An investigation of the 2012 Annual National Assessment Grade 6 mathematics instrument

by

Charlotte Madumelani Modzuka

Submitted in partial fulfilment of the requirements for the degree

Magister Educationis

in the Faculty of Education

University of Pretoria

Supervisor: Prof. M.C. Long

Co-supervisor: Prof. S.J. Howie

15th February 2017

DECLARATION

I declare that the thesis, which I hereby submit for the degree Master of Philosophy at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

________________________________ ________________________________

Charlotte Madumelani Modzuka Date

ETHICS CLEARANCE CERTIFICATE

The author, whose name appears on the title page of this thesis, has obtained, for the research described in this work, the applicable research ethics approval. The author declares that she has observed the ethical standards required in terms of the University of Pretoria's *Code of ethics for researchers and the Policy guidelines for responsible research.*

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Education

RESEARCH ETHICS COMMITTEE

CLEARANCE CERTIFICATE	CLEARANCE NUMBER: SM 13/08/03
DEGREE AND PROJECT	MEd An investigation of the 2012 Annual National Assessment Grade 6 mathematics instrument
INVESTIGATORS	Ms CM Modzuka
DEPARTMENT	Educational Studies Science, Mathematics and Technology Education
APPROVAL TO COMMENCE STUDY	9 October 2013
DATE OF CLEARANCE CERTIFICATE	13 February 2017

Please note:

For Master's application, Ethics Clearance is valid for 2 years.
For PhD application, Ethics Clearance is valid for 3 years.

CHAIRPERSON OF ETHICS COMMITTEE: Prof Liesel Ebersöhn

CC Ms Bronwynne Swarts
Prof C Long
Prof SJ Howie

This Ethics Clearance Certificate if issued subject to the following conditions:

1. A signed personal declaration of responsibility
2. If the research question changes significantly so as to alter the nature of the study, a new application of ethical clearance must be submitted
3. It remains the student's responsibility to ensure that all the necessary forms for informed consent are kept for future queries

Please quote the clearance number in all enquiries

ABSTRACT

The aim of this study was to investigate the quality of the Annual National Assessment (ANA) Grade 6 mathematics instrument including its design, with reference to a single education district. The main question that was investigated was: To what extent does the 2012 Annual National Assessment Grade 6 mathematics assessment instrument provide meaningful information for making appropriate interpretations on district level? The conceptual framework underpinning this study was drawn from the Queensland Studies Authorities Assessment Policy document. The research comprised a secondary analysis design applying mixed methods using the scripts of 546 learners in one district from 5 schools selected to represent a range of achievement. A content analysis of the instrument was undertaken, followed by a statistical item analysis applying the Rasch measurement model. These analytical methods were utilised to determine the quality of the ANA Grade 6 mathematics instrument. Content validity, construct validity and reliability was investigated in order to evaluate inferences that were made and actions that were taken based upon the mathematics performance of learners in Grade 6 in the Gauteng North District (GND) in the year 2012.

The investigation revealed that construct validity and content validity were largely achieved, as items were appropriately aligned to the 2012 ANA Grade 6 mathematics curriculum. However errors in mathematics and language formulation detracted from the validity of the instrument. In the case of some items, lack of clarity may have confused learners. As far as reliability is concerned the investigation revealed that the instrument had a reasonable person separation index, a measure of both item and person reliability. However, these conclusions are based on a relatively small sample from only one district and therefore has somewhat limited applicability but is nevertheless of educational consequence.

Keywords: assessment; Mathematics; test evaluation; cognitive levels; content analysis; Rasch analysis; inferences; statistical item analysis; validity; reliability.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ACARA	Australian Curriculum Assessment Reporting Authority
ANA	Annual National Assessment
CAPS	Curriculum Assessment Policy Statement
CTA	Common Task Assessment
DBE	Department of Basic Education
DoE	Department of Education
EFA	Education for All
FET	Further Education and Training
FFL	Foundations for Learning
FP	Foundation Phase
GG	Government Gazette
GDE	Gauteng Department of Education
GND	Gauteng North District
GET	General Education and Training Band
GP	Gauteng Province
IEA	International Association for the Evaluation of Educational Achievement
IIEP	International Institute for Educational Planning
IP	Intermediate Phase
LTSM	Learning and Teaching Support Material
LOLT	Language Of Learning and Teaching
MLA	Monitoring Learning Achievement
NAPLAN	National Assessment Program on Literacy And Numeracy
NQF	National Qualifications Framework
NCS	National Curriculum Statement

NEPA	National Education Policy Act
OBE	Outcomes-Based Education
OFSTED	Office of Standards in Education
RNCS	Revised National Curriculum Statement
SA	South Africa
SASA	South African Schools Act
SACMEQ	Southern and Eastern Africa Consortium for Monitoring of Educational Quality
SAOU	Suid Afrikaanse Onderwyseresunie
SAQA	South African Qualifications Authority
SBA	School Based Assessment
SD	Standard Deviation
SES	Socio-Economic Status
SGB	School Governing Body
SP	Senior Phase
SMT	School Management Team
TIMSS	Trends In Mathematics and Science Study
UCSMP	The University of Chicago School Mathematics Project
UNICEF	United Nations International Children’s Emergency Fund
UNESCO	United Nations Educational Scientific and Cultural Organisation
UP	University of Pretoria
USA	United States of America

ACKNOWLEDGEMENTS

I would like to convey my sincerest gratitude and appreciation to the following people for their support, contribution and guidance on my path to completion of this dissertation:

- I thank the Almighty God for giving me the strength, wisdom, guidance, unconditional love and protection.
- My supervisor, Prof. Caroline Long, and co-supervisor, Prof. Sarah Howie, for their expert guidance, motivation, support and tolerance.
- The Centre for Evaluation and Assessment (CEA) in the University of Pretoria for funding the data capturing process for my study.
- Cilla Dowse, coordinator, for her exceptional language editing and her professionalism in the early stages of my study.
- Mishack Tshele, computer manager, for assistance in data capturing, data cleaning and transferring data to RUMM 2020 program, a big thank you for the unconditional support during the whole process.
- Sister, Joy Sewanywa, for her significant input, assistance and good advice in preparing my research proposal for defence.
- Buffalo Aphane, Mathematics teacher, for his interest, professional input, support and understanding.
- Dr Masilo Machaba, Mathematics Subject Specialist during the process of data analysis and presentation of this study, for professional input, moral support and motivation.
- Maryna du Plooy, my friend, for her efforts and understanding, a big thank you for assistance in preparing my thesis for final submission.
- My dear son Nakedi and dear daughter Tsebo, for their love, inspiration and understanding.
- My family members, for their love, support and their continuous encouragement.
- Julius Untiedt, my supervisor at work, for the support, motivation and understanding.
- Rebone Ramatlo, Assessment Subject Specialist for Intermediate Phase during data collection of this study, for assistance, extraordinary input, understanding and support.

A SINCERE APPRECIATION AND THANKS TO YOU ALL!!!

I dedicate this dissertation to my two lovely children; son Nakedi and daughter Tsebo.

You will always have a special place in my heart, you are a blessing to me.

May the good Lord bless you abundantly in all facets of your lives.

I love you!

"Trust in the Lord with all your heart, and lean not unto your own understanding. In all your ways, acknowledge Him, and He will direct your path". Proverbs 3: 5-6

CHAPTER 1 – INTRODUCTION AND BACKGROUND TO THE STUDY

Introduction

The aim of this study was to investigate the quality of the 2012 Annual National Assessment Grade 6 mathematics instrument, the English version.

In investigating the quality of an assessment instrument it is important to define quality and what constitutes an effective, valid and reliable instrument. The criteria for an effective, valid and reliable instrument are discussed in Chapter 2; however, they are briefly stated here. Firstly, the instrument should be constructed by a team, which include both subject specialists, and people technically trained in quality testing and statistical methods (Matters, 2009). The items of such an instrument must have been thoroughly piloted ahead of the testing. Norms and standards of performance must accompany the test, enabling the interpretation of learner's scores in relation to other learner's scores (Matters, 2009). These criteria for effectiveness, validity and reliability serve as a guideline for investigating the quality of a test used for systemic purposes.

Preceding the Annual National Assessment, South Africa participated in international and regional assessments. At the time that South Africa participated in its first international study, the Third International Mathematics Science Study (TIMSS), in 1994-1995, the country was emerging from decades of international political and social isolation, accompanied by a legacy of inequity (Howie, 2012). There was an urgent need to establish the then current quality of the education system in the country, particularly given that there was no systemic overview of the education quality of the schooling system available in 1994 (Howie, 2012). Having an external assessment such as TIMSS 1995 provided unique opportunities for South Africa to gauge the performance of learners (Howie, 2012).

This TIMSS 1995 study was then followed by the international studies; TIMSS 1999, 2003, 2007, 2011 and 2015. Regional studies, including the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) II, III and IV also followed over these two decades (Howie, 2012). These studies provided South Africa with a set of educational quality indicators and benchmarks against which to review the performance of the system (Howie, 2012).

South Africa abstained in TIMSS 2007; some progress was indicated in TIMSS 2011 and 2015, however, the country's learners continued to display relatively low performances levels for mathematics and science subjects at the grade 8 and 9 level.

Although SACMEQ was not initially intended as a comparative study, comparisons across African countries were made and again South African learners performed relatively poorly as compared to some of their African peers (Howie, 2012).

As a result of these international test outcomes, the Department of Education introduced the Foundations for Learning (FFL) campaign in 2008. This campaign focused on equipping learners with basic mathematics, reading and writing skills at the Foundation and Intermediate Phases (DBE, 2005). One of the FFL campaign's main focii was to equip learners with the mastering of basic mathematical skills, with the hope that the FFL campaign would enable learners to improve learning in general.

The FFL intended to improve academic performance in both languages and mathematics. At that time the accepted means to measure improved academic performance formally was through the implementation of the systemic assessment. The Annual National Assessment was first implemented in 2011. The trial runs for this assessment had already been conducted in 2008 and 2009 at a few sampled schools. In 2011, however, it became compulsory for all public schools to administer the assessment with the purpose of informing decisions about improving learning in schools.

As will become clear in Chapter 2, this objective was not immediately reached, since the ANA Grade 6 mathematics administered in 2012 still indicated low performance at all levels of education, nationally, provincially, and at district level (DBE, 2013). This particular test, Grade 6 mathematics, was the focus of this investigation.

Two important aspects of the efforts of the DBE have to be put into perspective. On the one hand the FFL had been implemented from 2008 to improve learning in general. On the other hand the trials of ANA as a national systemic assessment had concurrently been implemented. The results of the 2012 ANA apparently decreased unexpectedly from 2011. This gave rise to the need to investigate factors that might have contributed to the outcomes of 2012 ANA not having improved. While these factors may hypothesized to be that the FFL was not achieving as expected and the other possibility is that the 2012 ANA instruments had increased in difficulty.

The study aimed to highlight only the test variables that might affect learner performance, however, in order to provide the context, I elaborate on a number of other factors affecting the assessment more broadly.

It is widely acknowledged in the literature that attainment in learning is influenced by internal and external factors. Some of the most important external factors are located in the Socio-Economic Status (SES) of learners, which affect learner performance (Reddy, 2006). In terms of external factors, evidence of the relationship between SES and mathematics performance was found in large-scale studies reports (Graven, 2014). These reports indicate a number of factors affecting learner performance, including: SES, gender, linguistic factors, teacher knowledge, teaching time and teacher absenteeism, poorly managed schools, poverty, including malnutrition and HIV/AIDS (Graven, 2014). Two correlating factors influencing mathematics attainment were found to be the language of learning being a language other than mother tongue, coinciding with poor Socio-Economic Status (Graven, 2014).

As far as the internal factors are concerned, some key variables identified that affect learner performance directly are discussed; for example, the teaching and learning of mathematics in the prior grades, access to resources and the management of assessment practices. The three above mentioned variables are considered important to enrich the understanding of systemic assessment generally although they are not the primary focus area in this study.

The first critical identified variable was teaching and learning of mathematics in prior grades. In South Africa, the Foundation Phase (FP) Grade 1-3, is the first phase in the primary school level, where learners are taught the subject mathematics. The Foundation Phase is considered to be the phase when the basic foundation of mathematics is laid. It is expected by the education system that a good foundation in numeracy should be established here. The skill that learners acquire in the Foundation Phase should support them to understand mathematics in the next phase, the Intermediate Phase (IP), progressing through Grades 4 and 5, and culminating in Grade 6 which is the grade of interest in this study.

Should it happen that the numeracy knowledge and skills are not developed to the level required at the preceding Foundation Phase; learners will have difficulties in understanding mathematics in the Intermediate Phase.

Learners with a good foundation in numeracy and particularly in the crucial transition stage experienced from Grades 3 to 4 should have a better chance of getting good marks in mathematics in the next phase, Grade 4-6. Generally, a sound education system, a good school, a proficient teacher and a well-disciplined learner all contribute to a competent foundation in all the subjects at school (Biesta, 2009). Provided that the standard is maintained, it is anticipated that in Grades 4-6, most learners are able to attain the adequate achievement level of 50-59% (see Table 3, page 27) because they should have mastered the basic mathematical knowledge in the lower grades (Foundation Phase).

The second variable, lack of access to resources, such as the Learning Teaching Support Material (LTSM), could be another factor contributing to low performance (DoE, 2000). It is almost impossible for proper teaching and learning to take place where learners do not have access to learning resources such as textbooks (DoE, 2000). Both the teachers and learners need to have access to resources to use regularly for classroom activities and to refer to when learning a concept. The availability of learner and teacher support material supports achievement in all spheres at school. The Department of Basic Education expects the schools to be guided by LTSM policy, to ensure that each learner is provided with the required basic resources per subject, and the teacher should have access to the teaching resources needed. Without Learning Teaching Support Material, teaching and learning will be affected drastically, leading to poor performance (DoE, 2005). Therefore, accessibility to teacher and learner LTSM resources should contribute to good attainment in the ANA.

The third variable, assessment, plays an important role in teaching and learning (Wiliam, 2009). At schools teaching and learning is required to take place every day. After teaching and learning has taken place, the teacher should assess the learners in order to monitor progress on what has been taught and what is expected to have been mastered in class. The School Based Assessment (SBA) plays a vital role; it provides the opportunity for feedback to the teacher on progress.

Effective feedback aligned with teaching and learning is important to the relevant stakeholders involved such as teachers, the district officials and parents; it provides a perspective on achievement and a diagnosis on what needs to be done to improve performance. According to Airasian (2005, p. 5), assessment is defined as “a continuous planned process of identifying, gathering and interpreting information about the performance of learners, this information is used to understand and assist the learner’s development and improve the process of learning and teaching”.

The assessment feedback is valid only when the assessment has been administered fairly (Black, 1998). Assessment can be regarded as not being fair, when learners are assessed on the content that has not been taught in class. It is expected that the teacher teaches the required content or scope prior to assessment (Wiliam, 2009). On completion of the required scope, the teacher can then plan the assessment, and learners should be informed about the content and the format of the assessment. If the assessment requires a rubric, then the marking criteria of the rubric should be discussed with the learners first, so that they should be aware of what is expected of them (Airasian, 2005). Assessment is used for different purposes, which include “assessing learners’ knowledge, assessing learners’ understanding and application of content knowledge, obtaining instructional feedback, and grading and monitoring growth in mathematical achievement” (Webb & Coxford, 1993).

