyd, S.C., Wittenstein, R., Sellers, K., & Swanson, B.C. (2011). *Improving student learning by supporting quality teaching*. Bethesda: Editorial Projects in Education.

Demir, O. (2012). Students' concept development and understanding of sine and cosine functions. (Unpublished doctoral dissertation), Amsterdam University, Amsterdam, The Netherlands. Retrieved from https://esc.fnwi.uva.nl/thesis/centraal/files/f107257570.pdf

Dempster, E. R., & Reddy, V. (2007). Item readability and science achievement in TIMSS 2003 in South Africa. *Science Education, 91*(6). doi: 10.1002/sce.20225

Department of Basic Education. (2011). *The Curriculum and Assessment Policy Statement for all the subjects listed in the National Curriculum Statement Grades R-12*. Pretoria: Government Printing Works.

Department of Basic Education. (2012). *National protocol for assessment in grades R-12.* Pretoria: Government Printing Works.

Department of Basic Education. (2015). *National Senior Certificate Examinations Report*. Pretoria: Government Printing Works.

Department of Basic Education. (2016). *Marks Adjustments for Schools that have administered the National Common Grade 10 and 11 Question Papers in Mathematics and Physical Sciences*. Pretoria: Government Printing Works.

Department of Basic Education. (2017). *Education in South Africa: Statistical Overview.* Pretoria: Government Printing Works Retrieved from https://www.education,gov.za/EducationiSA.aspx.

Department of Basic Education. (2018). *National Senior Certificate 2017 Diagnostic Report*. Pretoria: Government Printing Works.

Department of Basic Education. (2019). *National Senior Certificate 2018 Diagnostic Report.* Pretoria: Government Printing Works.

Department of Basic Education. (2020). *National Senior Certificate 2019 Diagnostic Report.* Pretoria: Government Printing Works.

Department of Education. (2003). *Revised National Curriculum Statement*. Pretoria.

Diseth, A. (2003). Personality and Approaches to learning as Predictors of Academic Achievement. *Journal of Personality, 17,* 143-155. doi: 10.1002/per.469

Douglas, P. (1982). *Teaching Mathematics in Primary Schools*: Oxford University Press.

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing Scientific Knowledge in the Classroom. *Educational Researcher, 23*, 5-12. doi: 10.3102/0013189X0230077005

Dune, T., Long, C., Craig, T., & Venter, E. (2012). Meeting the requirements of both classroom-based and systemic assessment of mathematics proficiency. The potential of Rasch measurement theory. *Pythagoras, 33*(3).

Edwards, A., & Alcock, L. (2010). Using Rasch analysis to identify uncharacteristic responses to undergraduate assessments. *Teaching Mathematics and its Applications, 29*(4), 165-175.

Edwards, T. G., Ozgun-Koca, A., & Barr, J. (2017). Interpretation of Boxplots: Helping middle school students to think outside the box. *Journal of Statistics Education, 25*(1), 21-28. doi: 10.1080/1069188988.2017.1288556

Engelhard, G. (2013). Invariant measurement: *Using Rasch models in the social, behavioural, and health sciences.* New York: Routledge.

Ercikan, K., McCreith, T., & Lapointe, V. (2005). Factors associated with mathematics achievement and participation in advanced mathematics courses: An examination of gender differences from an international perspective. *School Science and Mathematics, 105*(1), 5-14. doi: 10.1111/j.1949-8594.2005.tb18031.x

Ermeling, B. A., Hiebert, J., & Gallimore, R.G. (2015). "Best Practice" The enemy of better teaching. Educational leadership. *Journal of the Development of Supervision and Curriculum Development, N.E.A, 72*(8), 48-53.

Fieldt, L.S., & Brennam, R.L. (1989). Reliability. In R.L. Linn (Eds). *Educational Measurement* (3rd Ed.). New York: American Council on Education and Macmillan.

Firmender, J. M., Gavin, M.K., & McCoach, D.B. (2014). Examining the relationship between teachers' instructional practices and students' mathematics achievement. *Journal of Advanced Academics 25*(3), 214-236. doi: 10.1177/1932202X14538032

Fisher, D., & Frey, N. (2009). *Background knowledge: The missing piece of the comprehension puzzle*. Portsmouth: Heinemann.

Fleisch, B. (2008). *Primary education in crisis: Why South African school children under-achieve in reading and mathematics* Cape Town: Juta & Co.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *American Educational Research Association, 18*(9), 27-32.

Gardner, H. (2006). *Development and Education of the Mind: The Selected Works of Howard Gardner*: Routledge.

Garrison, C., & Ehringhaus, M. (2013). *Formative and Summative Assessments in the Classroom*. Florence: Best Practices.

Gibbons, M., & Borders, L.D. (2010). Perspective first-generation college students: A social-cognitive perspective. *The Career Development Quarterly 58*(3). doi: 10.1002/j.2161-0045.2010.tb00186.x

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K. I. Manktelow, & Over, D.E. (Ed.), *International library of psychology. Rationality: Psychological and philosophical perspectives*. 284-313: Taylor & Frances/Routledge.

Gellert, U. (1999). Prospective elementary teachers' comprehension of mathematics instruction. *Educational Studies in Mathematics, 37*, 23-43.

Gorard, P. S., & See, B.H. (2013). Do parental involvement and intervention increase attainment? *Nuffield Foundation*.

Grossman, P. L., Wilson, S.M., & Shulman, L.S. (1989). Teachers of substance: Subject matter for teaching. In M. C. Reynods (Ed.), *Knowledge Base for the Beginning Teacher* (pp. 23-36). New York: Pergamon Press.

Groth, R. E. (2007). Towards a conceptualisation of statistical knowledge for teaching. *Journal of Research in Mathematics Education, 38*(5), 427-437.

Guba, E.G., & Lincoln, Y.S. (1994). Competing paradigms in qualitative research. In N.K. Denzin & Y.S. Lincoln (Eds). *Handbook of qualitative research*. Sage Publications, Inc.

Guttman, L. (1950b). The problem of attitude and opinion measurement. In S. A. Stouffer & Others (Eds.), *Measurement and prediction*. New York: Wiley.

Hagquist, C., & Andrich, D. (2004). Is the Sense Coherence instrument applicable on adolescents? A latent trait analysis using Rasch modelling. *Personality and Individual Differences, 36*, 85-968.

Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. University of Michigan: Allyn and Bacon.

Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education, 5*(2), 93-116. doi: 10.1016/0898-5898(93)90026-7

Hannula, M. S. (2002). Attitudes towards mathematics: emotions, expectations and values. *Educational Studies in Mathematics, 49*, 25-46.

Hannula, M. S., Maijala, H., & Pehkonen, E. (2004). Development of Understanding and Self-Confidence in Mathematics; Grades 5-8. In the Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education. Vol 3, pp. 17-24. Bergen: PME

Harlen, W. (Ed.) (1994). Enhancing Quality in Assessment. London, Paul Chapman.

Healy, L., & Hoyles, C. (2000). A study of Proof Conception in Algebra. *Journal for Research in Mathematics Education, 31*(4). doi: 10.2307/749651

Hlalele, D. (2012). Exploring rural high school learners' experience of mathematics anxiety in academic settings. *South African Journal of Education, 32*(3), 267-278.

Henderson, J., & Wellington, J. (1998). Lowering the language barrier in learning and teaching science. *The School Science, 79*(188).

Hoadley, U. (2012). What do we know about teaching and learning in South African primary schools? *Education as Change, 16*(2), 187-202.

Hodgen, J., Kuchemann, D., Brown, M., & Coe, R. (2009). Lower secondary school students' attitudes to mathematics: Evidence from a large-scale survey in England. *Proceedings of the British Society for Research into Learning Mathematics, 37*(2), 1-6.

Holmes, D. (2005). *Psychology: The science of Behaviour and Mental Process*. New York: Thomson Custom Publishing.

Hubbard, D. W. (2010). *How to measure anything: Finding the value of intangibles in Business*. New Jersey: John Wiley & Sons.

Hudson, P., & Miller, C. (1997). The treasure hunt: strategies for obtaining maximum response to a postal survey. *Evaluation and Research in Education, 11*(2), 102-112.

Hugo, W. J. M., Jack, M., Wedekind, V., & Wilson, D. (2010). The state education in KwaZulu-Natal: A report to the Provincial Treasury. Pietermaritzburg. In J. Jackson (Ed.). Pietermaritburg: KZN Provincial Education.

Ije, S. B., & Onwu, G.O.M. (2013). What instructional skills and strategies do competent mathematics teachers use in statistics in school mathematics during PCK development *International Journal Education Science, 5*(4), 363-374.

Iliya, A. (2014). Formative and Summative Assessments in Education Enterprise. *Journal of Education and Practice, 5*(20).

Jacobs, M., Mhakure, D., Fray, R.L., Holtman, L., & Julie, C. (2014). Items difficulty analysis of high-stakes mathematics examination using Rasch analysis: The case of sequences and series. *Pythagoras, 35*(1). doi: https://doi.org/10.4102/pythagoras.v35il.220

Juan, A., Reddy, V., Zuze, T.L., Namome, C., & Hannan, S. (2016). Does it matter whether students enjoy learning science? : exploring student attitudes towards science in South

Africa. Human Sciences Research Council. Doi: http://hdl.handle.net/20.500.11910/9543

Kanfer, R., Ackerman, P.L., & Heggestad, E.D. (1996). Motivational skills & self-regulation for learning: A trait perspective. *Learning and Individual Differences, 8*(3), 185-209. doi: 10.1016/S1041-6080(96)90014-X

Kanjee, A., & Sayed, Y. (2013). Assessment policy in post-apartheid South Africa: Challenges for improving education quality and learning. *Assessment in Education: Principles, Police and Practice, 20*. doi: 10.10880/0969594X.2013.838541

Kanjee, A., & Moloi, M. (2016). A standards-based approach for reporting assessment results in South Africa. *Perspectives in Education, 34*(4). doi: 10.18820/2519593X/pie.v34i4.3

Ketterlin-Geller, L. R., Gifford, D.B., & Perry, L. (2015). Measuring middle school students' algebra readiness: Examining validity evidence for three experimental measures. *Assessment for Effective Intervention, 41*(1), 28-40. doi: 10.11777/1534508415586545

Khan, S. (2012). The rise of the tech-powered teacher. Education Week.

Khan, M., & Bansilal, S. (2012). Learning from previous experiments: readability in national mathematics assessments. *Acta Academia, 45*(1), 79-99.

Khan, S. (2014). Qualitative Research Methods: Grounded Theory. *International Journal of Business and Management, 9*(11). doi: 10.5539/ijbm.v9n11p224

Kriek, J., & Grayson, D. (2009). A holistic professional model for South Africa physical sciences teachers. *South African Journal of Education, 29*(2), 285-204.

Krosnick, J., & Presser, S. (2010). Question and Questionnaire Design. In J. D. Wright, & Marsden, P.V. (Ed.), *Handbook of Survey Research* (2nd ed.). England: Emarald Group.

Lacombe, A. (1985). *Mathematical learning difficulties in the secondary school: Pupil needs and teacher roles.* Philadellphia: University Press.

Lee, J. S., & Ginsburg, H.P. (2009). Early Childhood Teachers' Misconceptions about Mathematics Education for Young Children in the United States. *Australasian Journal of Early Childhood, 34*(4), 37-45. doi: 10.1177/183693910903400406

Lemmer, E. M., Van Wyk, N., & Berkhout, S.J. (2010). *Themes in South African Education*. South Africa: Pearson.

Lesage, J. P. (2003). Spatial Regression Models. In M. Altman, Gill, J., & McDonald, M.P. (Ed.), *Numerical Issues in Statistical Computing for the Social Scientists*: John Wiley & Sons.

Lim, S. K., Tay, L.Y., & Lim, C.P. (2011). A discussion of the factors affecting the implementation of one-to-one computing learning environment in primary school. In G. Williams, Statham, P., Brown.,7& Cleland. (Ed.). Habart: ProceedingsAscilit.

Linacre, J. M. (2013). Disconnected subsets, Guttman patterns and data connectivity. *Rasch Measurement Transactions, 27*(2), 1415-1417.

Linacre, J. M. (2016). *Winsteps. Rasch measurement computer program User's Guide*. Oregon: Winsteps.com.

Lincoln, Y. S., & Guba, E. (1985). *Naturalistic Enquiry*. Beverly Hills: Sage.

Long, C., Dunne, T., & Craig, T.S. (2010). Proficiency in multiplicative conceptual field: using Rasch measurement to identify levels of competence. *Africa Journal of Research in Mathematics and Technology Education, 14*(3), 79-91.

