Assessing the algebraic attainment of South African grade 9 learners: designing a test using Rasch analysis

Samantha Ehrlich

Assessing the algebraic attainment of South African grade 9 learners: designing a test using Rasch analysis

Samantha Ehrlich

A research report submitted to the school of Education, Faculty of Science, University of the Witwatersrand in partial fulfilment for the degree of Master of Science

JOHANNESBURG, 2017

ABSTRACT

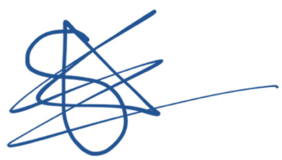
South African learners perform poorly in national and international mathematics assessments (Howie, 2004). A contributing factor to this poor performance is low mathematics knowledge of mathematics teachers in South Africa (Howie, 2003). One means of addressing this is professional development programs. The Wits Maths Connect Secondary Project runs such a program. A test is required by the project in order to assess whether learners are making learning gains after being taught by teachers who participated in this program. The focus of this study is the design of a test used to assess learners' algebraic attainment. The aim is to design an informative and fair test using Rasch analysis. A sample of 235 learners' responses to 47 questions was analysed using the Rasch model. In this study, the mean person measure was 2,87 (SD=1,38) logits, while the mean item measure was 0,41 (SD=2,25) logits, suggesting that overall, the test was too difficult. For the learners who wrote this test the person separation index is 1,78 and the person reliability 0,76. This implies that the test may not be not sensitive enough to distinguish between learners of high attainment from learners with low attainment. Various ways of improving the test are discussed.

KEYWORDS

Assessment, Rasch analysis, algebra, item response theory

DECLARATION

I declare that this research report is my own unaided work. It is being submitted for the degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



Samantha Anne Ehrlich

18th day of September in the year 2017

ACKNOWLEDGEMENTS

I would to gratefully acknowledge my supervisors Professor Mike Askew, and Dr Craig Pournara for their generous and unceasing support. I extend a special thanks to Dr Mike Linacre for his assistance via the Rasch online forum. Financial support of this work from the Education, Training and Development Practices Sector: Education and Training Authority is gratefully acknowledged. Many thanks for the multi-faceted and invaluable help from the team at the Wits Maths Connect Secondary Project (WCMS). The work of this thesis forms part of the research carried out by WCMS, which is supported by the FirstRand Foundation Mathematics Education Chairs Initiative, as well as the Department of Science and Technology, and administered by the National Research Foundation. The opinions, findings and or conclusions expressed in this thesis are my own and do no necessarily reflect the views of any of the institutions I have named above.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY OF TERMS OR ABBREVIATIONS

CAPS Curriculum and Assessment Policy Statement

DBE Department of Basic Education, South Africa

DIF Differential Item Functioning

IRT Item Response Theory

NSC National Senior Certificate

PD Professional Development

RM Rasch Model

RMT Rasch Measurement Theory

WMCS Wits Maths Connect Secondary Project

1. INTRODUCTION

1.1 Preamble

As a mathematics educator I keep in the forefront of my mind a quote from section 28(2) of the Constitution of South Africa: “A child’s best interests are of paramount importance in every matter concerning the child” (1996, p. 19).

The current state of mathematics education in South Africa is failing to uphold this right of the children of the nation. The continued lack of access to quality mathematics education perpetuates the injustice of the political past of South Africa.

Great change is the sum of small efforts and this research report is my small effort to effect some change in the current education landscape in this country.

I am gripped by the idea of Zalman Usiskin: “People can live without algebra, but as a result they cannot appreciate as much of what is going on around them. They cannot participate fully. They are more likely to make unwise decisions and will find themselves with less control over their lives. They live in the same world, but they do not see or understand as much of its beauty, structure, and mystery” (2004, p. 150). Indeed, I am convinced that proficiency in algebra should be a skill as ubiquitous and as fundamental a human right as basic numeracy.

This past year I have seen profound change in my life as I gave birth to my first child. This research report is my small contribution to South Africa, mathematics education research, and the world. My hope is that through my small efforts documented in this research report, and collaboration with the pre-existing efforts of others, as well as organisations such as the

Wits Maths Connect Secondary project, that the world that my daughter finds herself in in the future will be one where all people can better see and understand its "beauty, structure, and mystery".

1.2 Main argument

Assessment scholar Dr Douglas J. Eder once said, "If you don't know where you are headed, you'll probably end up someplace else." (Eder, 2004). To begin, I must make clear the need for the algebra test that I will be discussing throughout this report.

Learners in South Africa are required to write assessments on a regular, sometimes even daily basis (DBE, 2011). Such assessments, an example of which is the ANA (Annual National Assessments), reveal little about the algebraic attainment of learners (DBE, 2014). This is due to the flooring effect of such tests – the minimum standard scores do not distinguish learners with low algebraic attainment.

The test I will be writing about in this research report is different from existing tests. The aim is to design a test that can effectively measure algebraic attainment of learners, and distinguish learners of various degrees of algebraic attainment.

This brings me to the main argument of my research report: that it is possible to design an algebra test that can be used to assess the algebraic attainment of South African grade 9 learners, and that Rasch analysis is an appropriate and useful tool that can contribute to improvement of the test design. It does this by contributing information about the targeting of the test, the extent to which items are functioning according to what they mean to do, and whether items are multidimensional (Reckase, 1997).

In the following paragraphs I will give a brief overview and definition of each of the three terms that I have used in my title and in my argument, these are: *assessment*, *algebraic attainment*, and *South African grade 9 learners*.

1.2.1 Assessment

The National Council on Measurement in Education defines assessment as "a tool or method of obtaining information from tests or other sources about the achievement or abilities of individuals". Globally, much time, effort and resources are spent on assessing students (National Research Council, 2011). Indeed, as long as humans have been educating one another, they have been assessing one another (Graves, 1914).

The aim of educational assessment is to determine how well students are learning and provide feedback to various stakeholders, i.e. the students themselves, parents, policymakers, educators and the public at large (National Research Council, 2011). Assessment in South Africa is especially important as it relates to mathematics and poor performance (Howie, 2003; Fleisch, 2008).

The significance and importance of assessment to the improvement of education will be discussed further in chapter 2. Next I will discuss what is the focus of the assessment referred to in this research report, namely *algebraic attainment.*

1.2.2 Algebraic attainment

To lay a foundation of what I mean by algebraic attainment, I would like to start with the origin and definition(s) of algebra and then move on to define "algebraic attainment".

Algebra is a word stemming from the Arabic *"al-jabr"* meaning *the reunion of broken parts* (Simpson & Weiner, 2009). In the most general definition, algebra is a study of mathematical symbols and the rules for manipulating these symbols (Boyer, 1991). More detailed definitions of algebra will be provided in chapter 2.

Algebra is used in almost all spheres of mathematics, which adds to its significance in school mathematics (also discussed in chapter 2) even further. Elementary algebra forms school curriculums in the world and includes processes like solving equations and simplifying algebraic expressions (Kilpatrick & Izsák, 2008). Elementary algebra is generally considered fundamental to the study of mathematics, engineering, science and has application to medicine and economics (Usiskin, 1995). Equally important is the need for *elementary algebra* as a precursor to the study of *abstract algebra* - a major area of advanced mathematics.

Attainment is the action or fact of achieving a goal towards which one has worked (Simpson & Weiner, 2009). This term is used in particular to refer to a skill learnt or an educational achievement. The goal of learning algebra in school is to be able to use the knowledge and skills learnt. An example of using an algebraic skill is solving an algebraic equation like $3x + 5 = 3 + x$ correctly. Thus algebraic attainment can be reflected in the performance of a learner on test questions. In other words, the learner is able to apply their knowledge of algebra to provide correct responses to test questions.

The students who are the focus of my research are South African grade 9 learners. In the following paragraph I will discuss these students and their particular context.

1.2.3 South African grade 9 learners

Elementary algebra is taught in schools worldwide (Kilpatrick & Izsák, 2008) and South Africa is no exception. South Africa is a country with a complex history and a young democracy, as well as unique idiosyncrasies. It is necessary to take into account the influence of these characteristics when designing an algebra test. Indeed, one such example of a South African idiosyncrasy is the term “learner” given to a school student (DBE, 1996) and hence the use of this term in the title of this thesis. Therefore in my research report I will use the word *learners* to refer to *students* at school.

The curriculum document that is followed by the teachers in South Africa is called CAPS, which is an acronym for Curriculum and Policy Statements. The Department of Education (DBE) in South Africa publishes these policy documents. They outline what knowledge and skills learners must know and demonstrate.

Learners in South Africa are first introduced to elementary algebra at the start of grade 7 when they are 12-13 years old (DBE, 2011). Grades R through to 9 are referred to as the General Education and Training (GET) phase of a learner's school life. The algebra learnt in grade 9 is fundamental to more advanced algebra learnt in grade 10, 11 and 12, and culminates in a high stakes examination at the end of grade 12, called the National Senior Certificate (NSC), which is colloquially referred to as the "matric exam".

Grade 9 is an important year in the life of a South African learner. It is an "exit" year, which means learners are permitted to leave the formal schooling system upon completion of the GET certificate at the end of grade 9 (DBE, 2011). They may join the workforce, and/or take up studies and practical training in a Further Education and Training (FET) college or vocational schools. With grade 9 being such an important year in the life of a learner the focus of this research report is thus grade 9 learners.

The following paragraphs will give further detail to the background and context of this study, and will also give the reasons for the need for a test to be designed to assess the algebraic attainment of South African grade 9 learners.

1.3 Background and Context

1.3.1 The Wits Maths Connect Secondary Project[1]

This study forms part of a research and professional development project called the *Wits Maths Connect Secondary Project* (WMCS). One of the goals of this project is to develop professional development (PD) models for mathematics teachers in South Africa. This is to strengthen the teachers' relationship with mathematics, with the ultimate goal of learning gains at all levels of secondary schooling (Pournara, Hodgen, Adler & Pillay, 2015).

Such a goal requires that a professional development program is conceptualised, designed and implemented. It also requires research to be carried out regarding learner attainment. Thus the research of the WMCS Project focuses on learning gains of learners. *Learning gains* refers to an increase in skills and knowledge over a given period of time. This study fits into the larger WMCS project by contributing to the design of an algebra test to assess the algebraic attainment of learners.

1.3.2 The algebra test

The focus of this study is the design of an algebra test to assess algebraic attainment of learners. In order to assess the efficacy of the WCMS PD model, a balanced, well-performing test is required to accurately assess learning gains.

[1] Further detail on WMCS is available at www.wits.ac.za/WitsMathsConnect.

[2] See also http://www.rasch.org/software.htm.

[3] http://www.winsteps.com/a/ftutorial2.pdf

[4] Average measure of items or persons

[5] Standard errors of the measures

[6] See notes at http://www.winsteps.com/winman/reliability.htm

[7] The ratio of sample or test standard deviation (corrected for estimation error) to the average estimation error.

An algebra test is an assessment tool that includes questions that require the learner to use their understanding of algebra to answer the questions. The algebra test used in this study covers content up to a grade 9 level in the South African syllabus outlined in the CAPS for Mathematics (DBE, 2011). Further detail of the content of this algebra test is given in chapter 3.

1.3.3 Test design

Test design is a lengthy and multi-faceted process. In chapter 2, I will describe the various considerations taken into account in the design of the algebra test referred to in this report.

Briefly put, test design includes which questions are included in a test, as well as how the test is administered, length of the test (in terms of minutes and scores), who sits the test, how the test is scored, and how the results are analysed and used (Downing & Haladyna, 2006 and Reynolds, Livingstone & Willson, 2007). These activities will be described in more detail in chapter 2.

1.3.4 Rasch analysis

Results of tests can be analysed statistically. One means of statistically analysing results is making use of Rasch analysis.

This paragraph serves as a brief and simplified introduction to Rasch analysis, or Rasch Measurement Theory (RMT). RMT is a useful statistical tool for assessing responses to questions. Simply put, RMT can be used to identify questions that are far too easy or far too difficult for respondents, as well as questions that are not performing as expected. For example, Rasch analysis can help identify questions that are not functioning as they *should,* that is, respondents with high ability (attainment) getting "easy" questions wrong, while respondents with low ability (attainment) get them right.

In the following paragraphs I will outline core elements of this study in order to give an appropriate grounding to the literature review that follows. These core elements are the research questions, the context and the sampling of the study.

1.4 Elements of the Study

1.4.1 Research questions

I this study I seek to answer the following main questions, with related subsidiary questions listed underneath each main question.

1. How can we go about designing a test that is fit for purpose and useful in terms of assessing if learners have increased in learning gains? Subsidiary questions include:
    a) What kinds of questions should be included in an algebra test?
    b) What should the length of the test be in terms of number of questions and time?
    c) Who should sit the test?
    d) Who should and how should the test be marked?
    e) What is the best way to analyse the test statistically?
2. How does Rasch analysis help us to design such a test?
    a) What does the analysis suggest about the difficulty of items in the test?
    b) What does the analysis suggest about the algebraic attainment of the learners?

The aim of this study is to contribute meaningfully to the WCMS project, keeping in mind the larger aim of investigating whether learners of Mathematics benefit when their teachers participate in a PD course offered by the WMCS. This "benefit" is an increase in algebraic attainment, and thus the construct to be measured by the test is algebraic attainment.

1.4.2 Sample

This study focuses on 3 schools in the Johannesburg area, 1 of which is located in the inner city and the other two in townships. All three schools are government (public) schools and co-educational. Further details of these schools will be given in chapter 3. The unit of analysis of this study is the responses of 235 individual learners to 47 questions.

1.5 Outline of the structure of the research report

Lastly, to finish up the introduction to this report, I will provide brief outlines to the five chapters that follow.

In chapter 2 I review relevant literature and theory regarding mathematics education in the world, with particular emphasis on the teaching and learning of algebra in South Africa. This is followed by a description of the work of WMCS, as well as situating this study within that work.

Following this I will outline steps for effective test development. Downing (2004) provides 12 activities that he argues must take place in the design of an assessment such that sufficient evidence is gathered to assert validity of measurement of a construct.

A brief history of practical additive measurement follows as a preamble to a description of item response theory (IRT) and Rasch measurement theory (RMT). The rationale behind using Rasch analysis in the design of the algebra test is presented, including examples of the use of RMT in mathematics education research around the world, and in South Africa.

Chapter 3 outlines the methods used in this research, beginning with a description of how sampling was carried out, and details of the algebra test. RMT is then presented (including the mathematical model developed by Rasch) and data analysis techniques described.

Findings are presented in chapter 4, and then analysed, interpreted and discussed in chapter 5. Finally the report concludes in chapter 6 with implications of this research and suggested directions for future studies.

2. LITERATURE REVIEW

2.1. Overview of chapter

In this chapter I situate my research in the relevant contemporary literature, by focusing on broad topics.

The first broad topic is the learning of algebra. I will begin by trying to answer the question*, what is algebra*? By considering various definitions put forth in the literature, I will try to define algebra.

Following this will be a review of studies that show in what ways school algebra differs from the academic discipline. Examples of how algebra is taught in schools globally will be given, with particular emphasis on algebraic activity in schools in South Africa.

Thereafter will be a review of research on the learning of algebra, and how learners conceptualise algebra, as well as their thinking and reasoning. Also considered will be the relationship between arithmetic and algebra, as well as the problems associated with the transition between the two in school learning and teaching. How algebra and algebraic activity is defined in the South African curriculum is also considered.

Subsequent to this I will look at the reasons such emphasis is given to learners learning algebra, and will transition into arguments for the importance of learning algebra.

The second broad topic is the teaching of algebra in South Africa, including what it means to be a teacher of mathematics in this country, leading into the professional development program offered by WMCS, and the subsequent need for a test to measure learning gains of learners whose teachers attend the course.

The third broad topic is assessment and test design, including an overview of the 12 steps for effective test development by Downing (2004).

The final topic reviewed in this chapter is Rasch analysis. Beginning with a discussion of measurement in the human sciences, I will give a brief history of practical additive measurement, including item response theory and finally Rasch measurement theory (RMT). I will include examples from around the world where RMT has been used successfully in the design of various assessment tools.

2.2. Learning algebra

2.2.1 What is "algebra"?

There are various definitions of algebra debated in the literature.

Derbyshire (2006) gives a broad and more philosophical view of algebra as abstract thought and the modelling of reality; arguing that algebra has a distinctive quality that sets it apart as a discipline by itself. Algebra is the "abstractions of abstractions", which help us to model the real world to some degree (Derbyshire, 2006).

Mason, Graham & Johnson-Wilder (2005) gives a definition of algebra as investigating and solving problems by generalising and using techniques; noting that it is unhelpful to describe algebra as *generalised arithmetic,* Generalised arithmetic means the expression of general arithmetical rules using letters. At the very heart of algebra is the expression of generality and that algebraic symbols are a language for expressing such generality.

Watson (2009) describes algebra as the way in which we express generalisations about numbers, quantities, relations and functions. Related to success in using algebra is a thorough understanding of the connections between numbers, quantities and relations.

All of the definitions above frame algebra as a discipline in and of itself. It is useful for modelling reality and solving problems. The latter two definitions help us to understand the means by which this is done – namely generalising and using algebraic symbols.