In summary, the above three internal variables might contribute to the low performance of learners, although the primary focus in this study is neither on the external nor on the internal variables affecting performance, but on the quality of the instrument used to assess performance.

The reason for a focus on the instrument used to assess the performance of learners in the subject mathematics at Grade 6 is because assessment can never be separated from teaching and learning (Black, 1998). Assessment in mathematics can be defined as the comprehensive accounting of learners’ knowledge (Webb & Coxford, 1993). Schoenfeld (2007, p. 74) stated that “knowledge plays an important role, but an individual’s ability to employ problem-solving strategies, the individual's ability to make good use of what they know, and their beliefs and dispositions; are also critically important". These aspects of mathematical proficiency explained by Schoenfeld are further elaborated in Chapter 2.

If teachers are capable of producing learners with good mathematical proficiency, all aspects of mathematical knowledge should be assessed on a continuous basis so as to reinforce the knowledge content taught and learned. Furthermore, multiple assessment types should be used to produce balanced evidence of proficiency. It is therefore expected that learners' performance will be assessed in a more valid and reliable way if the ANA instrument is designed and implemented in addition to continuous School Based Assessment (SBA).

A phenomenon that occurs in the South African education context is that a discrepancy exists between the outcomes of SBA and the ANA. It may be possible that learners in general, perform well during the School Based Assessment which can either be formative or summative, for example, tests, projects, examinations, and assignments, but they perform poorly in the systemic assessment. This phenomenon is an indication that learners do show some understanding of the content knowledge. The fact that the ANA results do not reflect the same level of understanding is a problem to be investigated.

In the ANAs, the learners' performance was very low; this outcome contradicts the reported performance in SBAs'. The Suid-Afrikaanse Onderwyserunie (SAOU) raised questions on the big differences between the scores of School Based Assessments, the Annual National Assessments and matric results (Pretoria News, 17 October, 2015). According to the SAOU, learners performed much better in School Based Assessments and matric results than in the ANAs. A possible reason for such discrepancies it could be the fact that the total mark obtained in the ANAs hides the fact that there are some areas of strength. The discrepancy between SBA and the ANA results is critical, but it is not within the scope of this study. The question in focus here is therefore: What characteristics of the assessment might influence the low performance in the ANA?

The observed low performance, inferring low proficiency in mathematics, is a significant problem that needs to be investigated; a necessary step is to assess how the quality of the instrument itself may contribute to low performance. Learners' ability to understand the questions and answer appropriately may be another variable contributing to the low performance.

Therefore, the aim of this study was to investigate the instrument itself, by analysing the content validity, the construct validity and the reliability of the 2012 ANA Grade 6 mathematics assessment instrument.

The findings of this investigation respond to the main research question "To what extent does the 2012 ANA Grade 6 mathematics assessment instrument provide meaningful information for making appropriate interpretations on district level?" Many factors impact on performance, as mentioned previously. The approach taken in this study was to make inferences based on test results, in the light of both the assessment environment and other learner factors such as curriculum coverage.

The findings will also inform stakeholders about the limitations and strengths of the ANA instrument and the validity and reliability thereof. The Department of Basic Education could consider the findings for future use in the Annual National Assessment.

The statement of purpose is outlined in this section. The South African education context is discussed. The systemic assessment and the Annual National Assessment in South Africa are then explained. The problem statement is expanded, followed by the rationale of the study. The main research question guiding the study and the sub-questions are presented. The chapter closes with an outline of the dissertation.

1.1 STATEMENT OF PURPOSE

As already discussed in the previous section, South Africa has participated in systemic assessments at international and regional level. The Annual National Assessment results of the year under investigation 2012 confirm the patterns and levels of learner performance that were observed in the international and regional assessments. In this regard, Howie (2012, p. 88) reports as follows:

> "Difficulties in mathematics in SA schools were pointed out in the Trends in International Mathematics and Science Study (TIMSS) results and in the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) results".

The ANA results of Grade 6 therefore confirm that in the subject mathematics, learners are struggling to perform at the expected level both as stipulated in the national curriculum and beyond.

In 2008 and 2009, trial runs of the Annual National Assessment "tests were distributed and conducted across the country with most schools participating in the programme" (DBE, 2011, p. 17). In the first two years, ANA was administered in a few sampled schools per district in key subjects such as English and mathematics in Grades 3, 5 and 9. However, the "decision was taken for a major implementation of the ANA at the beginning of 2011", when participation was compulsory for all public schools, independent schools and private schools receiving subsidies (DBE, 2011, p. 18). The ANAs were administered for literacy and numeracy in the Foundation Phase (Grades 1-3), in the Intermediate Phase (Grades 4-6) and the Senior Phase (Grade 9) for languages and mathematics.

The results of the ANA in the year 2012 showed that learners generally performed "poorly". There may be different reasons for this low performance in the Annual National Assessment as mentioned in the previous section.

From an international perspective, Matters (2009) identified two major variables that contribute to learners' achievement in systematic tests that are conducted in many countries. She argued that "achievement is influenced by factors internal to learners as well as those imposed by features of the assessment environment" (Matters, 2009, p. 220). Factors that are associated with the learner, according to Matters (2009, p. 221), are "motivation, test anxiety, academic self-concept and attributive style". The factors related to the "assessment environment include the assessment instrument itself, the preparation" involved with the assessment process and the conditions under which the assessment is applied (Matters, 2009, p. 220).

An important question arises as to what the main factor is considered to have contributed to the low performance in the ANA. This question prompted this enquiry into the quality of the ANA instrument by analysing the validity and reliability of the design. In agreement with Matters (2009), the assessment instrument itself needs to be thoroughly investigated in order to assess this fact in the assessment programme. Locally the studying of the national assessment instrument has not been given enough attention, hence the need for this study.

1.2 SOUTH AFRICAN CURRICULUM REFORM AND LARGE SCALE ASSESSMENT

There is a need to look at the South African education system to establish the relevance in terms of the implementation of curriculum and assessment policies.The White Paper on Education and Training of Basic Education (1995) emphasised "the need for major changes in education and training in South Africa in order to normalise and transform teaching and learning, stressing the need for a shift from the traditional aims and objectives approach to outcome-based education" (DoE, 2000, p. 15). "A vision of a prosperous, truly united, democratic and internationally competitive country with literate, creative and critical citizens leading productive, self-fulfilled lives in a country free of violence, discrimination and prejudice was promoted" (DoE, 2000, p. 3).

Large-scale testing in South Africa can be traced back to the leadership of Nelson Mandela in 1994 (Howie, 2012). Attempts to improve the functioning of the education system in South Africa have come into greater focus since 1994 with the move to a democratic South Africa and one unified education system. The key focus for reform within the general and further education band was to ensure equity in education. This reform was introduced with Curriculum 2005, the curriculum implemented in 1998.

With the new dispensation came the outcomes-based education approach (Jansen, 1997) and a new curriculum aiming to provide all children in South Africa with equitable, quality access to education as mentioned in the previous section. In "order to monitor changes occurring in the system formally, a system of systemic evaluations (internationally known as national assessments) was introduced in 2000 for Grades 3, 6 and 9 for large-scale testing (sample based) in mathematics, science and language" (Howie, 2012, p. 87).

According to the National Education Policy Act 27 of 1996, "monitoring and evaluation should be done with the objective of assessment progress (at system level) that corresponds with the provisions of the Constitution of the Republic of South Africa and the National Education Policy" (Howie, 2012, p. 86). In addition to the Act mentioned above, provision for conducting systemic evaluation "on a nationally representative sample of learners and learning sites was made in section 48 of the Assessment Policy for General Education and Training Band" (DoE, 2001, p. 56). This implementation of systemic assessment is done in "order to evaluate all aspects of the school system and learning programmes contained therein; that is, to assess the effectiveness of the entire system and assess the extent to which the vision and goals of the transformation process are being realised" (Howie, 2012, p. 87).

Thus, transformation of the education system started with the changes to the intended curriculum (the formal curriculum policy documents that reflect what society wants the children to learn). This change impacted on the implemented curriculum (what is being implemented in the schools and classrooms) and is clearly having an impact on the attained curriculum (what has been learned), how children are performing on the assessments based upon the curriculum as intended and implemented (van den Akker 2003).

The effectiveness of the implementation of the attained curriculum and the implementation of a systemic testing system for the monitoring of the system has been the focus, based on the grounds that systemic type assessment will support a positive change in the system. As mentioned in the previous section, prior to the introduction of the systemic evaluation, a number of international comparative studies had taken place, also inspired by the changes taking place in the system, but also because of the international recognition of the "political changes taking place in South Africa and the reintegration of South Africa into the international community" (Howie, 2012, p. 86).

As decreed by the DBE, "systemic tests at various levels have been put in place in the South African education system to track progress towards achieving the adequate levels of literacy and numeracy in schools (see Table 1) on curriculum reform and systemic assessment" (Mason, 2005, p. 56). The learner performance is used as a yardstick to measure the standard of the education system (Maree & Fraser, 2008). There is a "demand for schools and teachers to use assessment information to improve learner achievement and for education departments to enhance educational systems" (Matters, 2009, p. 221). Table 1, adapted from Mason (2005), illustrates the Curriculum Reform and Systemic Assessment Timelines in South Africa.

Table 1

Curriculum Reform and Systemic Assessment Timelines in South Africa (Adapted from Mason, 2005)

DATE	REFORM	SYSTEMIC EVALUATION
1995	**National Education and Training Forum** prepares the way for the development of a core interim syllabus.	TIMSS 1995
1995	**South African Qualifications Authority** (SAQA) enacted and established the **National Qualifications Framework** (NQF).	
1996	**Constitution of the Republic of South Africa** outlines a new vision for South Africa and directly influences all legislation and policies. **The South African Schools Act -** establishes a national education system; makes schooling compulsory for children aged 7 – 14; School Governing Bodies (SGBs) established.	
1996/7	Process of designing Curriculum 2005 begins.	
1997	**Curriculum 2005**, including **Outcome Based Education** (**OBE)**, becomes national policy.	
1998	Implementation of Curriculum 2005 begins. Problems soon become apparent with the design of the curriculum and the implementation process.	
1999		TIMSS 1999 and MLA 1999.
2000	Ministry of Education commissions a	Systemic Evaluation of Grades 3, 6 and 9.

DATE	REFORM	SYSTEMIC EVALUATION
	review of Curriculum 2005.	SACMEQ II and Systemic Evaluation Grade 3
2001		.
2002	The review committee publishes its findings in a report South African Curriculum for the Twenty- First Century – review committee on Curriculum 2005. The report identifies both strengths and flaws in C2005 and makes specific recommendations. C2005 is revised and appears as the **Revised National Statement Curriculum Statement Grades R – 9 (Schools).** The RNCS becomes policy.	
2003		Systemic Evaluation of Grades 3, 6 and 9 TIMSS 2003.
2004	Implementation of the RNCS begins (Foundation Phase).	Systemic Evaluation of Grade 6
2007		SACMEQ III TIMSS 2007 – South Africa decides not to participate
2008		ANA trial of Grades 3, 6 and 9
2009		ANA trial of Grades 3, 6 and 9
2011	Process of amendment of the NCS to CAPS begins.	TIMSS 2011 and ANA full scale
2012	Implementation of CAPS in Foundation Phase begins.	ANA of Grades 1-3, 4 -6 and 7 -9
2013	Implementation of CAPS in Intermediate Phase begins.	ANA of Grades 1-3, 4 -6 and 7 -9
2013		SACMEQ IV
2014	Implementation of CAPS in Senior Phase and Further Education and Training begins.	ANA of Grades 1-3, 4 -6 and 7 -9
2015		Systemic Evaluation TIMSS 2015. South Africa participated.

Changes from the NCS to CAPS with regard to assessment in the General Education and Training band included the replacement of the Common Task Assessments (CTA) in Grades 3, 6 and 9 by the Annual National Assessment focusing on finding evidence of learner achievement (Du Plessis, 2013).

The focus of this study is on the national assessment; therefore it is of importance to highlight the comparison with TIMSS, SACMEQ and ANA as systemic assessments. TIMSS is conducted internationally in countries around the world, SACMEQ is conducted regionally in the Southern Eastern countries of Africa and the ANA is conducted nationally in South Africa.

The similarities identified of these systemic assessments is that mathematics and science subjects are assessed with the common goal of generating information that can be used to improve the quality of the education system and to make informed decisions about how to improve teaching and learning in mathematics and science (SACMEQ , 1995). The structure of subject content used in the TIMSS, SACMEQ, and the ANAs for the subjects; language and mathematics indicate similar components.

In brief, the purpose of TIMSS is to measure trends in mathematics and science achievement, as well as monitoring curricular implementation and identifying promising instructional practices around the world (Olson, 2007).

The SACMEQ's main mission has been to undertake integrated research and training activities that will expand opportunities for educational planners and researchers in ministries of education in order to gain the technical skills required to monitor and evaluate the general conditions of schooling and quality of their own basic education system, and undertake research that generates evidence-based information which can be used by decision-makers to plan improvements in the quality of education (SACMEQ 1995-2013).

An important aim in the Annual National Assessment was to contribute towards improved learning in schools, and this would be possible if the assessment could serve as a systemic tool and also as an instrument to diagnose and identify areas of strength and weakness in teaching and learning (DBE, 2013).

1.3 OVERVIEW OF SYSTEMIC ASSESSMENT

Systemic assessment is a "quantifiable measure used to describe and monitor the education process" (Kellaghan & Greaney, 2001, p. 52). It is understood that a systemic assessment is intended to estimate and explain learners' attainment of the curriculum (Greaney & Kellaghan, 2008). This estimate in turn is used to guide curriculum reform and improved curriculum implementation (Howie, 2012). A detailed description of the purposes of systemic assessment is provided in Chapter 2.

If systemic assessment is implemented in a valid and reliable way, the effectiveness of any aspect of the education system can be determined externally (Van Rooyen & Prinsloo, 2003). In this way learners' performance can be compared with the national indicators of learner performance (Du Plessis, 2007).

The primary purpose of Systemic Evaluation is to address problems in the external environment that may have an influence on learner performance like resources and facilities, leadership, management and communication, governance and relationships, curriculum provision and teacher characteristics (DoE, 2003). In the South African context systemic assessment had the additional purpose of gauging whether the education transformation process has been achieved (DoE, 2003).

According to Greaney, Kellaghan and Murray (2009) the main purpose of national assessments is to determine:

a) How well learners are learning in the education system (with reference to general expectations, aims of the curriculum, and preparation for further learning and for life);
b) Whether there is evidence of particular strengths and weaknesses in learners; knowledge and skills;
c) Whether particular subgroups in the population perform poorly;
d) Which factors are associated with learner achievement;
e) Whether government standards are being met in the provision of resources;
f) Whether the achievements of learners change over time (Greaney, Kellaghan & Murray, 2009, p. 1).