Long, C. (2011). *Mathematical, cognitive and didactic elements of the multiplicative conceptual filed investigated within a Rasch assessment and measurement framework.* PhD, University of Capetown, Cape Town.

Long, C., Bansilal, S., & Debba, R. (2014). An investigation of Mathematical Literacy assessment supported by an application of the Rasch measurement. *Pythagoras, 35*(1).

Luneta, K. (2014). Foundation phase teachers' (limited) knowledge of geometry. *South African Journal of Childhood Education, 4*(3).

Luneta, K. (2015). Understanding students' misconceptions: A analysis of the final Grade 12 examination questions in geometry. *Pythagoras, 3*(1). doi: 10.4102/Pythagoras.v366i1.261

Ma, X. (1999). A meta-analysis of the relationship between anxiety towards mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30*(5), 520-540.

Ma, X., & Xu, J. (2004). Determining the casual ordering between attitude towards mathematics and achievement in mathematics *American Journal in Mathematics, 110*, 256-280.

Mainbashian, A., Huon, A., & Bird, K. (2004). Approaches to studying and academic performance in short-essay exams. *Higher Education 47*(2), 161-176. doi: 10.1023/B:HIGH.0000016443.43594.d1

Makeleni, N. T., & Sethusha, M.J. (2014). The experience of foundation phase teachers in implementing the curriculum. *Mediterranean Journal of Social Sciences, 5*(2), 103-109.

Makgato, M., & Mji, A. (2006). Factors associated with the high school learners' poor performance: A spotlight on mathematics and physical science. *South African Journal of Education, 26*(2), 253-266.

Maor, D. (1998). *How does one evaluate students' participation and interaction in an internet based unit?* Paper presented at the Annual Teaching Learning Forum, Perth.

Marais, I., & Andrich, D. (2008). Formalising dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement, 9*(3), 200-215.

Massof, R. W. (2011). Understanding Rasch and Item Response Theory Models: Application to the Estimation and Validation of Interval Trait Measures from Responses to Ratings Scales *Ophthalmic Epidemiology, 18*(1), 1-19.

Matters, G. (2009). A problematic leap in the use of test data: From performance to inference. In C. Wyatt-Smith, & J.J. Cumming (Eds.). *Educational assessment in the 21st century*: *Connecting theory and practice*. Dordrecht: Springer. http://dx.doi.org/10.1007/978-1-4020-9964-9-11

McDonough, S. (2007). Motivation in ELT. *ELT Journal, 61*(4), 369-371. doi: 10.1093/etl/ccm056

McGill, M.T. (2009). *An investigation of unidimensional testing procedures under Latent Trait Theory using principal component analysis.* Unpublished doctoral dissertation. Virginia Polytechnic Institute and State University. Virginia Tech. Doi: http://hdl.handle.net/10919/29699.

McLeod, D. B. (1992). Research on affect i mathematics education: a reconceptualization. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Learning and Teaching* (pp. 575-596). New York: MacMillan.

McLeod, D. B. (1994). Research on affect and mathematics learning in JRME: 19970 to the present. *Journal for Research in Mathematics Education, 24*, 637-647.

McMillan, J. H., & Schumacher, S. (2006). *Research in Education: evidence-based inquiry* (6th ed.). Boston: Pearson/Allyn and Bacon.

Meece, J. L., Wigfield, A., & Eccles, J.S. (1990). Predictors of maths anxiety and its influence on young adolescents' course enrolment intentions and performance in mathematics. *Journal of Educational Psychology, 82*(1), 60-70. doi: 10.1037/0022-0663.82.1.60

Merriam, S. B. (2009). *Qualitative research. A guide to design and implementation.* (2nd ed.). San Francisco: Jossey-Bass.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational Measurement* (3rd ed.). Washington DC: American Council on Education.

Middleton, J.A. (1995). A study of intrinsic motivation in the mathematics classroom: A personal construct approach. *Journal for research in mathematics education, 26*(3), 254-29.

Mkhwanazi, T., Ndlovu, Z., Ngema, S & Bansilal, S (2018). Exploring Mathematics teachers' usage of the Curriculum Planner and Tracker. In Secondary Schools in King Cetshwayo and Pinetown Districts. In P. Christie and M. Manyokolu (eds.), *Learning about sustainable change in education in South Africa*, (pp.195-224) Johannesburg: SAIDE.

Modzuka, C., Long, C., & Machaba, F. (2019). Rasch analysis of South Africa's Grade 6 Annual National Assessment. *Africa Journal of Research in Mathematics and Technology Education, 23*(1), 86-99. doi: 10.1080/18117295.20199.1600893

Moila, M.M. (2006). The use of educational technology in mathematics teaching and learning: An investigation of a South African rural secondary school. Mini-dissertation Master of Education. University of Pretoria.

Montague-Smith, A., & Price, A.J. (2012). *Mathematics in early years education* (3rd ed.). Lodon: Routledge.

Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics, 27,* 249-266. doi: 10.1007/BF01273731

Morrow, W. (2007). *Learning to teach in South Africa*. Capetown: HSRC Press.

Morse, J. M., & Niehaus, L. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research, 40*, 120-123.

Morse, J. M., & Niehaus, L. (2009). Mixed Methods Design: Principles and Procedures. *Qualitative Social Research, 12*(1).

Moss, C. M., & Brookhart, S.M. (2009). *Advancing formative assessment in every classroom: A guide for instructional leaders*. Alexandria: Association for Supervision and Curriculum Development.

Mudaly, V. (2006). Marking matric scripts is a learning process. Learning and Teaching Mathematics. *A Journal of AMESA, 3,* 41-46.

Mulkeen, A. (2006). *Teachers for rural schools: A challenge for Africa*. Working paper. Biennale on Education in Africa. Association for the Development of Education in Africa.

Muller, F.H., & Louw, J. (2004). Learning environment, motivation and interest: Perspectives on self-determination theory. *South African Journal of Psychology, 34*(2), 169-190.

Mullis, I. V., Martin, M.O., Foy, P., & Drucker, K.T. (2012). *PIRLS 2011 International Results in Reading*. United States of America: Boston College.

Mutodi, P., & Ngirande, H. (2014). Exploring mathematics anxiety: Mathematics students' experiences. *Mediterranean Journal of Social Sciences, 5*(1).

Mustamir, A. (2019). Application of Demonstration Methods to Improve Learning Achievement in Cultural Arts Subject and Skills of Filter Art Graphic Materials in Class IX E Students of SMP Negeri 3 Surabaya. *Indonesian Journal of Contemporary Education, 1*(1), 15-17.

Mvududu, N., & Kanyongo, G.Y. (2011). Using real examples to teach abstract statistical concepts. *Teaching Statistics, 33*(1), 12-16.

Naidoo, J., & Kapofu, W. (2020). Exploring female learners' perceptions of learning geometry. *South African Journal of Education, 40*(2).

Nardi, E., & Steward, S. (2003). Is mathematics T.I.R.E.D? A profile of quite disaffection in the secondary mathematics classroom. *British Education Research Journal, 29*(3), 345-367.

National Education Evaluation and Development Unit. (2013). State of our education system: National report 2012: the state of literacy teaching and learning in the foundation phase.

Ngirishi, H., & Bansilal, S. (2019). An exploration of high school learners' understanding of geometric concepts. *Problems of Education in the 21st Century, 77*(1), 82-95.

Nopiah, Z.M., Osman, M.H., Razali, N., Ariff, F.H.M., & Asshaari, M.F. (2010). How good was the test set up? From Rasch analysis perspective. *International Journal of Scientific and Engineering Research, 2*(4), 116-129.

North, D., Gal, I., & Zewotir, T. (2014). Building capacity for developing statistical literacy in a developing country: Lessons learnt from an intervention. *Statistics Education Research Journal, 13*(2), 15-27.

Oberdorf. C.D., Taylor-Cox., J. (1999). Shape up! *Teaching Children Mathematics, 5*(6), 340-345.

O'Connor, J., & Geiger, M. (2009). Challenges facing primary school educators of English second language (or other) learners in the Western Cape. *South African Journal of Education, 29*, 253-269.

Pajares, F., & Schunk, D.H. (2001). Self-beliefs and school success: self-efficacy, self-concept, and school achievement. In S. G. Rayner (Ed.), *International perspectives on individual differences* (Vol. 2, pp. 239-265).

Pajares, F. S., & Schunk, D.H. (2002). Self and self-belief in psychology and education: A historical perspective. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education*. Washington DC: Academic Press.

Prins, F., & Ulijn, J. (1998). Linguistic and cultural factors in the readability of mathematics texts. *Journal of Research on Reading, 21*(2), 139-159.

Qamar, M. R., Ahmad, A., & Niaz, K. (2005). Learning through small group discussion versus didactic lectures. *Pak Armed Forces Medicine Journal, 65*(3), 386.

Rahman, F., Khali, J.K., Jumani, N.B., Ajmal, M., Malik, S., & Sharif, M. (2011). Impact of discussion method on performance. *International Journal of Business and Social Science, 2*(7), 84-95.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Education Research.

Reddy, V. (2005). State of mathematics and science education: schools are not equal. *Perspectives in Education, 23*(3), 125-138.

Reddy, V. (2006). *Mathematics and Science achievement at South African schools in TIMSS 2003*. Cape Town: Human Sciences Research Council.

Reddy, V., Isdale, K., Juan, A., Visser, M., Winnaar, L., & Arends, F. (2016). TIMSS 2015: Highlights of mathematics achievement of Grade 5 South African learners: Human Sciences Research Council.

Reddy, V., Visser, M., Winnaar, L., Arends, F., Juan, A., Prinsloo, C.H., & Isdale, K. (2016). TIMSS 2015: Highlights of Mathematics and Science achievement of Grade 9 South African Learners.: Human Sciences Research Council.

Reid, W. A., Duval, E., & Evans, P. (2007). Relationship between results and approaches to learning and studying in Year Two medical students. *Medical Education, 41*(8), 754-762. doi: 10.1111/j.1365-2923.2007.02801.x.

Reus, S. (2010). A case of constructivist instructional strategies for adult online learning. *British Journal of Education Technology, 141*(5), 706-720.

Reynolds, C. R., Livingston, R.B., & Wilson, V. (2010). *Measurement and Assessment in Education* (2nd ed.). New York: Pearson.

Rohimah, S. M., & Prabawanto, S. (2019). Students' difficulty identification in completing the problem of equation and trigonometry identities. *International Journal of Trends in Mathematics Education Research, 2*(1), 34-36. doi: 10.33122/ijtmer.v2i1.50

Roman, S., Cuestas Diaz, P.J., & Fenollar, P. (2008). An examination of the interrelationships between self-esteem, others' expectations, family support, learning approaches and academic achievement. *Studies in Higher Education, 33*(2), 127-138. doi: 10.1080/03075070801915882

Romberg, T. A. (1983). A common curriculum for mathematics. In G. D. Fenstermacher, & Goodlad, J.I. (Ed.), *Individual Differences and the Common Curriculum: Eighty-second Yearbook of the National Society for the Study of Education, Part 1*. Chicago: University of Chicago Press.

Rosebery, A. S., Warren, B., & Conant, F.R. (1992). Appropriating scientific discourse: Findings from language minority classrooms. *The Journal of the Learning Sciences, 2*(1), 61-94.

Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American Educator, 36*(1).

Ross, J. A., McDougall, D.E., & Lesage, A. (2003). A survey measuring elementary teachers' implementation of standards-based mathematics teaching. *Journal for Research in Mathematics Education, 34*(4), 344-363. doi: 10.230/30034787

Roux, C. D. (2003). Playing games with religion in education. *South African Journal of Education, 23*(2), 130-134.

Roux, C. D. (2014). *Science Assessment Review Report*. Michael and Susan Dell Foundation (MSDF).

Salako, E. C., Eze, I.R., & Adu, E.O. (2013). Effects of corporative learning on junior secondary school students' knowledge and attitudes to multicultural education concepts in social studies. *Journal of Education, 133*(3), 61-67.

Sambell, K., McDowell, L., & Montgomery, C. (2013). *Assessment for Learning in Higher Education* (1st ed.): Routledge.

Saritas, T., & Akdemir, O. (2009). Identifying factors affecting the Mathematics Achievement of Students for Better Instructional Design. http://www.itdl.org/journal/dec-09/artile03.htm.Retrieved on September 16 2018

Schunk, D. H. (2001). Social cognitive theory and self-regulated learning. In B. J. S. Zimmerman, D.H. (Ed.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 125-151): Lawrence Erlbaum Associates Publishers.