Formal definitions often differ to everyday uses of a word, and a prime example of this is the word algebra. School mathematics and school algebra is a selection and adaptation from the discipline of mathematics with different goals. This leads me to discuss in what ways school algebra (the focus of this research report) is different from algebra as a discipline.

Proficiency in algebra is required for other content areas of the mathematics curriculum (DBE, 2011). Areas such as trigonometry and analytical geometry seem like vastly different subjects to algebra, yet they require deep conceptual understanding of simplification and solving of equations, as well as the concept of equivalence.

Algebra is seen as a "gate-keeper" to various technical vocations (Stinson, 2004). Thus there is a view that due to the fact that elementary algebra is a gateway to higher mathematics, all school learners should have access to quality education in this topic of mathematics in particular (Usiskin, 2004).

2.2.2 What is school algebra?

There are contradictory and competing views on what constitutes school algebra. A variety of views have been debated as mathematics education undergoes a reform globally (Shifter & Fosnot, 1993).

As school curricula have developed since the beginnings of formal mass education during the industrial revolution of the 1800s, the view of algebra as a tool for manipulating symbols and solving problems has persisted (Kieran, 2007).

Indeed, the perspective of school algebra as algebraic activities has been proposed: whereby all that is considered algebra in school could be categorised into representational activities (e.g. showing a linear relationship using tables and graphs), transformational activities (e.g. changing the appearance of expressions), generalising (e.g. finding the general term of a number pattern), and justifying activities (e.g. solving geometry riders) (Kilpatrick, Swafford & Findell, 2001).

Watson (2009) contrasts two views of school algebra: a 'bottom-up' developmental approach (focussing on what learners can do and how their generalising and use of symbols develop) with a 'top-down' hierarchical approach (a view which states what is required in order to do higher mathematics).

The 'top-down' view sees school algebra as a list of techniques that need to be fluent. This view results in research into errors made by learners and studies designed to mitigate these. Such research gives insight into obstacles that need to be overcome in the development of understanding, as well as revealing how learners think (Watson, 2009).

The 'bottom-up' view, however, focuses on *algebraic thinking*. Algebraic thinking is the expression and use of general statements about relationships between variables (Watson 2009). Algebraic thinking is an intentional shift from context to structure and arises when people are detecting and expressing structure (Lins, 1990).

These two views complement one another as the 'bottom-up' view takes into consideration the development of the natural ability learners have to detect patterns and generalise them, and the 'top-down' view relates this to the increase in competence in understanding and using symbols. This leads onto the idea of the content of school algebra as the development of algebraic reasoning (Thomas & Tall, 2001). The development of algebraic reasoning is the shift between procedure, process/concept, generalised arithmetic, and manipulations. In this view manipulation is the generation and transformation of equivalent expressions.

I see school algebra as a way of thinking that needs to be taught such that learners "see" the various symbols and notations as objects, and thus are able to manipulate such objects and carry out meaningful procedures using these objects as tools to solve problems. I derive my view from learning algebra in South Africa, as well as teaching algebra in South Africa.

Learning algebra in South Africa will be discussed in the following paragraphs when research into how learners learn algebra is discussed.

2.2.3 The learning and teaching of algebra in South Africa

Poor performance and failure in school mathematics is a problem in South Africa and other developing countries (Pournara, Hodgen, Sanders & Adler, 2016). The vast majority of South African learners are not coping in school mathematics (Spaull, 2013).

If one takes the view that algebra is learnt by participating in a discourse, such as is argued by Sfard & Linchevski (1994) in the theory of commognition, or the communicational framework of mathematics, then it is not surprising that there would be low performance of South African learners. If the teachers lack the knowledge and interlocutory skills to present algebra as more than just examples and exercises, then learners will be unable to appreciate and use algebra in the context in which they find themselves.

In addition, the ideas of Thorndike (1922) persist: that success in a subject results from formulaic practice of said subject – rather than conceptual understanding combined with the learner being enabled to make meaning out of what is being learnt and thus being able to articulate that in their own way.

This results in state resources being poured into thick mathematics textbooks replete with exercises, more than investing into teacher development (DBE, 2014). Indeed, this brings further incentive into researching whether the WMCS professional development intervention is effective. If it is shown using empirical data that the intervention is effective to increase learning gains, then with increased investment more teachers could have access to the course, and thus the positive effects on the learning of algebra in South Africa could abound still more.

Due to the abstract nature of algebra, there is an inherent difficulty to be found when learning algebra (Sweller & Chandler, 1994). Learners have considerable difficulties with adopting the conventions of algebra (Watson, 2009). There are obstacles to be overcome in order to understand the meaning of letters and expressions and to use them.

This difficulty is increased all the more due to the various social and educational challenges faced by South African youth. These challenges include poverty, poor service delivery, lack of adequate resources, and the knowledge deficit of parents and the community as a legacy of bantu-education under the pre-1994 apartheid regime. These all have far reaching effects, including the inability to achieve academically at school (Fleisch, 2008).

Moreover, most learners do not learn in their home language. Generally, South African learners perform poorly in national and international mathematics assessments (Howie, 2004, Simkins & Paterson, 2005). It could be argued that a significant contributing factor to this is that international assessments are written in the language of the hegemony, and not the average South African learner. This further disadvantages them.

Added to this, is inadequate teaching training. It has been established that poor teacher knowledge is a contributor to low performance of learners (Howie, 2003). In some areas of South Africa, up to 79% of grade 6/7 mathematics teachers have content knowledge below a grade 6/7 level (Venkat & Spaul, 2015). Similarly, some grade 12 teachers are unable to answer up to 45% of the final school leaving examination (Bansilal, 2015).

Various government interventions have been put in place to address the lack of content knowledge of teachers. Professional development (PD) courses seek to address teachers' lack of content knowledge (Borko & Putnam, 1996). This inadequate teacher training is further incentive for quality professional development interventions, such as the one offered by WCMS.

2.3 Effective test development

In the past, issues concerning the design of tests have received little scholarly or scientific attention in the academic discipline of educational measurement. Indeed, much research focuses on the statistical issues of testing (Downing, 2006).

The big question in test design is one of validity – are the questions that are being asked *actually* measuring the construct that the test is designed to measure. Item response theory, can be incorporated into answering this question (Downing, 2006).

In addition to this sound testing practices must be considered. Which activities form part of sound test design, and how do they contribute to effective test development? Answers to these questions are complex and varied, leading some to describe test design as simultaneously an art and a science (Downing, 2006).

Downing (2006) provides activities associated with designing tests. The 12 steps he describes can be carried out to increase the probability that a test has sufficient validity evidence to support the intended score interpretations. In chapter 5 of this research report I present a comprehensive, coherent, scholarly (yet pragmatic) discussion of all the issues regarding how a test was designed, administered and analysed.

Psychometric theory is the foundation for all the activities required in test design. Effective test development requires a systematic, detail-oriented approach based on sound theoretical educational measurement principles. In order to gain sufficient validity evidence to support the proposed inferences from test scores, it is imperative that the process of test deign is systematic and well organised (Downing, 2006).

2.3.1. Steps for effective test design

Downing (2006) argues that there are twelve discrete steps that must be accomplished in the design of tests. This maximises the validity evidence for the intended test score interpretation. Test design begins with detailed planning, through to discussions of what the content will be, then creating these questions (items), and then to how the test is administered, scored and reported on. All these procedures are integral to effective test design. All of the procedures must be well executed in order to a test to be produced that estimates examinee achievement with a high degree of validity.

Each test development activity needs to receive sufficient attention – this is to maximise the probability that the test is an effective measure of the construct of interest. However, the technical sophistication of each activity depends on the resources at hand: the human resources and the financial resources. Some of the suggestions in the table that follows may be too costly or too time consuming. Table 1. lists the twelve steps of test development and a brief summary of the tasks, activities and issues associated therewith. Although the steps appear as a linear model (or sequential timeline) they are however open to be modified or occur simultaneously. There is a discrete beginning and a final end point, but the tasks in between may be happening altogether.

Table 1. Twelve steps for effective test development (adapted from Downing 2006, p. 5).

	Suggestions of test design tasks
1. Overall plan	Systematic guidance for all test development activities construct; desired test interpretations; test format(s); major sources of validity evidence; clear purpose.
2. Content definition	Sampling plan for domain/universe; various methods related to purpose of assessment; essential source of content-related validity evidence; delineation of construct.
3. Test specifications	Operational definitions of content; framework for validity evidence related to the systematic, defensible sampling of content domain; norm or criterion referenced; desired item characteristics.
4. Item development	Development of effective stimuli; formats; validity evidence related to adherence to evidence-based principles; training of item writers, reviewers; effective item editing; construct-irrelevant variance (CIV) owing to flaws.
5. Test design and assembly	Designing and creating test forms; selecting items for specified test forms; operational sampling by planned blueprint; pretesting considerations
6. Test production	Publishing activities; printing or computer-based testing (CBT) packaging; security issues; validity issues concerned with quality control
7. Test administration	Validity issues concerned with standardisation; issues relating to those examinees with special needs/disabilities/learning difficulties; security issues; timing issues
8. Scoring test responses	Validity issues: quality control; key validation; item analysis
9. Passing scores	Establishing defensible passing scores; relative vs. absolute; validity issues concerning cut scores; comparability of standards: maintaining constancy of score scale (equating, linking)
10. Reporting test scores	Validity issues: accuracy; quality control; timely; meaningful; misuse issues; challenges; retakes
11. Item banking	Security issues; usefulness, flexibility; principles for effective item banking
12. Test technical report	Systematic, thorough, detailed documentation of validity evidence, 12-step organisation, recommendations

I will return to and elaborate further on these twelve steps in my discussion and critique of the test that is the focus of this study in Chapter 5.

2.3.2 The purpose of the test

First and foremost the stated purpose of the test must be known. The purpose of this test is to assess the algebraic attainment of learners taught by teachers who have attended a professional development course. Indeed, this setting out of purpose is fundamental as all major decisions such as defining what content will be covered in the test, and which construct hypothesised to be measured by the test are all directly associated with the stated purpose of the test (Downing, 2006).

The first decision to be made when beginning the design of a test is what construct is to be measured. For this test it was decided the construct to be measured was algebraic attainment – and whether there was a change in this attainment. In other words, were knowledge and skills gained by learners who were taught by teachers who attended the WMCS PD course, over and above learners who were not taught by such teachers?

The next decision is what score interpretations are desired. For example, for this test reported on more full in chapter 4, the score interpretation was dichotomous i.e. zero for missing or wrong and 1 for correct.

Following this is what test format or combination of formats will be used (i.e. selected response or constructed response/performance). This test was constructed response, i.e. learners had to write down their answers. After this the test administration modality is decided. The mode of this test was paper and pen as opposed to a computer based modality.

In chapter 5 I will discuss the process of test design undertaken for the test that was studied, and compare the steps taken to the steps Downing proposes. Part of the test design was the use of statistical analyses to interpret the functioning of not only the learners who wrote the test, but also the items (questions) in the test itself. This leads to the next part of this chapter, which is an introduction to the statistical analysis of assessments, and ending with the model that was used in this research, namely the Rasch model of measurement.

2.4 Test design theory

We have all experienced a test that we thought was not fair – perhaps we considered a test too easy and a waste of our efforts, or too difficult and thus we were not able to demonstrate how knowledgeable and skilful we actually were. Enter the statistical analysis of tests as we seek after an unbiased, objective estimation of the difficulty of a test. In essence, a *scientific* way of telling how *fair* a test really is.

For the last part of this literature review I will begin with the debate regarding measurement in the human sciences. Following this will be a brief introduction to classical test theory, the precursor to item response theory (IRT). It is under item response theory that Rasch analysis is categorised.

2.4.1 Why is measurement fundamental?

The concept of using a scale to measure our weight, or a thermometer to measure temperature is so familiar to us that we take for granted the centuries of thought and study that was required to come up with scales and units to quantify these physical phenomena (Bond & Fox, 2012). There was a time when kilograms, and degrees Celsius did not exist. Thus when we think about quantifying *constructs* (that is, an idea or theory containing various conceptual elements) we may be quick to think it will be impossible, forgetting that the same was once thought of for physical constructs such as weight and temperature.

Measurement is fundamental to living and understanding the world around us. The need to quantify things has been a pursuit of humans since the dawn of time – how to measure quantities such that one unit added is the same as another no matter how much is already present (Bond & Fox, 2012). We could even go so far as to agree with the McNamara fallacy (also known as the *quantitative* fallacy): that if you cannot measure it, it does not exist (Fischer, 1970). Such ideas will be described more as the application of measurement to the human sciences is considered in the paragraphs below.

2.4.2 Social science measurement

"Is psychological measurement possible?" was first debated in the 1930s (Linacre, 2016a). So-called "hard" scientists such as physicists responded in the negative. From a positivist perspective, measurement requires a *deliberate action.*

An important idea when considering how to go about measuring "things" is what it means to concatenate them: how to link them in a series such that each successive unit is as far from the next unit as it is from the one before. Thus a *concatenation* is a deliberate action and one cannot concatenate the minds of people (Linacre, 2016b). Examples of concatenation are putting sticks end to end to measure length or piling bricks one of top of the other to measure weight.

In contrast "soft" scientists could not answer this question. Thus a different definition of measurement was devised i.e. measurement is the assignment of numbers to objects or events according to a rule (Stevens, 1946). Indeed, social scientists call "measures" whatever numbers they happen to have acquired (Linacre, 2016b).

This caused confusion as scientists were using the same word to mean different things. For social scientists, any number is a measurement provided it follows a given rule. For other scientists, measures had to conform to strict objective criteria (Linacre, 2016b).

The aim of measurement in the human sciences is to quantify constructs such as attainment or understanding. Furthermore we seek to understand and improve the reliability of psychological tests. Mathematics tests are a form of psychological tests, as the examinee is required to think through and respond to questions posed, as opposed to standing on a scale to be weighed, for example. This leads us to classical test theory, which deals with predicting the outcomes of psychological tests.

2.4.3 Classical test theory

When scientists first started endeavouring to predict outcomes of psychological testing it was thought that a person’s score on a test was the sum of a true score and an error score. This idea is part of a body of related psychometric theory called *classical test theory* (Novick, 1966).

This theory was used to predict outcomes of psychological tests such as the difficulty of items or the ability of the examinees. The term *item* is generic and used to refer to all kinds of informative items: multiple choice questions which have correct and incorrect responses; statements of questionnaires that allow respondents to indicate their level of agreement; questions to which the response could be present/absent, yes/no etc.

The aim of classical test theory is to understand and improve the reliability of psychological tests (Allen & Yen, 2002). Classical test theory contrasts with the more recent psychometric theories collectively referred to as item response theory (IRT).

2.4.4 Item response theory

The name of this theory comes from the focus of the theory on the item, as opposed to the test-level focus of classical test theory. Item response theory models the response of each examinee of a given ability to each item in a test.

Item response theory is thus a paradigm for the design, analysis and scoring of tests that measure constructs. For example a questionnaire measuring a person's perception of a political party, or, as is the case in this study, a test measuring a learner's algebraic attainment.

Item response theory is based on the relationship between an individual's performance on an item and the individual's levels of performance on an overall measure of the ability that that item was designed to measure. Furthermore, item response theory is based on the idea that the *probability* of a correct response to an item is a mathematical function of the person parameters and item parameters.

The person parameter is a single latent trait (or dimension) such as intelligence or the strength of an attitude. Parameters on which items can be characterised include difficulty. This means an item can be located on a scale of difficulty. Thus item response theory is based on the application of mathematical models to testing data. IRT is thus regarded as superior to classical test theory (Hambleton, Swaminathan & Rogers, 1991)

As ability varies, the probability of a correct response to the item also varies. The probability of a person with low ability responding correctly is correspondingly low, and the probability of a person with high ability responding correctly is correspondingly high (and approaches 1 asymptotically as ability increases).

This is shown graphically using item characteristic curves (ICCs), examples of which are shown in figure 1 (Hambleton, Swaminathan & Rogers, 1991).

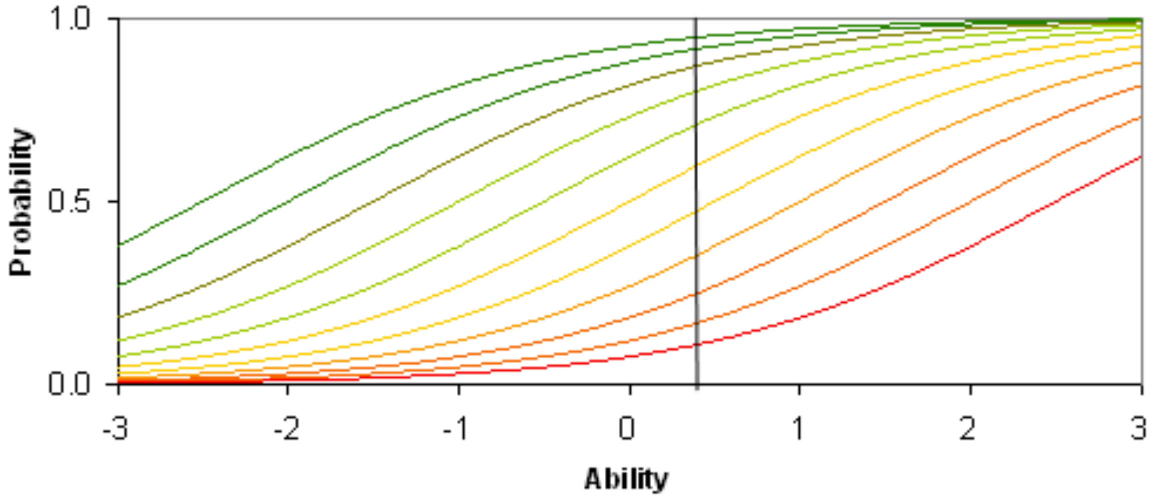


Figure 1. Item Characteristic Curves (Item response theory, 2017). Persons are represented on the horizontal axis (with low ability on the left to high ability on the right). The probability of a correct response is represented by the vertical axis (from 0 to 1).