In summary, learner performance is used as one of the key indicators of education performance, alongside the "contextual factors that impact on teaching and learning" (DoE, 2001, p. 18).

1.3.1 International and Regional Assessments

This section focuses on the international and regional studies that have been conducted in South Africa on mathematics.

There are two advantages of taking part in these international assessment studies. The first advantage is a practical one, if a country does not already have a national assessment programme, important development costs can be saved by making use of internationally available instruments (Kellaghan, Greaney & Murray, 2009). The second potential advantage is the opportunity to compare national performance levels to international standards (Scheerens et al., 2003).

South Africa also took part in regional studies and projects at different grade levels. The "Monitoring Learning Achievement (MLA) project began internationally in 1992 and was a joint project of the United Nations Education Scientific and Cultural Organisation (UNESCO) and United Nations Children's Education Fund (UNICEF)" (Graven, 2014, p. 4). The aim was to examine the effectiveness of the basic education provision in terms of actual learning attainment (Chinapah, 1997). The project formed part of the Education for All (EFA) initiative. Learning attainment of Grade 4 learners was "measured by means of achievement scores in literacy, numeracy and life skills, and a criterion-referenced approach was used" (Tshabalala, 2008, p. 22). The results showed that "learners in South Africa performed poorly in numeracy" (Howie, 2002, p. 30).

The Department of Education in South Africa conducted a Systemic Evaluation survey at the Intermediate Phase (Grade 6) level in 2004. The purpose of the survey was to establish baseline data on learner achievement and conditions of "teaching and learning at the Intermediate Phase (Grade 6) level" (DoE, 2005, p. 41). The survey involved approximately 35 000 learners from about 1000 schools (about 7% of the mainstream public schools with Grade 6) that were selected randomly from all the nine provinces for this purpose (DoE, 2005).

The respondents in the survey were the "Grade 6 learners, parents / guardians, teachers of the affected subjects, principals and district officials, who all completed specially-designed questionnaires on the contextual factors that may affect teaching and learning in the school system" (DoE, 2005, p. 50).

The main findings showed that in general, learners obtained the lowest score for numeracy (national mean 30%) relative to other subjects of both test difficulty and learner proficiency. The distribution was skewed heavily towards the lower scores, with a high concentration of learners obtaining scores between zero and 40%. These results confirm those of international and regional studies that were conducted in South Africa (Tshabalala, 2008). A trend of low performance from some of the primary schools and secondary schools has been indicated by various studies; for example, the international comparative study, the Trends in International Mathematics and Science Study (TIMSS), and a regional study, the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ).

TIMSS 1995 "provided South Africa with the first national representative overview of how South African learners were performing in mathematics" (Howie, 1997, p. 23). Learners in "Grades 8 and 12 showed lack of understanding of mathematics questions, and inability to communicate their answers in instances where they did understand the questions" (Howie, 2012, p. 90). This group of "learners performed particularly badly in questions requiring a written answer" (Howie, 2012, p. 90). South Africa participated in 1999 and 2003 with similar poor results, with the 2003 result indicating a decline in performance from an already low base in 1995. The progression towards participation in the TIMSS in 2011 has already been discussed earlier.

SA "participated in SACMEQ II in 2000 SACMEQ III in 2007 and in SACMEQ IV in 2013; an improvement was noted in overall performance of South African learners in the third SACMEQ and the fourth SACMEQ studies" (SACMEQ, 2011, p. 1). The results indicated "that, for South Africa as a whole, the mean score for reading increased by 3 points, from 492 points in 2000 to 495 in 2007" (SACMEQ, 2011, p. 2), meaning that the performance has improved, however, not significantly. Also, "for mathematics, there was a minimal increase of 9 points in the national mean score, that is, from 486 points in 2000 to 495 points in 2007, compared to a set SACMEQ mean score of 500 for reading and mathematics" (SACMEQ, 2011, p. 2). In "both reading and mathematics, the percentage of learners achieving at the higher SACMEQ levels was significantly low in relation to other countries, learners were still underperforming in both reading and mathematics" (SACMEQ, 2011, p. 2).

The "SACMEQ II (and SACMEQ III) tests were constructed carefully so as to ensure that the structure of the learner tests was congruent with the content and skills derived from detailed analyses of the curricula, syllabi, examinations and textbooks used in the SACMEQ countries" (SACMEQ, 2011, p. 3). The SACMEQ IV results revealed that learners in South Africa for the first time achieved a mean score above the centre point with 558 in reading and 587 in mathematics (SACMEQ, 2010 - 2014). Across all SACMEQ countries, there was a notable improvement in the overall reading and mathematics scores of learners.

In local projects, regional studies and Systemic Evaluation conducted in the country at different levels and in different subjects, learner performance was low, although some improvements have been recorded, and efforts are required to understand this phenomenon. SACMEQ tests were designed to measure trend and can be compared from study to study.

1.3.2 The Annual National Assessment

An important aim of in the ANA was to contribute towards improved teaching and learning in schools. The six objectives of the ANA have been identified in the Annual National Assessment Report (DBE, 2011) as follows:

1. ANA serves as a diagnostic tool;
2. ANA should encourage learners to perform at an acceptable level / adequate achievement (Level 4, 50 – 59%) in languages and mathematics;
3. Teachers are to assess learners, using appropriate standards and methods;

4. There is to be better targeting to support schools by the districts;
5. To enable the celebration of success in schools;
6. To encourages parent involvement (DBE, 2011, p. 6).

The above Annual National Assessment objectives are considered worthwhile for gauging the health of the system. The question remains, whether these purposes are achievable within districts, such as the area for investigation in this study, for example the Gauteng North District.

The ANA has achieved to some extent its purpose in terms of the first objective. It serves as a diagnostic tool. A detailed analysis was conducted to further explore whether these objectives have been met. After the administration of the ANA, the relevant subject adviser as well as the mathematics teachers meet and analyse each question, explore the content subtopic and item format that might influence the result, and together come up with strategies on how to enable improvement in the teaching and learning of mathematics. This process takes place at cluster meetings in different areas (refer to APPENDIX J) for the ANA question analysis template. What remained to be done as the focus of this study was to provide an in-depth analysis of the assessment instrument.

The second key objective has not been met. ANA has not succeeded in getting learners in general to perform at an acceptable level, meaning the learners are expected to achieve Level 4 which is the (adequate achievement see Table 3, page 27) in mathematics in the country, generally, and neither in the Gauteng North District. In some instances gains may have been observed, but in general the performance has been low.

However as observed earlier, ANAs were not designed (in the initial stage) to measure trends and therefore gains cannot be claimed. This study explores characteristics of the Grade 6 tests that may improve their functioning and so provide more reliable information, and which may, given attention to associated factors, lead to better education.

The third objective has been met in part, that is an attempt to present a model of good assessment to teachers. It is the robustness of the test that supports this objective. However the problem is that until recently these assessments did not achieve this as several mistakes were identified in the content analysis of the ANAs (refer to Chapter 4, page 94). This discussion is taken further in the presentation of results on the extent to which the ANAs serve as good models of assessment (refer to Chapter 5, page 136). Details of these lacks in the 2012 Grade 6 instrument are the focus of this study.

With regard to the fourth objective, the DBE are moving towards achieving this objective. The districts categorise their schools based on the Annual National Assessment results. The schools that did not manage to get an average of 50% on the ANA results are categorised as the underperforming schools and receive continuous support from the district. This action is believed to improve teaching and learning. However, the proviso is that the type of support given enables the development of mathematics proficiency.

With regard to the fifth objective, the celebration of ANA success is rare; this is a longer - term objective for many schools as very few schools manage to achieve the adequate achievement. The focus district Gauteng North District is under a lot of pressure because of the continuous low performance by a number of schools. For the focus district to succeed the factors impacting the poor performance need to be identified hence the need for this study.

There is no evidence of the final sixth objective being met in the Gauteng North District.

In support of the objectives, the 2012 ANA assessments were developed by panels comprised teachers currently teaching the Grade and the subject advisors and curriculum specialists who were appointed as test developers (DBE, 2013). As stated by the DBE (2013, p. 44) the composition of the panels encompassed the following; two assessment developers, moderators and editors. The duties of the panel included to develop the assessment prototype in English, and a panel of 'translators' were responsible for translating the English assessment to the other 10 official languages in the case of the Foundation Phase (Grades 1-3) and into Afrikaans for the Intermediate and Senior Phases (DBE, 2013).

The Subject Specialists in home languages were used for the process of translation of the ANA tests. After the translation process was complete, the assessment instrument was then edited by the relevant editor for each official language (DBE, 2013, p. 45). Editors of the "different languages sat together to discuss the changes as a standardisation measure and to ensure that there was no compromise to the assessment frameworks" (DBE, 2013, p. 45). The process for the development of the ANAs remained the same over time.

This study only focused on the mathematics Grade 6 instrument and the data from Grade 6 learners in a Gauteng district, who wrote mathematics in English and not those, who wrote it in Afrikaans, as stated earlier.

1.4 PROBLEM STATEMENT

The demands for democratic change and transformation in education in South Africa are embraced in the Constitution, made statutory by the National Education Policy Act (NEPA), Act No. 27 of 1996. These pieces of legislation require that scientific research studies be undertaken to obtain baseline information on the performance of the education system towards meeting the set goals.

The complexity of the old system of education in South Africa which included the Bantu Education and Christian National Education during the years 1953 to 1994, had challenges of accountability, there was lack of transparency, and a lack of proper assessment practices. All the factors contributed to the inefficiencies in the learning system (DoE, 1998). With the unification of education departments, curriculum transformation occurred from the years 1995 to 2011 in the South African Education system.

Curriculum change "started immediately after the election in 1994, when the National Education and Training Forum began a process of syllabus revision and subject rationalisation. The purpose of this process was mainly to lay the foundations for a single national core syllabus (DoE, 2000). "For the first time, curriculum decisions were made in a participatory and representative manner, the emphasis being placed on lifelong learning as the major curriculum design for the now democratic South Africa" (DoE, 2005, p. 39).

The on-going National Curriculum Statement (NCS) implementation challenges, identified as problems with the curriculum, resulted in the most recent change to the curriculum with the introduction of the Curriculum Assessment Policy Statement (CAPS, DoE, 2011).

The implementation of the ANAs was an attempt to find objective evidence of learner achievement that would aid in improving learner competence in key subjects such as mathematics. However, the success of assessment programmes such as the Annual National Assessment depended on the excellence of the curriculum, the value of teaching and learning and the quality of the instrument if they were to inform teaching and learning. South Africa faces the challenge of providing sound mathematics education (DoE, 2005). Furthermore, in some cases learners do not necessarily lack the appropriate content knowledge, but lack mathematical skills, in other words, learners' lack of mathematical abilities would influence their current and future performance in mathematics (DoE, 2000).

A number of factors have been discussed earlier pertaining to the low performance of learners in mathematics. These include: 1) Teaching and learning in prior grades; 2)

The difficulty of mathematics; 3) Lack of Learning Teaching Support Material resources; and 4) The lack of coherence and correlation between the School Based Assessments and the Annual National Assessments.

The low results of the ANA, administered in the year 2011, led to the introduction of the Gauteng Primary Language and Mathematics Strategy 2012 (GPLMS) which aimed to increase the pass rate from the current average of 35% (mathematics) and 40% (languages) to at least 60% by 2014 (DBE, 2012). The DBE expected a drastic improvement in ANA results within one year. However, this good expectation was not realistic. The suggestion that the pass rate be increased is commendable; however, what is needed is rather the understanding of mathematics. A pass rate can be changed by making it an easier test (DoE, 2000).

The first year (2011) of implementation of the ANA did not produce good results therefore the quality of the instrument is a key variable to investigate. However, despite the introduction of these strategies by the DBE, which aimed at increasing the learner achievement in mathematics, the results were very low for ANA Grade 6 for the year 2012.

Table 2 illustrates the results of the Annual National Assessment of the year 2012. The ANA, conducted in 2012, is the focus year for this study, it is important to distinguish how learners performed in that particular year. The ANA averages at national level and the district of interest GND are similar but more than 3% lower than the Gauteng Province average. Gauteng Province performed better than the national average and the district of interest.

Table 2

Grade 6 Mathematics: National, Provincial and District Average (Adapted from the Department of Basic Education, 2013)

	2012
National Average	27,0%
Gauteng Province Average	30.9%
Gauteng North District	27.2%

The Gauteng Department of Education (GDE) consists of 15 district offices of which Gauteng North District (GND) is one. Gauteng North District is responsible for schools in Cullinan, Bronkhorstspruit, Donkerhoek, Sokhulumi, Hammanskraal, Delmas, Moloto, Leeuwfontein and Ekangala. It covers 33 public primary schools of which five are farm schools, 15 public secondary schools and 18 independent schools. This district is considered one of the poorest districts in terms of the residents' economic background.

Of particular interest in this study, are the results of the Grade 6 mathematics in Gauteng North District. Grade 6 is the focus grade in this study because it is the exit grade from the Intermediate Phase to the Senior Phase and is one of the three grades tested nationally. The Senior Phase prepares learners for the last phase in school education the Further Education and Training (FET). Gauteng North District was amongst the low performing districts in ANA in Gauteng Province.

The learners in this district are sampled for the current study. The Gauteng North District learner mean performance was at Level 1 (0 – 29%), meaning "not achieved" according to Curriculum Assessment Policy Statement (CAPS) (see Table 3). These assessment results provide some indication of poor performance; the learners in Gauteng

North District were unable to get the adequate achievement (pass average) which is Level 4. Again we asked the question, “What exactly does this result mean in terms of content and skills?” In order to answer this question we conducted an analysis of the instrument of this particular year, 2012.

According to the Department of Basic Education, the adequate achievement is defined as the average pass percentage in Curriculum Assessment Policy Statement, that is, a Level 4 on a 7- point scale. “The national codes and their descriptions provided in Table 3 are used for recording and reporting performance of learners in Grades R-12, a 7-point scale is used to assist teachers to grade learners at the correct level” (DoE, 2011, p. 17). The table also illustrates that the level 4 is considered adequate achievement whilst level 3 is moderate achievement.

Table 3

Rating Codes and Achievement Description in Grade 1-12 (Adapted from the National Protocol for Assessment Grade R-12, 2011)

RATING CODE	ACHIEVEMENT DESCRIPTION	MARKS %
7	Outstanding Achievement	80-100
6	Meritorious Achievement	70-79
5	Substantial Achievement	60-69
4	Adequate Achievement	50-59
3	Moderate Achievement	40-49
2	Elementary Achievement	30-39
1	Not achieved	0-29

Possible factors for not meeting the basic adequate achievement in mathematics are investigated in various studies; however, this study explores the extent to which the instrument used is appropriate for measuring learners’ performance at the attained level. Did the instrument in any way contribute to the poor results because of some inadequacies or is the instrument valid and reliable for its purpose of providing consistent, valid and reliable data for the district use? To date, no study has been undertaken to answer those questions regarding the assessment instrument.

In order to determine the extent of the validity and reliability of the ANA Grade 6 mathematics instrument, an analysis of the data gathered, from both the professional judgement perspective and the statistical analysis, can be used to establish the development of a theoretically based, empirically tested instrument to measure mathematical knowledge and skills for the Grade 6 learners.