Sebastian, J., Moon, J.M., & Cunningham, M. (2017). The relationship of school-based parental involvement with student achievement: a comparison of principal and parent survey reports from PISA 2012. *Educational Studies, 43*(2), 1-24.

Setati, M., & Adler, J. (2012). Between languages and discourses: Language practices in primary multilingual mathematics classrooms in South Africa. *Educational Studies in Mathematics, 43*(3), 243-269. doi: 10.1023/A:1011996002062

Shannon-Baker, P. (2016). Making Paradigms Meaningful in Mixed Methods Research. *Journal of Mixed Methods Research, 10*. doi: 10.1177/1558689815575861

Shenton, A. K. (2004). Strategies for Ensuring Trustworthiness in Qualitative Research Projects. *Education for Information, 22*(2), 63-75. doi: 10.3233/EFI-2004-22201

Shulman, L. (2004). *The wisdom of practice: Essays of teaching, learning and learning to teach*. San Fransisco: Jossey-Bass.

Sibaya, P.T., & Sibaya, D.C. (1996). Black secondary school pupils' problems with mathematical concepts. *South African Journal of Education, 16*, 32-37.

Simon, M. (2011). *Analysis of qualitative data. Dissertation and Scholarly Research: Recipes for Success*. Seatle: WA: LLC.

Siyepu, S. W., & Mtonjeni, T. (2014). *Geometrical concepts in real-life context: A case of South African traffic road signs.* Paper presented at the 20th Annual National Congress of Association for Mathematics Education of South Africa, South Affrica.

Siyepu, S. W. (2015). Analysis of errors in derivatives of trigonometry functions. *International Journal of STEM Education, 2*(1). doi: 10.1186/s40594-015-0029-5

Smith, L. J. (1999). US Patent No. Derwent Innovations Index: U. S. P. Office.

Smith, R. M. (2002). Rasch Measurement Transactions. *Journal of Applied Measurement, 16*(2), 877-881.

Smith, R. M., & Plackner, C. (2009). The family approach to assessing fit in Rasch measurement. *Journal of Applied Measurement, 10*(4), 424-437.

Spaull, N. (2011). A preliminary analysis of SACMEQ 111 South Africa *Stellenbosch Economic Working Papers* (Vol. 11). Stellenbosch.

Stiggins, R. J. (2005). *Assessment for Learning Defined* Paper presented at the ETS Assessment Training Institute's International Conference: Promoting Sound Assessment in Every Classroom, Portland, OR.

Stols, G. (2013). An investigation into the opportunity to learn that is available to Grade 12 mathematics learners. *South African Journal of Education, 33*(1), 1-18. Retrieved from: http://www.sajournalofeducation.co.za.

Stols, G., Long, C., & Dunne, T. (2015). An application of the Rasch measurement theory to an assessment of Geometric thinking levels. *African Journal of Research in Mathematics and Technology Education*. doi: 10.1080/10288457.2015.1012909.

Tal, E. (2017). Measurement in Science. Retrieved June 23 2018, from The Stanford Encyclopaedia of Philosophy. https://plato.stanford.edu/archives/spr2017/entries/measurement-science/

Tapia, M. (2004). The relationship of mathematics anxiety and gender. *Academic Exchange Quarterly, 8*(2).

Taylor, N. (2011). *The National School Effectiveness Study: Summary for the synthesis report*. Johannesburg: JET.

Taylor, S., & Yu, D. (2009). The importance of Socio- Economic in determining educational achievement in South Africa (Vol. Working Papers ER). Stellenbosch: Stellenbosch University.

Tennant, A., & Conaghan, P.G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied and what should one look for in a Rasch paper? *Arthritis Rheumatology, 57*(8), 1358-1362. doi: 10.1002/art.23108

Tennant, A., & Pallant, J.F. (2007). DIF Matters: A practical approach to test if differential item functioning makes a difference. *Rasch Measurement Transactions, 20*(4). 1082-1084.

Thompson, S. J., Johnstone, C., & Thurlow, M.L. (2002). *Universal Design Applied to Large Scale Assessments. Synthesis Report*. Mineapolis: University of Minnesota Retrieved from http://cehd.umn.edu/...

Thompson, A. G. (1992). Teachers' beliefs and conceptions: Asynthesis of the research. In D. A. Grouws (Ed.), *Handbook of Research in Mathematics Teaching and Learning*. (pp. 127-146). New York: Macmillan.

Tobias, S. (1990). Math anxiety: An update. *NACADA Journal, 10*(1), 47-50.

Umugiraneza, O., Bansilal, S., & North, D. (2017). Exploring teachers' practices in teaching mathematics and statistics in KwaZulu-Natal schools. *A Journal of Education, 37*(2).

Umugiraneza, O., Bansilal, S., & North, D. (2016). Teachers’ Confidence and Beliefs in Teaching Mathematics and Statistics Concepts. *International Scientific Researches Journal, 72*(09), 1-16.

Usman, M.H., & Hussaini, M. M. (2017). Analysis of Students’ Error in Learning Trigonometry Among Secondary School Students in Zaria Metropolis, Nigeria. *Journal of Mathematics, 13*(2).

Van der Berg, S. (2007). Apartheid’s enduring legacy: Inequality in education. *Journal of African Economies, 16*(5), 849-880.

Van der Sandt, S. (2007). Research Framework on mathematics teacher behaviour: Koehler and Grouws' framework revisited. *Eurasia Journal of Mathematics, Science & Technology Education, 3*. doi: 10.12973/ejmste/75413

Van Laren, L. (2012). Mathematics teachers at the intersection of policy interventions: What practising teachers offer as challenges in teaching trigonometry. *Education as Change, 16*. Doi: 10.1080/16823206.2012.745726

Van Wyke, J., & Andrich, D. (2006). *A typology of polytomously scored items disclosed by the Rasch model: Implications for constructing a continuum of achievement of achievement.* Perth: Murdoch University.

Vansteenkiste, M., Lens, W., & Deci, E.L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: another look at the academic motivation. *Educational Psychology, 41*(1*)*, 19-31.

Visser, M., Juan, A., & Feza, N. (2015). Home and school resources as predictors of mathematics performance. *South African Journal of Education, 35*(1). doi: 10.15700/201503062354

Visser, Y. L., & Visser, J. (2000). *The learning stories project.* Paper presented at the International Conference of the Association for Educational Communications and Technology.

Watters, D. J., & Watters, J.J. (2007). Approaches to learning by students in the biological sciences: Implications for teaching. *International Journal of Science Education, 29*(1), 19-43.

Weber, A. (2004). Problem-Based Learning. *Swiss Journal of Education*, *26*(3). doi: 10.24452/sjer.26.3.5360

Wigfields, A., & Eccles, J.S. (1997). Change in children's comptence beliefs and subjective task value across the elementary school years: A 3-year study. *Journal of Education Psychology*, *89*(3), 451-469. doi: 10.1037/0022-0663.89.3.451

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership, 46*(7), 41-47.

Wilder, S. (2014). Effects of parental involvement on academic achievement: a meta-synthesis *Educational Review, 66*(3), 377-397. doi: 10.1080/00131911.2013.78009

Wilkins, J. L. M. (2008). The relationship among elementary teachers' content knowledge, beliefs and attitudes. *Journal of Mathematics Teacher Education, 11*(2), 139-164. doi: 10.1007/s10857-007-9068-2

Wilson, B. G. (1996). *Constructivist learning environments: Case studies in instructional design.* Englewood Cliffs: Educational Technology Publications.

Wilson, M. (1990). Measuring a van Hiele geometric sequence: A reanalysis. *Journal for Research in Mathematics Education. 21*, 230-237.

Wilson, M. (2005). *Construction Measure: An item Response Modelling Approach*. New York: Psychology Press.

Winter, G. (2000). A comparative discussion of the notion of validity in qualitative and quantitative research. The Qualitative Report, 4 (3 and 4) Retrieved from http://www.nova.edu/ssss/QR/QR4-3/winter.html.

Wong, E.M.L., & Li, S.C. (2008). Framing ICT implementation in a context of educational change: a multilevel analysis. *School effectiveness and school improvement, 19*(1), 99-120.

Wood, E., Levinson, M., Postlethwaite, K., & Black, A. (2011). Equity matters. Retrieved August 23, 2018 http:www.download.ei ie.org/Docs/WebDepot/Equity%20Matters.pdf

Wright, B., & Stone, M. (1979). *Best test design*. Chicago: MESA Press.

Wright, B., & Stone, M. (1999). *Measurement Essentials*. Wilmington: Wide Range.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Eduactional Measurement, 14*, 97-116.

Yiga, S.I.N., Khoarai, L., Khosana. T., Lesupi, R., Mduli, J., Shadwell, T., Botes, J., & Joubert, G. (2019). Profile of factors influencing academic motivation among grade 6 and 7 learners at state school. *South African Journal of Childhood Education, 9(*1).

Yin, R. K. (2003). *Case Study Research: Design and Methods* (3rd ed.). London: Sage Publications.

Yulandari, T. I. (2012). Pengembangan Lembar Kerja Siswa Materi Trigonometri untuk Siswa SMA Kelas X dengan Metode Penemuan Terbimbing. *Journal Online Universitas Negeri Malang, 1*(2).

Zuze, L., Reddy, B.V., Visser, M., Winnaar, L., & Govender, A. (2017). TIMSS 2015 Grade 9 national report: Understanding mathematics and science achievement amongst Grade 9 Learners in South Africa. South Africa: Human Sciences Research Council.

APPENDIX SECTION

Appendix A: Turnitin Certificate

Exploring the use of Rasch analysis in improving the functioning of a mathematics assessment tool

ORIGINALITY REPORT

10%	8%	6%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	sajs.co.za Internet Source	1%
2	repository.up.ac.za Internet Source	1%
3	researchspace.ukzn.ac.za Internet Source	1%
4	David Andrich, Ida Marais. "A Course in Rasch Measurement Theory", Springer Science and Business Media LLC, 2019 Publication	1%
5	mafiadoc.com Internet Source	1%
6	uir.unisa.ac.za Internet Source	<1%
7	www.journalijdr.com Internet Source	<1%
8	hdl.handle.net Internet Source	<1%

Appendix B: Ethical Clearance University of KwaZulu-Natal

UNIVERSITY OF ™
KWAZULU-NATAL
INYUVESI
YAKWAZULU-NATALI

12 December 2017

Mr Harrison Ngirishi 209528282
School of Education – Mathematics, Science and Technology Education
Edgewood Campus

Dear Mr Ngirishi

Protocol reference number: HSS/2209/017D
Project title: Exploring the use of the Rasch Analysis in contribution to the improvement of an assessment instrument for grade 12 mathematics in South Africa.

Expedited Approval

In response to your application dated 21 November 2017, the Humanities & Social Sciences Research Ethics Committee has considered the abovementioned application and the protocol have been granted **FULL APPROVAL.**

Any alteration/s to the approved research protocol i.e. Questionnaire/Interview Schedule, Informed Consent Form, Title of the Project, Location of the Study, Research Approach and Methods must be reviewed and approved through the amendment/modification prior to its implementation. In case you have further queries, please quote the above reference number.

Please note: Research data should be securely stored in the discipline/department for a period of 5 years.

The ethical clearance certificate is only valid for a period of 3 years from the date of issue. Thereafter Recertification must be applied for on an annual basis.

I take this opportunity of wishing you everything of the best with your study.