Figure 1 represents ICCs for a number of items. Learners are represented on the horizontal axis from low ability on the left to high ability on the right. The ICCs have been given various colours to highlight the change in the probability of a correct response for a person whose ability is located at the vertical line. There is a high probability that this person will respond correctly to the easiest items (green hues), about a 50:50 chance of responding correctly to the items shown by the yellow curves, and a low probability of responding correctly to the more difficult items (orange hues).

Several statistical models can be used to represent both the characteristics of each item as well each individual examines. This makes item response theory more sophisticated than classical test theory, as it is not assumed each item is equally difficult. Indeed scales can be created and response evaluated by treating the difficulty of item (represented graphically as item characteristic curves) as information to incorporated into the creation of a scale of the items. The model put forward by Georg Rasch is an example of such a statistical model.

2.4.5 Georg Rasch's "models for measurement"

Returning to the introductory discussion on practical additive measurement, where the debate regarding whether measurement was possible in the human sciences, Georg Rasch showed how the strict criteria of the physical scientists could be applied to social science by means of Rasch models. Georg Rasch called these models "Models for Measurement." (Rasch, 1961).

Rasch devised a way of "concatenating heads" in a psychological sense in a manner that parallels "concatenating rods" in a physical sense. The Rasch model implements "additive" measurement, that is, adding one more unit means the same amount extra no matter how much there is already (Linacre, 2016b)

Rasch measurement theory 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 that 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 Rasch model 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, 2016b)

In simplified terms, Rasch analysis sheds light on which items the learners are *actually* experiencing as difficult, rather than what examiners *expect* learners to find difficult. Applying the Rasch model when analysing tests helps to identify items 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 β_n and item difficulty is denoted by δ_i. These two constructs may be represented on the same scale. The equations from Rasch measurement theory used to calculate the probability of a correct and incorrect response to a given item are outlined in chapter 3.

2.4.6 Uses of the Rasch Model

The Rasch model is a useful psychometric model used to analyse categorical data. The usefulness of the model comes in 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. Rasch measurement theory (RMT) is a useful statistical tool for the improvement of tests. Indeed, the application of RMT to algebra tests can be particularly helpful (Edwards & Alcock, 2010), as is outlined in the following paragraphs.

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, Rasch measurement theory can help identify questions that should be excluded or included in an assessment based on the functioning of the item (Anderson, Alonzo & Tindal, 2012). The technical adequacy of the measures can be increased while at the same time increasing the accessibility of the items, especially for those learners that struggle with mathematics (Anderson et al., 2012).

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 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 is increasingly being 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, Mhakure, Fray, Holtman & Julie, 2014), as well as assessments of geometry thinking levels (Stols, Long & Dunne, 2015).

2.4.8 Using RMT in algebra research

Internationally, RMT is used in the analysis of Trends in International Mathematics and Science Study (Mullis, Martin, Foy & Arora, 2012), measuring the algebra readiness of US middle school students (Ketterlin-Geller, Gifford & Perry, 2015), and algebra reasoning of UK lower secondary students (Hodge, Brown, Coe & Küchemann, 2012).

Many researchers have used RMT when investigating the algebraic attainment of learners, as well as analysing how accurately an assessment is measuring this attainment. RMT can be used to (1) determine whether an instrument is unidimensional, (2) determine whether any items are *anomalous* and why, (3) develop a ranking of items in order of difficulty, and (4) measure the algebraic attainment of students within the tested population (Craig & Campbell, 2013). Similarly, RMT can be used to analyse the validity and reliability of algebra items (Nopiah, Osman, Razali, Ariff & Asshaari, 2010). Furthermore, RMT can aid in increasing the quality and reliability of algebra tests by reducing item gaps by identifying items that might need to be rephrased or replaced (Nopiah et al. 2010).

In summary, it is possible and beneficial to make use of Rasch measurement theory when designing mathematics tests. Such analyses help to improve questions and to test if questions are functioning differentially in a cohort (e.g. success or failure on an item differs according to gender) (Bond & Fox, 2012). Indeed, the potential for RMT in mathematics education research is considerable (Dunne, Long, Craig & Venter, 2012).

2.5 Summary of chapter

It is clear from the literature that the learning and teaching of algebra in South Africa is difficult, and the country has unique challenges to improving the performance of learners. We can infer from the evidence that an intervention to increase the knowledge and skills of mathematics teachers could be a means of addressing the low performance of learners. The WMCS offers such an intervention. In order to assess if the intervention is effective in increasing the learning gains of learners a well-designed and fit-for-purpose assessment is required. There are various steps that can be taken to increase the likelihood that the test is indeed assessing the construct it was designed to assess, with a high degree of validity. The sources suggest that Rasch analysis is a useful and practical tool for designing such an assessment.

In the next chapter, I will describe how a test was designed and Rasch analysis was used to analyse the test designed to assess the algebraic attainment of learners in schools around the Johannesburg area.

3. METHOD

3.1 Overview of chapter

This chapter is a description of when and how the test was administered, scored and analysed. Details regarding the learners who wrote the test is given. RMT is described, including the relevant mathematical models.

3.2 Administration of the test

In this study, the research instrument was an algebra test. The format of this test is constructed response, that is the examinees had to write down short answers. The mode of the test was traditional paper-and-pencil.

There was a preliminary test administered in February 2016. This test was reviewed and refined to give rise to the test that was administered in June 2016. It is this June 2016 test that is the focus of this study, and the results of which were analysed using Rasch analysis. The timeline of testing is described in more detail in the following paragraphs.

3.3. Analysing the results of the February tests

The first iteration of tests was carried out in February 2017. There were 65 questions in this test. To see if the number of questions, type of questions or order of questions influenced the responses, the questions were "packaged" into 4 tests.

After these tests were marked, the data was organised into a *scalogram* (Bond & Fox, 2012). A scalogram is a table. The items (the columns) are sorted by item score (left to right in descending order), and the persons (the rows) by person score (top to bottom in descending order). This scalogram was used to identify any questions that were very similar in terms of structure and score.

The 65 questions were reduced to 47 questions described in the paragraph 3.4. Decisions were made to include or exclude questions for a variety of reasons. Generally there were too many of questions involving algebraic expressions of similar structure. For example both these questions appeared in the February test: simplify $b - 6 + 6$ and simplify $a + b + a - b$. These two questions were similar in format and both had an average of 17,9%. An average for a question was calculated by dividing the number of learners who got the question correct by the number of learners who responded to the question. In the case of the two previously mentioned questions, the decision to only include the latter question was made – as it only included 'letters'.

3.4 Details of the June test

3.4.1 Test layout

The 47 questions in the June test were ordered two different ways. The test was printed on blank, white A4 pages. There was about 5cm of blank space underneath each question for the learners to write down calculations or working out.

3.4.2 Content of the test

The questions in the test covered content up to a grade 9 level in the South African syllabus outlined in CAPS (DBE, 2011). Some questions were adapted from previous studies (e.g. Hart (1981), whereas others were created by various members of WMCS. Table 1 shows each question, the number of sub-questions as well as the topic.

Table 2. Details of the algebra test that was analysed using Rasch analysis. (Note that question 3 has been omitted, with reasons for this given in the text.)

Question		Topic
1	None	Ordering of integers
2	2	Substitution
4	3	Equality (from Hart, 1981)
5	10	Simplifying algebraic expressions
6	6	Simplifying algebraic fractions
7	3	Products
8	4	The linear function
9	6	Solving algebraic equations
10	3	Factors
11	3	Input-output flow diagram
12	3	Factors and simplification
13	3	Patterns
TOTAL:	47	

3.4.3 Description of questions

The test is included in entirety in appendix G. Below is a description of each question and reasons as to why they were included in terms of what intended with the design of the test.

The focus of the content area *numbers, operations and relationships* is the development of number sense that includes the relative size of different numbers (DBE, 2011). Therefore question 1 was included in the test. This question involved the ordering of positive and negative numbers, ("Write these numbers in order from smallest to largest: 30; -35; -2; -500; -10; 4). This question was used to test if learners had knowledge of the magnitude of integers.

The rest of the questions could be categorised under the content area *patterns, functions and algebra* (DBE, 2011). The focus of this area is the description of patterns and relationships through the use of symbolic expressions, graphs and tables; and identification and analysis of regularities and change in patterns, and relationships that enable learners to make predictions and solve problems. (DBE, 2011).

Question 2 involved substituting in integers in place of a variable and read, "If $a = 2,\ b = -5,\ c = 3,$ evaluate the following. Show all your working. a) $ab + 2c$ b) $4 + (a - b)$.

Question 3 was made up of 11 sub-questions. I have omitted these questions from the analysis as they involved arithmetic, that is, the adding and subtracting of integers. Furthermore, during the test some learners had access to calculators. This meant that the learners were able to get all 10 sub-questions of question 3 correct.

Developing algebraic manipulative skills that recognize the equivalence between different representations of the same relationship is also a skill required by CAPS (DBE, 2011). Therefore question 4 was included in the test, and was adapted from Hart (1981). The instruction for question 4 read, "Write down the missing number in the space provided." followed by these three sub-questions:

a) $7+5=\_+2$ b) $4747+3945=\_\_+3943$ c) $4747+n\ =\ \_\_\_\ +\ (n-2)$

The notion of equality tested here, as well as the impact of the larger numbers in equations in b) and the generality in c).

Question 5 was made up of 10 sub-questions, which mirrored question 3 in terms of asking the learners to simplify the expressions, and included the use of different letter and brackets:

a) $2a+5a=$ b) $2a+5b+a=$ c) $(a+b)+a=$ d) $(a\ +\ b)b=$ e) $b(a-b)=$

f) $b(a-b)=$ g) $a+4+a-4=$ h) $3a-b+a=$ i) $5-a+a=$ j) $a\ +\ b\ +\ a\ -\ b=$

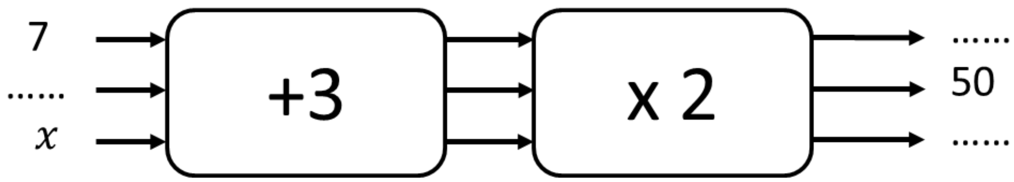
Question 6 was a mix of sub-questions with algebraic fractions involving using the laws of exponents and adding like/unlike terms. A common theme was the use of a vinculum in each sub-question indicating the operation division. The instruction read, "Simplify the following. No denominator is zero".

a) $\frac{3\times5^9}{5^7}$ b) $\frac{a\times b^9}{b^7}$ c) $\frac{3b^7}{b^9}$ d) $\frac{15x+9x}{8}$ e) $\frac{15x+9x}{3xy}$ f) $\frac{16xy+8x^2y}{4xy}$

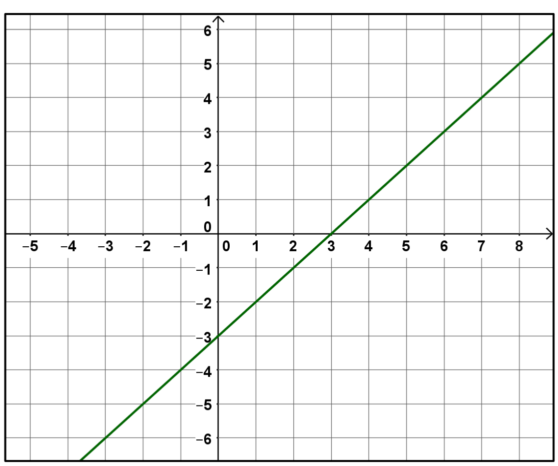
In grade 8 (DBE, 2011) learners are taught the commutative property of multiplication for the multiplication of variables. Question 7 was designed to test this, and involved the simplification of algebraic expressions that would result in products. The instruction read, "Multiply out".

a) $(2x+1)(x+4)$ b) $3(x+2)^2$ c) $8(p-5)(p+5)$

Learners are introduced to the concept of function beginning with *spider diagrams* (see the example of a spider diagram in question 11). The concept of function as a machine when numbers are put in ("input numbers") and after undergoing transformation, a resulting number comes out ("output numbers"), is though to lead to an understanding of the graphing of functions such as the linear graph (Howse, Molina, Taylor, Kent & Gil, 2001). Question 11 included such a spider diagram, and read "The diagram below show inputs and outputs for the machine diagram. Work out the missing information and write your answers in the spaces provided."



Learners are required to understand various representations and descriptions of situations in algebraic language, formulae, expressions, equations and graphs (DBE, 2011). Question 8 included characteristics of a straight-line graph, and read, "Look at the diagram below which shows a straight-line graph".



a) Write down the x-intercept of the graph.

b) Write down the y-intercept of the graph.

c) Write the equation of the straight line in the form $y = mx + c$.

d) Sketch the graph of $y = -x + 3$ on the set of axes given above.

One of the most fundamental skills to be learnt in elementary algebra is the solving of equations (Filloy & Rojano, 1989). Question 9 required the learners to solve linear equations and equations involving algebraic fractions:

a) $3x - 1 = 5$ b) $3x - 1 = 4 + x$ c) $1 - 3x = 5 - x$

d) $2p(p - 4) - 8 = 2p^2 - 7p + 3$ e) $\frac{a+1}{3} = 2$ f) $\frac{x+1}{3} + \frac{2x-1}{6} = \frac{x+2}{2}$

Learners are introduced to factorising algebraic expressions in grade 9 (DBE, 2011). Question 10 involved factorising of the following expressions:

a) $7x - 28$ b) $7 - 28x$ c) $7x^2 - 28$

Question 12 was a compound question, in that it combined factorisation of expressions, and the simplification of an algebraic fraction that included those expressions:

a) Factorise fully: i) $x^2 - 4x$ ii) $x^2 - 2x - 8$

b) Simplify. The denominator is not equal to zero. $\frac{x^2-4x}{x^2-2x-8}$

The purpose of this question was to see if learners saw the link between factorising the expressions of a) which are repeated as the numerator and denominator of b).

Learners are required to investigate numerical and geometric patterns to establish the relationships between variables, and must be able to express rules governing patterns in algebraic language or symbol (DBE, 2011). Therefore question 13 involved a pattern: which read, “Matchsticks are arranged as shown”

Figure 1 Figure 2 Figure 3

a) How many matchsticks are required for Figure 36?
b) Which Figure would need exactly 51 matches? Explain how you got your answer.
c) Give an expression for the number of matches required for the n^{th} figure.

3.5 Test implementation

The instrument was administered to the learners under test conditions, and learners were given 1 hour to write the test. The principals of the schools were given detailed feedback of the results so that they could be informed of areas that could be improved. The 235 learners were in grade 9 in 3 high schools (sample sizes from each school were 158, 45 and 32). These high schools are situated in areas surrounding Johannesburg, South Africa. Two of the schools are located in townships and one in the inner city. The medium of instruction of each of the schools is English.

3.6 Data analysis

<u>3.6.1 Coding of responses</u>

The learners' responses were all coded by the author, and moderated by senior project staff of WMCS. Only final answers were considered i.e. missing and incorrect responses were coded 0, correct answers were coded 1. Issues related to only coding final answers are discussed in chapter 5.

Codes were captured and combined into a scalogram, and subsequently imported into WINSTEPS®[2] (Linacre, 2016) for analysis.

3.6.2 Fit statistics of the data to the Rasch model

In order to claim measurement is working within the framework of RMT it is important that the data fits the model. Indeed the properties of Rasch measurement are only applicable according to how well the data fits the requirements of the model (Bond & Fox, 2012). When data fits the Rasch model, raw total scores for persons and frequencies of correct/incorrect responses for each question can be appropriately transformed (Dunne et al, 2012). This allows estimates for learner attainment and question difficulty. These estimates can then be represented on the same scale (or linear dimension) i.e. adding one unit increases the total by the same amount, no matter how much there was to start with (Linacre, 2016b). This allows inferences to made based on interpreting the results in the given context.