The data gathered from the instrument can be used to provide meaningful information to refine the ANA Grade 6 mathematics instrument in terms of achieving the validity and reliability required to inform district-level interventions.

1.5 RATIONALE FOR THE STUDY

Worldwide, there is an increasing emphasis on the importance of mathematics and science. Globally, mathematics is regarded as an important subject because it opens doors for a wide choice in career opportunities. Mathematics is perceived as “the foundation for further study in a number of school subjects, most notably the sciences; and mathematics problem solving builds logical reasoning skills that can be applied in many situations” (Howie, 1997, p. 28).

Two studies, namely Third International Mathematics and Science Study (TIMSS) (1999, 2003, and 2011) and the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) (2004, and 2007) which compares the performance of South African learners with that of learners from other countries, indicate that South African learners perform poorly in mathematics and science, even in comparison with other developing countries. South Africa’s own Grade 6 Systemic Evaluation Report, which established a baseline at the Intermediate Phase level and released in December 2005, also reveals low performance about educational quality in the country.

In 2012, Grades 1-6 in the primary schools wrote the ANA. The aspects of the test performance of the Annual National Assessment Grade 6 mathematics assessment instrument written in the year 2012 will be investigated further. Investigation into the 2012 ANA instrument using the Rasch measurement model, will highlight the properties of the assessment instrument itself, reflect on the results within one district and offer some observations that may impact on the low results and how the results are used.

A rigorous scientific investigation of the ANA mathematics Grade 6 instrument has been undertaken. A sample of the Gauteng North District data has been interrogated using Rasch measurement theory. Performance on the ANA in mathematics of this particular district was investigated in detail, in order to inform the inferences that can be made as a result of the assessment instrument.

The Rasch measurement model was applied to analyse the mathematics instrument by verifying assessment validity and for aligning item difficulty and person proficiency on the same scale. This ordering on item difficulty and person proficiency will be of interest to the department officials interested in providing some feedback on the teaching and learning of mathematics at Grade 6.

The application for the Rasch model served to determine the extent of the validity and reliability of the ANA Grade 6 mathematics instrument. The data gathered from this study can be used to establish the development of a theoretically based empirically tested instrument to measure mathematical knowledge and skills for the Grade 6 learners.

Secondly, the data gathered from the instrument can be used to provide meaningful information to refine the ANA Grade 6 mathematics instrument in terms of achieving its validity and high reliability to inform district-level interventions.

These findings can be compared with the overall Annual National Assessment findings.

1.6 MAIN RESEARCH QUESTION

The main question that is investigated by this study is: "To what extent does the 2012 Annual National Assessment Grade 6 mathematics instrument provide meaningful information for making appropriate interpretations on district level?"

The main question is based on the investigation of the sub-questions which follow:

1. What is the purpose of the Annual National Assessment Grade 6 instrument?

This question seeks to explore what the purpose of the original implementation of the DBE was with regard to the Grade 6 mathematics instrument, what it intended to achieve;

2. To what extent does the instrument fulfil the criteria for good assessment design?

This question seeks to investigate the design of the ANA instrument by analysing the structure of a quality national assessment, and to evaluate whether the instrument meets the conditions of a good national assessment. To this end, the criteria applied for good assessment design were, representation of knowledge and skills, the level of development of learners, the language use, and assessment results as means of monitoring change (refer to Chapter 2, page 50), a presentation of a conceptual framework in Chapter 3;

3. To what extent does the instrument have construct validity and content validity?

To ascertain the extent to which the instrument provides valid and reliable information. Validity is investigated with regard to construct validity and content validity and reviews the scope assessed;

4. Are the items on the mathematics assessment instrument functioning as expected? Specifically do the items fit the requirements of the Rasch measurement model?

This question proposes to investigate the degree of correspondence between ideal item performance and actual item performance. The Rasch measurement model seeks to investigate to what extent the items on the instrument are functioning as expected in terms of assessing the difficulty level and to confirm whether the items fit the requirements of the model or do the items misfit.

5. How well are the items distributed along the continuum of the variable?

The Rasch model output aligns the learner proficiency and item difficulty on the same scale. The results are analysed through an interactive process including quantitative method verification, specifically for the district concerned for the purpose of checking assessment construct validity and reliability. Reliability ascertains the extent to which one might expect to render the same results if the instrument were re-administered on the same population and in the same conditions.

6. How well is the assessment instrument targeted to the abilities of the learners?

The Rasch measurement model investigates in more detail the problem conditions, evident in the functioning of the ranges within which firstly, the items difficulty and person's ability falls.

The purpose of this study was to investigate and understand the presumed conditions by conducting a secondary analysis design applying mixed methods. A content analysis undertaken (applying professional judgement of the teachers and subject advisers) of the Annual National Assessment is conducted, followed by a statistical item analysis applying the Rasch measurement model. These analytical methods are utilised to determine the quality of the ANA Grade 6 mathematics instrument, in terms of its construct validity, content validity and reliability in order to evaluate the potential inferences that could have been made and actions taken based upon the mathematics performance of learners in Grade 6 in Gauteng North District.

Secondary analysis design applying mixed methods of the ANA instrument and data is the research procedure in this study. Sampling for analysis was conducted as follows:

Five schools were selected according to a range of performance and comprising a selection from the highest achieving to the lowest achieving school from the Gauteng North District. As mentioned in the previous section, only English data was analysed.

Gauteng North District has a small number of schools, the schools are scattered because of the nature of the environment, a crucial factor being that the area is largely rural. A total of 546 learners' scripts were captured from the five schools. The mark allocation and the item type on the instrument are described in APPENDIX E.

1.7 THE STRUCTURE OF THE DISSERTATION

This study is divided into five chapters.

- Chapter 1 provides an introduction and the background to the study, and outlines the statement of purpose, problem statement and rationale for this study. It encompasses a discussion of the South African education context, the systemic assessment including international, regional and the annual national assessment. It highlights the objectives of the study and the main research questions in the study. It gives a brief discussion of the methodology and outline of the dissertation.

- Chapter 2 is composed of a detailed discussion based on a literature review of systemic assessment, the role of assessment, reasons for undertaking systemic assessment throughout the world as well as the comparison of international, regional and national studies. The purpose of systemic assessment and of the ANA are discussed, validity and reliability of the assessment instrument and the design of a quality national assessment are discussed. A review of previously conducted studies on learner attainment in mathematics is also discussed.
- Chapter 3 presents and argues the choice of the research design and methods. The latter includes; the population and sample, as well as the instrument evaluated in this study; the analytical methods; content analysis and statistical item analysis, applying the Rasch measurement model and the techniques for investigating the validity and reliability of the instrument.
- Chapter 4 presents findings, addressing the specific research questions based upon the literature and analyses undertaken.
- Chapter 5 presents a summary of the study and its findings, reflections about methodology and conceptual framework and concludes with main conclusions to the study and the recommendations regarding policy, practice and research.

CHAPTER 2 – LITERATURE REVIEW

Introduction

In Chapter 1 the background of the study was presented by specifying the rationale, purpose and key question guiding the study. A summary of the South African education context and issues of systemic evaluation contributing to this study were also introduced.

Chapter 2, the literature review, is presented and sources are selected on the basis of the main question in this study. This literature review then leads to the conceptual framework for this study (refer Chapter 3) that seeks to frame the main research question.

The main research question for this study is: To what extent does the 2012 Annual National Assessment Grade 6 mathematics instrument provide meaningful information for making appropriate interpretations on district level? This question required that a broad literature review be done on learner performance in systemic evaluation of mathematics.

Research on key variables that were found to influence learner performance in mathematics was explored.

Numerous studies have investigated problems that have an effect on learner attainment in mathematics; as such a focused literature review is presented to provide information with factors that have been found to have an effect on this performance.

The structure of the literature review is as follows: Firstly, there is discussion on international studies on systemic evaluation on mathematics and learner attainment, and the presentation of other studies, which focused on factors that were found to have an effect on learner attainment in mathematics internationally, regionally and locally.

2.1 THE ROLE OF SYSTEMIC ASSESSMENT IN SOUTH AFRICA

In this section, different perspectives of authors writing about systemic assessment are considered. Furthermore the roles that systemic assessments play and the way they are positioned in different countries is briefly touched.

The concept "systemic assessment" as utilised in this study is explained as follows. A "systemic assessment is designed to describe the achievement of learners in a curriculum area aggregated to provide an estimate of the achievement level in the education system as a whole at a particular age or grade level" (Anderson & Morgan, 2009, p. 10) and is conducted either on a sample or a whole population of learners. Systems assessment (also known as systemic evaluation in South Africa) is primarily concerned with quality in education and is a dynamic concept (Ross & Genevois, 2006).

Here systemic assessment is taken to mean assessing the "adequacy or appropriateness of objects or processes for the purposes for which they were intended" (Anderson & Morgan, 2009, p. 9). In national assessments, "the major focus when considering quality is on cognitive outcomes of the educational process, that is, what learners have learned with a view to developing strategies to improve those outcomes" (Schreenes et al., 2003, p. 340).

> "There are a number of different purposes that systemic assessment serves, for instance at the learner level, it can be used to describe learners' learning and diagnose learning problems, whilst at systemic level, the main purpose would be to reach a judgement on the effectiveness of an education system or part thereof, which is primarily the interest of governments and policymakers" (Howie, 2012, p. 87).

In light of this, Van Rooyen and Prinsloo (2003) assume that through systemic evaluation, the effectiveness of any aspect of the education system can be determined externally. Du Plessis et al. emphasised that the focus is on quality assurance, for example, the education system can be monitored by comparing the achievement of learners' performance with the national indicators of learner performance (Du Plessis et al., 2007).

In a national assessment, performance may be considered as "an indicator, which is a quantifiable measure, used to describe and monitor the education process" (Kellaghan & Greaney, 2001, p. 52). It can include monitoring a year on year comparison of learners' average achievements at the regional or system level. The investigation of the Annual National Assessment conducted in South Africa is the focus area of this study; thus, it is important to understand the definitions around the concept systemic assessment.

In South Africa systemic assessments aim to "measure the effectiveness of the education system by assessing the components of the education system at Grades 3, 6 and 9" (Howie, 2012, p. 87). Assessment programmes consist of achievement tests that are designed "to monitor acceptable levels of performance" in the core school subjects in a country (Scheerens, 2003, p. 34). The policies on systemic assessments have the stated aim to evaluate the performance of the entire education system (DoE, 2001). Locally, the systemic assessments "provide and implement a national framework for the evaluation of the education system and to develop benchmarks from which performance can be interpreted" (Howie, 2012, p. 88).

According to the existing legislature, the Minister of Education in South Africa is responsible for monitoring provision, delivery and the performance of education standards (Howie, 2012). Systemic assessments endeavour both to measure learner performance, and monitor the teaching and learning context (Howie, 2012). It therefore assesses the extent to which the education system achieves the desired social, economic and transformational objectives (Howie, 2012).

The systemic assessment on the learner-achievement component pursues the acquisition of knowledge, skills, values and attitudes by learners at different exit points in the education system (Howie, 2012). Learners should be equipped to become responsible citizen through the acquirement of the knowledge, skills, values and attitudes in the environment that is contributing to learning. "The contextual component is to provide insight into the environment in which teaching and learning take place and to establish the performance of the education system with respect to the principles of access, redress, equity and quality" (Howie, 2012, p. 88). In South Africa, systemic assessment is generally guided by the following principles (DoE, 2001):

- Incorporation with other quality assurance initiatives;
- Practicality of the design of the programme;
- Collaboration between the provincial and national departments of education;
- Collecting and using information to improve education provision and delivery;
- Ensuring inclusivity through the active participation of learners with special education needs (DoE, 2001, p. 51).

2.2 INTERNATIONAL AND REGIONAL STUDIES ON MATHEMATICS ATTAINMENT

South Africa has participated in two cross-country studies focusing on the performance of learners on the subjects' mathematics and literacy: Trends in Mathematics Science Studies (TIMSS) and SACMEQ. The message that came from these two studies was clear: the country compared poorly when compared with many of its less economically strong neighbours, and very poorly in relation to developing countries in other parts of the world (Tshabalala, 2008).

2.2.1 Trends in International Mathematics and Science Study (TIMSS)

In this section some of the characteristics of international assessments are explained as well as reasons why countries might decide to participate in the studies.

There was a need for international comparative assessments in view of a lack of standards of achievement that could be regarded as internationally valid. Individual countries needed to compare the performance of their learners with the performance of the learners of other countries (Anderson & Morgan, 2009). According to Anderson and Morgan (2008, p. 33), “a unique feature of international assessments is that they provide an indication of where learners’ achievements in a country stand relative to the achievements of learners in other countries, particularly in countries that may be regarded as economic competitors”.

International and national assessments have many procedures in common. International assessments however, are designed to allow administration across countries (Greaney & Kellaghan, 1996). “Instead of representing the curriculum of only one education system, the instruments have to be considered appropriate for use in all participating systems” (Greaney & Kellaghan, 1996, p. 20).

Husen and Postlethwaite (1996) describe the goals of international assessments as providing data for comparisons across countries; capitalising on the variability in education systems; exploiting different international school structures and curricula under one umbrella; describing the existing conditions of education in various countries; and suggesting further possibilities for education development.

As far as the comparative goal for international assessments is concerned, Beaton et al., (1999, p. 77) states that, international assessments are able to investigate “the distribution of achievement in schools throughout the country or the distribution in other countries, and the possible reasons for differences”.

An outcome rather than a goal of international assessments that can be perceived as an advantage is that “it tends to attract a lot of political and media attention” (Kellaghan & Greaney, 2001, p. 45). The publicity in the media can be useful if it results in the public having a greater appreciation for the value of education, and if it increases public support for educational expenditure (Anderson & Morgan, 2009).

A further advantage of participation in an international assessment is that participating countries are introduced to various experiences of large scale research studies, such as rigorous sampling, item review, supervision, analysis, and report writing (Anderson & Morgan, 2009). This exposure to international assessments can benefit participating countries that have not been exposed up to now to a tradition of empirical educational research and the associated technologies. Of particular importance is the developmental value of being involved in survey methodology and construction of assessments instruments (Kellaghan & Greaney, 2001).

The requirements for designing assessment instruments used for the international studies include that they should comply with the required standard. Firstly, there has to be an agreement about the grade and age at which the instruments are to be administered, because it impacts on the type and level of assessment items. This is also important because the selection of schools and learners in international studies has always been based on samples of learners and those samples have to be specified by grade and age group (Anderson & Morgan, 2009). Furthermore, taking into account the grade and age of learners in the sample allow for an examination of diverse phenomena as the significance of age of starting school, which is applied in various ways from country to country (Beaton et al., 1999).

2.2.2 The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ)

In this section regional studies conducted in different countries in Africa are discussed.

A number of ministries of education in Southern and Eastern Africa took the initiative during the 1990s to collaborate with the International Institute for Educational Planning (IIEP) to empower educational planners. This empowerment included training in the technical skills that are required for the monitoring and evaluation of basic education systems (SACMEQ, 1995). The collaboration also resulted in a further initiative by a group of educational planners to propose the creation of an association that would later be known as the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ). This suggestion received a positive reaction from ministries of education and the association was officially launched in 1995 (SACMEQ, 1995). During the period between 1995 and 1998, seven education ministries collected valuable information in SACMEQ I.