Yours faithfully

[redacted]

Prof Shenuka Singh (Chair)

/px

cc Supervisor: Prof Sarah Bansilal
cc Academic Leader Research: Dr SB Khoza
cc School Administrator: Ms T Khumalo and Ms P Ncayiyana

Humanities & Social Sciences Research Ethics Committee
Dr Shenuka Singh (Chair)
Westville Campus, Govan Mbeki Building
Postal Address: Private Bag X54001, Durban 4000
Telephone: +27 (0) 31 260 3587/8350/4557 **Facsimile:** +27 (0) 31 260 4609 **Email:** ximbap@ukzn.ac.za / snymanm@ukzn.ac.za / mohunp@ukzn.ac.za
Website: www.ukzn.ac.za

1910 - 2010
100 YEARS OF ACADEMIC EXCELLENCE

Founding Campuses: Edgewood, Howard College, Medical School, Pietermaritzburg, Westville

Appendix C: Ethical Clearance DoBE

education

Department:
Education
PROVINCE OF KWAZULU-NATAL

Enquiries: Phindile Duma	Tel: 033 392 1041	Ref.:2/4/8/1342

Mr H Ngirishi
71 Pardy Road
Isipingo Hills
4133

Dear Mr Ngirishi

PERMISSION TO CONDUCT RESEARCH IN THE KZN DoE INSTITUTIONS

Your application to conduct research entitled: **"THE USE OF THE RASCH ANALYSIS TO IMPROVE THE VALIDITY OF AN ASSESSMENT INSTRUMENT FOR GRADE 12 MATHEMATICS",** in the KwaZulu-Natal Department of Education Institutions has been approved. The conditions of the approval are as follows:

1. The researcher will make all the arrangements concerning the research and interviews.
2. The researcher must ensure that Educator and learning programmes are not interrupted.
3. Interviews are not conducted during the time of writing examinations in schools.
4. Learners, Educators, Schools and Institutions are not identifiable in any way from the results of the research.
5. A copy of this letter is submitted to District Managers, Principals and Heads of Institutions where the Intended research and interviews are to be conducted.
6. The period of investigation is limited to the period from 22 September 2017 to 09 July 2020.
7. Your research and interviews will be limited to the schools you have proposed and approved by the Head of Department. Please note that Principals, Educators, Departmental Officials and Learners are under no obligation to participate or assist you in your investigation.
8. Should you wish to extend the period of your survey at the school(s), please contact Miss Connie Kehologile at the contact numbers below
9. Upon completion of the research, a brief summary of the findings, recommendations or a full report/dissertation/thesis must be submitted to the research office of the Department. Please address it to The Office of the HOD, Private Bag X9137, Pietermaritzburg, 3200.
10. Please note that your research and interviews will be limited to schools and institutions in KwaZulu-Natal Department of Education.

UMlazi District

[redacted]

Dr. EV Nzama
Head of Department: Education
Date: 26 September 2017

...Championing Quality Education - Creating and Securing a Brighter Future

KWAZULU-NATAL DEPARTMENT OF EDUCATION
Postal Address: Private Bag X9137 • Pietermaritzburg • 3200 • Republic of South Africa
Physical Address: 247 Burger Street • Anton Lembede Building • Pietermaritzburg • 3201
Tel.: +27 33 392 1004/41 • **Fax.:** +27 033 392 1203• **Email:**Phindile.Duma@kzndoe.gov.za •**Web:**www.kzneducation.gov.za
Facebook: KZNDOE....Twitter: @DBE_KZN....Instagram: kzn_education....Youtube:kzndoe

Appendix D: Editor's Letter

Leverne Gething, M.Phil., t/a WHIZZ@WORDS
PO Box 1155, Milnerton 7435; cell 072 212 5417
e-mail: leverne@eject.co.za

19 October 2021

Declaration of editing of a PhD thesis for UKZN

TITLE: The application of Rasch measurement theory to improve the functioning of a mathematics assessment instrument.

I hereby declare that I carried out language editing of the above thesis on behalf of Harrison Ngirishi.

I am a professional writer and editor with many years of experience (e.g. 5 years on *SA Medical Journal*, 10 years heading the corporate communication division at the SA Medical Research Council), who specialises in Science and Technology editing - but am adept at editing in many different subject areas. I have edited a great deal of work for various academic journals, universities and publishers.

I am a full member of the South African Freelancers' Association as well as of the Professional Editors' Association.

Yours sincerely

[redacted]

LEVERNE GETHING leverne@eject.co.za

Appendix E: Consent Letter for School Governing Body and Principal

School of Education
Private Bag X03
Ashwood, 3605, South Africa
Telephone +27 (0) 031 2603919/3440/3436
Facsimile +27 (0) 031 260 7594
Emails: Ncayiyanap@ukzn.ac.za/Khumalot9@ukzn.ac.za/Ngcobom4@ukzn.ac.za

Gate keepers informed consent form

Chair of School Governing Body/Principal

Academic research: Request for permission to conduct a research study in your school.

My proposed research title is:

<u>Exploring the use of Rasch Analysis in contribution to the improvement of an assessment instrument for grade 12 mathematics in South Africa.</u>

My name is Mr H Ngirishi from the School of Education, University of KwaZulu-Natal, Edgewood campus. I am an educator who is currently studying towards a PhD Degree in Education in the field of Mathematics Education at the University of KwaZulu Natal. Your school has been identified through voluntary inclusion process as a possible site of research for this project to produce some data that will help us understand situated realities of schooling and its impact on learner performance.

The learners of your school are being invited to participate in this research project that is aimed at investigating the assessment of mathematics using the Rasch Measurement theory. Learners' mathematics trial examination paper 2 scripts will be photocopied and used in the study.

The purpose of this study is to explore the use of Rasch Analysis in contribution to the improvement of an assessment instrument for grade 12 mathematics by investigating the functioning of the items, allocation of marks, the grading of proficiency along a continuum, the outcomes of both the initial scoring memorandum and the revisions and together with the educational considerations.

I humbly seek permission to conduct the above mentioned research study in your school.

I would like to conduct this research from October-December 2017. The school and learners who partake in this study will do so, on a voluntarily basis and confidentiality and anonymity will be ensured. The participants have no obligation to participate in this research and may withdraw from it any point. I also hereby undertake that the name of your school or the learners will not be mentioned in the subsequent theses. I will ensure that normal learning and teaching is not disrupted in any way whatsoever whilst conducting this research study.

I will to share my findings and feedback on this research with you and your staff members and the learners. The information acquired from this research study, will be accessible to the Department of Education, as well as school managers. I hope that the information gained from this research will be of great help to you and your learners and that together we might find solutions for our current problems we face in the teaching and learning of mathematics.

For further information regarding this study, feel free to contact my supervisor. My supervisor is Prof Sarah Bansilal who lectures at the University of KwaZulu Natal, Edgewood Campus, who can be contacted on 031- 2603451.
The reply could be sent to me by e-mail at: ngirishih@gmail.com

Thank you for your assistance in this matter

Yours faithfully
H. Ngirishi (Mr)   Student number: 209528282

Contact No's : 083 923 6816

DECLARATION

I………………………………………………………………………… (Full names of the Chairperson of the School Governing Body/Principal) hereby confirm that I understand the contents of this document and the nature of the research project.
I do/do not consent for my school to participate in the research project.

I understand that I am at liberty to withdraw my school from participating in the project at any time, should I so desire.

SIGNATURE OF CHAIRPERSON OF SGB/PRINCIPAL
DATE:__________________

Appendix F: Consent Letter for Teachers

School of Education
Private Bag X03
Ashwood, 3605, South Africa
Telephone +27 (0) 031 2603919/3440/3436
Facsimile +27 (0) 031 260 7594
Emails: Ncayiyanap@ukzn.ac.za/Khumalot9@ukzn.ac.za/Ngcobom4@ukzn.ac.za

Dear Mathematics Teacher
Informed Consent (TEACHERS)

Title of the Project: Exploring the use of Rasch Analysis in contribution to the improvement of an assessment instrument for grade 12 mathematics in South Africa.

I am a PhD (Mathematics) student at the University of KwaZulu Natal. I am currently conducting a research on the use of the Rasch measurement theory to improve the validity of assessment in mathematics at grade 12. Participation in this study is voluntary. The study involves photocopying learners' scripts for their trial examinations (mathematics paper 2), recording the data from the scripts and analyzing it using Rasch analysis. The results from the Rasch Analsyis will be used as a baseline for semi-structured interviews with the teachers to get more insight and understanding of the trends and patterns which will arise from the analysis. The interviews will be tape-recorded.
Confidentiality is assured as names will not be used in the analysis or the report.
Queries or concerns can be kindly directed to the following persons:

Researcher: Mr H. Ngirishi Cell: 0839236816 Email: ngirishih@gmail.com	Supervisor: Prof S. Bansilal Tel: 031-260 3451 Cell: 083 279 5916 Email: Bansilals@ukzn.ac.za
HSSREC contact details:	P. Ximba Tel: 031 260 3587 Email: 260 3587

I ……………………………………………………………………(full names of participant)
Hereby confirm that I understand the contents of this document and the nature of the research project, and consent to participation in the research project.
I do/do not consent to be audio recorded.
I understand that I am at liberty to withdraw from the project at any time, should I so desire.

Signature of Participant　　　　　　　　Date
…………………………　　　　　　　……………………………………

Appendix G: Consent Letter for Learners

School of Education
Private Bag X03
Ashwood, 3605, South Africa
Telephone +27 (0) 031 2603919/3440/3436
Facsimile +27 (0) 031 260 7594
Emails: Ncayiyanap@ukzn.ac.za/Khumalot9@ukzn.ac.za/Ngcobom4@ukzn.ac.za

Dear Mathematics Learner
Informed Consent (Learner)

Title of the Project: Exploring the use of Rasch Analysis in contribution to the improvement of an assessment instrument for grade 12 mathematics in South Africa.

I am a PhD (Mathematics) student at the University of KwaZulu Natal. I am currently conducting a research on the use of the Rasch measurement theory to improve the validity of assessment in mathematics at grade 12. Participation in this study is voluntary. The study involves collecting learners' scripts for their trial examinations (mathematics paper 2), recording the data from the scripts and analyzing it using Rasch analysis. The results from the Rasch Analsyis will be used as a baseline for semi-structured interviews with the teachers to get more insight and understanding of the trends and patterns which will arise from the analysis. The interviews will be tape-recorded.
Confidentiality is assured as names will not be used in the analysis or the report.
Queries or concerns can be kindly directed to the following persons:

Researcher: Mr H. Ngirishi Cell: 0839236816 Email: ngirishih@gmail.com	Supervisor: Prof S. Bansilal Tel: 031-260 3451 Cell: 083 279 5916 Email: Bansilals@ukzn.ac.za
HSSREC contact details:	P. Ximba Tel: 031 260 3587 Email: ximba@ukzn.ac.za

---

I ……………………………………………………………………(full names of participant)
Hereby confirm that I understand the contents of this document and the nature of the research project.
I do/do not consent for the use of my script in the research project.

I understand that I am at liberty to withdraw from the project at any time, should I so desire.

Signature of Participant ………………………… Date ……………………………………

Appendix H: Summary of Rescoring

Summary of Rescoring							
Item	**Item with comments**	**Fit residual (FR)**	**Decision run 1**	**Results of rescoring 1 on FR**	**New score/ comment**	**Results of rescore 2**	**DIF**
2	1.2 (4 Marks) ICC show extreme misfit Disordered thresholds; Chi square is 0.00, less than significant value p<0.05	4.087	Rescore 0,1-0 2,3-1 4-2	F.R = 2.144 ICC show better fit Chi square = 0.428, indicating a better fit	2 Marks		
7	2.2.2 (2 Marks) ICC not aligned to theoretical curve Disordered thresholds Category 1 not working	2.962	Rescore 1,2-1	F.R =3.981 which outside the recommended range ICC still not aligned to theoretical curve	1 Mark	F.R = 0.77 0,1-0	
18	4.1.4 (4 Marks) Disordered thresholds	1.1474	1,2,3-1 4-2	F.R =-0.358	2 Marks		
21	5.1 (5 Marks) Chi square 0.00, less than the significant value ICC show misfit Disordered thresholds	4.386	1,2,3-1 4,5-2	F.R = -0.218 Ordered thresholds	2 Marks		
25	6.1.1 (3 Marks) Disordered thresholds Chi square 0.002 less than the significant value	-2.657	2,3-2	F.R = -2.316 Category 1 not working Disordered thresholds	Left as it is		

	ICCs show observed proportions not aligned to theoretical curve						
27	6.2 (6 Marks) Disordered thresholds Chi square 0.000 less than the significant value ICC show misfit	2.841	0,1-1 2,3-2 4,5-2 6-3	F.R = 1.833 Chi square = 0.477 indicating a better fit ICC show better ift All categories now working as intended	3 Marks		
29	7 (9 Marks) Scores of 6,7,8 not working	2.216	1,2-1 3,4-2 5,6-3 7,8,9-4	F.R = 0.739 Ordered thresholds,	4 Marks		
10	3.2 (2 Marks) Disordered thresholds, Score of 1 not working ICC shows mis-fit .Over discrimination/response dependence	-2.817	Leave as it is	F.R = -1.967 Without any rescoring the FR has improved as a result of changes to other items	Left as is		
11	3.3 (2 Marks) Disordered thresholds Chi square probability=0.00 ICC show over discrimination	-4.084	Leave as it is	F.R = -3.179	Left as it was		
15	4.1.1 (5 Marks) Disordered thresholds Chi-square probability =0.00. ICC shows misfit	6.1176	Leave as it is	F.R =6.774	Leave as it is		

	Observed proportions flatter than theoretical curve						
16	4.1.2 (2 Marks) Disordered thresholds Chi-probability =0.00 ICC shows misfit, observed proportions are steeper than theoretical curve	-2.955	1,2-1	F.R = -1.199	1 mark		
31	8.2 (3 Marks) Disordered thresholds Chi-square =0.006 Steeper observed proportion when compared to theoretical curve	-2.811	Leave as it is	-2.296 Without any rescoring the FR has improved as a result of changes to other items	Leave as it is		
32	8.3 (4 Marks) Disordered thresholds Chi-square probability = ICCs show mis-fit	-3.076	Leave as it is	-2.397	2 Marks		
	Total score = 150			Rescored total = 132			

Appendix I: Teacher Questionnaire

Teacher Questionnaire

Follow up to the Trial Examination 2017 Mathematics Paper 2 Findings

1. From the analysis done, learners found this trial examination to be generally difficulty. Can you please comment on this and give reasons why it might be so?