Fit statistics give an indication of discrepancies between data collected and the prescriptions of the Rasch model. An indication of how well data fit the Rasch model is given by chi-square statistics[3]. The expected mean of a chi-square distribution is its "degrees of freedom" (d.f.), the number of independent squared unit-normal distributions it represents. If we divide a chi-square value by its degrees of freedom, then we have a mean-square value. The total item chi-square for this test was 5481,00 and total d.f. = 5488,00. Dividing these two values yields a mean-square value of 0,998, which is very close to the expected value of a mean-square i.e. 1,0. This indicates that the data is productive for measurement and fits the Rasch model.

[2] See also http://www.rasch.org/software.htm.

[3] http://www.winsteps.com/a/ftutorial2.pdf

3.6.3 Analysis of the test data using Rasch measurement theory

Once it was established that the data could be analysed using Rasch measurement theory, the test was analysed. The paragraphs that follow describe how the Rasch model was used to analyse the data.

3.7 Initial analysis

WINSTEPS® software was used to run the Rasch analysis. The software is deigned to test whether the data is close to the theoretical pattern predicted by the model. WINSTEPS® reports fit statistics in terms of person (learner) and item (question) fit residual statistics. These fit residual statistics are an indication of the differences between their actual and expected responses. In addition, WINSTEPS® reports item-trait interaction chi-square statistics. Item-trait interaction chi square statistic is a reflection of the property of invariance across the trait.

In chapter 4 the results of the initial Rasch analysis are presented. This analysis generates fit statistics, a person-item (learner-question) location distribution as well as a person-item (learner-question) map.

3.7.1 The Rasch dichotomous model

The mathematical theory underlying the Rasch model is a special case of *item response theory*. Rasch presents several mathematical models (Rasch, 1980). His model for dichotomous responses (e.g. right or wrong, yes or no, present or absent). by persons to items has become known as *the Rasch Model* (Linacre, 2016b).

The Rasch dichotomous model states the equation for the log-odds of person n succeeding on item I as: $\log_e\left(\frac{P_{ni}}{1-P_{ni}}\right) = B_n - D_i$ where B_n is the ability of person n and D_i is the difficulty of item i. P is the probability of success and $1 - P$ is the probability of failure. Since either success or failure must always happen, when we add their probabilities they must sum to 1: i.e. (probability it does happen) + (probability it does not happen) $= P + 1 - P = 1$ (Rasch, 1960).

Rasch (1980) provides the simple logistic (symbolic logic) model for dichotomous items, where learner ability is denoted by β_n and item difficulty is denoted by δ_i. These two constructs may be represented on the same scale. The equations from Rasch measurement theory used to calculate the probability of a correct and incorrect response to a given item are as follows. Let $X_{vi} = x \in \{0,1\}$ be a dichotomous random variable. $x = 1$ denotes a correct response and $x = 0$ denotes an incorrect response to a question. In the Rasch model for dichotomous data, the probability of the outcome $X_{vi} = 1$ (success) is given by: $P\{X_{vi} = 1\} = \frac{e^{\beta_v - \delta_i}}{1+e^{\beta_v - \delta_i}}$. Similarly, the probability of $X_{vi} = 0$ (failure) is given by:

$$P\{X_{vi} = 0\} = 1 - \frac{e^{\beta_v - \delta_i}}{1+e^{\beta_v - \delta_i}} = \frac{1+e^{\beta_v - \delta_i} - e^{\beta_v - \delta_i}}{1+e^{\beta_v - \delta_i}} = \frac{1}{1+e^{\beta_v - \delta_i}}$$

Where P is the probability of a correct answer, X_{vi} is the item score variable allocated to a response of person v, β_v is the ability of person v, i is a dichotomous item and δ_i is the difficulty of item i

3.7.2 Making use of the Rasch model

These equations relate the ability of persons and the difficulty of items. It 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).

If a person v is placed at the same location on the scale as the item i, then $\beta_v = \delta_i$ (i.e. $\beta_v - \delta_i = 0$) and the probability is equal to 0,5 or 50%. This is because $\frac{e^{\beta_v - i}}{1+e^{\beta_v - \delta_i}} = \frac{e^0}{1+e^0} = \frac{1}{1+1} = \frac{1}{2} = 0{,}5$

Therefore one can say that 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 the item's difficulty level is above a person's ability location, then the person will have a less than 50% chance of obtaining a correct answer.

If the item's difficulty level is below a person's ability location, then the person will have a greater than 50% chance of obtaining a correct answer (Bond & Fox, 2012). For example, suppose a learner has an ability of 0,4 logits, and in a test such a learner meets a question which is a difficulty of 0,4 logits, it would be predicted that such a learner would have a 50/50 chance of getting that question correct.

Figure 2. shows the probability of responding correctly and incorrectly to an item of particular difficulty. The point on the ability scale where the curves of 0 and 1 intersect is the location of the item. This is the point at which the probability of an incorrect response (0) and a correct response (1) are equally like i.e. 50% for either response. Around this point, the probability of a correct response decreases as ability decreases, and increases as ability increases.

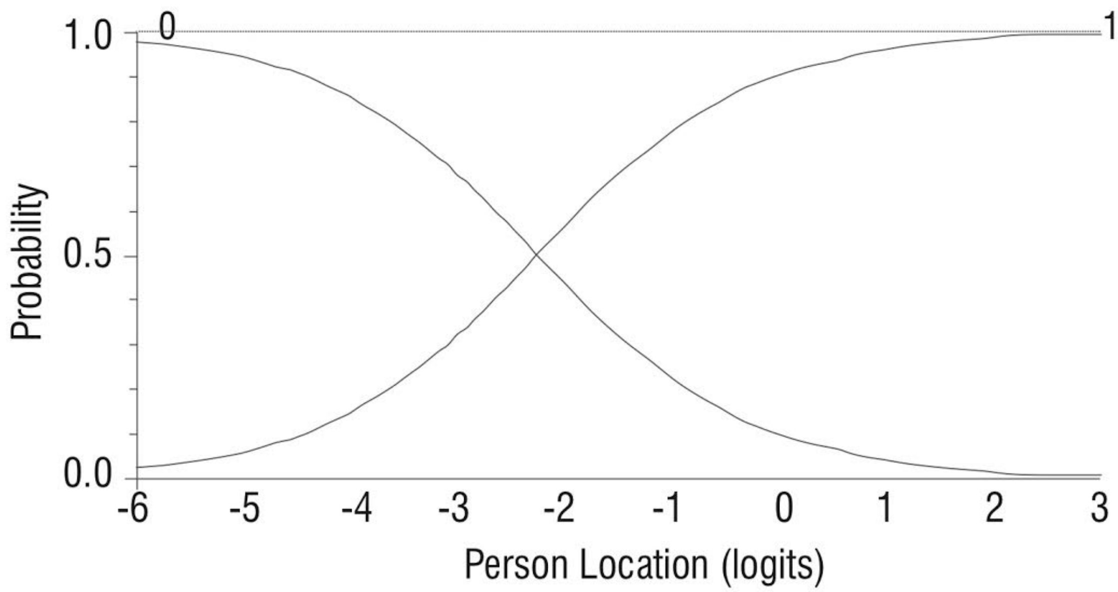


Figure 2. Category Probability Curves. These curves represent the probabilities of scores 0 and 1 on a single item as a function of ability (Bansilal, 2015).

3.8 Summary of chapter

In this chapter I outlined how the test was administered and to whom. I also included details on the content of the test and how it was coded. The Rasch model was presented and how it was used in this study explained.

In the next chapter I will describe the results of the test as well as what the Rasch analysis revealed about the difficulty of the questions in the test.

4. FINDINGS

4.1 Overview of chapter

In this chapter I present the findings of the test that was used in this study. I present summary statistics of the results of the June test, before the results of the Rasch analysis are reported in the subsection *individual item analysis.* The statistics presented here give an indication of which questions were functioning as predicted and which were not. In other words, the statistics help us to identify questions that are easy for learners of low attainment, but difficult for learners of high attainment, and vice versa.

Next, the test statistics are presented and the person-item (learner-question) map is discussed. These are graphical representations of the position of the algebraic attainment of learners relative to one another, and relative to the questions, as they are presented along one scale.

Lastly, the locations of the questions are presented to indicate types of questions that the learners found easy, and others that they were generally unable to answer. How learners may have used arithmetic for the former, and the algebra necessary to answer the latter is briefly discussed before elaborated on further in chapter 5.

4.2 Descriptive statistics

Generally learners performed poorly on this test, with an average of 21,87% (SD=7,14%) and a range 73% (MIN= 2%, MAX=75%). Indeed, the results of the top 9 learners are considered outliers when considered with the results of the rest of the sample (see figure 3.).

Figure 3. is a box and whisker plot of the results of the 235 learners who wrote the test. The median was 20%, which indicates half the learners scored below 20% and half above 20%, with the interquartile range being 14%.

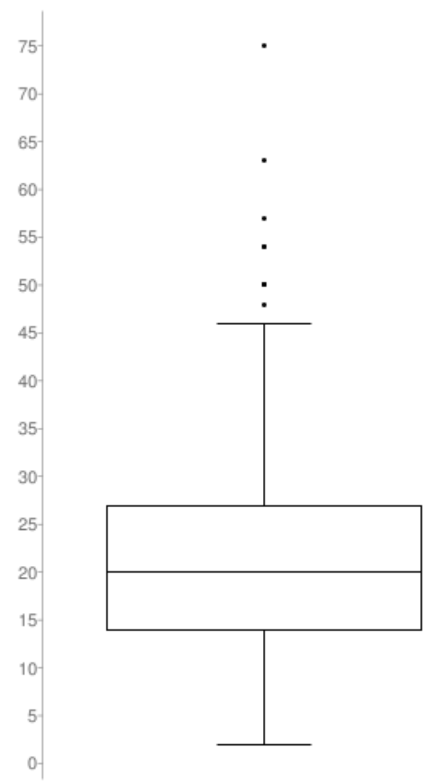


Figure 3. Box and whisker plot of the results of 235 learners who wrote the test in June. Box indicates the results of 50% of the learners and outliers are indicated by solid squares.

The top 9 learners (who scored 48%, 50%, 50%, 50%, 54%, 54%, 57%, 63%, and 75%) are considered statistical outliers in comparison with the results of their peers. 226 out of the 235 learners scored 45% and below for this test.

These results give an indication that overall the learners performed poorly in the test. Further analysis using Rasch measurement theory, such as the results presented in the next paragraph, allow more detail to be seen with regards to the difficulty of the individual questions.

4.3 Results of the Rasch analysis

Rasch analysis enables us to "see" far more from the data than basic descriptive statistics can show us. Rasch analysis enables us to measure the difficulty of each question on a scale. This enables us to compare the difficulty of the questions with the attainment of the learners (Bond & Fox, 2012).

4.3.1 Initial analysis

The model standard error is the precision of the Rasch measures when the data fit the model. The fact that the data fit the model and the degree to which this is true was discussed previously in chapter 3. The initial analysis of all 47 questions is summarised in table 3.

Table 3. Initial summary statistics. The mean and standard deviation of the locations and standard errors of the learners and question are shown. Units are logits.

	Questions [n=47]	Learners [n=235]
	Location (Standard error)	Location (Fit residual)
Mean	0,42 (0,47)	-2,87 (0,60)
SD	1,5 (0,48)	1,38 (0,24)

For the 47 questions in this test, the mean measure[4] is 0,41 (SD=1,5) logits. The model standard error for the questions is 0,47[5] (SD=0,48). For the 235 persons, the mean measure is -2,87 (SD=1,38). The model standard error for the persons is 0,60 (SD=0,24).

This indicates that the mean attainment of the learners is lower than the mean difficulty of the questions. This means that learners generally found the test difficult, which echoes the results of the descriptive statistics presented previously. The average learner in the sample has algebraic attainment lower than a question of average difficulty on this test.

The standard deviation for the question measure is 2,25 (which is above the ideal value of 1), while the standard deviation of the learner measure is 1,38 (closer to the ideal of 1). This suggests that the distribution of the learner measure is spread out, rather than clustered together. This leads me to discuss in the next paragraph the idea of learner separation and why it is relevant.

[4] Average measure of items or persons

[5] Standard errors of the measures

4.3.2 Separation of measures and reliability[6]

Learners are classified using *person separation.* The person separation index is an estimate of the internal consistency of the scale. Separation is considered low when the person separation index[7] is less than 2 and the person reliability is less than 0,8.

Table. 4. Separation of measures. The separation index and reliability statistics give an indication of how well learners and questions are separated along the common scale.

	Separation index (Reliability)
Questions [n=47]	3,21 (0,91)
Questions [n=45]	3,22 (0,91)
Learners [n=235]	1,78 (0,76)

Item hierarchy is verified using *item separation.* Item separation is considered low when the item separation index is less than 3 and the item reliability is less than 0,9. For the questions in this test the item separation index is 3,21 and the item reliability 0,91. This implies that the sample size is sufficient to confirm the question difficulty hierarchy (that is, the construct validity) of the test (Linacre, 2016b).

Point-measure correlation is a measure we can be used to determine which questions are not 'acting' as they should. In other words, learners with high attainment are getting these

[6] See notes at http://www.winsteps.com/winman/reliability.htm

[7] The ratio of sample or test standard deviation (corrected for estimation error) to the average estimation error. This is the number of statistically different levels of performance that can be distinguished in a normal distribution with the same "true" standard deviation as the current sample. When separation = 2, then high measures are statistically different to low measures.

questions incorrect, while learners with low attainment are getting them right. It helps us to answer the question “Do the responses to this question align with the attainment of the learner?” Negative correlations indicate that the responses to the question contradict the direction of the latent variable. This would indicate a need to check the question for reversed question wording, and rescore (Linacre, 2016b).

None of the questions in this test had zero or negative correlation. However, 2 questions were identified as having misfit statistics outside the recommended limits of -2,5 logits to 2,5 logits (Bond & Fox, 2012):

Question 7c) Multiply out: $8(\mathrm{p}-5)(\mathrm{p}+5)$

Question 5i) Simplify: $5-\mathrm{a}+\mathrm{a}$

Subsequent to the removal of these questions the learner-question location distribution was generated, as well as the learner-question map, reducing the number of questions from 47 to 45. Once these two questions were removed from the analysis, the question separation increased to 3,22; while the question reliability remained the same (i.e. 0,91).

Once the samples size is shown to be sufficient, the statistics regarding the learners can be considered. A reliability of 0,8 is necessary to reliably distinguish between learners with high attainment and learners with low attainment. For the 235 learners who wrote the test the separation index is 1,78 and the reliability 0,76. This implies two things: the test may not be sensitive enough to distinguish between learners of high attainment from learners with low attainment, and perhaps more questions in the test are needed in order to distinguish them (Linacre, 2016b).

Person reliability depends on sample ability variance; number of categories per question (more categories means higher person reliability); as well length of test (Linacre, 2016b). The lower the variance in the attainment of learners in a sample the higher the person reliability statistic. The greater the number of categories per question also increases the person reliability statistic. The rating scale length influences this statistic as well. In other words, the longer the test (in terms of the number of questions it has), the higher the reliability.

4.3.3 Learner-question location distribution

How learners and questions are distributed relative to one another is shown in the learner-question distribution (figure 4). This figure allows a visual comparison of the learners with varying degrees of algebraic attainment (the upper histogram) with questions of varying degrees of difficulty (the lower histogram).

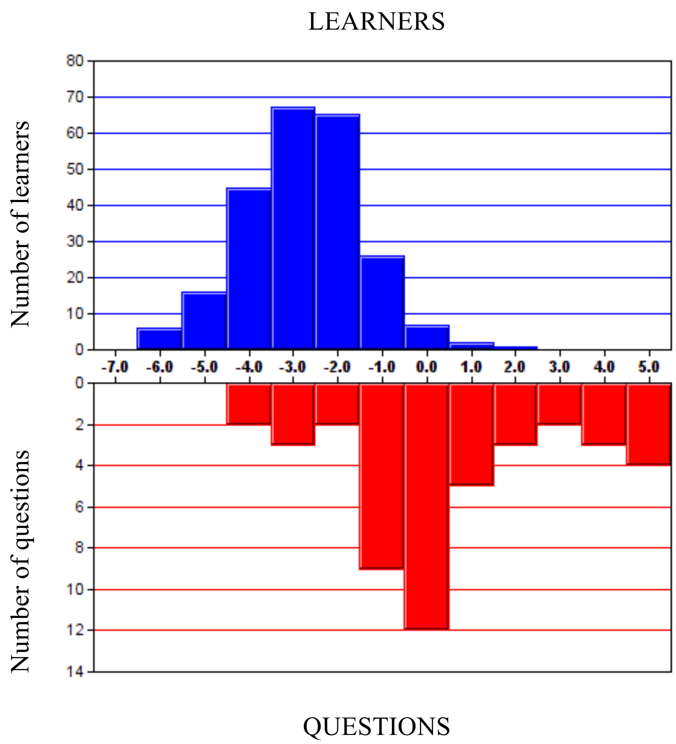


Figure 4. Learner-question location distribution. The upper histogram is the frequency of learners at various intervals of algebraic attainment, and the lower histogram indicates the number of questions at each location of difficulty.