This regional assessment provided some baseline indicators for educational inputs, as well as general conditions of schooling. It also influenced the equity assessment and the allocation of human and material resources. Most importantly, however, it informed educational authorities about the literacy levels of Grade 6 learners (SACMEQ, 1995). The participation in SACMEQ continued to grow with the result that about fifteen ministries were participating in SACMEQ II between 1998 2011 and 2017 (Ross et al., 2000).

Since its inception in 1995, SACMEQ has undertaken three research projects.

SACMEQ I took place in 1995-1998, SACMEQ II was conducted in 1998-2004, SACMEQ III was administered in 2005-2010, and SACMEQ IV was administered in 2010-2014.

It was possible through these projects to investigate the conditions of schooling, as well as the quality of education at the primary school level, specifically in the subjects' numeracy and literacy (SACMEQ, 1995-2013).

The SACMEQ policies on presenting comparative data were set down by the participating countries ahead of the SACMEQ I study. In this policy it was agreed that only national reports would be published, not including cross-national reports. There were however, some indications that would be reviewed for the SACMEQ II study (SACMEQ, 1995). Even though the report on SACMEQ I did not include comparative cross-national data, it was possible to use this data for comparisons across countries (SACMEQ, 1995). The research questions, sampling procedures, the assessment instruments, the selection of target population, and the analyses of data are common to all participating countries (SACMEQ, 1995).

Interestingly, the SACMEQ developments coincided with the rapid development in Latin America and Caribbean countries to establish regional assessments during the 1990s. These assessments assisted in the provision of baseline data for educational development and reform. Practically all ministries of education in the mentioned region have to date incorporated regional assessments into their educational agendas (SACMEQ, 1995).

Apart from the regional assessments that were supported by UNESCO-UNICEF and IIEP, various countries, for example, Columbia, India, Namibia, Nepal, Uganda and Zambia, have conducted assessments with the support of the World Bank and other agencies (Kelly & Kanyika, 2000).

When regional assessments are compared, considerable similarity, rather than uniformity, can be observed as far as the assessed curriculum areas are concerned. In all of these assessments, language and mathematics are included. In about half of them, science, social studies or history and geography are assessed. Most of the regional assessments were conducted at a primary school level. Census and sample procedures have been used across these studies (SACMEQ, 1995).

2.3 COMPARISON OF SOUTH AFRICAN MATHEMATICS CURRICULUM AND UNIVERSITY OF CHICAGO SCHOOLS MATHEMATICS PROJECT

If the assessment results of South African learners at a specific level are to be compared with the learners in other countries, it is of importance that the curriculum offered in Grade 6 and the annual assessment administered in the same grade are compared to international standards. This section compares the structure of the mathematics content taught in Grades 4-6 of schooling in South Africa with the University of Chicago Schools Mathematics Project (UC[1]SMP), which is regarded as a global standard.

The reason for including this overview of topics was to confirm that the curriculum is generally aligned with international standards. The comparison is done with the aim of exploring the similarities on mathematics curriculum offered locally and internationally. Further elaboration is not within the scope of this study.

The comparison of the topics designed for the subject mathematics in the Intermediate Phase (Grade 4-6) learners in South Africa and in the (Grade 4-6) level in the University of Chicago Schools Mathematics Project (USCMP) is provided in this section.

[1] The United States of America (USA) comprised the states that are fairly autonomous but to date there has been no national curriculum.

The Department of Basic Education emphasis that, “this curriculum aims to ensure that learners acquire and apply knowledge and skills in ways that are meaningful to their own lives” (DBE, 2013, p. 18) as it provides an outline of mathematics topics that are designed to be taught at Grade 6 level.

This comparison will contribute in establishing whether the items that appear on the ANA instrument mathematics of the year 2012 correspond with the designed topics in the curriculum. It seems that the South African curriculum is in line with that of UCSMP as the major five domains from the table below indicate that similar domains appear (see Table 4). The topic depths are also equivalent. The only difference in the content is that UCSMP includes the topic of Reference Frames on the topic maps, which in SA is done in another subject which is Social Sciences in the same Grade. In the UCSMP, Reference Frames refers to map topics which include activities that range from using maps to estimating distances between cities and towns (Bell, Bell & Flander, 1998).

Table 4

Comparison between South African National Curriculum Statement, Curriculum Assessment Policy Statement and the United States of America University of Chicago School Mathematics Project (Adapted from: DoE, 2000; DBE, 2011; and Bell, Bell & Flander, 1998)

NCS CAPS DOCUMENT GRADE 4-6 (South Africa)	**GRADE 4-6 EVERYDAY MATHEMATICS (UCSMP, USA)**
▪ Numbers, operations and relationships	▪ Numeration and order ▪ Operations
▪ Patterns, functions and algebra	▪ Patterns, functions and sequence ▪ Algebra and uses of variables
▪ Space and shape (Geometry)	▪ Geometry and spatial sense
▪ Measurement	▪ Measures and measurement
▪ Data handling	▪ Exploring data and chance
	▪ Reference Frames (maps).

Table 4 (DoE, 2000; DBE, 2011 & Bell, Bell & Flander, 1998) draws comparisons on the structure of the content taught in the Grades 4-6 of schooling according to the NCS CAPS in South Africa, (derived from DoE, 2000 & DBE, 2011), and in a curriculum designed by the UCSMP (derived from Bell et al., 1998). It is important to be aware of the mathematics topics taught in the Intermediate Phase (Grade 4-6) because the focus area of this study was to investigate the mathematics content and the length of scope in Grade 6 which is expected to be covered by schools before the administration of the ANA.

2.4 DEFINITION OF ASSESSMENT AND THE PURPOSE OF SYSTEMIC ASSESSMENT

This section focuses on assessment and the purposes of systemic assessment.

Two definitions of assessment were considered for use in this study. According to Dreyer (2008, p. 5), assessment is the “process of gathering and discussing information from multiple and diverse sources in order to develop a good understanding of what learners know, understand, and can do with their knowledge as a result of their educational experiences”. McMillan conceptualises assessment “as the gathering, the interpretation and the use of information to aid and guide all the stakeholders involved including the teacher, with decision-making concerning learners’ progress” (McMillan, 2007, p. 37).

For the purpose of this study the second definition was found to be appropriate in guiding the investigation of 2012 ANA Grade 6. The important aspects of this definition (gathering information, interpreting information, and using information to guide stakeholders) are systematically explored in this study. Dreyer’s definition (2008) is considered more appropriate for School Based Assessment.

The literature is reviewed on the research done about systemic assessment and the theory underpinning such assessments. By way of introduction reference is made to the policy documents of South Africa with a view to compare and contrasts the literature with the policy in South Africa.

The main purpose of systemic evaluation introduced in 2000 in South Africa by the Department of Education was to measure the effectiveness of the entire educational system. Furthermore, it was meant to evaluate the extent to which the vision and goals of the educational transformation process in the country have been achieved (DoE, 2003). The systemic assessment would produce information that would impact on the system, for example, on the resources and facilities, on leadership, communication and management, school governance and internal relationships, curriculum design, and teacher- and learner performance (DoE, 2003).

The above purposes of systemic assessment in South Africa resonate with what is internationally accepted, namely; the “main purposes of an assessment of an institution or system are to reach a judgement about the effectiveness of a school, and to reach a judgement about the adequacy of the performance of an education system or part of it” (Kellaghan & Greaney, 2001, p. 59). The information gained from the assessment programmes can lead to “adaptations in the curriculum in the sense of goals (standards) or means (curriculum contents) and all conditions that have impact on the performance in a particular subject for example, teacher training in the particular subject matter area” (Scheerens, Glas & Thomas, 2003, p. 35).

In pursuit of these objectives, procedures have to be established to collect data from learners and other stakeholders in the education system (Kellaghan, Greaney & Vincent, 2009). The data collected should result in improved transparency of outcomes of the educational management and practice in a country, which would in turn inform improvement of the system (Kellaghan & Greaney, 2008). Furthermore, the extent to which the objectives of a national assessment have been attained, have implications with respect to access, quality, efficiency, and equity within the system (Anderson & Morgan, 2009).

In view of the fact that the educational and political power structures of various countries differ significantly, it can be deduced that their individual stakeholders will have different interests in a national assessment (Kellaghan & Greaney, 2001). This fact has an influence on the type of information needed by the stakeholders and also on the procedures that are chosen for an assessment. Other factors that influence the differences between systemic assessments in different countries are the level of sophistication in policy making, the different perspective about the nature of education and educational accountability, and also the country's capacity for policy implementation (Kellaghan & Greaney, 2001).

The specific aims of systemic assessments identified by Anderson and Morgan (2009) are to determine:

- Whether learning in the education system is taking place effectively, as far as general expectations, the aims of the curriculum, and preparation for life are concerned;
- Any evidence of particular strengths and weaknesses in individual learners, knowledge and skills;
- Whether any subgroups of the population perform poorly as compared to other, for example, disparities between the performance of boys and girls, of learners in rural and urban areas, of learners from different language and ethnic groups, and of learners in different regions of the country;
- Any factors that can be associated with learner achievement. Also, to what extent is achievement related to the learning environment, for example,

resources, teacher competence, and the location of the school or with learners' home and community circumstances;

- Whether the achievements of learners show change over time.

Assessment programmes consist of educational achievement tests that are meant to monitor acceptable levels of performance in the basic school subjects in a country.

National assessment tests in a particular subject need not be administered each year, rather it is advisable to test each subject every six years (Scheerens, Glas & Thomas, 2003). Locally, however, the ANA is administered each year. This may be perceived as a weakness in South Africa's national assessment, as it is argued that teachers are not given enough time to adapt to the curriculum, and to focus on strategies that are meant to remedy conditions that have contributed to performance.

2.5 VALIDITY AND RELIABILITY ON THE DESIGN OF A QUALITY NATIONAL ASSESSMENT INSTRUMENT

This section explores the design of a quality national assessment instrument. In an educational setting, the term "quality" applies to various aspects of learners' educational experiences. The safety of the environment, the resources, curricula, instructional practices, teacher competency and learners' learning can all be judged according to quality criteria. The focus in this section however, is on the quality of a national assessment instrument.

In national assessment studies, there is a major focus on the cognitive outcomes of the educational process and on what learners have learned in order to develop strategies to improve those outcomes (Kellaghan et. al., 2009). Thus "this emphasis is in keeping with the Target 6 of the Dakar Framework for Action, which highlights improving the quality of education, so that recognised and measurable learning outcomes are achieved by all, especially in literacy, numeracy and essential life skills" (Anderson & Morgan, 2009, p. 85). A comparison of the Dakar Framework for Action and the Queensland Framework used in this study shows that Dakar Framework outlines action the Queensland Framework provides information on the systemic assessment process that is comprehensive and detailed. The Queensland Framework served the particular need in the case of this study.

Before considering four specific conditions to ensure quality of national assessments it has to be mentioned briefly that any assessment instrument has to be both valid and reliable.

In this study a content analysis and an item analysis were conducted on 2012 ANA Grade 6 mathematics instrument. Within that analysis the validity and reliability of the mentioned content and items are explored in addressing the research sub-questions.

Messick defines validity broadly as "an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (Messick, 1989, p. 5). In a similar vein Anderson and Morgan (2009) state that, test validity refers to a broad concept that involves making appropriate interpretations and uses of scores or test information. In this study the focus is on two types of validity namely on content validity and construct validity.

According to Babbie "content validity is used as a theoretical concept that focuses on the extent to which an assessment instrument shows evidence of fairly and comprehensive coverage of the domain of items that it supposed to cover" (2007, p. 2). The areas in focus for content validity, according to Babbie are its relevance to the curriculum; its focus on what was taught; the comprehensiveness of content coverage; the proportion of the scope of learning; and the sampled potential content (p. 2).

In Messick's view, construct validity is an all-encompassing idea. It means that the content of the test must be appropriate and, meaningful, so that the inferences based on the scores can be useful. This also means that the strongest quality of a test must be the trustworthiness of empirically grounded score interpretation (Messick, 1989).

According to Anderson and Morgan (2009), test reliability refers to the extent to which the "evidence collected is sufficient for generalisations. In addition, **r**eliability is the extent to which a "test measures consistently" (Scheerens et al, 2003, p. 101). The "reliability refers to the inherent consistency of an assessment task to produce the same results if repeated under the same condition" (Popham, 2003, p. 10). An assessment task is considered reliable if it is measuring something consistently (Messick, 1989).

The above definitions have been selected as they contain the elements of validity and reliability that are investigated in this study, notably Babbie's areas of focus in content analysis and Messick's criteria for validity in item analysis .The focus of this study was to investigate the quality of the ANA Grade 6 instrument and the design. The construct validity, the content validity and the reliability of the 2012 ANA Grade 6 mathematics instruments is explored so as to examine whether the ANA measured what it was intended to measure.

In recognition of the important role conferred to learner learning in a national assessment, this section describes four conditions that should be met to ensure that the test that is used, actually accurately represents the learner achievements and also that the information obtained serves the needs of the users (Beaton & Johnson, 1992).

Table 5 presents a summary of the four conditions for accurate learner achievement and detailed explanation follows later.

Table 5

The Four Conditions for Accurate Learner Achievement (Adapted from Anderson and Morgan, 2009)

	Conditions
1.	Knowledge and Skills
2.	Level of development
3.	Language use
4.	Results to monitor change

2.5.1 Representation of Knowledge and Skills

The first condition for ensuring quality of a national assessment is that the knowledge and skills of learners are tested by adequately representing what they have learned. It would have been ideal to test everything that they have learned but a test can only measure a part of the knowledge and skills specified in a curriculum (Messick, 1989). Moreover, "test items should exhibit curricular importance, cognitive complexity, linguistic appropriateness, and meaningfulness" for learners (Kellaghan et al., 2009, p. 15). Hence, it is not necessary to limit a test measuring single skill components or isolated items of knowledge that require straight recall of facts. This is a feature (and a weakness) of many national assessments, according to Kellaghan et al., (2009). However, the goal of the education system is to develop higher levels of cognitive skill, including the reasoning capacity, the ability to identify and solve problems in a logical way, and the ability to perform non-routine tasks (Kellaghan et al., 2009).

When developing a test, according to Anderson and Morgan (2009), it must be considered that the outcomes of the test must provide a basis for policy and decisions that may influence curriculum development and effect instructional changes. This, in turn may "foster the development of valued knowledge and skills" (p. 20). It is therefore important to include an adequate number of items within a test, to ensure an adequate representation of the subject content, for example, including content strands or skills in mathematics (Anderson & Morgan, 2009). Kellaghan et al., (2009) reason that, "the small number of items in some national assessments must raise questions about their adequacy in this respect (p. 17).