   ____
   ____
   ____

2. Despite the statistics questions (Items/Questions 1.1, 1.3.1 and 1.3.2) being placed among the easy questions (levels 1 and 2 on the cognitive levels), the results showed that they were among the top 7 most difficult questions. Why do you think these statistics questions were harder than expected?

   ____
   ____
   ____

3. The results also showed that the 6 easiest items/questions were dominated by Analytical Geometry questions (Questions 3.1, 3.2, 3.4 and 3.6) which were level 2 questions on the cognitive levels. Can you give reasons on what you might have done that made learners to perform better in these items.

   ____
   ____
   ____

   In general the analysis showed that many items (questions) which were at the cognitive levels 1 and 2 were not always found to be easy by the learners. What are the possible causes of this?

   ____
   ____
   ____

4. In general the analysis also showed that many items (questions) which were at the cognitive levels 3 and 4 were not found to be always difficult by the learners (For example Questions 3.3; 4.1.1; 4.1.4; 5.3; 6.3 and 9.2). What do you think are some reasons for this?

   ____
   ____
   ____

5. There were many instances where marking was not consistent with the marking guideline, for example, very few educators in their marking were using the continuous

accuracy (CA) factor hence depriving learners from getting potential marks (Refer to Question 2.2.2 for example).
Can you give your views on this issue and why teachers sometimes do not adhere to this aspect?

6. In cases where an item carries 4 marks or ha 4 categories and all the marks were accuracy marks (A), for example Question 4.1.4, results showed that the majority of learners will either get zero or get all marks. Very few learners got marks in between. Do you think this is a fair or unfair situation? Please explain.

7. The analysis showed that some questions, learners who had the same total scores, from different schools did not perform in the same way on most of the items. For example learners from school A did better on particular items than those from school B, while for some items it was the other way around. What are some of the contributing factors for this observation?

8. The analysis indicated items where learners who used English as the first home language, performed better than the English second language speakers (Items 2.1, 2.2.1, 2.2.3, 3.5 and 6.1.1). However a different scenario was also observed for Questions 4.1.1, 5.1, .1 and 10.1, where the English second language speakers seemed to be performing better than the English first language speakers.
What are some of the factors contributing to these findings?

Appendix J: KZN Mathematics Trial Paper 2 Question Paper 2017

Education

KwaZulu-Natal Department of Education

REPUBLIC OF SOUTH AFRICA

MATHEMATICS P2

PREPARATORY EXAMINATION

SEPTEMBER 2017

NATIONAL SENIOR CERTIFICATE

GRADE 12

MARKS: 150

TIME: 3 hours

This question paper consists of 12 pages, including information sheet.

Please Turn Over

INSTRUCTIONS AND INFORMATION

Read the following instructions carefully before answering the questions.

1. This question paper consists of 10 questions.
2. Answer ALL the questions in the SPECIAL ANSWER BOOK.
3. Clearly show ALL calculations, diagrams, graphs, et cetera that you have used in determining your answers.
4. Answers only will not necessarily be awarded full marks.
5. An approved calculator (non-programmable and non-graphical) may be used, unless stated otherwise.
6. If necessary, answers should be rounded off to TWO decimal places, unless stated otherwise.
7. Diagrams are NOT necessarily drawn to scale.

QUESTION 1

The table below shows the heights of palm trees in a park.

HEIGHT IN CM	NUMBER OF PALM TREES
$120 < x \leq 135$	1
$135 < x \leq 150$	15
$150 < x \leq 165$	45
$165 < x \leq 180$	28
$180 < x \leq 195$	1

1.1 Determine the estimated mean height of the palm trees in the park. (2)

1.2 Draw an ogive to represent this data. (4)

1.3 Use your ogive curve to estimate the:

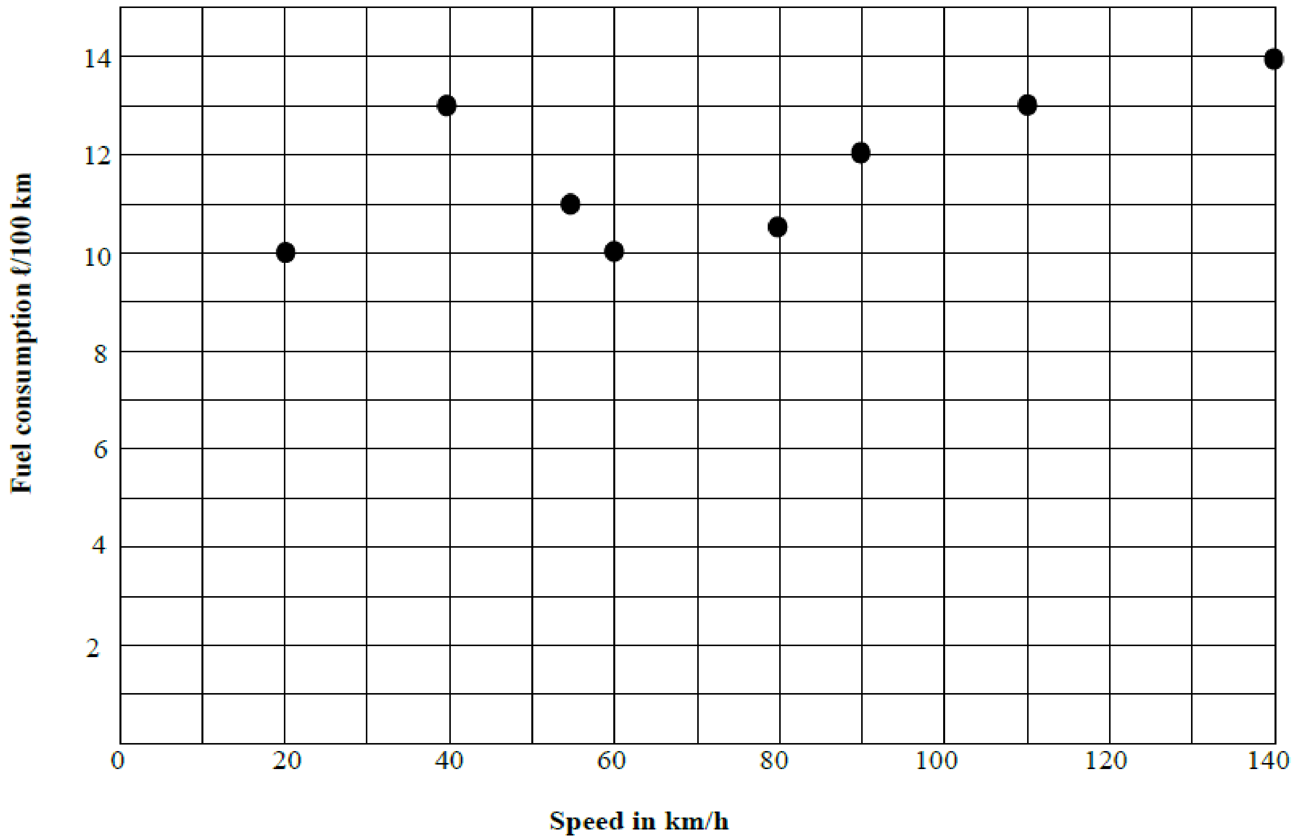
1.3.1 median height of the palm tree. (2)

1.3.2 interquartile range (IQR). (3)

[11]

QUESTION 2

The scatter plot below shows the fuel consumption versus the speed of a motor car.



2.1 Identify an outlier. Write down its co-ordinates. (1)

2.2 Determine:

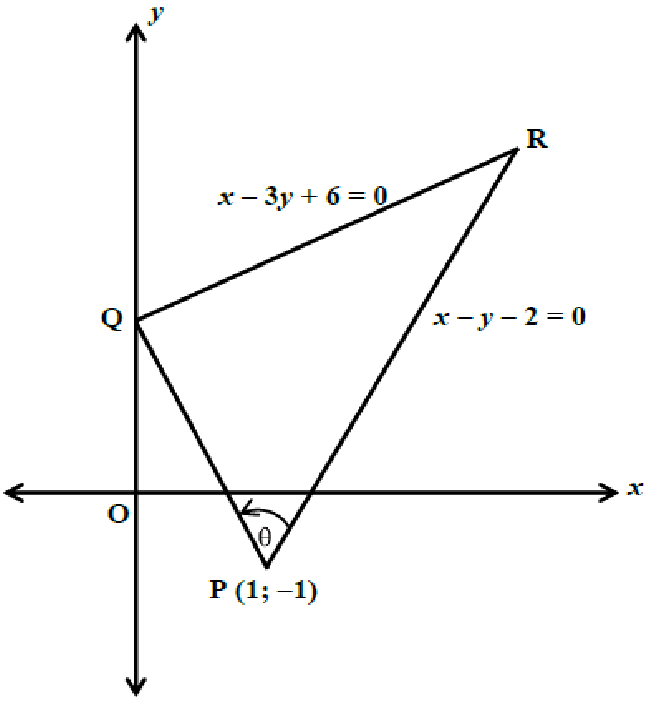
2.2.1 the equation of the regression line excluding the outlier. (3)

2.2.2 the correlation coefficient excluding the outlier and explain the type of correlation. (2)

2.2.3 the average fuel consumption of the motor car. (2)
[8]

QUESTION 3

In the figure below, PQR is a triangle with P$(1\ ;-1)$. Q is a point on the y-axis. The equations of QR and PR are $x-3y+6=0$ and $x-y-2=0$ respectively. Given $Q\hat{P}R=\theta$.

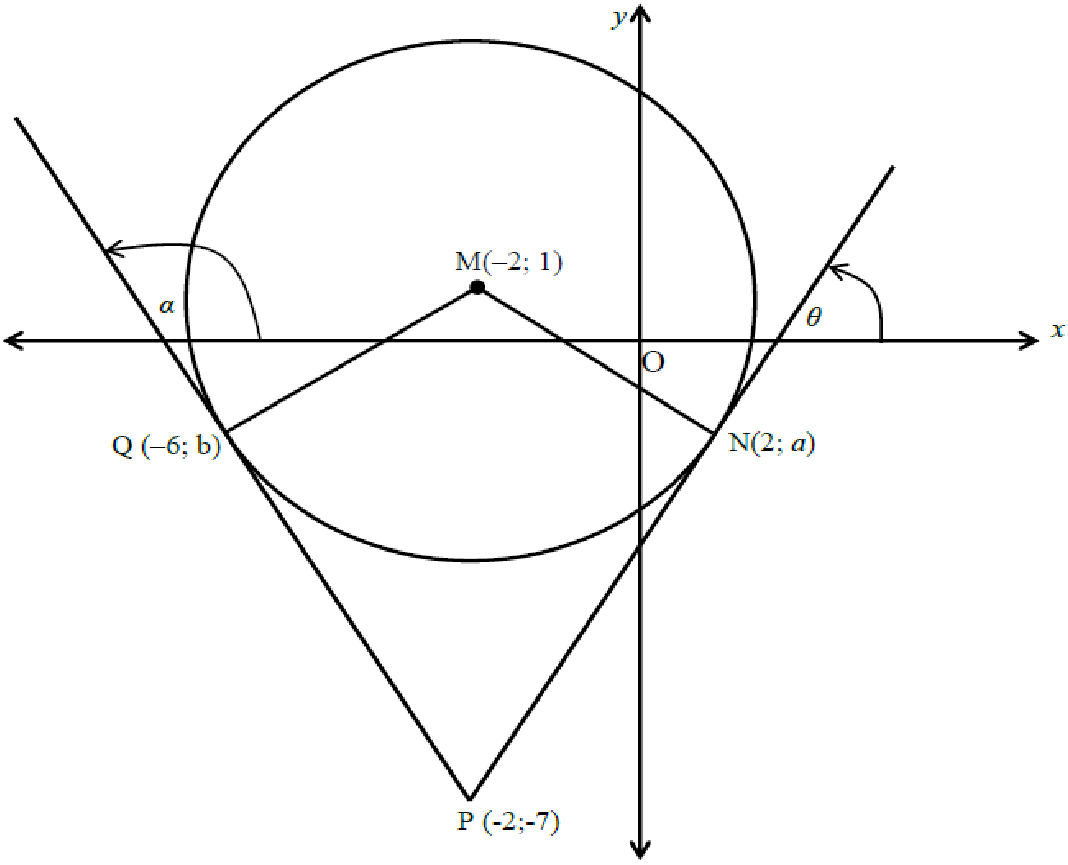


3.1	Show that the co-ordinates of Q are $(0\ ;2)$.	(2)
3.2	Write down the gradient of QR.	(2)
3.3	Prove that $P\hat{Q}R=90^\circ$.	(2)
3.4	Calculate the co-ordinates of R.	(3)
3.5	Calculate the area of Δ PQR.	(4)
3.6	Calculate the length of PR. (leave your answer in the simplest surd form).	(2)

[15]

QUESTION 4

4.1 In the diagram below, MN is a radius of a circle with centre M(–2 ; 1). The co-ordinates of N are (2 ; a) and $a < 0$. The co-ordinates of P are (-2'-7). PQ and PN are tangents to the circle at Q and N respectively.
The coordinates of Q is (–6 ; b). PM is parallel to the y – axis.