The highest number of learners were located at -3,0 logits. The left-skewedness of the lower histogram contrasts with the right skewedness of the upper histogram. This indicates the learners have low algebraic attainment compared to the high difficulty of the questions. In other words, generally the difficulty of the questions is located above the attainment of the learners. This global view is described in more detail in the table 5.

Table 5. Location of questions and learners. Range and mean of locations of questions and learners.

	Range	Mean (SD)
Questions [n=45]	8,65	0,43 (2,29)
Learners [n=235]	8,45	-2,85 (1,41)

The locations of the questions range from -3,77 logits to 4,88 logits, with a mean of 0,43 (SD=2,29). The learner locations are estimated between -6,46 logits to 1,99 logits, with a mean of -2,85 (SD=1,41). The fact that the question location is higher than the mean of the question location, suggests that this algebra test was too difficult for this group of learners.

4.3.4 An analysis of the item location distribution

Rasch analysis enables the algebraic attainment of the learners to be compared with the difficulty of the questions are they are spread out along a common scale. This is represented in Figure 5.

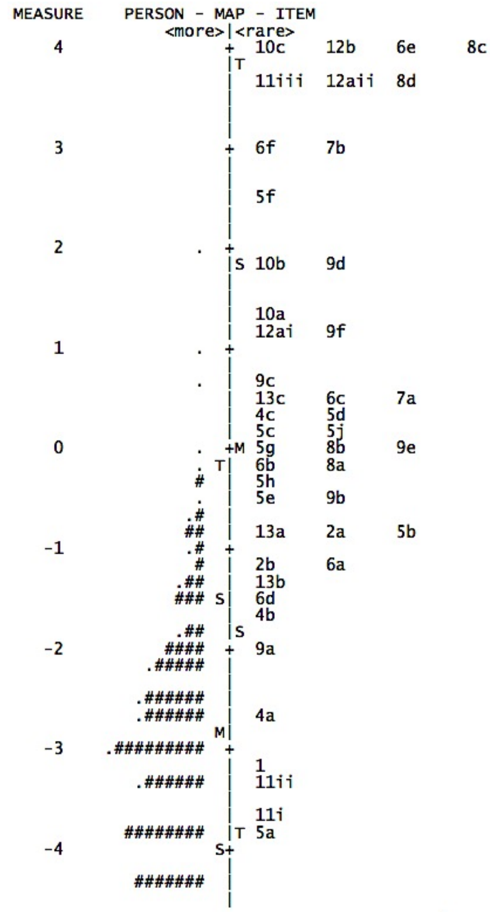


Figure 5. Learner-question map. This map approximates the algebraic attainment of the learners and the difficulty of the questions on a common scale. (# = Up to 2 learners. M = mean of learner or question distribution. S= 1 SD from the learner or question mean. T=2 SD from the learner or question mean.)

This map echoes the distribution in that the locations of the algebraic attainment of the learners (indicated by the hash marks) are generally below the location of questions to the right of the scale. Some questions have the same or similar difficulty and therefore might be considered unnecessary in terms of distinguishing learners. These questions and their respective locations are outline in the table below.

Table 6. Non-discriminating questions. Questions with the similar location i.e. of similar level of difficulty are listed in the same row with the respective location.

Question		Location (logits)
6a	Simplify $\frac{3\times5^9}{5^7}$	-1,23 and -1,10
2b	If $a = 2,\ b = -5,\ c = 3$, evaluate the following: $4 + (a - b)$	
2a	If $a = 2,\ b = -5,\ c = 3$, evaluate the following: $ab + 2c$	-0,86; -0.90 and -0,79
5b	Simplify: $2a + 5b + a =$	
13a	How many matchsticks are required for Figure 36?	
5e	Simplify $b(a - b)$	-0,51
9b	Solve $3x - 1 = 4 + x$	
6b	Simplify $\frac{a\times b^9}{b^7}$	-0,23 and -0,13
8a	Write down the x-intercept of the graph.	

When questions have the same or a similar location on the learner-question map it means that they have the same or a similar level of difficulty. Therefore they are not discriminating between learners with different algebraic attainment well, and one or more question may be unnecessary. One or more questions could possibly be removed from the test, although considerations other than location might be considered important. For example the algebraic structure of a question, or content covered by a question e.g. solving a linear equation, or simplifying an expression that included brackets. Further considerations for the inclusion of questions in this test will be discussed in chapter 5.

In the paragraphs that follow I will be considering patterns of attainment, i.e. which questions were of low difficulty for learners and similarities in these questions. We can see various patterns when considering the locations of questions in this test. Questions can be group in certain ways according to different aspects of algebra is assessed.

In Table 7 the location as well as the score of 11 questions are presented. These 11 questions were the 11 easiest questions for the learners. The questions were all located at -1,10 logits or less, the easiest being 5a, which asked the learners to simplify $5a + 2a$; followed by the 'spider' diagram question (11i and 11ii).

The questions are better described as arithmetic questions. This is because they all require knowledge of the four basic operations (addition, subtraction, multiplication and division) and are reminiscent of the structure of primary school mathematics questions. Further nuances of these 11 questions discussed in chapter 5.

Table 7. Arithmetic questions. The 11 easiest questions of the test are presented including the location and score of each question, as well as a description.

Questions (N=11)		Location (logits)	Score (and %)
5a	Simplify $2a + 5a$	-3,77	156 (66)
11i	$7 \rightarrow +3 \rightarrow \times 2 \rightarrow ?$	-3,62	150 (64)
11ii	$? \rightarrow +3 \rightarrow \times 2 \rightarrow 50$	-3,32	138 (59)
1	Write these numbers in order from smallest to largest: 30; -35; -2; -500; -10; 4	-3,16	131 (56)
4a	Write down the missing number in the space provided. $7 + 5 =$ ______ $+ 2$	-2,69	111 (47)
9a	$3x - 1 = 5$	-2,07	85 (36)
4b	Write down the missing number in the space provided: $4747 + 3945 =$ __ $+ 3943$	-1,69	70 (30)
6d	Simplify $\frac{15x+9x}{8}$	-1,50	63 (24)
13b	Which Figure would need exactly 51 matches?	-1,32	57 (24)
6a	Simplify $\frac{3\times5^9}{5^7}$	-1,23	54 (23)
2b	If $a = 2,\ b = -5,\ c = 3$, evaluate the following: $4 + (a - b)$	-1,10	50 (21)

Following these questions are 19 questions which less than 20% of the learners got correct. These questions were located between -0,90 and 0,68 logits. They involve elementary algebra, such as simplifying algebraic expressions including more than 1 variable (e.g. 5b Simplify: $2a + 5b + a$), and the product of two binomials (e.g. 7a

Multiply out $(2x + 1)(x + 4)$.) The common characteristics of these 19 questions are discussed in chapter 5.

Table 8. Grade 8 questions. Questions testing algebraic attainment to a grade 8 level are presented including the location and score of each question, as well as a description.

Question (N=47)		Location (logits)	Score (and %)
5b	Simplify: $2a + 5b + a =$	-0,90	44 (19)
2a	If $a = 2,\ b = -5,\ c = 3,$ evaluate the following: $ab + 2c$	-0,86	43 (18)
13a	How many matchsticks are required for Figure 36?	-0,79	41 (17)
5e	Simplify $b(a - b)$	-0,51	34 (15)
9b	Solve $3x - 1 = 4 + x$	-0,51	34 (15)
5h	Simplify $3a - b + a$	-0,37	31 (13)
6b	Simplify $\frac{a \times b^9}{b^7}$	-0,23	28 (12)
8a	Write down the x-intercept of the graph.	-0,13	26 (11)
5g	Simplify $a + 4 + a - 4 =$	-0,07	25 (11)
8b	Write down the y-intercept of the graph.	0,04	23 (10)
9e	Solve $\frac{a+1}{3} = 2$	0,04	23 (10)
5c	Simplify $(a + b) + a$	0,16	21 (9)
5j	Simplify $a + b + a - b$	0,16	21 (9)
4c	Write down the missing number in the space provided: $4747 + n =$ _____ $+ (n - 2)$	0,36	18 (8)
5d	Simplify $(a + b)b$	0,36	18 (8)
6c	Simplify $\frac{3b^7}{b^9}$	0,44	17 (7)
7a	Multiply out $(2x + 1)(x + 4)$	0,44	17 (7)
13c	Give an expression for the number of matches required for the n^{th} figure.	0,51	16 (7)
9c	Solve $1 - 3x = 5 - x$	0,68	14 (6)

Lastly, the 15 questions that tested algebraic attainment at a grade 9 level are grouped together in table 9. These include equations with fractions (e.g. 9f Solve $\frac{x+1}{3} + \frac{2x-1}{6} = \frac{x+2}{2}$);

factors (e.g. 12ai Factorise $x^2 - 4x$); and straight-line graphs (e.g. 8d Sketch the graph of $y = -x + 3$). Less than 4% of the learners got 9f correct, with locations ranging from 1,10 to 4,88.

Table 9. Grade 9 questions. Questions testing algebraic attainment to a grade 9 level are presented including the location and score of each question, as well as a description.

Question (N=15)		Location (logits)	Score
9f	Solve $\frac{x+1}{3} + \frac{2x-1}{6} = \frac{x+2}{2}$	1,10	10 (4)
12ai	Factorise $x^2 - 4x$	1,10	10 (4)
10a	Factorise $7x - 28$	1,36	8 (3)
9d	Solve $2p(p-4) - 8 = 2p^2 - 7p + 3$	1,91	5 (2)
10b	Factorise $7 - 28x$	1,91	5 (2)
5f	Simplify $b(a - b) =$	2,48	3 (1)
6f	Simplify $\frac{16xy+8x^2y}{4xy}$	2,92	2 (1)
7b	Multiply out $3(x+2)^2$	2,92	2 (1)
8d	Sketch the graph of $y = -x + 3$.	3,65	1 (0)
11iii	$x \to +3 \to \times 2 \to ?$	3,65	1 (0)
12aii	Factorise $x^2 - 2x - 8$	3,65	1 (0)
6e	Simplify $\frac{15x+9x}{3xy}$	4,88	0 (0)
8c	Write the equation of the straight line in the form $y = mx + c$.	4,88	0 (0)
10c	Factorise $7x^2 - 28$	4,88	0 (0)
12b	Simplify $\frac{x^2-4x}{x^2-2x-8}$	4,88	0 (0)

4.4 Summary of chapter.

Generally the learners performed very poorly on the test. The data suggests that the learners struggled to answer most of the questions in the test, and that the vast majority of questions were located beyond the algebraic attainment of the learners in this sample.

In chapter 5 I will be discussing the findings and presenting possible reasons for the patterns presented in section 4.3. Also discussed are ways in which the implementation, format and content of the test could be changed based on the findings in chapter 4.

5. DISCUSSION

5.1 Overview of chapter

I begin the chapter with a discussion of the descriptive statistics before moving onto individual item analysis. The algebra involved in particular questions will be discussed. Following this I will discuss the design of this test before finishing off with recommendations for possible changes that could be made to the test in the future.

5.2. Descriptive statistics

Generally the learners performed very poorly on the test, as indicated by a median of 20%, and an average of 22% (figure 3). However, this average is higher than the national average (11%) for grade 9 mathematics (DBE, 2014). Nationally 90% of South African grade 9 learners function at the not achieved level (0-29%) in grade 9 mathematics and about 1% of learners function at high achievement levels (80-100%) (DBE, 2014).

This result is unfortunate, and does not give one hope for the learners advancing to grade 10 and who eventually want to complete matric with mathematics as a subject. That being said, it does give further incentive for PD programs to be developed, such as the one offered by WMCS, in order to equip teachers address the low algebraic attainment.

The learners' results are important, nevertheless the main aim of the study is the design of the algebra test, and especially the individual questions. The results of the Rasch analysis give greater insight into the functioning of the questions of the test. These results are discussed in the paragraphs that follow.

5.3. Results of the Rasch analysis

The mean measure of the algebraic attainment of learners (1,38) is lower than the average measure of the questions (2,25) (table 3). This indicates that learners generally found the test difficult. In other words, the average learner has attainment lower than a question of average difficulty on this test. This means that the majority of the learners did have sufficient algebraic attainment to answer most of the questions in the test.

An important question in the design of the test was whether the test was measuring the desired construct (i.e. algebraic attainment) and if there were any questions that were not functioning as they should.

The test did measure the desired construct, and there were no questions with a negative correlation. A negative correlation would indicate learners with low attainment were getting the question correct while learners with high attainment were getting the question incorrect. The separation index of the questions (3,21) was sufficient to confirm the hierarchy of question difficulty (table 4), and increased when two items that had misfit statistics were removed from the analysis.

The general rule of assessment is all questions must be about the same thing (our intended latent variable i.e. what we are trying to measure), but then be as different as possible, so that they tell us different things about the latent variable (Linacre, 2016b) A test which shows all participants are doing very well or very badly (which Bond & Fox, 2012 call “coarse-grained”) is unlikely to inform the researchers about the spread of the attainment of the learners (Bertram & Christiansen, 2015).

The test included very few questions that were easy for the learners (figure 4). Most of the questions were located beyond the attainment of most of the learners. This is a contributing factor to the low performance of learners on the test (figure 3).

The question of least difficulty was 5a that read, “Simplify $2a + 5a$”. However, it is important to note that the location of 5a is -3,77 logits, a position on the scale above the algebraic attainment of 65 (27,66%) learners in the sample (N=235).

It is concerning that the *easiest* question of the test is located above the algebraic attainment of almost 30% of the learners (figure 5). In addition, almost one quarter of the test was made up of questions that were located above the attainment of all the learners (figure 5). The majority of these questions tested grade 9 algebra, and will be discussed in more detail in paragraphs that follow.

5.4 Questions involving arithmetic and algebra

I categorised questions in table 7 as questions that can be answered by using arithmetic. By this I mean that the learners can use arithmetic reasoning to answer the questions. For example the question 5a which read “Simplify 2a+5a” could be successfully answered by adding the numbers 2 and 5 followed by an ‘a’, rather than seeing the letter ‘a’ as an unknown. A similar train of thought could be used to answer question 6d “Simplify $\frac{15x+9x}{8}$” by adding and then dividing the natural numbers and ignoring x.

Generally, learners coped better with questions that involved the use of arithmetic (table 7), however they struggled to answer questions testing grade 8 algebra (table 8), and grade 9 algebra (table 9). These questions involved algebraic structures such as letters, numbers and exponents. Watson (2009) argues that in order for learners to understand algebraic symbolisation, they must have an understanding of the underlying operations, and be fluent with the notational rules. For example, the difference between the operations and notation of expressions that are similar in appearance, such as 2^x and x^2. The question of least difficulty (5a that read, "Simplify $2a + 5a$") is structurally simple, and mirrors the bond $2 + 5$.

The meaning and the symbol are essentially two kinds of learning, and seem to be most successfully learnt when learners know what is being expressed and also have sufficient time to become fluent at using the notation (Watson, 2009). In order to answer such questions, learners need to understand the nature and role of letters. Mapping symbols to meaning is learnt through repeated experience (Watson, 2009).

It appears that generally the learners have not successfully learnt how to read what is being expressed in algebraic structures, nor have they become fluent in the notation required to answer questions required algebraic knowledge and skill. The questions that involved the use of arithmetic were generally better answered. Despite the fact that such questions are not necessarily testing algebra, it is important that they are included in an algebra test as the responses to such questions give insight not only into whether learners are able to compute arithmetically, but also whether they have successfully transitioned from arithmetical thinking to algebraic thinking. This is a critical step in learning algebra.

5.4.1 Transition from arithmetic to algebra

The 11 easiest questions in the test ranged from -3,77 to -1,10 logits (table 7). These questions included expressions that could be described as structurally simple, and almost all of them involve arithmetic, and very little, if any, algebra.

The transition from arithmetic to algebra is not easy for most learners (Watson, 2009). Learners meet new concepts when they start learning algebra such as equations, formulae, functions, variables and parameters (Vergnaud, 1998). Symbols are higher order objects and become mathematical objects in and of themselves (Sfard & Linchevski, 1994).

Expressing generalities; solving equations; and working with functions are three approaches to enabling learners to transition from arithmetic to algebra (Watson, 2009). All three approaches have shortcomings. There is a strong commitment to arithmetic in the United States, particularly an emphasis on proficiency in fractions; which is seen as an essential precursor to algebra (Watson, 2009). The South African curriculum parallels the syllabus in the United Kingdom, in which secondary algebra is not taught separately from other mathematics. Rather than fractions leading to algebra, fractions are seen as particular instances of algebraic structures that can be calculated. Fraction calculations can be seen as enactments of relationships between rational structures, with those generalised enactments being expressed as algorithms (Watson, 2009). The difficulties learners had with the questions in the test involving algebraic fractions is discussed later in the chapter.

Part of the transition from arithmetic to algebra is learners realising for themselves the advantages of using algebra over arithmetic. Allowing learners to have the mind-set that any method that gives the right answer is as good as any other locks learners into methods that might be inappropriate, such as counting on (additive procedures) where multiplicative ones would be more appropriate (Watson, 2009). Question 6a, which read "Simplify $\frac{3\times5^9}{5^7}$" in an example of where learners were locked into multiplicative methods, rather than using exponential methods that were by far more appropriate in this context. Many learners used long multiplication (sometimes unsuccessfully) to compute 5^9 and 5^7, rather than using the law of exponents to simplify the expression.