2.5.2 Level of Development of Learners.

The second condition is that a test should assess knowledge and skills at the level that is appropriate for learners at that development stage. In realistic terms, it can be expected that only a few learners would get all the items right or all the items wrong (Anderson & Morgan, 2009). If a test is based completely on curriculum documents, and if the curriculum sets unrealistic expectations for learners' achievement, it might happen that the test will be much too difficult for lower-achieving learners (Kellaghan et al. 2009). The "solution lies in taking into account in test development not just the standard of the intended curriculum, but also what is known of the actual achievement of learners in schools" (Kellaghan et al., 2009, p. 18).

The actual achievement of learners can be established by involving practising teachers in the development and selection of test items. Carefully field-trialing of the items before the main assessment takes place is also important. The field trial test has to make use in a sample that is a representative of the variation in schools of the target population (Kellaghan et al., 2009).

Assessment should be reasonable, and respond equitably to learner difference. MacMillan (2007) provided a list of possible factors that may influence the fairness of an assessment task. These factors, together with a brief explanation of each, are summarised (see Table 6).

Table 6

Factors that May Influence Fairness of an Assessment Task (Adapted from MacMillan, 2007)

Factor	**Explanation**
Learner knowledge of learning targets and assessments	Learners need to have clarity and be informed about expected performances (learning outcomes) and the nature of the assessment task.
Opportunity to learn	Assessment tasks should be aligned with teaching and learning by making provision for sufficient time, resources and conditions to assist learners to demonstrate their achievement.
Prerequisite knowledge and skills	Provision should be made to ensure that learners have adequate knowledge and skills to complete the assessment task successfully. For example, will learners be able to read, understand and interpret the questions? Are learners aware that the time allocation of the assessment task is crucial for planning their responses?
Avoiding learner stereotyping	Prejudice based on preconceived notions of group of learners or individual learners should be eliminated, since this interferes with objectivity.
Avoiding bias in assessment task and procedures	Avoid offensive assessment content and unfair penalisation that may disadvantage learners.
Accommodating special needs	When assessing learners who are experiencing any barriers, assessment tasks should be modified to avoid disabling traits that interfere with learners' performance.

2.5.3 Language Use

The third condition is that a test should provide valid information on learners' knowledge and skills in a particular subject or curriculum domain and that the competence is not influenced by their competence in other domains (Messick, 1989). For instance, a mathematics test should not contain so much language that the learner's performance depends on the language ability to read rather than on the mathematics ability (Kellaghan et al., 2009). This problem is worse when a learner's skill in reading is not adequate to understand or interpret complicated questions (Anderson & Morgan, 2009).

2.5.4 Assessment Results to Monitor Change

The fourth and final condition is that, in order for assessment results to monitor change over time, the instruments must be comparable (Kellaghan et al., 2009). The ideal situation would be that the same test is used every round and that it is kept secure between administrations (Anderson & Morgan, 2009). Practically that is almost impossible because the test content can also be spread by word of mouth. Best practice involves that the subset of items are carried over from test to test to effect a strong to link a previous tests (Kellaghan et al., 2009). Moreover, it is ideal that learner samples and test procedures be the same. If conditions differ over which there is no control, for example, response rates, it is necessary to take those factors into account when comparing learners' achievement at different points in time (Anderson & Morgan, 2009).

To sum up, the assessor needs to be aware of the possible pitfalls of written assessment. Table 7 presents a summary of the four pitfalls for learner achievement.

Table 7

The Four Pitfalls for Learner Achievement (Adapted from Chamber, 2002)

	Pitfalls
1.	Style and context
2.	Questions
3.	Language
4.	Time

Chamber distinguished four pitfalls that may in a way reduce the reliability of the results:

- The style or context of the question may appeal to some learners more than others;
- Questions may be unintentionally ambiguous;
- Small details of language can affect the facility of the question;
- The time needed to complete the test may be underestimated (Anderson & Morgan, 2009, p. 47).

The above discussed four conditions to ensure quality on the national assessments are included in this study as evaluation criteria for what constitutes an effective, valid, reliable mathematics ANA instrument for the Grade 6 level.

2.6 TYPE OF ASSESSMENT

Once the purpose of assessment is determined, the designer should consider the best applicable type of assessment which would be applicable to fulfil the purpose. The "potential for use of the information derived from an assessment depends on the characteristics of the assessment" (Kellaghan et al., 2009, p. 30).

In this study, the Annual National Assessment attempts to achieve both types of assessment: systemic assessment and diagnostic assessment. This dual aim is given for the following reasons (DBE):

- to determine the performance of the education system;
- to diagnose the causes of possible learning barriers by learners; and
- to address or remedy the problem of underperformance (DBE, 2009, p. 58).

The results of the 2012 ANA Grade 6 mathematics instrument are investigated based on the above reasons specified by the Department of Education.

The two types of assessment relevant to this study are defined as follows:

2.6.1 Systemic Assessment

As quoted earlier, the main purposes of an assessment a system are to make a judgement about “the effectiveness of a school”, and “the adequacy of the performance of an education system or part of it” (Kellaghan & Greaney, 2001, p. 70).

2.6.2 Diagnostic Assessment

Diagnostic assessment helps with the decision about which applicable intervention and supportive strategies must be applied to eliminate the relevant barriers to learning (Van Rooyen & Prinsloo, 2003).

As argued earlier in the study, there is a gap between the ANA and the SBAs and it is of importance to recognise the types of assessments that are applicable in the classrooms even though some of them are not applicable to the ANA. There is a relationship between the assessment types; the assessment approach and the assessment purpose (see Table 8).

Table 8

The Relationship between the Assessment Type, Approach and Purpose (Adapted from Meyer et al., 2010)

Assessment type	Assessment approach	Assessment purpose
Baseline assessment	Informal/ formal	To determine the entry level of learners to a new learning experience
Diagnostic assessment	Informal/ formal	To determine the nature of possible learning barriers experienced by learners
Formative assessment	Predominantly informal	To determine teaching success and learner development, with the aim of the advancement of learning during the teaching and learning situation
Summative assessment	Predominantly formal	To determine learning success at the end of culminated learning experience

In this study the focus is to investigate the two types of assessments, that is, the systemic assessment and the diagnostic assessment as mentioned earlier.

2.7 SOME PROBLEMS THAT AFFECT LEARNER ATTAINMENT

In this section, a review of the literature accessed from other similar and previous studies on mathematics attainment, both locally and internationally, is conducted. The purpose is to extract lessons from these on problems that were found to have an effect on learner attainment in mathematics. This will enrich this study with experiences of what other researchers have done previously.

2.7.1 Language

In South Africa, the Language of Learning and Teaching (LOLT) is the official language of instruction in a school decided upon by the School Governing Body (SGB) (Tshabalala, 2008). “It is argued that mother tongue is the best medium of education at school because it is the language the child knows well, and in which they can express meanings” (Pattanayak, 2003, p. 96).

The language of instruction and testing at the Foundation Phase (Grade 1-3) is the learner’s home language, as specified by the school. All eleven official languages in South Africa are used as LOLT in the Foundation Phase (Tshabalala, 2008).

This situation is changed however when learners proceed to the Intermediate Phase, where the LOLT of the school may be a language other than a learners’ home language, that is either English or Afrikaans. This situation might have influenced learner attainment in the ANAs which were also conducted in English and Afrikaans. An analysis to determine the effect of language of instruction used on the Annual National Assessment instrument and on learner attainment in mathematics would be valuable as a future research study project.

2.7.2 Prior Knowledge, Skills and Attainment

Prior knowledge and skills mastered, contributes to learner attainment in mathematics and all the subjects that are offered at school. Kurdek and Sinclair (2001, p. 30) conducted research in the United States of America focused on preceding reading and mathematics in Grade 4 children investigation “kindergarten readiness scores”.

Their research examined “the age in verbal and visual-motor skills at kindergarten” the later attainment in reading and mathematics, and the “link between skills at kindergarten and later attainment” cited by (Tshabalala, 2008, p. 99). Kurdek and Sinclair’s study found that “younger children in kindergarten perform less well than their older counterparts, and with controls for age” and their “verbal skills uniquely predicted later mathematics attainment” (Kurdek & Sinclair, 2001, p. 31). They concluded that, “readiness in the specific areas of auditory memory, number skills, and visual discrimination” was a good predictor for later mathematics achievement. They also found that age plays a vital role in determining learner readiness for starting school (Kurdek & Sinclair, 2001, p. 32).

Matters pointed out that other background characteristics such as “gender, type of school attended and ethnicity” often influence learner attainment (Matters, 2009, p. 219).

The above factors might affect performance, but they are not the focus area of this study. This study explores the quality of the assessment instrument and hypothesises the effect that it might have on mathematics attainment in Grade 6 level.

A number of different issues related to learners’ performance in mathematics were reviewed in this section. The role of systemic assessments, the reasons for undertaking systemic assessments throughout the world and the purpose of systemic assessments were explored. South Africa has participated in a number of international studies namely, MLA, TIMSS and the SACMEQ. The results of all these studies on the whole showed that learners performed comparatively poorly in mathematics. The results of the Systemic Evaluation at Grade 6 level confirmed findings of earlier studies.

The comparison of the South African and the UCSMP curriculum looked at the content offered in mathematics at Grade 6 level in both the countries. The comparison provided similar characteristics; there are no gaps observed.

The concept assessment was discussed broadly, including the definition of assessment and purpose of systemic assessment; validity and reliability on the design of a quality national assessment instrument, and the type of assessment for the ANA. Problems that were found to have an effect on learner performance in mathematics were discussed.

The broad discussion on systemic assessment lays a the foundation towards the conceptual framework and research methodology presented and is discussed in the next chapter.

CHAPTER 3 – CONCEPTUAL FRAMEWORK AND RESEARCH METHODOLOGY

Introduction

The aim of the study was to investigate the quality of the Annual National Assessment instrument including its design, with specific reference to the Grade 6 mathematics assessment.

In this chapter, the conceptual framework, as well as the rationale for the adoption of the conceptual framework on which this study is based and the research methodology are discussed. Under research methodology, the research design applied in this study is presented, and the research questions used in this study to address the design are presented and discussed. The population and sample employed in this study are described as well as the instrument used. The content analysis and the statistical item analysis applied to this study for data collection are discussed. Finally, the methodological norms (validity and reliability) that guide this study and the ethical clearance requirements for conducting this study are discussed.

3.1 CONCEPTUAL FRAMEWORK FOR THIS STUDY

The conceptual framework was informed by the main research question and the sub-research questions, which addresses the quality of the 2012 ANA Grade 6 mathematics instrument, including the design thereof, and to inform the potential inferences that can be made on the mathematics performance of learners in Grade 6 in the Gauteng North District. The conceptual framework guides the researcher in this study in identifying aspects that are required in the quality and design of an assessment instrument, though this study is confined to an analysis of the instrument. The conceptual framework is discussed from a theoretical point of view. In addition, the strategy of inquiry and the procedure for obtaining collected data and the analysis of data are described in this study.

As far as the structure of the conceptual framework is concerned the Queensland Studies Authorities Assessment Policy document (2009) inspired the design and development of a model for quality systemic assessment proposed in the study. The Queensland Studies Authorities Assessment Policy document is implemented in Australia. Australia administers the National Assessment Program-Literacy and Numeracy (NAPLAN) similar in some respects to the ANA. The NAPLAN assessments were first implemented in May 2008 when national assessments were held in literacy and numeracy for all learners in Australia in Years 3, 5, 7 and 9 (ACARA, 2008). In South Africa, the ANA was introduced in 2009 for assessing learners in Grade 1, 3, and 6 in key subjects such as literacy and numeracy. From 2012 onwards all the learners in Grade 1 to 6 and 9 were included in the ANA (DBE, 2012). In the present model however the emphasis is not generic but specifically on systemic assessment.

The framework that is presented in Figure 1 is a model for quality systemic assessment with specific reference to the performance of the Grade 6 on the ANA mathematics written in the year 2012 (English data only) in the Gauteng North District, Pretoria, South Africa. The model was chosen for this study because it is highly recommended, and it is comprehensive. As such it can be used as a comparison to look at our curriculum broadly. The main purpose of the conceptual framework was to direct the course of this investigation in answering the main research question, namely: “To what extent does the 2012 Annual National Assessment Grade 6 mathematics instrument provide meaningful information for making appropriate interpretations on district level?”. The study mainly focused on the five highlighted blocks of the conceptual framework.

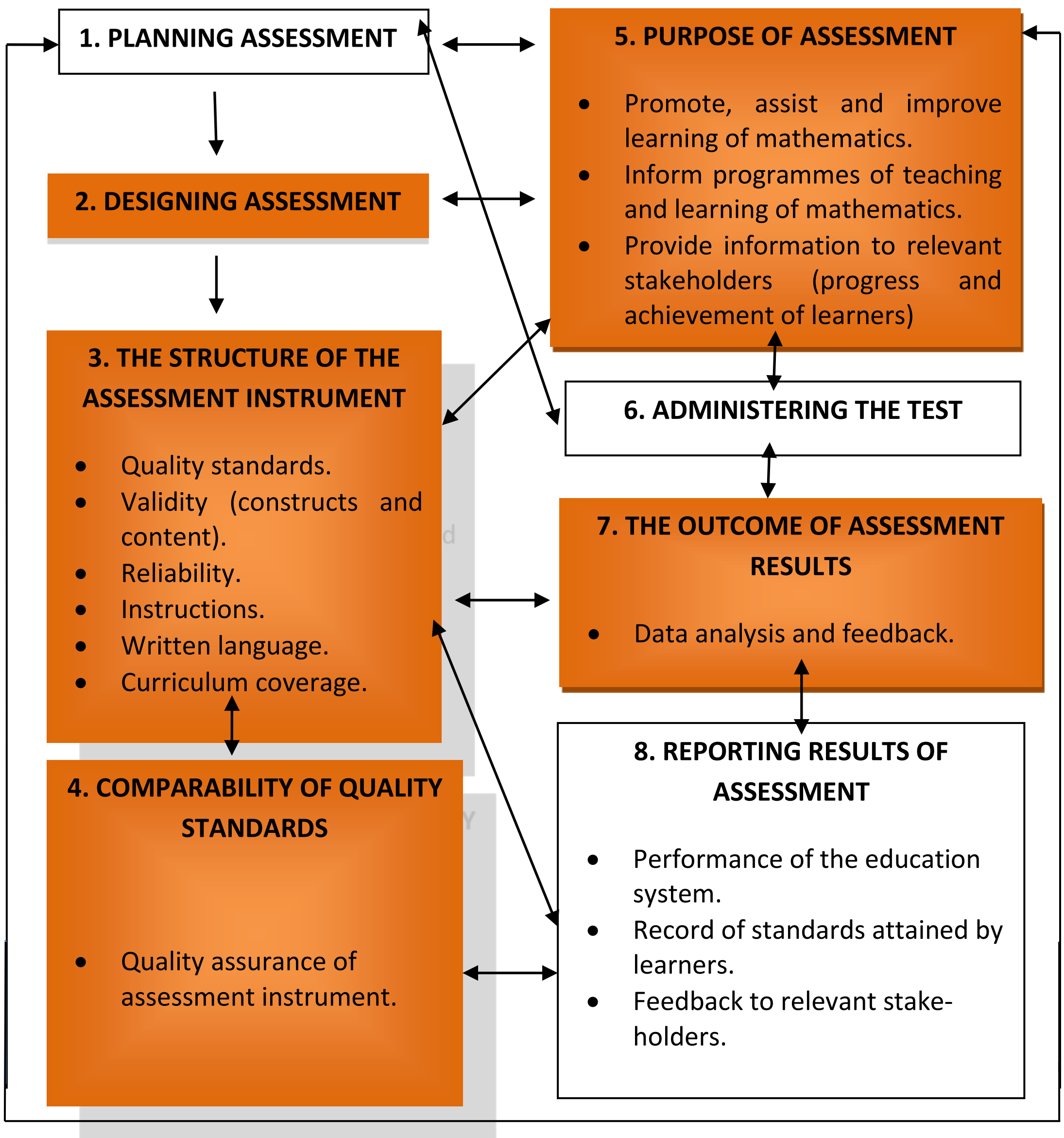


Figure 1. A Model for quality systemic assessment

The eight components of the above conceptual framework can be explained as follows:

1. The first step of the assessment process is planning. This step involves the type of assessment to be administered, the objective of that assessment and the time at which the assessment will be administered. At the end of the teaching and learning process, assessment should take place in order to find out whether there was progress on what was done in the classroom. Generally, assessment is used as a yardstick to reinforce what was taught in class. According to Dreyer (2013, p. 5)

assessment is the process of gathering and discussing information from multiple and diverse sources in order to develop a good understanding of what learners know, understand, and can do with their knowledge as a result of their educational experiences.