4.1.1 Deduce that $a = -3$. Show all your workings. (5)

4.1.2 Determine the equation of the circle. (2)

4.1.3 Calculate the gradient of the tangents at Q and N. (4)

4.1.4 If the angle of inclination of lines PN and PQ are θ and α respectively, without using a calculator, show that $\tan^2 \alpha + \tan^2 \theta = 2$. (4)

4.2 The circle defined by $(x+1)^2 + (y-1)^2 = 16$ has centre C and circle defined by $x^2 + y^2 - 2y = 8$ has centre D.

4.2.1 Show that the two circles touch each other internally. (5)

4.2.2 Determine the equation of the common tangent to the circles. (2)

[22]

QUESTION 5

5.1 Show, without using a calculator, that

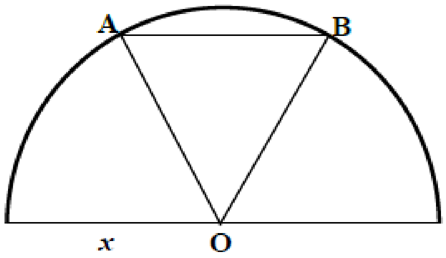
$$\sqrt{2}\cos(-45°) + \cos 210° - \tan 840° = \frac{2+\sqrt{3}}{2}.$$ (5)

5.2 If $\sin\theta = \frac{2n}{n^2+1}$, $n > 1$ and $0° < \theta < 90°$, prove that $\frac{1+\sin\theta}{\cos\theta} = \frac{n+1}{n-1}$. (7)

5.3 Prove the identity:

$$\frac{\sin 2x}{\cos x\,(1-\cos 2x)\left(1+\frac{1}{\tan^2 x}\right)} = \sin x$$ (5)

5.4 In the figure below, semi-circle with centre O has radius x. Points A and B are on the circumference of circle. Calculate in terms of x the maximum area of ΔAOB.



(3)

[20]

QUESTION 6

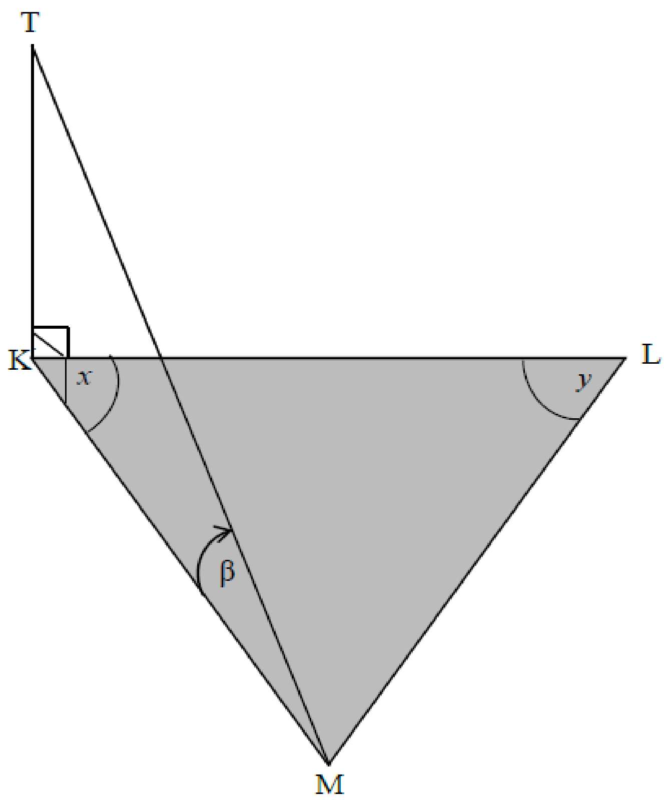
6.1 6.1.1 Write down an expansion for $\sin(x+30^\circ)$. Leave your answer in surd form. (3)

6.1.2 Hence, solve the equation:

$$2\cos x = \sin(x+30^\circ) \text{ for } x \in [-180^\circ; 270^\circ]$$ (7)

6.2 On the axes provided, in your answer book sketch the graphs of $f(x) = 2\cos x$ and $g(x) = \sin(x+30^\circ)$ for the interval $x \in [-180^\circ; 270^\circ]$. (6)

6.3 TK is a pole with K in the same horizontal plane as L and M. The angle of elevation of T from M is β. $L\hat{K}M = x$ and $K\hat{L}M = y$.

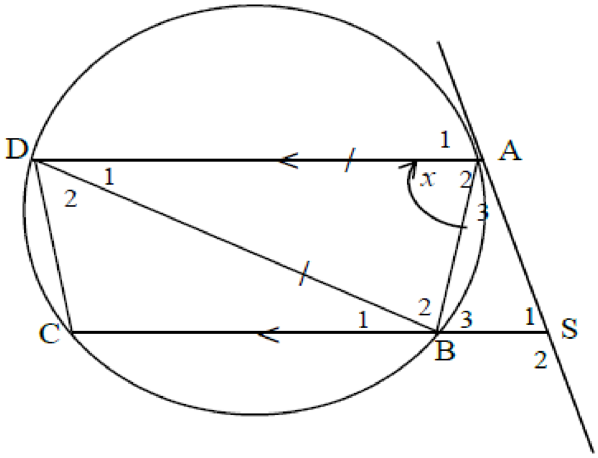


Show that $KT = \dfrac{KL \sin y \,.\, \tan\beta}{\sin(x+y)}$ (5)

[21]

QUESTION 7

7. Refer to the figure below. ABCD is a cyclic quadrilateral. AS is a tangent to the circle at A. CB is produced to S. AD || SBC; AD = BD; $\hat{A}_2 = x$.

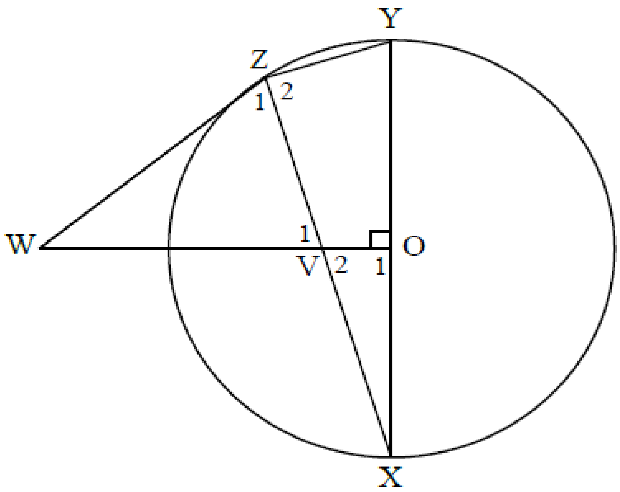


Write down, with reasons, FIVE other angles each equal to x. (9)

[9]

QUESTION 8

In the figure below, O is the centre of the circle ZYX. WO intersects XZ at V and WZ is a tangent to the circle at Z. WO ⊥ XY.



8.1 Prove that VOYZ is a cyclic quadrilateral. (3)

8.2 Prove that ΔWVZ is isosceles. (3)

8.3 Prove that ΔXOV ||| ΔXZY. (4)

8.4 Calculate VO, if XZ = 16 units, ZY = 12 units and the radius of the circle is 10 units. (3)

[13]

QUESTION 9

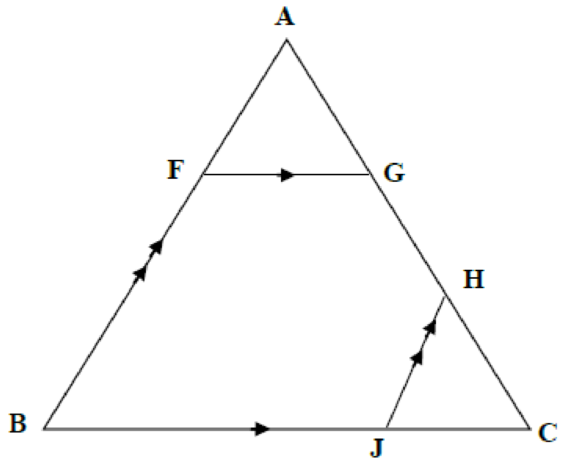
In the diagram below FG || BC, HJ || AB.
FA = 3 units
FB = 9 units
AG = 2 units

CJ: JB = 1:3

Calculate (stating reasons) the lengths of:

9.1 GC (4)

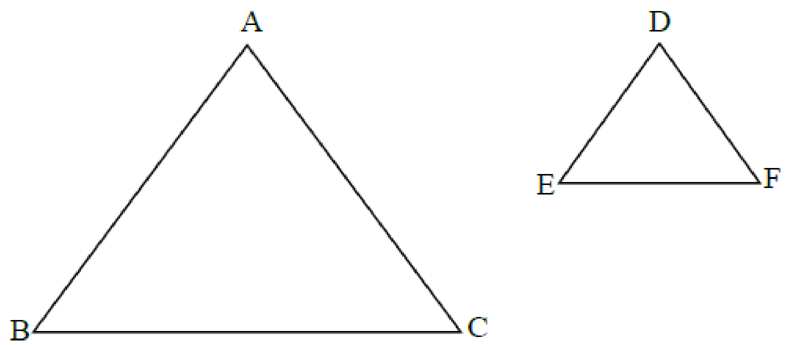
9.2 GH (5)



[9]

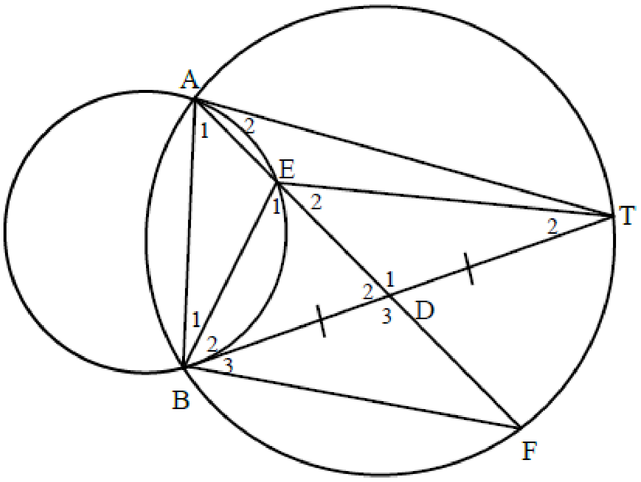
QUESTION 10

10.1 ΔABC and ΔDEF with $\hat{A} = \hat{D}$; $\hat{B} = \hat{E}$ and $\hat{C} = \hat{F}$ are shown below.



Prove that: $\dfrac{AB}{DE} = \dfrac{AC}{DF} = \dfrac{BC}{EF}$ (7)

10.2 In the figure below, two circles intersect at A and B. TB is a tangent to the smaller circle at B. The line through D and A cuts the circles at E and F such that BD = DT. AB, BE and EA are joined.



10.2.1 Prove that $\Delta TDA \parallel\!| \ \Delta FDB$. (4)

10.2.2 Prove that $TB^2 = 4FD \cdot AD$. (2)

10.2.3 Prove that $BD^2 = DE.AD$. (4)

10.2.4 Deduce that ET = BF. (5)
[22]

TOTAL MARKS: 150

Appendix K: KZN Mathematics Trial Examination Paper 2 Marking Guideline

Education

KwaZulu-Natal Department of Education

REPUBLIC OF SOUTH AFRICA

MATHEMATICS P2

MARKING GUIDELINE

PREPERATORY EXAMINATION 2017

NATIONAL
SENIOR CERTIFICATE

GRADE 12

MARKS: 150

N.B. This marking guideline consists of 14 pages.