Algebraic understanding of exponent laws is required for questions involving algebraic fractions. A misleading oversimplification is the view that algebraic understanding is solely a generalisation of arithmetic. Arithmetic involves the four basic mathematical operations, and not the laws of exponents. Indeed, number sense does precede formal algebra, but algebraic understanding is not wholly based on number sense (Watson, 2009).

The relationship between arithmetic and algebra is not a direct conceptual hierarchy or necessarily helpful. Learners can use arithmetic to solve some questions involving algebra, but arithmetic is limited, as are test-and-guess methods. Learners have to see for themselves the algebraic methods required for generalising, as well as solving equations where the unknown appears on both sides of the equals sign. This is described as the cognitive gap.

5.4.2 The cognitive gap

As previously mentioned, problems arise for the teaching and learning of algebra when algebra is seen solely as generalised arithmetic (Watson, 2009). Generalised arithmetic means the expression of general arithmetical rules using letters. The term 'cognitive gap' is used to describe difficulties seen as learners transition from arithmetic to algebra (Filloy & Rojano, 1989; Herscovics & Linchevski, 1994).

An example of a question that can be solved using arithmetic is question 9a, which read "Solve $3x - 1 = 5$". This question was answered correctly by 36% of the learners, and was located at -2,07 logits. This question is an example of an equation of the form $ax + b = c$. Learners tend deal arithmetically with such equations, by using inverse operations on the number to complete the arithmetical statement. This approach is considered arithmetical because it depends only on using operations to find a hidden number (Filloy & Rojano, 1989). That being said, the location of this question was above the attainment of more than two-thirds of the learners. Few learners were able to answer this question correctly; indicating that they were did not have sufficient arithmetic nor algebraic attainment to do solve for the unknown.

In contrast are questions such as 9b Solve $3x - 1 = 4 + x$ and 9c Solve $1 - 3x = 5 - x$. These questions are examples of equations of the form $ax + b = cx + d$ where the unknown appears on both sides of the equals sign. Learners have to use mathematical operations to maintain the equation by manipulating the expressions on either side of the equals sign. Adding to the difficulty learners experience in answering these questions is the appearance of the negative signs. Negative signs cannot be related to concrete understanding (Filloy & Rojano, 1989), and require a major shift from concrete models to a focus on structure (Vlassis, 2002).

I have included the following examples of possible solutions to question 9a to elucidate what I mean by an arithmetic means of answering the question, as opposed to an algebraic method that relies on an previously established concept of structure.

Arithmetic ("How can I add/subtract/multiply/divide to find the answer?")	Algebraic ("How can I use the structure of the equation to find the unknown?")
$3x - 1 = 5$	$3x - 1 = 5$
Learners may use mental maths to think through what x could possibly be, often choosing natural numbers, and starting with one. If they played around with the idea that x was one that would have calculated that the answer is 2 and thus try the "next" number i.e. 2, which just so happens to be the correct answer to this question. Thus a learner with low algebra attainment is able to get the correct answer as they are able to use their knowledge of arithmetic. Another method is for them to ask themselves, "What times 3 and then minus 1 will equal 5". This is an example of mental maths strategy.	Importance here is the maintenance of equivalence $3x - 1 + 1 = 5 + 1$ $3x = 6$ $\frac{3x}{3} = \frac{6}{3}$ $x = 2$

The transition from arithmetic to algebra requires learners to observe structure, and deal with abstract concepts such as negative numbers. Added to this are letters appearing in expressions and equations, and these letters have new and various uses and meanings.

5.4.3 The meaning of letters

This shift from arithmetic to algebra includes a change in the meaning of notation. Learners have to learn to tell the difference between similar notations used for arithmetic and algebra (Wong, 1997). For example, $3(4+5)$ is an expression that can be calculated. However an algebraic expression that looks similar, such as $a(b+c)$ is a structure of operations (Watson, 2009). When there is a mixture of letters and numbers in an expression such as $3(b+5)$, students assume this can be calculated and confusion can occur (Wong, 1997). Question 6e, which read "Simplify $\frac{15x+9x}{3xy}$" is an example of such a question. The structure of such an expression was particularly confusing to the learners as no learner was able to simplify the expression to $\frac{8}{y}$.

Added to the change of the meaning of notation is the fact that the algebra of unknowns is different to the algebra of variables. The algebra of variables is about expressing and transforming relations between numbers. The 'variable' view is dependent on the idea that expressions linked by the equals sign are not only equal numerically but also are also equivalent (Watson, 2009). That being said, leaners still need to retain the concept of 'unknown' when setting up and solving equations that have finite solutions. For example, $10-5=5(2x-1)$ is an equivalent statement where x is a variable. In contrast $10x-5=2x+1$ defines a value for a variable in which this equality is true (Watson, 2009).

In the question involving equality in the algebra test, half the learners were able to write down the missing number in the statement $7+5=\_\_+2$, whereas less than a third were able to do this for the statement $4747 + 3945 = \_\_ + 3943$. This demonstrates the potential for confusion between equality and equivalence as it relates to finding unknowns and expressing relationships between variables. Equivalence is when graphs coincide and equality is when graphs intercept. The equals sign has different uses – sometimes it means equality and sometimes equivalence – and learners have to learn the difference. It appears that generally the learners in this sample have yet to do so.

When simplifying algebraic expressions, learners need to understand that algebraic structures can be represented in equivalent forms, as well as which kind of manipulations can be carried out in order to do this. The performance of learners with regards to successfully simplifying algebraic expressions such as $a(a + b)$ is of particular importance to gain insight into algebraic attainment.

5.4.4 Simplifying algebraic expressions

As discussed in chapter 2, school algebra can described as: the manipulation and transformation of symbolic statements; generalisations of laws about numbers and patterns, the study of structures and systems abstracted from computations and relations; rules for transforming and solving equations; learning about variables, functions and expressing change and relationships; as well as modelling the mathematical structures of situation within and outside mathematics (Watson, 2009). Given this description, the sample of learners knows very little school algebra.

Question 5 included short (one, two or three terms) algebraic expressions that required the learners to simplify. The fact that many learners struggled with question 5 indicates that basic algebra remains a challenge for them. Therefore when they attempt questions involving the *application* of these basic algebraic principles, such as straight-line graphs (question 8), they are unable to succeed at answering them. The learners' performance regarding straight-line graphs will be discussed later on in the chapter.

In algebra the equals sign has an expanding meaning (Watson, 2009). In the arithmetic learners are exposed to in primary school it comes to mean ‘calculate’. In contrast, the algebra of high school using the equals sign to mean ‘is equal to’ or ‘is equivalent to’. This leads onto ideas of algebra being about expressing and transforming equivalent forms – not simply finding an answer or computing (Watson, 2009).

Many experiences are required to help learners recognise that an algebraic equation or equivalence is a statement about the relation between quantities of a combination of operations on quantities (Watson, 2009). Learners often reject ‘answers’ that are made up of more than one term as they are unfamiliar with such expressions. In other words, students want ‘closure’ and achieve this by compressing algebraic expressions into one term or into a whole number (Hart, 1981). This demonstrates a lack of understanding of what is being expressed (Watson, 2009)

Learners have to understand the meaning of the equals sign depends on the context of the question, as well as the order of operations that need to be carried out in order to manipulate the expression. BEDMAS is an acronym (i.e. brackets, exponents, division, multiplication, addition and subtraction) that is taught to help learners decide on the order of carrying out operations. Despite this, learners often simplify by combining terms by reading operations from left to right, or by ignoring notations such as brackets (Watson, 2009). Learners often do not understand the purpose of conventions and notations in algebra.

An example of this is learners not seeing the need for brackets when there are multiple operations. In algebra, brackets are sometimes necessary and at other times irrelevant. Associativity is the property that $a + (b + c)$ is the same as $(a + b) + c$. This property applies to multiplication as well. In such instances BEDMAS is unnecessary. Students get confused as to how to ‘undo’ such related operations, and how to undo other paired operations that are not associative (Brown & Coles, 1999). New rules such as BODMAS can be misused and do not effectively replace old rules based on familiarity, habit and arithmetic.

The learners performed poorly on questions requiring the use of the associative property. When it comes to algebraic expressions, learners may react to the visual appearance, without thinking about the meaning. An example of an expression with structure is $97 - 49 + 49$. If the emphasis is on computing then the learners will first attempt to subtract (potentially getting that wrong) and then add, rather than recognizing $-49 + 49 = 0$ and therefore the answer is 97. This may have happened in what would have seemed the simplest questions in the test i.e. $a + 4 + a - 4 =$; $5 - a + a =$ and $a + b + a - b =$. Furthermore, the inclusion of an equals sign after these questions may have made the learners think they must compute rather than observe structure (Watson, 2009). Perhaps this equals sign needs to be removed in subsequent iterations of this test to avoid this possibility.

The difficulty learners have simplifying algebraic expression demonstrate the five inherent difficulties in making direct shifts between arithmetic and algebra (Kieran 1981, 1989, 1992). Firstly, in algebra the focus is on relations rather than calculations. Secondly, students must not only understand operations but also the inverses of operations. Thirdly, learners need to convert the written word in an equivalent algebraic statement. Fourthly, letters and numbers are used together such that it could be required that the numbers be treated as symbols in a structure and therefore not evaluated. Lastly, in algebra the equals sign has expanded meaning. In arithmetic it is taken to mean ‘calculate’, whereas in algebra it more often means ‘is equal to’ or ‘is equivalent to’.

Such difficulties are not easily overcome. Function machines, substitution and patterns are all used to help learners transition from arithmetic to algebra.

5.4.5 Function machines, substitution and patterns

In the test, the question involving the “spider diagram” or “function machine” was well answered in comparison with the other questions. Function machines can lead to better understanding of what an equation is and the variable nature of x (Vergnaud, 1998), and appear in primary school mathematics (DBE, 2011). Questions involving the use of the function machine, and reversing it had similar locations (-3,62 and -3,32). Reversing the function machine was more difficult as learners who understand inversion might not understand that, when inverting a sequence of operations, the inverse operations cannot just be carried out in any order i.e. the order in which they are carried out influences the result. Learners seem reluctant to make use of brackets to indicate priority of operation in an expression.

Another algebraic activity involving the use of brackets is substitution. Many of the textbooks I have come across while teaching begin the topic of algebra in grade 8 with many exercises involving substituting values in place of letters in algebraic expressions. Overemphasis on substitution can mean that learners become preoccupied with arithmetical meaning and rules, rather than being able to recognise structure. Furthermore, substitution is problematic as it may contribute to the idea that expressions must be calculated, rather than helping learners to see the structure of such expressions, and that such structure has meaning in and of itself (Kieran, 1983). The focus should be on how structure is expressed rather than focusing on substituting in values so that the expression can be calculated. Although substitution can understand and verify relationships, it does not help learners understand what an expression means (Kieran, 1983). In addition, substitution may reinforce the idea that a letter can have only one value in one situation, and that different letters must have different values. Substitution is useful for learners to explore the equivalence of expressions but it may be a distraction to learners developing their algebraic thinking. Thus the poor performance of the learners on question 2 (the substitution question in this test) might not be telling us very much about their algebraic thinking, and because of this could possibly be removed from the algebra test in future.

A precursor to algebraic thinking is explaining a general result or structure using full sentences (Watson, 2009). Some questions in the algebra test asked for an explanation of an answer, however this was not scored. A verbal description enables students to bridge between observing relations and writing them algebraically. Instead of finding the general term perhaps the question should rather ask the learners to write the generality of the pattern of matchsticks in words. Furthermore, the use of a diagram could be adding difficulties to an already difficult subject. Indeed term-to-term descriptions are far easier when data is expressed sequentially such as in a table (Reed, 1972)

One approach to address inherent difficulties in algebra is to draw on our natural inclination to observe patterns and to impose patterns on disparate experiences (Reed, 1972). Like the need to find the general term for a pattern in order to find terms such as T_{100}, the expectation is that this generates a need for algebraic symbolization, similar to the need to use algebra to solve algebraic equations, where test-and-guess is not helpful and time-consuming.

5.4.6 Solving algebraic equations

One of the most fundamental skills of algebra is solving for an unknown in a linear equation such as $3x - 1 = 4 + x$. Often part of solving equations is simplifying expressions of each side of the equals sign. Rules for transforming expressions and solving equations are often confused, misapplied or forgotten by learners (Watson, 2009). Furthermore, learners may try to apply arithmetical meanings to algebraic expressions. For example for question 6d Simplify $\frac{15x+9x}{8}$, some learners assumed x to be 1 and gave an answer of 3 and not $3x$. Answers such as theses are associated with notational manipulation, or generalised arithmetic, being over-emphasised, resulting in learners seeking to get as concise an answer as possible. An understanding of inverse operations (i.e. addition is the opposite of subtraction, and multiplication is the inverse of division, for example) is mostly notably demonstrated in the solving of equations. The equations in this test were very poorly answered. When solving equations, many students guess and check, rather than “undoing: the algebraic structure (Watson, 2009).

This inability to not structure and manipulate expressions is also demonstrated by the poor performance of learners on questions involving algebraic fractions. Students can be confused by expressions that combine numbers and letters such that learners need to learn to 'read' expressions structurally even when numbers are involved (Watson, 2009).

5.4.7 Simplifying algebraic fractions

There are inevitable confusions that arise in symbolic conventions. Therefore division and rational structures are problematic such as questions like 6b simplify $\frac{a \times b^9}{b^7}$ than 6c simplify $\frac{3b^7}{b^9}$. Learners may have found question 6c to be more difficult as 6b is very clearly in the realm of algebra and rules about letters as it only includes letter. (Wong, 1997). Learners are confused by expressions that include numbers and letters (Watson, 2009). Expressions such as these are required to be read structurally, even when numbers are involved. Learners have to shift from seeing $\frac{1}{4}$ as 'one pizza shared amongst four family members' to seeing $\frac{1}{4}$ (and fractions in general) as expressions with meaningful structure. Rational structures are a specific class of objects indicating a particular quantitative relationship (Watson, 2009). Learners have to come to realise that division is a tool for constructing a rational expression, and not solely about sharing.

So far the algebraic expressions considered have been for questions that are at a grade 8 level (DBE, 2011), or can be solved using arithmetic. If learners are struggling with questions such as these it is not surprising that they struggle with questions involving grade 9 algebra, especially products, factors and straight-line graphs.

5.4.8 Grade 9 algebra

If learners' main experience of algebra is to simplify expressions, the shift to using the new kinds of transformations afforded by algebra is hindered (Dettori, Garutti, and Lemut, 2001). To understand algebraic notation requires an understanding that terms made up of additive, multiplicative and exponential operations e.g. $4a^3b - 8a$ are variables rather than instruction to calculate, and have a structure and equivalent forms.

The poor performance of learners on questions requiring them to multiply or factorise algebraic expressions means the learners are not seeing expressions as structures (Carpenter and Levi, 2000). These need to be seen as relations, by combining operations and inverses. This seeing relationship seems to depend on the ability to discern details (Piaget, 1969) and an application of an intelligent sense of structure (Wertheimer, 1960), as well as knowing when to handle specifics and when to stay with structure. The most difficult questions were questions involving expressions that included rational structures, parameter, and the need to manipulate into equivalent forms:

Question 6e Simplify $\frac{15x+9x}{3xy}$

Question 8c Write down the equation of the straight line in the form $y = mx + c$

Question 10c Factorise $7x^2 - 28$

Simplify 12b Simplify $\frac{x^2-4x}{x^2-2x-8}$

Indeed, none of the learners in the sample got any of the aforementioned questions correct. These learners have not developed a structural perspective on algebraic expressions.

The root of algebraic development is learners understanding of the meaning of letters, and how one goes about using letters to express mathematical relationships. School algebra involves the use of different nature and roles of *letters*, that is, as unknowns, variables, constants and parameters (Watson, 2009). For example, 8c asked the learners to write the equation of the graph into a standard form $y = mx + c$. This requires that the learners know that m and c are parameters i.e. the gradient and the ordinate of the point of interception of the graph with the vertical axis. A parameter is a value that defines the structure of a relation. For $y = mx + c$, variables are x and y, while m and c define the relationship. The parameters have to be fixed before we can consider the covariation of x and y. It is difficult to explain how to know the difference between a parameter, a constant and a variable. Interpretation is relation to the context: an algebraic equation, expression, equivalence, function or other relation. The learners have not yet undergone the critical shift from seeing a letter representing an unknown, or 'hidden', number defined within a number sentence such as $3 + x = 8$ to seeing it as a variable, as in $3 + x = y$. Understanding x as some kind of generalised number which can take a range of values is seen as a bridge from the idea of unknown to that of variables (Bednarz, Kieran and Lee, 1996).

That being said, a central issue is that in most contexts for a letter to represent anything, the learner must understand what is being represented, yet it is often only by the use of a letter that what is being represented can be understood. This is an essential shift in abstraction. This shift takes multiple experiences. The learners may not have had sufficient exposure to dealing with the various representations of products, factors and straight-line graphs by the time this test was written. The timing of the test is discussed of the next section of this chapter, as well issues relating to sampling, questions, coding and administration.