2. Once planning is done, the designing of the assessment follows. Designing assessment involves making a decision on the form of assessment and the type of task needed for a particular content. The first major consideration in assessment design is how much assessment is needed and how much time should be spent on it (Anderson & Morgan, 2009). Assessment should be reasonable, and should respond equitably to learner difference. This implies that for an assessment to be fair, it should be unbiased and provide all learners with equal opportunities to demonstrate achievement (Meyer et al., 2010). The negative factors that may influence the assessment experience should be eliminated. The instrument should be time-efficient and manageable.
3. The structure of the assessment instrument involves a number of elements that need to be considered during the construction of an assessment instrument. The assessment instrument should be of an acceptable standard, meaning; the quality standards, validity, reliability, instructions, written language and curriculum coverage should be taken in to consideration. The style or context of the questions used on the assessment instrument should be appealing to learners. The assessment instrument should be valid and reliable, cater for different cognitive styles and varied skills, and include a variety of questions in order to cater for these differences. The items included on a test should be a representation of the curriculum and the cognitive domains, and be language appropriate and relevant for learners (Kellaghan et al., 2009, p. 15). The instructions should be clear and

simple. The language of the items should be clear, simple and appropriate to the level of development of learners. A mathematics test should not contain so much language that the learner's performance depends on the language ability to read rather than on the mathematics ability (Kellaghan et al., 2009). The assessment instrument should be a true representation of the curriculum and scope expected. This will ensure that learners have a fair opportunity to exhibit their knowledge and to complete the assessment task successfully. In order to ensure successful assessment, the assessment instrument should be balanced, comprehensive and varied (Dreyer, 2013).

4. Comparability of standards involves the quality assurance of the assessment instrument. In the education system emphasis is given to the comparability of standards. Through quality assurance; the education system can be monitored by comparing the achievement of learners' performance with the national indicators of learner performance (Du Plessis, et al., 2007). Achievement and progress should be recognised through the comparability of standards. In order for assessment results to monitor change over time, the instruments must be comparable (Kellaghan et al., 2009). The ideal situation would be that the same test is used every round and that it is kept secure between administrations (Anderson & Morgan, 2009).

5. The purpose of the assessment should be established. Assessment is administered for different purposes and in different situations. Whilst assessment is seen as an integral part to teaching and learning, it should promote, assist and improve the learning of mathematics. Assessment should inform programmes of teaching and learning of mathematics. Assessment should provide information to the relevant stakeholders. Some assessments may be more useful for one purpose than another, while some may be useful for both. This will depend on how they relate to the

learners' work and to the way the results are interpreted and used (Black, 1998). Assessment should improve the quality of learning, and should be a direct and clear link with the outcomes.

6. Administering the test involves procedures to be adhered to during the process. The assessment environment should be strictly controlled. The assessment instrument should be available or received on the day of the assessment. There should be a schedule of invigilation, marking and moderation for the whole process. The conditions of the classroom, noise level, light and ventilation should be taken into considered when administering the test.
7. The outcome of the assessment results involves the analysis of results and giving feedback. Then communication of the results should be done at all levels.
8. Reporting results of assessment involves the use assessment results as a measure to monitor the effectiveness of the system. The information drawn from the assessment results may be used to reach a judgement about the performance of an education system (Kellaghan and Greaney, 2001). For safe keeping, the results of the assessment task should be recorded. These results can be used to evaluate whether learners have gained the necessary skill and knowledge and achieved their educational goals. Finally, feedback should then be given to the relevant stake-holders. Assessment feedback involves gathering information, interpreting and using information to guide and assist stakeholders with decision-making (McMillan, 2007).

The model is linked to the main research question, in that it incorporates the design, structure and comparability of standards of the instrument as well as the purpose and outcomes of the systemic assessment. In this way a judgement about the meaningfulness of the instrument is made possible. On the basis of this conceptual framework it was possible to phrase six sub-research questions to address each component of the model, as follows:

1. What is the purpose of the Annual National Assessment Grade 6 instrument?
2. To what extent does the instrument fulfil the criteria for good assessment design?
3. To what extent does the instrument have construct validity and content validity?
4. Are the items on the mathematics assessment instrument functioning as expected? Specifically do the items fit the requirements of the Rasch measurement model?
5. How well are the items distributed along the continuum of the variable?
6. How well is the assessment instrument targeted to the abilities of the learners?

These questions seek to explore the quality of the ANA Grade 6 mathematics instrument, in terms of its construct validity, content validity and reliability in order to evaluate the potential inferences that could have been made and actions taken based upon the mathematics performance of learners in Grade 6 in the GND in the year 2012.

In the literature review that served as the basis for the conceptual framework, international studies on systemic evaluation on mathematics and learner attainment were reviewed in Chapter 2.

Direction was gained from the way that the outcomes of other studies were presented, with a focus on factors that probably had an effect on learner attainment in mathematics internationally, regionally and locally. The purpose of systemic evaluation generally, and the purpose of the ANA specifically were also investigated.

The most prominent gain from the literature review which directed the conceptual framework of the study was the idea of four conditions to ensure the quality of a national assessment, as explained by Kellaghan et al. (2009). These conditions are: firstly, the instrument must be a representative of the knowledge and skills that learners have gained in class; secondly, the level of learner development should correspond with the level of questioning; thirdly, the language used should not hinder the demonstration of learners' mathematics knowledge; and finally, the assessment results should be used to monitor change. These four conditions guided the first three highlighted sections of the conceptual framework.

Throughout the literature review, the role of systemic assessments and the reasons for undertaking systemic assessments throughout the world were explored. South Africa has participated in a number of international studies, which were also reviewed. Although not directly incorporated in the framework, this information served as a backdrop against which the 2012 ANA was investigated. What was used directly from the literature review, was the purpose of systemic assessments (refer to Figure 1 page 61, the fourth highlighted block).

What was of particular importance to motivate the present study, was that the results of international studies showed that learners performed comparatively poorly in mathematics, a notion that was confirmed by the results of the Systemic Evaluation at Grade 6. Realising that various assessments cannot be compared directly, the international studies provided results that indicated trends, but could not be used as a yardstick to compare the ANAs.

Aspects from the literature review that further informed the conceptual framework were the concept of assessment, including the definition of assessment and purpose of systemic assessment; validity and reliability on the design of a quality national assessment instrument, and the type of assessment for the ANA.

In Figure 1, (Chapter 3 and page 61) the five components of assessment namely; purpose of assessment, designing assessment, the structure of the assessment instrument, the outcomes of the assessment results and the comparability of quality standards are linked within the assessment process. The model explains in detail the procedure followed in the process of assessment from the planning stage to the final stage of reporting. The arrows in the model point to the direction of the interaction amongst components of the assessment process.

These linkages are meant to illustrate how the elements presented are interlinked in terms of answering the main research question and the sub-questions as mentioned earlier in this chapter.

3.2 RESEARCH METHODOLOGY

The Annual National Assessment conducted by the Department of Basic Education in 2012 focused on learners in Grades 1 to 6 (Foundation and Intermediate Phase) and Grade 9 (Senior Phase) in both Language and Mathematics as explained previously.

The Department of Basic Education compiled a diagnostic report "after the administration of the ANA tests, in which an analysis was made to provide evidence that would inform and direct appropriate interventions" to the relevant stakeholders (DBE, 2013, p. 4).

The ANA Diagnostic Report of 2012 by the Department of Basic Education presented "critical aspects of knowledge and skills highlighting the inadequacies that were identified from random samples of Grades 3, 6 and 9 learner scripts in the Language and Mathematics tests of the 2012 Annual National Assessment cycle" (DBE, 2013, p. 6).

> "As part of verifying the quality and consistency of marking of the ANA scripts which was done at individual school level, the Department of Basic Education and the Provincial Education Departments (PEDs) collected random samples of marked

scripts from each school that had learners in Grades 3, 6 and 9 and arranged centralised re-marking of the scripts in district centres" (DBE, 2013, p. 7).

Based on this background an appropriate methodology had to be selected to investigate the quality of the 2012 ANA Grade 6 mathematics. This chapter discusses the research design that was adopted for the study, including the phases of development that unfolded in the cause of the investigation. The analytical methods of research (content analysis and statistical item analysis) are discussed as it complies with the acceptable methodological norms. The method of research required sampling and evaluation of the instrument, which is also discussed in this chapter. The processing and analysis of data are then discussed, using the Rasch measurement model. Finally, the ethical considerations for this study are briefly mentioned.

3.3 RESEARCH DESIGN

The study is rooted in a pragmatic paradigm. The great educational pragmatist was Dewey, who viewed education as a process for improving the human condition in life. Dewey saw schools as specialised environments within the larger social environment. He emphasised problem solving and the scientific method. In contrast to the traditional theories, pragmatism is based on change, process, and relativity (Ornstein & Hunkins, 2009).

This study fits into the pragmatism paradigm because from the data gathered from this study; change, process and relativity can be utilised in the design of a quality assessment instrument in future. Pragmatism construes knowledge as a process in which reality is constantly changing (Ornstein & Hunkins, 2009). This study is fully dependent on existing ANA Grade 6 mathematics instrument administered in 2012 which could not be changed in any way in the existing study, but recommendations could be made. To pragmatists, in a real world, curriculum is based on a child's experiences and interests and prepares the child for life's affairs. Both the learner and the environment are constantly changing; pragmatists reject the idea of unchanging universal truths. Since learning occurs as the person engages in problem solving, this is transferable to a wide variety of subjects and situations.

To pragmatists, teaching should focus on critical thinking, because teaching is more exploratory than explanatory and the subject matter is interdisciplinary (Ornstein & Hunkins, 2009). The study of the 2012 ANA Grade 6 mathematics finds itself embedded in a pragmatic paradigm that focuses on the responses of the learner. Working with the pragmatic paradigm is consistent with a secondary analysis design where the study has to constantly adjust to the origins of the original data. In the case of this study the approach is limited to the analysis of the test instrument.

The research was conducted applying a secondary analysis design and applying mixed methods. The secondary analysis allowed the researcher to investigate the previously administered ANA Grade 6 mathematics instrument. In order to gain the understanding of low performance by learners in the year 2012, this study focused on the quality and the design of the ANA instrument and as such applied the Rasch measurement model to investigate possible attributes that contributed to the performance of learners. A limitation for this study is that the focus was on the quality design and excluded the external factors that might have contributed to the low performance of the learners.

A content analysis was conducted (applying professional judgement) of the Annual National Assessment instrument, followed by a statistical item analysis applying the Rasch measurement model. These analytical methods were utilised to determine the quality of the ANA Grade 6 mathematics instrument, in terms of its content validity, construct validity and reliability in order to evaluate the potential inferences that could have been made and actions taken based upon the mathematics performance of learners in Grade 6 in the GND in the year 2012.

The research design is anchored around the main question and the sub-questions investigated in this study as mentioned earlier in this chapter.

3.3.1 Phase One

The first three sub-questions are:

1. What is the purpose of the Annual National Assessment Grade 6 instrument? The literature review explored the concept systemic assessment, the role of the systemic assessment in other countries, the purpose of the systemic assessment and the purpose of the annual national assessment in South Africa. In this question, the various factors that might have an effect on learner attainment in mathematics were identified in other studies in the literature review. The literature review, particularly the purposes for good assessment, underpins the conceptualisation of the overall research questions used to investigate whether the ANA Grade 6 mathematics instrument had met the purposes as expected.

2. To what extent does the instrument fulfil the criteria for good assessment design? The design of a quality national assessment instrument was discussed in the literature review. To address this question, the criteria in the literature review were used to evaluate the ANA Grade 6 mathematics instrument. The elements of the conceptual framework informed the evaluation of the ANA Grade 6 mathematics instrument and the extent to which it met the criteria for good assessment design.

3. To what extent does the instrument have construct validity and content validity? This question explored the validity of the assessment instrument that was, whether

the ANA Grade 6 mathematics measured what it intended to measure (both construct and content) of the national assessment instrument. To address this question the content analysis and the item analysis was used to evaluate the ANA Grade 6 mathematics instrument.

The first three sub-questions are addressed with reference to the literature review, the test framework. The analysis of the first sub- questions is addressed in Chapter 4, and the results of the analysis are presented in Chapter 5.

3.3.2 Phase Two

The Rasch measurement model is used in Phase Two to explore the last three sub-questions. The last three sub-questions guide the research in this phase:

4. Are the items on the mathematics assessment instrument functioning as expected? Specifically do the items fit the requirements of the Rasch measurement model? To address this question, the Rasch analysis was applied to investigate whether the items were consistent with the way other items were functioning on the assessment as a whole. Item fit was used to determine whether the ANA Grade 6 mathematics items contributed to the measurement of one construct.
5. How well are the items distributed along the continuum of the variable? This question is related to the distribution of items along the continuum to investigate the range of person abilities. The alignment of persons and items on the continuum was used to explore whether there were enough ANA Grade 6 mathematics items spread along the continuum, as opposed to clumps of them, and enough spread of ability among persons.
6. How well is the assessment instrument targeted to the abilities of the learners? This question investigated the distribution of items that were equivalent to the

distribution of persons. To address this question, item difficulty and item analysis served the purpose to explore the difficulty level of the ANA Grade 6 mathematics items and the person ability.

The last three sub-questions are addressed with reference to the application of the Rasch measurement model. The analysis of the first three sub- questions is addressed in Chapter 4 and the results of the analysis (4^{th}, 5^{th} and 6^{th} question) are presented in Chapter 5.

3.4 RESEARCH METHODS

In this section, the assessment instrument, data collection and research procedure for the study are described and discussed. A content analysis was conducted (applying professional judgement) of the Annual National Assessment instrument, followed by a statistical item analysis applying the Rasch measurement model.