Please Turn Over

QUESTION 1

1.1

	f	Midpoint(x)	$f.x$
$120<x\leq 135$	1	127.5	127.5
$135<x\leq 150$	15	142.5	2137.5
$150<x\leq 165$	45	157.5	7087.5
$165<x\leq 180$	28	172.5	4830
$180<x\leq 195$	1	187.5	187.5
TOTAL	90		14370

159,6

OR

$$Mean\ \ height = \frac{14370}{90}$$
$$= 159{,}67$$

Marks:

Answer only – full marks

✓A table method

✓CA answer (2)

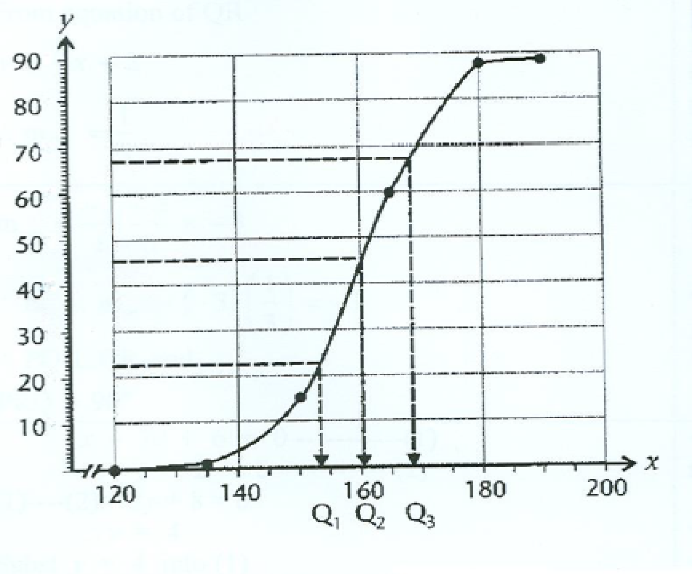
OR

✓14370

✓ answer

(2)

1.2



Marks:

✓ A shape
✓ A (120;0)
✓ A (60;165)
✓ A for any other correct plotted points

(4)

1.3

1.3.1 median height = 161 cm

Marks:

✓✓ A A Answer
(Accept : 160 – 163)

(2)

1.3.2

IQR = $Q_3 - Q_1$
= 169 – 152
= 17 cm

Marks:

✓ CA Q_3
✓ CA Q_1
✓CA answer
(Accept 1 mark deviation for Q_1 and Q_3)

(3)

[11]

QUESTION 2

2.1	(40; 13)	✓ A (40; 13) (1)
2.2.1	$y = 0{,}04\,x + 8{,}64$	✓CA $0{,}04x$ ✓ CA 8,64 ✓ CA equation (3)
2.2.2	$r = 0{,}91$ strong, positive correlation	✓ CA 0,91 ✓ CA justification (2)
2.2.3	The car will be using about 11,69 *l*/100 km.	A✓CA✓ average 11,69 *l*/100 km (2)
		[8]

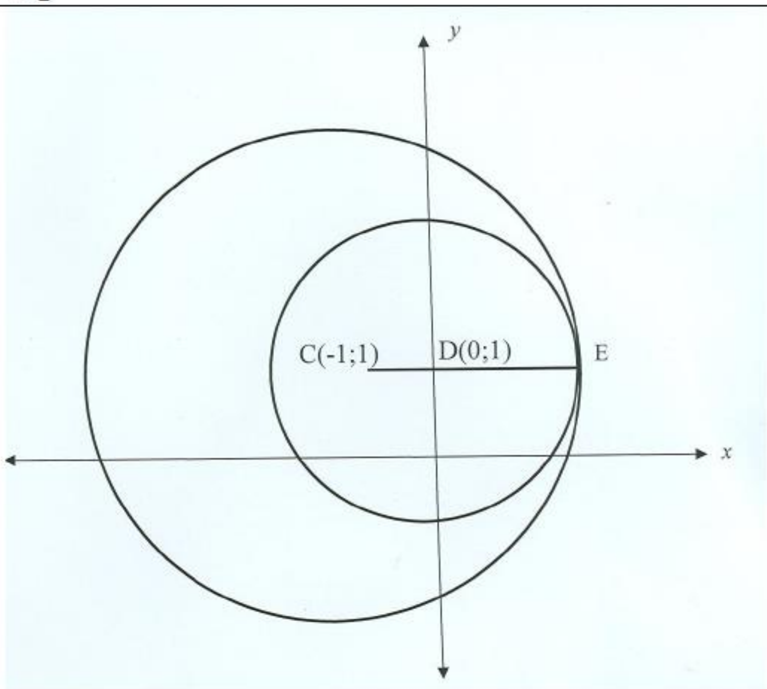
QUESTION 3

3.1	$0-3y+6=0$ $-3y=-6$ $y=2$ $\therefore$ Q(0;2)	✓A subst $x=0$ ✓A $y=2$ (2)
3.2	$y=\frac{1}{3}x+2$ $\therefore m_{QR}=\frac{1}{3}$	✓ A writing in standard form ✓ A answer (2)
3.3	$m_{PQ}=\frac{-1-2}{1-0}=-3$ $\therefore m_{PQ}\times m_{QR}=(-3)\left(\frac{1}{3}\right)=-1$ $\therefore$ PQ $\perp$ QR Thus $P\hat{Q}R=90°$	✓ A gradient of PQ ✓ A products of gradients (2)
3.4	$x-3y+6=0$........(1) $x-y-2=0$...........(2) $(1)-(2): -2y+8=0$ $y=4$ subst $y=4$ into (1) $x-3(4)+6=0$ $x-12+6=0$ $x=6$ R(6;4)	✓ M solving both equations simultaneously ✓ CA substituting $y=4$ ✓ CA $x=6$ (provided R is in first quadrant) (3)

3.5	$QR=\sqrt{(0-6)^2+(2-4)^2}=2\sqrt{10}$ $PQ=\sqrt{(0-1)^2+(2+1)^2}=\sqrt{10}$ Area of $\Delta PQR=\frac{1}{2}QR\times PQ$ $=\frac{1}{2}(2\sqrt{10})(\sqrt{10})$ $=10$ square units	✓ CA QR value ✓ A PQ value ✓ CA correct substitution into area formula ✓CA answer (4)
3.6	$PR^2=(6-1)^2+(4-(-1))^2$ $=5^2+5^2=50$ $\therefore PR=\sqrt{50}=5\sqrt{2}$ units	✓CA correct subst. into distance formula ✓ CA answer (2)
		[15]

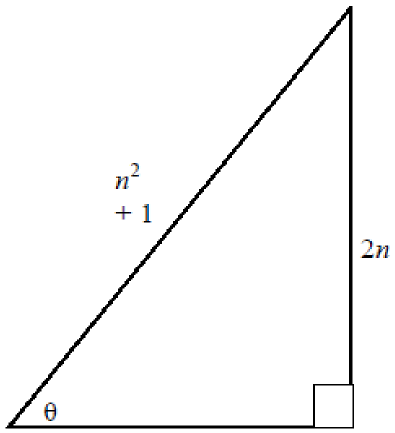
QUESTION 4

4.1.1	$m_{MN}\times m_{NP}=-1$ Radius ⊥ Tangent $\left(\frac{a-1}{2+2}\right)\times\left(\frac{a+7}{2+2}\right)=-1$ $\left(\frac{a-1}{4}\right)\left(\frac{a+7}{4}\right)=-1$ $\frac{a^2-a+7a-7}{16}=-1$ $a^2+6a-7=-16$ $a^2-6a+9=0$ $(a+3)(a+3)=0$ $\therefore a=-3$	✓ A $m_{MN}\times m_{NP}=-1$ ✓ A substitution ✓A simplification ✓A standard form ✓A factorization (5)
4.1.2	$MN^2=r^2=(-2-2)^2+(1+3)^2=32$ $(x+2)^2+(y-1)^2=32$	✓ A value of radius ✓CA equation of circle (2)
4.1.3	M(− 2; 1) N(2 ; − 3) Q(− 6; − 3) $m_{MN}=\frac{1+3}{-2-2}=-\frac{4}{4}=-1$ $\therefore m_{PN}=1$ [tangent at N] $m_{MQ}=\frac{1+3}{-2+6}=\frac{4}{4}=1$ $\therefore m_{PQ}=-1$ [tangent at Q]	✓A gradient of MN ✓CA gradient of PN ✓ A gradient of MQ ✓CA gradient of PQ (4)

4.1.4	$\tan\theta = 1$ $\tan\alpha = -1$ $\tan^2\alpha + \tan^2\theta$ $= (1)^2 + (-1)^2$ $= 2$	✓A $\tan\theta = 1$ ✓A $\tan\alpha = -1$ ✓A Substitution (1) ✓A Substitution (-1) (4)
4.2.1	 Circle centre C: C(−1 ; 1) and radius = 4 units Circle centre D: $x^2 + (y-1)^2 = 9$ D(0;1) and radius = 3 units Difference of radii = 4 − 3 = 1 unit $CD^2 = (-1-0)^2 + (1-1)^2 = 1$ $\therefore CD = 1$ unit ($CD < r_1 + r_2$ therefor the circles touch internally) Therefore the two circles touch each other internally	✓A centre and radius of circle centre C ✓ A equation of circle center D ✓ CA centre and radius of circle centre D ✓A Difference of radii = 1 unit ✓length of CD (5)

4.2.2	$x^2+2x+1+y^2-2y+1=16$ $x^2+2x+y^2-2y=14$..........(1) $x^2+y^2\quad -2y=8$ ....(2) $(1)-(2)$ $2x=6$ $x=3$ $\therefore$ The equation of the common tangent is: $x=3$	✓M solving simultaneously ✓A answer (2)
		[22]

QUESTION 5

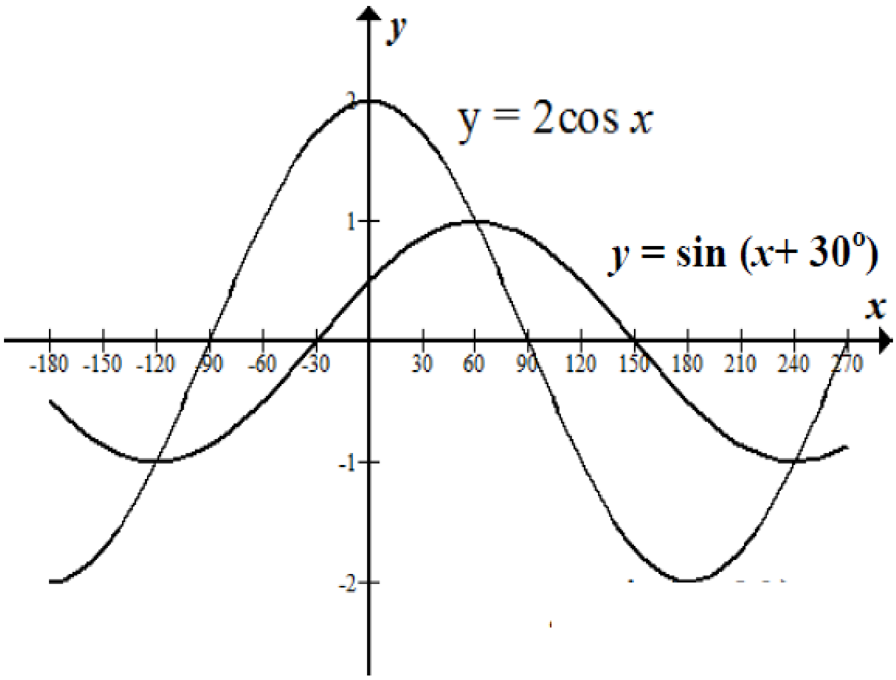
5.1	$\sqrt{2}\cos\left(-45^{\circ}\right)+\cos 210^{\circ}-\tan 840^{\circ}$ $=\sqrt{2}\cos 45^{\circ}+(-\cos 30^{\circ})-(-\tan 60^{\circ})$ $= \sqrt{2}\cdot\frac{\sqrt{2}}{2}-\cos 30^{\circ}+\tan 60^{\circ}$ $= 1-\frac{\sqrt{3}}{2}+\frac{\sqrt{3}}{1}$ $= \frac{2-\sqrt{3}+2\sqrt{3}}{2}$ $= \frac{2+\sqrt{3}}{2}$	✓A $\cos 45^{\circ}$ ✓ A $-\cos 30^{\circ}$ ✓ A $-\tan 60^{\circ}$ ✓A substitution of special angle values ✓ A simplification (5)
5.2	 $\sqrt{(n^2+1)-(2n)^2}$ $=\sqrt{n^4-2n^2+1}$ $=\sqrt{(n^2-1)^2}$ $=n^2-1$ $\frac{1+\sin\theta}{\cos\theta}=\frac{1+\frac{2n}{n^2+1}}{\frac{n^2-1}{n^2+1}}$ $=\frac{n^2+1+2n}{n^2+1}\times\frac{n^2+1}{n^2-1}$ $=\frac{(n+1)^2}{n^2-1}$ $=\frac{(n+1)(n+1)}{(n-1)(n+1)}=\frac{n+1}{n-1}$	✓ A Drawing the sketch ✓ A calculating remaining side ✓ A substitution sinθ value ✓ A substitution cosθ value ✓numerator simplification ✓factorization : n^2+1+2n ✓factorization : n^2-1 (7)