5.5. Recommendations

5.5.1 Sample

The Rasch analysis revealed that the learner separation index is 1,78 and the learner reliability 0,76. This implies that the test may not be sensitive enough to distinguish between learners of high attainment from learners with low attainment (Linacre, 2016b).

There are two possible changes that could be made for the next iteration of the test. One is that learners from a greater range of attainment sit the test. The second change is that the test could be lengthened (i.e. more questions are included). However, the aim of the February test was to pilot questions in order to minimise the amount of questions in the June test. Thus it would not be consistent with this aim to lengthen the test by adding more questions. Later on in this discussion, I will speak to a possible solution to this problem.

5.5.2 Types of questions

It is imperative that a test be constructed with questions ranging in difficulty. This test had too many difficult questions (figure 4), and perhaps more questions that would be considered at a grade 6 or 7 level could be included.

Considering the learner-question map in figure 5, including questions targeting the attainment levels of the learners located between 1 (which read, "write these numbers in order from smallest to largest: $30; -35; -2; 500; -10; 4$") and 4a ("write down the missing number in the space provided $7\ + 5 + = \_\_\ +\ 2$") as well as 4a and 9a ("Solve: $3x - 1 = 5$") would be beneficial to differentiating learners at this a lower level of algebraic attainment.

The test needs more middle difficulty questions and fewer very hard questions. That being said, in the case of this test, "very hard" questions are indeed questions testing knowledge and skills required at a grade 9 level. It would be counterproductive to exclude questions solely on the basis that learners found them difficult. Such questions must be included in order to see if grade 9 learners are indeed learning algebra that is expected at a grade 9 level. Furthermore, if the purpose of the test is to gauge if the learners have increased in learning gains, then it is necessary to have questions that they might find difficult at the beginning of grade 9, but which they are able to master at the end of grade 9.

This leads me to argue that the test told us more about what the learners *do not* know, that about what they *do* know. If the majority of questions are located above the algebraic attainment majority of learners, we are essentially being told very little about their algebraic attainment.

The questions that were in fact located at a similar position to the majority of the learners have a fairly even spread (compared to the rest of the questions) and what appears to be a linear progression (figure 5). When there are few questions close to the algebraic attainment of learners a question-targeting problem exists (Linacre, 2016b). For dichotomous tests, no nearby questions means less precision of measurement. This in turn means larger learner standard errors (Bond & Fox, 2012).

Questions from a grade 4,5,6 level could be included as precursors to questions at a 7,8,9 level. For example, moving from whole number to monomial to binomial expressions in questions in involving the simplification of algebraic fractions. Consider a suggested formatting of a question involving division is given below:

Simplify the following

a) $\frac{3}{24}$

b) $\frac{x}{2x}$

c) $\frac{y^2+y}{y}$

A similar approach could be taken to multiplication

a) 3×24

b) $2x(x)$

c) $5(x+1)$

If, for example, we have learners responding correctly to questions in the example above labelled a, we can assume they are at a grade 6 level; if they respond correctly to questions labelled b, we can conclude they are at a grade 8 level etc. This is an oversimplification, but my main point is to gather questions that tell us more about what learners *do* know, than what they *do not* know.

5.5.3 Purpose of the test

There may possibly be confusion as to the purpose of this test. Although the main aim of administering the test is to measure algebraic attainment of learners, there is incentive for research to include questions, such as ones involving brackets and other symbolic structures. Such questions are useful for insight into topics of interest such as the conceptual understanding by learners of the solving of equations.

Still further is the obligation on the part of the project to give feedback to principals of the schools involved in the sampling as to which topics learners need remediation. Indeed the test must be of some resemblance to what the teachers and learners are used to in terms of a typical grade 9 mathematics test. This may lead to a confusion of aim and an overload of some topic areas (such as equations previously mentioned) and a neglect of others.

What is necessary, in my opinion, is a distillation of the key algebraic skills required for successful progression to the higher grades, and then for the precursors to such skills to be identified and tested, alongside these skills.

That being said, the learners may not have had sufficient time by the time the test was written to have multiple experiences of working with algebraic expressions, particularly those involving products, factors and the straight-line graphs. This leads to a critique on how the test was administered, as well as possible improvements.

5.5.4 Administration of the test

This test was written by grade 9 learners in the middle of their grade 9 year. Some skills tested in this test might not have yet been taught in class yet, and might not have been mastered in grade 8.

This could be particularly true for the question involving the straight-line graph and factorisation. These sub-questions were above the attainment of the majority of learners, indicating they were possibly unfamiliar with such questions regarding graph interpretation and using factorisation to simplify algebraic fractions. An example from my experience of marking these scripts was that if learners had just learnt factorisation in class they tended to do very well at the questions involving factorisation, yet they also applied these skills to questions involving products, leading to unnecessary and irrelevant solutions. Perhaps the timing of this test should be at the end of school holiday, when skills learnt very recently in class are not applied inappropriately.

The previous example highlights the issues that occur when a method of coding is used. Although there was constant error due to there being only one coder, the coding process is not without fault. Only the final answer was considered in order to allocate a 0 or a 1 score. Coming back to the previous example where learners applied factorisation skills to products questions, this is essence means that although the learners could demonstrate the skill that they could multiply out, they would have received a 0 score, as their final answer would be in factored format.

If the test will be coded in this way in future, I would consider it consistent to make the majority of questions have very little to no working to get the final answer. Multiple choice could also be a considered format for the test. However, this would depend on whether the learners have been exposed to such a manner of test-taking in the past, as naturally learners take multiple-choice to mean “multiple-guess” and tend to apply very little of their time or their thinking in determining the correct answer.

In addition, to increase depth of insight it is suggested that future iterations of this test make use of questions of partial credit (Masters, 1982), instead of the test solely consisting of dichotomous questions. Indeed this would require minimal change in the data collection method, but could give significant insight into how rescoring questions can lead to a fairer outcome (Bansilal, 2015). This could also increase the learner separation and reliability statistics. However, allocating partial credit means a more sophisticated and time consuming marking process.

5.6. Overview of chapter

In this chapter the descriptive statistics, and the results of the Rasch analysis were discussed. The degree of algebraic attainment of the learners was reviewed in light of research into the relationship between arithmetic and algebra. Suggestions for changes to the questions, timing and coding of the tests were made in order to possibly improve the depth of insight of the results.

6. CONCLUSION

We can go about designing a test that is fit for purpose and useful by making use of a variety of strategies for effective test development (Downing, 2006), including statistical analysis of the results of the test. Rasch analysis is a useful tool for analysing test questions. Rasch analysis of this algebra test has revealed that generally this sample of grade 9 learners have low algebraic attainment. Indeed, the attainment of the average learner in this sample is located below questions of average difficulty. Applying the Rasch Model to this assessment instrument has contributed to its improvement by shedding light on question hierarchy and separation. Although questions were well separated, they were too difficult for the learners in this sample. Means of improving this test could include increasing the length of the test, making use of partial credit questions, including learners of higher attainment and including questions that would be considered at grade 6 or 7 level.

In this research report I discussed the need for an algebra test to be designed for the purposes of assessing the algebraic attainment of learners whose teachers attend a professional development course run by the WCMS. Algebraic thinking can be counter-intuitive at times (Watson, 2009). Fluency in using algebra requires good understanding of the symbol system, and abstract meanings of letters that change according to context. The process involved in shifting from an arithmetic view to an algebraic view, and the reification of new ideas, are unlikely to arise naturally, and thus require the deliberate action of teachers and teaching (Filloy & Sutherland, 1996). Development of algebraic reasoning can happen in deliberately designed educational contexts (Brown & Coles, 1999)

Teaching algebra by offering situations in which symbolic expressions make mathematical sense (e.g. through multiple representations, expressing generality, and equating functions) is more effective in leading to algebraic thinking and skill rather than the teaching of technical manipulation and solution methods as isolated skills. Fluency in understanding symbolic expressions seems to develop through use, and also contributes to effective use – essentially a two way process. Algebraic understanding takes time, multiple experiences and clarity of purpose. Learners need support in shifting to representations of generality, understanding relationships and expressing these in conventional forms. Such methods need to occur in conjunction with complex pedagogy.

It is my hope that this research report has contributed meaningfully to the program run by the WCMS that seeks to train teachers in this complex pedagogy, with the ultimate aim being the increase in learning gains of South African grade 9 learners of the future.

APPENDICES

A: Parent Information Form

Date

Dear Parent

My name is Samantha Ehrlich and I am a masters student in the School of Education at the University of the Witwatersrand. I am doing research on Learning Gains in Mathematics.

In our research we want to find out whether learners of Mathematics benefit when their teachers participate in a professional development course offered by the Wits Maths Connect Secondary project. To do this research we first need to design a test for learners. We need to check that the test is fair and that it tests the maths concepts that we think it tests.

I was wondering whether you would give consent for your child to be part of the group who will write a draft version of the test. Your child's test responses will help us to see whether the questions make sense to learners and whether they interpret the questions in the same we expected them to interpret the questions.

The test will take place at a time that is agreed with the school and will take place on the school property.

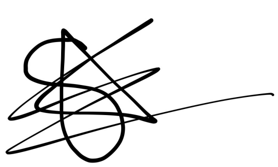
Although we talk about a *test*, this test is not for marks and learners are not expected to study for it. Your child's participation is voluntary, which means that s/he doesn't have to do it. Also, if your child decides halfway through that s/he would prefer to stop, this is completely his/her choice and will not affect him/her negatively in any way.

We will not be using your child's own name but I will make one up so no one can identify your child. All information about your child will be kept confidential in all our writing about the study. Also, all collected information will be stored safely and destroyed 5 years after we have completed the project.

Your child has also been given an information sheet and consent form. At the end of the day it is your child's decision to join us in the study.

Please feel free to contact me if you have any questions.

Thank you



Mrs Samantha Ehrlich
Masters Student
Wits Education Campus
1512315@students.wits.ac.za
072 542 5212

B: Parent Consent Form

Please fill in and return the reply slip below indicating your willingness to allow your child to participate in the research project called Learning Gains in Mathematics

I, ________________________ the parent of ______________________

Permission for questionnaire/test	**Circle one**
I agree for my child to write a test for this study.	YES/NO

Informed Consent
I understand that:

- My child's name and information will be kept confidential and safe and that my child's name and the name of the school will not be revealed.
- My child does not have to answer every question and can withdraw from the study at any time.
- All the data collected during this study will be destroyed 5 years after completion of the project.

Sign_____________________________ Date___________________________

C: Learner Information Form

Date

Dear Learner

My name is Samantha Ehrlich and I am a masters student in the School of Education at the University of the Witwatersrand. I am doing research on Learning Gains in Mathematics.

In our research we want to find out whether learners of Mathematics benefit when their teachers participate in a professional development course offered by the Wits Maths Connect Secondary project. To do this research we first need to design a test for learners. We need to check that the test is fair and that it tests the maths concepts that we think it tests.

I was wondering whether you would be part of the group who will write a draft version of the test. Your test responses will help us to see whether the questions make sense to learners and whether you interpret the questions in the same we expected you to interpret them.

The test will take place at a time that is agreed with your school and will take place on the school property.

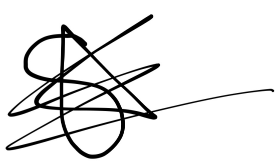
Although we talk about a *test*, this test is not for marks and you are not expected to study for it. Your participation is voluntary, which means that you don't have to do it. Also, if you decide halfway through that you prefer to stop, this is completely your choice and will not affect you negatively in any way.

We will not be using your own name but I will make one up so no one can identify you. All information about you will be kept confidential in all our writing about the study. Also, all collected information will be stored safely and destroyed 5 years after we have completed the project.

Your parents have also been given an information sheet and consent form, but at the end of the day it is your decision to join us in the study.

I look forward to working with you! Please feel free to contact me if you have any questions.

Thank you



Mrs Samantha Ehrlich
Masters Student
Wits Education Campus
1512315@students.wits.ac.za
072 542 5212

D: Learner Consent Form

Please fill in the reply slip below if you agree to participate in the study called: Learning Gains in Mathematics

My name is: ________________________________ Grade and class: ____________________

Permission for questionnaire/test **Circle one**
I agree to write a test for this study. YES/NO

Informed Consent
I understand that:

- my name and information will be kept confidential and safe and that my name and the name of my school will not be revealed.
- I do not have to answer every question and can withdraw from the study at any time.
- all the data collected during this study will be destroyed 5 years after completion of the project.

Sign______________________________ Date___________________________

E: School Consent Letter

Date

Dear Mr/Mrs/Miss/Ms/Dr XXX

I write to you to invite XXX Secondary School to participate in research that is being conducted through the School of Education at the University of the Witwatersrand.

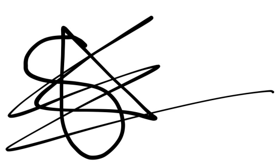
The research focuses on learning gains and I'd like to give you some background to the project. In 2013 we investigated the impact of the Transition Maths 1 and 2 courses on learners' gains in Mathematics over one year. We invited 5 project schools to participate and we tested approximately 800 Grade 10 learners in February and October of that year. There were 21 teachers who participated in the study, some of whom participated in a TM course and some who didn't. The results showed that the learners taught by teachers who participated in a TM course made larger gains than those taught by teachers who did not participate in the course. However, we treat these results as indicative evidence rather than conclusive evidence that the TM courses have an impact on learning gains.

In 2016 to 2019 we wish to conduct a more rigorous study of the impact of the TM1 course on learning gains. The first step of this research is to develop a new test for learners. In February 2016 approximately 120 Grade 10 Mathematics learners write a trial version of the test in February 2016. This trial version is being refined and in May/June 2016 I would like to pilot the new test with Grade 9 learners.

I was wondering whether you would agree to allow me to pilot the test with one of your classes of Grade 9 learners in May/June 2016. This will take approximately 45 min at a time convenient to the school. The learners will not be advantaged or disadvantaged in any way. They will be reassured that they can withdraw their permission at any time during this project, including during the test, without any penalty. There are no foreseeable risks in participating in this study. The participants will not be paid for this study. The names of the research participants and identity of the school will be kept confidential at all times and in all academic writing about the study. Your individual privacy will be maintained in all published and written data resulting from the study. All research data will be destroyed 5 years after completion of the project.

Please let me know if you require any further information. I look forward to your response.

Yours sincerely,



Mrs Samantha Ehrlich
Masters Student
Wits Education Campus
1512315@students.wits.ac.za
072 542 5212

F: Ethics

Ethics clearance has been granted to WMCS for this project under the protocol number and 2016ECE003S. The protocol number for my study is 2016ECE009M. I have attached ethics clearance from the University of the Witwatersrand for my study and the Gauteng Department of Education for the project on the following pages.

Wits School of Education

WITS UNIVERSITY

27 St Andrews Road, Parktown, Johannesburg, 2193 Private Bag 3, Wits
2050, South Africa. Tel: +27 11 717-3064 Fax: +27 11 717-3100 E-mail: enquiries@educ.wits.ac.za
Website: www.wits.ac.za

26 April 2016

Student Number: 1512315

Protocol Number: 2016ECE009M

Dear Samantha Anne Ehrlich

Application for Ethics Clearance: Master of Science

Thank you very much for your ethics application. The Ethics Committee in Education of the Faculty of Humanities, acting on behalf of the Senate has considered your application for ethics clearance for your proposal entitled:

A Rasch analysis of an algebra test written by Grade 9 learners.

The committee recently met and I am pleased to inform you that **clearance was granted**. However, there were a few small issues which the committee would appreciate you attending to before embarking on your research.

The following comments were made:

Information Letters:

- minor errors – insert in this line, "the questions in the same [way] we expected you to interpret them"
- Approximately 120 Grade 10 Mathematics learners write a trail [trial] version of the test in February 2016.

Please use the above protocol number in all correspondence to the relevant research parties (schools, parents, learners etc.) and include it in your research report or project on the title page.

The Protocol Number above should be submitted to the Graduate Studies in Education Committee upon submission of your final research report.

All the best with your research project.