5.3	$\dfrac{\sin 2x}{\cos x(1-\cos 2x)(1+\dfrac{1}{\tan^2 x})} = \sin x$ LHS $= \dfrac{\sin 2x}{\cos x(1-\cos 2x)(1+\dfrac{1}{\tan^2 x})}$ $= \dfrac{\dfrac{2\sin x\cos x}{\cos x}}{(1-(1-2\sin^2 x))(1+\dfrac{\cos^2 x}{\sin^2 x})}$ $= \dfrac{2\sin x}{(2\sin^2 x)(\dfrac{\sin^2 x+\cos^2 x}{\sin^2 x})}$ $= \dfrac{2\sin x}{2\sin^2 x \cdot \dfrac{1}{\sin^2 x}}$ $= \sin x$ $=$ RHS	✓ A $\sin 2x = 2\sin x\cos x$ ✓ A $\tan x = \dfrac{\sin x}{\cos x}$ ✓ A $\cos 2x = 1-2\sin^2 x$ ✓ A simplification - numerator ✓ A simplification - denominator (5)
5.4	Area Triangle AOB= $\frac{1}{2}$ ab sin $A\hat{O}B$ $= \frac{1}{2}x \cdot x \sin 90°$ (Area maximum if $A\hat{O}B = 90°$) $= \frac{1}{2}x^2 \cdot 1$ $= \frac{1}{2}x^2$	✓ A correct substitution into formula ✓ A Max area if $A\hat{O}B = 90°$ ✓A Answer (3)
		[20]

QUESTION 6

6.1.1	$\sin(x+30^\circ)$ $= \sin x \cos 30^\circ + \cos x \sin 30^\circ$ $= \frac{\sqrt{3}}{2}\sin x + \frac{1}{2}\cos x$	✓ A expanding ✓ A $\cos 30^\circ = \frac{\sqrt{3}}{2}$ ✓ A $\sin 30^\circ = \frac{1}{2}$ (3)
6.1.2	$2\cos x = \frac{\sqrt{3}}{2}\sin x + \frac{1}{2}\cos x$ $\frac{3}{2}\cos x = \frac{\sqrt{3}}{2}\sin x$ $\frac{3}{2}\cos x = \frac{\sqrt{3}}{2}\sin x$ dividing both sides by $\cos x$; $\cos \neq 0$ $\frac{\sin x}{\cos x} = \frac{3}{2} \text{ x } \frac{2}{\sqrt{3}}$ $\tan x = \frac{3}{\sqrt{3}}$ $= \frac{3\sqrt{3}}{3}$ $= \sqrt{3}$ $\therefore \quad x = 60^\circ + k.180^\circ, k \in \mathbb{Z}$ $\therefore \quad x \in \{-120^\circ; 60^\circ; 240^\circ\}$	✓ CA $\frac{3}{2}\cos x$ ✓ A $\tan x$ ✓ CA $\sqrt{3}$ ✓ CA $x = 60^\circ + k.180^\circ, k \in \mathbb{Z}$ ✓ CA 120° ✓ CA 60° ✓ CA 240° (7)

6.2



✓A✓A x - intercepts of EACH graph
✓A✓A - maximum and minimum values of EACH graph
✓A✓A - shape of EACH graph

(6)

6.3	$\frac{KT}{KM} = \tan\beta$ $\therefore$ KT = KM tanβ ....(1) $\frac{KM}{\sin y} = \frac{KL}{\sin(180^\circ - (x+y))}$ $\therefore \quad KM = \frac{KL \sin y}{\sin(x+y)}$ Substituting KM in equation (1) give $KT = \frac{KL \sin y . \tan\beta}{\sin(x+y)}$	✓ writing $\tan\beta = \frac{KT}{KM}$ (A) ✓ A making KT the subject of the formula. ✓ M applying sine rule ✓ A sin (x + y) ✓ A substituting in equation 1 (5)
		[21]

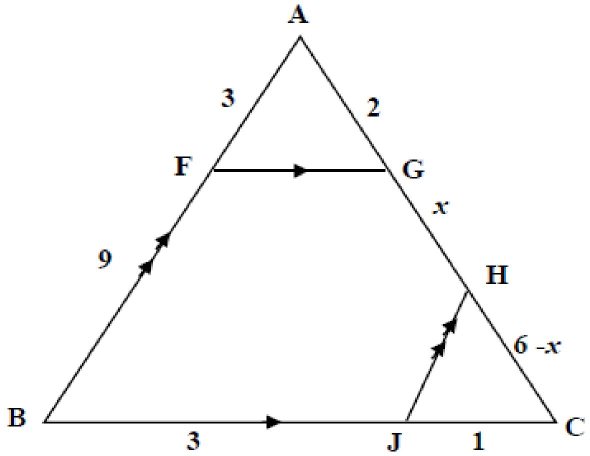
QUESTION 7

7	1. $\hat{B}_2 = x = \hat{A}_2$ (∠s opp = sides) 2. $\hat{A}_1 = x = \hat{B}_2$ ....(tan chord theorem) 3. $\hat{A}_1 = \hat{S}_1 = x$ ....(corres ∠s, DA ∥ CS) 4. $\hat{B}_3 = x$ (alternate angles, DA ∥ CS) 5. $C\hat{D}A = x$ ....(Exterior angle of a cyclic quad)	S✓/R✓ S✓/R✓ S/R✓ S✓/R✓ S✓ R✓ **(All Accuracy)** (Penalize once for not stating parallel lines) (9)
		[9]

QUESTION 8

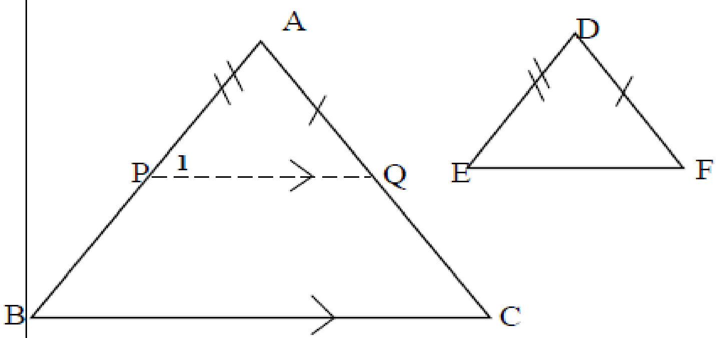
8.1	$\hat{Z}_2 = 90°$ $\hat{Z}_2 = \hat{O}_1$ $\therefore$ VOYZ is a cyclic quadrilateral ... (Converse opp angles of a cyclic quad)	✓S ✓R angle in semi circle each $= 90°$ ✓Reason (3)
8.2	$\hat{Z}_1 = \hat{Y}$ (tan. chord theorem) $\hat{V}_1 = \hat{Y}$ (ext cyclic quad) $\hat{Z}_1 = \hat{V}_1$ $\therefore \Delta$ WVZ is isosceles (two equal angles)	✓S/R ✓ S/R ✓R (3)
8.3	In Δ XOV and ΔXZY $\hat{X}$ is common $\hat{O}_1 = \hat{Z}_2 = 90°$ (ext cyclic quad) $\therefore \hat{V}_2 = \hat{Y}$ (remaining angles) $\therefore \Delta$ XOV ||| ΔXZY $\angle\angle\angle$	✓ S ✓S /R ✓ S/R ✓ R ($\angle\angle\angle$) (4)
8.4	$\therefore \frac{XO}{XZ} = \frac{VO}{ZY}$ (||| Δs) $\frac{10}{16} = \frac{VO}{12}$ $VO = \frac{10 \times 12}{16}$ $\therefore$ VO = 7,5 units	✓S $\frac{XO}{XZ} = \frac{VO}{ZY}$ ✓substitution ✓answer (3)
		[13]

QUESTION 9



9.1	$\frac{FA}{FB} = \frac{AG}{GC}$ $(Prop\ intercept\ theorem, FG \parallel BC)$ $\frac{3}{9} = \frac{2}{GC}$ $3GC = 18$ $\therefore GC = 6$	✓S/✓R ✓Substitution ✓ Answer (4)
9.2	Let GH $= x$ $\therefore$ HC $= 6 - x$ $\frac{6-x}{x+2} = \frac{1}{3}$... (HJ // AB prop theorem) $18 - 3x = x + 2$ Now $-4x = -16$ $\therefore x = 4$ $\therefore$ GH $= 4$ units	(CA applies in the question) ✓M S✓R✓ ✓Simplification ✓ Answer (5)
		[9]

QUESTION 10

10.1	Given: $\hat{A} = \hat{D}$; $\hat{B} = \hat{E}$ and $\hat{C} = \hat{F}$ RTP: $\frac{AB}{DE} = \frac{AC}{DC} = \frac{BC}{EF}$ Construction: mark P on Q so that AP = DE and AQ = DF  Proof: $\Delta\ APQ \equiv \Delta\ DEF$ … (S; ∠, S) $\therefore \hat{P}_1 = \hat{E} = P\hat{B}C$ PQ//BC …. Corresponding ∠'s $\hat{P}_1 = P\hat{B}C$ $\therefore \frac{AB}{AP} = \frac{AC}{AQ}$ (prop. theorem PQ//BC) $\therefore \frac{AB}{DE} = \frac{AC}{DF}$ Similarly, Mark P and R on AB and AC Such that BP = ED and BR = EF $\frac{BA}{BP} = \frac{BC}{BR}$ and $\frac{BA}{ED} = \frac{BC}{EF}$ $\therefore \frac{AB}{DE} = \frac{AC}{DF} = \frac{BC}{EF}$	✓ Construction ✓ S/✓R ✓ S ✓ S/R ✓ S/R ✓ S (7)
10.2.1	In ΔTDA and ΔFDB $\hat{A}_2 = \hat{B}_3$ (angles in same segment) $\hat{F} = A\hat{T}D$ (angles in same segment) $A\hat{D}T = F\hat{D}B$ (Vertically Opposite angles) $\therefore \Delta TDA \mid\mid\mid \Delta FDB$ (∠∠∠)	✓S/✓R ✓S/R ✓3 ∠'s (4)

10.2.2	$\Delta TDA \parallel \Delta FDB$ $\therefore \frac{AD}{BD} = \frac{TD}{FD}$ (||| Δs) $\therefore AD.FD = BD.TD$ $= BD^2$ $\therefore TB^2 = (2BD)^2$ $TB^2 = 4BD^2$ $TB^2 = 4AD.FD$	✓S ✓ S (2)
10.2.3	ΔBDE and ΔADB 1. $\hat{D}_2$ is common 2. $\hat{B}_2 = A_1$ (tan chord theorem) 3. $\hat{E}_1 = A\hat{B}D$ (3rd $\angle$ Δ) $\therefore \Delta BDE \parallel \Delta ADB$ ($\angle\angle\angle$) $\frac{BD}{AD} = \frac{DE}{BD}$ (||| Δs) $BD^2 = DE.AD$	✓ S ✓ S/R ✓ S/R ✓ S (4)
10.2.4	$\left(\frac{1}{2}TB\right)^2 = DE.AD$ $TB^2 = 4DE.AD$ $4AD.FD = 4AD.DE$ $\therefore FD = DE$ In Δ DET and ΔDFB 1. FD = DE 2. $E\hat{D}T = B\hat{D}F$ (vert opp $\angle$s) 3. BD = DT $\therefore \Delta DET \equiv \Delta DFB$ (SAS) $\therefore ET = FB$ ($\equiv \Delta s$)	✓S ✓S ✓S ✓all three statements ✓S/R (5)
		[22]

TOTAL MARKS: 150