Yours sincerely,

MMaseta

Wits School of Education

011 717-3416

Cc Supervisor: Professor Mike Askew and Dr Craig Pournara

For administrative use: **Reference no. D2016 / 397 G** Enquiries: Diane Buntting (011) 843 6503

GAUTENG PROVINCE

EDUCATION

REPUBLIC OF SOUTH AFRICA

GDE GROUP RESEARCH APPROVAL LETTER

Date:	**19 February 2016**
Validity of Research Approval:	**19 February 2016 to 30 September 2016**
Names of Researchers:	**Dr C. Pournara; Askew M; Sanders Y. Ntow F; Alshwaik J. and Owusu J.**
Name of Supervisor/s:	**Prof. K. Brodie**
Address of 1st Researcher:	**P.O. Box 4531; Pinegowrie; 2123**
Telephone/ Fax Number/s:	**011 717 3253; 082 696 8381; 076 158 3758; 072 325 5304; 063 332 9080; 060 883 8678; 078 033 8863; 082 696 8381**
Email addresses:	**craig.pournara@wits.ac.za; Michael.askew@wits.ac.za; Yvonne.gubler@gmail.com; Forster.ntow@wits.ac.za; Jehad.alshwaik@wits.ac.za; James.owusu@wits.ac.za.**
Research Topic:	**Development of test instrument to measure learning gains of Grade 9 learners in Mathematics**
Number and type of schools:	**TWO Secondary Schools**
District/s/HO	**Johannesburg East**

Re: Approval in Respect of Request to Conduct Research

2016/02/22

G: The research instrument

1) Write these numbers in order from smallest to largest
   $30 \quad -35 \quad -2 \quad -500 \quad -10 \quad 4$

2) If $a = 2,\ b = -5,\ c = 3$, evaluate the following. Show all your working.
   a) $ab + 2c$
   b) b) $4 + (a - b)$

3) *omitted here*

4) Write down the missing number in the space provided.
   a) $7 + 5 =$ ______ $+ 2$
   b) $4747 + 3945 =$ _________ $+ 3943$
   c) $4747 + n =$ ________ $+ (n - 2)$

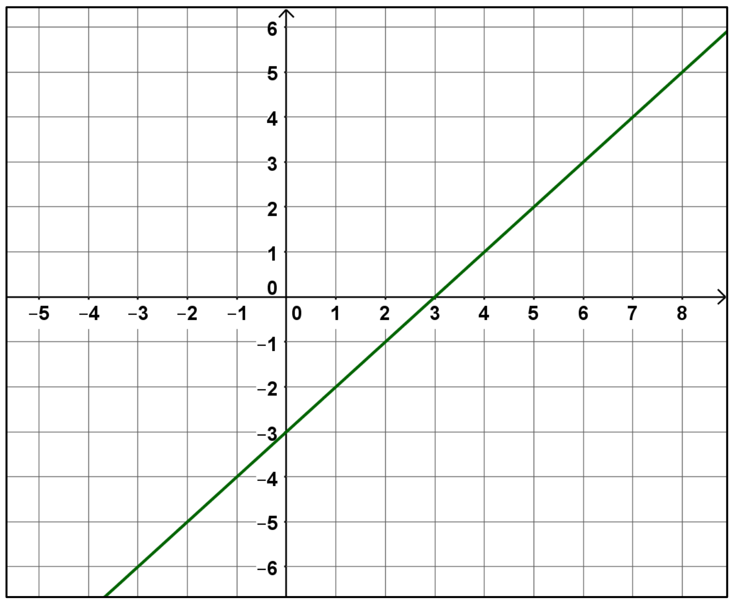
5) Simplify
   a) $2a + 5a =$
   b) $2a + 5b + a =$
   c) $(a + b) + a =$
   d) $(a + b)b =$
   e) $b(a - b) =$
   f) $3a - (b + a) =$
   g) $a + 4 + a - 4 =$
   h) $3a - b + a =$
   i) $5 - a + a =$
   j) $a + b + a - b =$

6) Simplify the following. No denominator is zero.

   a) $\frac{3\times5^9}{5^7}$
   b) $\frac{a\times b^9}{b^7}$
   c) $\frac{3b^7}{\mathrm{b}^9}$
   d) $\frac{15x+9x}{8}$
   e) $\frac{15x+9x}{3xy}$
   f) $\frac{16xy+8x^2y}{4xy}$

7) Multiply out:
   a) $(2x+1)(x+4)$
   b) $3(x+2)^2$
   c) $8(p-5)(p+5)$

8) Look at the diagram below that shows a straight-line graph.

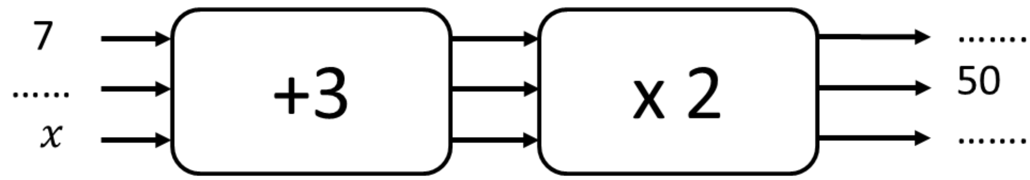


   a) Write down the x-intercept of the graph.
   b) Write down the y-intercept of the graph.
   c) Write the equation of the straight line in the form $y = mx + c$.
   d) Sketch the graph of $y = -x + 3$ on the set of axes given above.

9) Solve for the unknown:
   a) $3x - 1 = 5$
   b) $3x - 1 = 4 + x$
   c) $1 - 3x = 5 - x$
   d) $2p(p-4) - 8 = 2p^2 - 7p + 3$
   e) $\frac{a+1}{3} = 2$
   f) $\frac{x+1}{3} + \frac{2x-1}{6} = \frac{x+2}{2}$

10) Factorise fully
   a) $7x - 28$
   b) $7 - 28x$
   c) $7x^2 - 28$

11) The diagram below shows inputs and outputs for the machine diagram. Work out the missing information and write your answers in the spaces provided.



12) a) Factorise fully:

i) $x^2 - 4x$

ii) $x^2 - 2x - 8$

b) Simplify. The denominator is not equal to zero.

$$\frac{x^2-4x}{x^2-2x-8}$$

13) Matchsticks are arranged as shown:

Figure 1 Figure 2 Figure 3

d) How many matchsticks are required for Figure 36?

e) Which Figure would need exactly 51 matches? Explain how you got your answer.

f) Give an expression for the number of matches required for the n^{th} figure.

REFERENCE LIST

Allen, M. J. & Yen, W. M. (2002). *Introduction to measurement theory.* Long Grove, IL: Waveland Press.

Anderson, D., Alonzo, J. & Tindal, G. (2012). *easyCBM CCSS Math Item Scaling and Test Form Revision (2012 – 2013) Grades 6 – 8* (Technical Report No. 1313). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Bansilal, S. (2015). A Rasch analysis of Grade 12 test written by mathematics teachers. *South African Journal of Science. 111(5/6),* 1-9.

Bednarz, N., Kieran, C. & Lee, L. (1996). Approaches to algebra: perspectives for research on teaching. In N., Bednarz, C., Kieran, and L. Lee (Ed.). *Approaches to algebra: perspectives for research on teaching,* 3-14*, Dordrecht: Kluwer.*

Bertram, C. & Christiansen, I. (2015). *Understanding Research: An Introduction to Reading Research.* Pretoria: Van Schaik Publishers.

Bond, T. G. & Fox, C. M. (2012). *Applying The Rasch Model: Fundamental Measurement in the Human Sciences* (2nd ed.). New York, NY: Routledge.

Booth, L. (1984). *Algebra: Children's strategies and errors, a report of the strategies and errors in the secondary school project.* NFER-Nelson, London.

Borko, H., & Putnam, R. (1996). Learning to teach. In Borko, H. & Putnam, R. (Ed.), Handbook of educational psychology (pp. 673–708). New York, NY: Macmillan.

Boyer, C. B. (1991) *A History of Mathematics* (2nd ed.), Hoboken, NJ: John Wiley & Sons, Inc.

Britt, M. & Irwin, K. (2007). Algebraic thinking with and without algebraic representation: a three-year longitudinal study. *ZDM Mathematics Education, 40,* 39-53.

Brown, L. & Coles, A. (1999). Needing to use algebra – a case study. In O. Zaslavsky (ed.) *Proceedings of the 23rd annual conference of the International Group for the Psychology of Mathematics Education,* Technicon:Haifa, 2, 153-160.

Callingham, R. & Bond, T. (2006). Research in Mathematics Education and Rasch Measurement. *Mathematics Education Research Journal, 18(2),* 1-10.

Carpenter, T. & Levi, L. (2000). *Developing conceptions of algebraic reasoning in the primary grades. Research report,* Madison, WI: National Center for Improving Student Learning and Achievement in Mathematics and Science. www.wcer.wisc.edu/ncisla/publications

Craig, T. and 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 (Ed..), *Proceedings of the 19th Annual Congress of the Association for Mathematics Education, 1,* 59-69. Cape Town: AMESA.

DBE (1996). *South African Schools Act (Act No. 84 of 1996).* Pretoria: DBE.

DBE (2011). *Curriculum and assessment policy statement (CAPS): Mathematics Grades 7-9.* Pretoria: DBE.

DBE (2014). *Report on the annual national assessment of 2014 grades 1 to 6 & 9.* Pretoria: DBE.

Derbyshire, J. (2006) *Unknown quantity: A real and imaginary history of algebra*. Penguin: New York.

Dettori, G., Garutti, R., & Lemut, E. (2001). From arithmetic to algebraic thinking by using a spread sheet, in R. Sutherland, T. Rojano, A. Bell and R. Lins (Ed.). *Perspectives on School Algebra,* p. 191-208, Kluwer: Dordrecht.

Downing, S. M. (2006). Twelve Steps for Effective Test Development, in *Handbook of Test Development* ed. Steven M. Downing and Thomas M. Haladyna. Abingdon: Routledge.

Dunne, T., Long, C., Craig, T. & Venter, E. (2012). Meeting the requirements of both classroom-based and systemic assessment of mathematics proficiency. *Pythagoras, 33(3),* Art #19, 16 pages. http://dx.doi.org/10.4102/pythagoras.v33i3.19

Eder, D. J. (2004). General education assessment within the disciplines. *The Journal of General Education, 53(2),* 135-157.

Edwards, A. and Alcock, L. (2010). Using Rasch analysis to identify uncharacteristic responses to undergraduate assessments. *Teaching Mathematics and Its Applications.* doi:10.1093/teamat/hrq008.

Filloy, E. and Rojano, T., (1989). Solving equations: The transition from arithmetic to algebra. *For the learning of mathematics*, *9*(2),19-25.

Filloy, E. and Sutherland, R. (1996). Designing curricula for teaching and learning algebra, in A. Bishop, K. Clements, C. Keitel, J. Kilpatrick and C. Laborde (Ed..) *International Handbook of Mathematics Education* Part 1, Chapter 4, Kluwer, Fordrecht, 139-160.

Fischer, D. H. (1970). *Historians' fallacies: toward a logic of historical thought.* (1st ed.). Harper torchbooks: New York, Harper Collins.

Fleisch, B., 2008. Primary education in crisis: Why South African schoolchildren underachieve in reading and mathematics. Pretoria: Juta and Company Ltd.

Graves, F.P., (1914). A history of education (Vol. 3). NY: Macmillan.

Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory.* Newbury Park, CA: Sage Press.

Hart, K. (1981). *Children's Understanding of Mathematics: 11-16.* London: John Murray.

Herscovics, N. & Linchevski, L. (1994). A cognitive gap between arithmetic and algebra. *Educational studies in mathematics, 27,* 59-78.

Howie, S. (2003). Language and other background factors affecting secondary pupils' performance in Mathematics in South Africa. *African Journal of Research in Mathematics, Science and Technology Education, 7(1)*, 1-20.

Howie, S. (2004). A national assessment in mathematics within an international comparative assessment. *Perspectives in Education, 22(2),* 149-162.

Howse, J., Molina, F., Taylor, J., Kent, S. and Gil, J.Y., 2001. Spider diagrams: A diagrammatic reasoning system. *Journal of Visual Languages & Computing*, *12*(3), 299-324.

Item response theory. (2017). Retrieved March 4, 2017, from https://en.wikipedia.org/wiki/Item_response_theory.

Jacobs, M., Mhakure, D., Fray, R. L., Holtman, L. & Julie, C. (2014). Item difficulty analysis of a high-stakes mathematics examination using Rasch analysis. *Pythagoras, 35(1),* 7 pages. doi: 10.4102/pythagoras.v35i1.220

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational studies in mathematics, 12,* 317-326.
Kieran, C. (1983). Relationships between novices' views of algebraic letters and their use of symmetric and asymmetric equation-solving procedures. In J. Bergeron and B. Herscovics (Ed..) *Proceedings of the International Group for the Psychology of Mathematics Education* pp.1.161-1-168, Montreal: Universite de Montreal.
Kieran, C. (1989). The early learning of algebra: a structural perspective. In S. Wagner and C. Kieran (Ed..) *Research issues in the learning and teaching of algebra.* pp.33-56, Reston, VA, NCTM.
Kieran, C. (1992). The learning and teaching of algebra, in D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning,* Macmillan, New York, p 390-419.
Kieran, C. (2007). Learning and teaching algebra at the middle school through college levels: Building meaning for symbols and their manipulation. In Lester, F. K. Jr (Ed.) *Second Handbook of Research on Mathematics Teaching and Learning: a Project of the NCTM*. Information Age Publishing. Charlotte, 707-714.
Kilpatrick, J. & Izsák, A., 2008. A history of algebra in the school curriculum. *Algebra and algebraic thinking in school mathematics*, *70*, 3-18.
Kilpatrick, J., Swafford, J., & Findell, B. (Ed..). (2001). *Adding it up: Helping children learn mathematics*. Washington DC: National Academy Press.
Linacre, J. M. (2016a). Winsteps® Rasch measurement computer program (Version 3.92.0). Beaverton, Oregon: Winsteps.com
Linacre, J. M. (2016b). Winsteps® Tutorial. Beaverton, Oregon: Winsteps.com
Lins, R. L. (1990). A framework of understanding what algebraic thinking is. In G. Booker, P. Cobb and T. Mendicuti (Ed..) *Proceedings of the 14th Conference of the International Group for the Psychology of Mathematics Education, 14(2),* 93-101. Oaxtepec: Mexico.
Mason, J., Graham, A. & Johnson-Wilder, S. (2005): Developing thinking in algebra. Paul Chapman Publishing: London.
Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174**.**
Mullis, I.V.S., Martin, M.O., Foy, P., & Arora, A. (2012). *Trends in International Mathematics and Science Study.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
National Research Council, 2001. *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academic Press.
Nopiah, Z. M., Osman, M. H., Razali, N., Ariff, F. H. M., & Asshaari, I. (2010). How good was the test set up? From Rasch Analysis Perspective. *Regional Conference on Engineering Education & Research in Higher Education 2010,* 1-8.
Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3(1),* 1-18.
Reckase, M., D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement, 21(1),* 25-36.
Simpson, J. & Weiner, E. (2009). *Oxford English Dictionary*. Oxford: Oxford University Press.

Piaget, J. (1969). *The mechanisms of perception.* London: RKP.
Pournara, C., Hodgen J., Adler, J. & Pillay, V. (2015). Can improving teachers' knowledge of mathematics lead to gains in learners' attainment in Mathematics? *South African Journal of Education*, 35(3), doi: 10.15700/saje.v35n3a1083.
Pournara, C., Hodgen J., Sanders, Y., & Adler, J. (2016). Learners' errors in secondary algebra: Insights from tracking a cohort from Grade 9 to Grade 11 on a diagnostic algebra test. *Pythagoras, 37*(1), a334. http://dx.doi.org/10.4102/pythagoras.v37i1.334
Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*.(Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.
Rasch, G. (1961). On general laws and the meaning of measurement in psychology, pp. 321–334 in Proceedings *of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, IV. Berkeley: University of Chicago Press, 1980.
Reed, S. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3*, 382-407.
Sfard, A. & Linchevski, L. (1994). The Gains and the Pitfalls of Reification: The Case of Algebra. *Educational Studies in Mathematics*, *26*(2/3), 191-228.
Spaull, N. (2013) *South Africa's education crisis: The quality of education in South Africa 1994 – 2011*. Johannesburg: Centre for Development and Enterprise.
Simkins, C. & Paterson, A. (2005*). Learner performance in South Africa: Social and economic determinants of success in language and mathematics.* Cape Town: Human Sciences Research Council.
Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680.
Stinson, D. W. (2004). Mathematics as "gate-keeper" (?): Three theoretical perspectives that aim toward empowering all children with a key to the gate. *The Mathematics Educator, 14*(1), 8–18.
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, Science and Technology Education, 19*(1)*, http://dx.doi.org/10.1080/10288457.2015.1012909*.
Sweller, J. & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and instruction*, *12*(3), 185-233.
Thorndike, E. L. (1922). The fundamentals of arithmetic. New York, NY: The MacMillan Company.
Usiskin, Z., (1995). Why is algebra important to learn. *American Educator*, *19*(1), 30-37.
Venkat, H. & Spaull, N. (2015). What do we know about primary teachers' mathematical content knowledge in South Africa? An analysis of SACMEQ 2007. *International Journal of Educational Development, 41,* 121-130.
Vergnaud, G. (1998). A comprehensive theory of representation for mathematics. *Journal of Mathematical Behavior, 17*(2), 167-181.
Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics*, *49*(3), 341-359.
Watson, A. (2009). Paper 6: Algebraic reasoning. Key understandings in mathematics learning. In T. Nunes, P. Bryant and A. Watson, University of Oxford: 1-42. Nuffield Foundation: London.

Wertheimer, M. (1960). *Productive thinking.* New York, NY: Harper.
Wong, M. (1997) Numbers versus letters in algebraic manipulation: which is more difficult? In. E. Pehkonen (ed.) *Proceedings of the annual conference of the International Group for the Psychology of Mathematics Education, 4,* 285-290, Lahti: University of Helsinki.

FINIS