The mapping of the items was done by the researcher with the assistance of the mathematics teacher and the subject advisor. The profile of the mathematics teacher who assisted the researcher is a follows: (gender: male, age: 47, and had experience in teaching the subject mathematics for 25 years). The profile of the subject advisor who assisted the researcher is as follows: (gender: male, age 42, and had experience in the subject mathematics for 22 years). The responsibilities of the mathematics teacher and the subject advisor were to verify the placing of the items in the relevant mathematics topic category and to give professional judgement on the quality of the items when requested by the researcher.

The role of the data manager was to assist the researcher in cleaning the data captured and importing the data to the RUMM 2020 program for Rasch measurement model analysis.

These analytical methods were utilised to determine the quality of the ANA Grade 6 mathematics instrument, in terms of its construct validity, content validity and reliability in order to evaluate the potential inferences that could have been made and actions taken based upon the mathematics performance of learners in Grade 6 in the GND in the year 2012 of the resulting data.

The main research question for this study is: To what extent does the 2012 Annual National Assessment Grade 6 mathematics instrument provide meaningful information for making appropriate interpretations at district level?

These analytical methods were applied in order to answer the main research question with the aim to determine the extent of the validity and reliability of the ANA Grade 6 mathematics instrument. The information gathered from this study will then be used to establish the development of a theoretically based, empirically tested instrument to measure mathematical knowledge and skills for the Grade 6 learners. At district level the information gathered from the instrument could be utilised for district interventions.

Arising from the literature review and the conceptual framework adapted for this study, a number of specific sub-questions for this study are fore grounded.

In summary, this study is divided into two phases. In Phase One, the focus is on the first three sub-questions to analyse the content of the study. The Content Analysis Framework (developed from the Annual National Assessment Grade 6 Mathematics Test Framework, 2012) and the Conceptual framework model supported the analysis of the content in this study. The second three sub-questions were answered through the Rasch analysis.

3.5 POPULATION AND SAMPLE

Gauteng Province comprises 15 districts, including the focus district in this study, namely Gauteng North District. Gauteng Province is one of the nine provinces in South Africa. Gauteng North District is the smallest district in the Gauteng Province. It is situated in the east of Pretoria; it is responsible mainly for the schools in the following areas: Ekangala, Bronkhorstspruit, and Cullinan. The GND has only 33 primary schools compared to other districts and it was one of the lowest performing districts in the 2012 ANA Grade 6 mathematics. The districts in Gauteng Province that performed slightly better than the Gauteng North District in the 2012 ANA Grade 6 mathematics are listed in Table 9, as well as the two districts that performed lower than GND. The reason for selecting this district was that it was conveniently situated and accessible for the researcher.

Table 9

Gauteng Province 2012 ANA Grade 6 Mathematics Average Achievements by District

(adapted from Annual National Assessment Report, DBE, 2012)

	District	**Average Percentage**
1.	Ekurhuleni South	37,0 %
2.	Sedibeng East	34,6 %
3.	Ekurhuleni North	33,4 %
4.	Tshwane South	32,4 %
5.	Gauteng East	31,8 %
6.	Johannesburg North	31,8 %
7.	Johannesburg East	31,6 %
8.	Gauteng West	30,8 %
9.	Johannesburg West	29,8 %
10.	Tshwane North	29,5 %
11.	Tshwane West	28,7 %
12.	Sedibeng West	28,5 %
13.	**Gauteng North**	**27,2 %**
14.	Johannesburg Central	26,9 %
15.	Johannesburg South	24,9 %

In South Africa the total number of learners registered for the Annual National Assessment Grade 6 mathematics was 944 397 in 2012. In Gauteng Province a total of 148 637 learners, were registered for the ANA. In Gauteng North District, which is the focus district in this study a total of 2 976 learners were registered for the ANA Grade 6 mathematics in the year 2012. The selected population of the study was the cohort of Gauteng Grade 6 learners that participated in 2012 ANA, which is the exit grade of the three-year Intermediate Phase in the primary school level and the exit point to the Senior Phase.

From the target population of 33 public primary schools in GND, five schools were drawn of which the data for the study was sampled. The 33 primary schools were ranked based on average percentage for the school, from higher to lowest. The five selected schools were purposively sampled across the range in proficiency levels, and represented higher, middle and lower performance. Only data from learners who wrote the test in English was selected for this study. This was done in order to avoid translation and because English medium schools are in the majority in GND.

The total number of 546 learner scripts was sampled for this study representing all Grade 6 learners in those five schools. The scripts had been moderated at school level by the relevant Head of Department. Although this sample is slightly on the low side, it is sufficient for the purpose of exploring the quality and the usefulness for the particular district (GND). Given the exploratory nature of the study and the spread of achievement results, a small sample of schools with 546 learners was sufficient to conduct the Rasch item analysis (Cohen, et al. 2008) as well as the content analysis. The sample is entirely adequate for establishing the validity of the test. However, the size and the nature of the sample require that one triangulates any findings, for example the view that the test was not properly targeted for this cohort may be checked against the views of experienced subject advisors and teachers. The results are not generalizable outside of the district but may provide insights.

3.6 ANA GRADE 6 MATHEMATICS ASSESSMENT INSTRUMENT

The instrument used for the study is the existing 2012 Annual National Assessment Grade 6 mathematics test and the marking memorandum (refer to Appendices C page 178, and D page 180). Only the English instrument and the corresponding data were analysed. The ANA instrument was compiled according to the DBE Test Framework (DBE, 2011). The cognitive levels in the mathematics curriculum encompass; a) factual knowledge of the basic concepts, b) the application of concepts, and c) the ability to do non-routine problem solving (DBE, 2012) refer to Table 11 (on page 96). The content areas include the Grade 6 mathematics topics; a) Number, Operations and Relationships, b) Patterns, Functions and Algebra, c) Shape and Space, d) Measurement and e) Data Handling (DBE, 2012, p. 6) refer to Table 11 (on page 96). Webb (1992) notes that a design framework provides the means to map what a learner knows within the content domain in order to track the competencies within that domain. Phase one addressed the first three research questions of this study.

The Department of Basic Education specified the required percentage on the Test Framework for the cognitive levels, and the content areas, which the Annual National Assessment should comply with. The mapped ANA items percentages were calculated to verify whether they met the requirement specified on the DBE test framework. All the mathematics topics that fall within the five content areas mentioned in the literature review were represented in the items of the 2012 ANA mathematics. The weightings allocated to these topics are reflected in Table 11 (Chapter 4, page 96).

Table 10 depicts the classification that was used for the content analysis codes. The first row specifies the difficulty level of the ANA Grade 6 mathematics assessment instruments; the second row explains the cognitive levels applied on the ANA Grade 6 mathematics assessment instrument. The figures represented in the table were derived from the DBE test framework (2012).

Table 10

Classification Table Codes for ANA Grade 6 Mathematics (Adapted from DBE Test Framework, 2012)

Difficulty levels	**Easy** **E**	**Moderate** **M**	**Difficult** **D**
% of Items	25	60	15
Cognitive levels	**Knowledge of basic concepts** **K**	**Application of concepts** **A**	**Non-routine problem solving** **N**
% of Items	25	60	15

The mapped DBE Test Framework is discussed in Chapter 4, and for the number of items topics per topic refer (Table 11 page 96).

The 2012 ANA Grade 6 mathematics instrument comprised 52 items. The items included both closed and open-ended questions for a total of 75 marks. A total of eight items were multiple questions, consisting of a question and four options for an answer, namely the three distracters and the correct answer. A total of 44 items were open-ended questions where learners were expected to complete a word or number sentence by circling, drawing, writing only one word or numerical response and one extended response, where learners were required to show their working to get to an answer.

3.7 DATA PROCESSING AND ANALYSIS

The re-marked scripts were obtained from the selected schools. The marks per item were captured on an Excel spread- sheet. The codes used for the capturing of ANA test items were determined by the researcher and the data manager (refer to APPENDIX E, page 181). The data capturing process was completed in three weeks. The cleaning of data was done by the researcher and the data manager. The captured data on excel spread sheet was imported into the RUMM 2020 program for Rasch measurement model analysis.

The problem with the results of ANA identified in this study is that most of the Grade 6 learners are unable to meet the basic minimum standards in Mathematics as discussed in Chapter 1. While there are a number of factors that might have impacted on this low performance, the validity of the instrument may be the factor. Content analysis of the instrument and the statistical item analysis using the application of the Rasch measurement model are presented in this section.

The first analysis of the data of the 2012 ANA Grade 6 mathematics was done by the Department of Basic Education after the administration of the 2012 Annual National Assessment cycle.

The "Department of Basic Education compiled the Diagnostic Report on the 2012 ANA, in which an analysis was made to provide evidence that would inform and direct appropriate interventions for (a) teaching and learning, (b) the management of curriculum implementation by School Management Teams (SMTs), (c) curriculum and management support at district level and (d) resource provision and monitoring at provincial and national levels" (DBE, 2013, p. 3).

The Department of Education explained the "purpose of the Diagnostic Report as follows; to highlight and present to teachers and School Management Teams (SMT) specific areas of Language and Mathematics knowledge and skills in which learners, who participated in the Annual National Assessment 2012, were found to be inadequately equipped" (DBE, 2013, p. 4).

The Department of Education (DBE, 2012) emphasised that the results of Annual National Assessment should:

- Assist in identifying areas where urgent attention is required in order to help improve the learning success levels of learners;
- Provide assistance to provinces to make informed decisions about which schools require urgent action aimed at improving learner performance in these subjects;
- Assist in informing the government and the South African community as to how well the schools are serving the country's children in literacy and numeracy;
- Provide teachers with the information about the language and mathematics capabilities of learners and thereby help them to make informed decisions when planning the year's programme;
- Inform individual teachers about how close or far they are towards or away from realising the target goals they have set for their teaching.

3.7.1 Content Analysis

In the first phase, the Test Framework (refer to Table 11 page 96) was used to map the curriculum elements covered in the ANA Grade 6 mathematics instrument of the year 2012. In the initial stage of this research the mathematics content of the ANA Grade 6 mathematics instrument was analysed.

Content analysis can be used to analyse educational document, by extracting numerical data from word based data; it involves the coding of data, the categorising of information, data comparison, concluding and drawing theoretical conclusions from the text (Miles & Huberman, 1994). In this study, content analysis involved the mapping of the assessment items to the curriculum in order to verify difficulty level, content and scope, and the cognitive level covered was done by the researcher with the assistance and expertise of the subject specialist for mathematics at district and the grade 6 mathematics teacher. The difficulty level categorises on easy, moderate (easy and difficult) and difficult parts of the items.

The Test Framework is divided into columns with a specified criterion used to identify cognitive domain, competencies, cognitive level and difficulty level. The 2012 ANA Grade 6 items were placed against the framework to justify the mathematics content that was used in the instrument (see Table 11 page 96 in Chapter 4).

3.7.2 Item Analysis Using the Rasch Measurement Model

In the second phase, the Rasch measurement model was applied using basic principles to verify the validity and reliability of the instrument within this particular frame of reference. The program RUMM 2020 was used to analyse the 2012 ANA Grade 6 mathematics data. The program allowed a detailed investigation from different perspectives. The following outputs were provided:

- Item Statistics (the location, residual, chi-square and probability)
- Person-Item Location Distribution (the overview of the ANA test as a whole)
- Person-Item Map (item difficulty and person ability)

The Rasch analysis aligns item difficulty and person proficiency on the same scale. This analysis provided greater insight into the functioning of each item and into the performance of mathematics of subgroups of Grade 6 learners in GND. This stage addressed the last three sub-research questions of this study.

It is important to discuss first the background to the Rasch measurement model and to understand the theory behind it.

3.7.2.1 The Rasch Measurement Theory

In the 1920s and 1930s the Danish government, implemented a reading intervention for children experiencing reading difficulties. Georg Rasch was requested to determine the effectiveness of the reading strategy, but in the process was confronted by the situation that learners at different grade levels had been assessed and different tests had been used.

The Rasch model was then developed in the 1950s by George Rasch with the purpose of solving this educational dilemma, namely; that of measuring the reading progress of learners over using different assessments. His solution resulted in the alignment of all learners and all tests on the same scale in order to examine both the assessment instruments that were being used to assess those learners and the learners themselves. His solution is reported in the text “Probabilistic Models for Some Intelligence and Attainment Test” (Rasch, 1960).

Over the past 60 years, researchers have been using the Rasch measurement theory (RMT) to investigate the functioning of assessment instruments as a whole as well as to preview the individual items guided by the model’s assumptions or requirements, and to provide information on existing assessment instruments. Equating and linking of assessments over time, initiated in the 1950s, are examples of the immense power of the Rasch model (Rasch, 1960/1980).

In this study, the Rasch measurement model was applied and some aspects of the data output were investigated. The Rasch measurement model is based on a small set of requirements. According to Bond and Fox (2007, p. 26),

> "the basic premises are that a) each person is characterised by ability, and b) each item by a difficulty which c) can be expressed by numbers along one line, and finally d) from the difference between the numbers, the probability of observing any particular scored response can be computed".

A central feature of the Rasch measurement model is a table of expected probabilities designed to address the key question: When a person with ability (number of test items correct) encounters difficulty (number of persons who succeeded on the item), what is the likelihood that this person gets this item correct? The likelihood or probability of a person getting the item correct depends on the difference between what the person is able to do and the difficulty of the item (Bond & Fox, 2007). Therefore, the Rasch measurement model item-person maps report the relations only between the two key variables of the estimated difficulty of the item and the ability of the person (Bond & Fox, 2007).

The implication is that if person A has a greater ability than person B, then person A should be more likely to solve any type of item. Similarly if item 1 is more difficult than item 2, it means that any person has the probability to solve item 2 more easily than item 1 (Rasch, 1960).

The ability of an individual can be estimated by using the total number of correct responses and calculating each person's successes in relation to their failures. The difficulty of an item can be estimated by looking at the total number of correct responses and by calculating the success against the failures within each item (Bond & Fox, 2007).

The basic idea that underpins the Rasch measurement model, by which the ability of a person is measured by performance on an assessment, is that appropriate assessment items are combined in a coherent way in the design of the assessment (Bond & Fox, 2007). Therefore, every item is relevant to the aims of the assessment and it provides a part of the information about the ability that is measured (Bond & Fox, 2007).

A well-designed assessment instrument can therefore give the assessor a detailed image of an individual's performance. Not only that, but the instrument can also inform external stakeholders on the condition of the educational system (Long, 2015). Hence requirements of the Rasch measurement model are resonating with the requirements of a good educational practice (Dunne et al., 2012).

Of importance to this study is to recognise that the DBE uses the Rasch model practically in ANAs. The Rasch is used in the test construction and refining stage and with test developers but it does not reach as far as the districts. The value of providing information and outputs from the Rasch analysis could provide districts with meaningful information.

It is of importance to highlight other countries that use the Rasch analysis for the purpose of this study; for example, Australia. Australian education departments use the Rasch model to refine the assessment instrument used to measure the mathematical knowledge and skills and to determine the reliability of the instrument (Clements, Sarama and Liu, 2015).

3.7.2.2 Basic Principles of the Rasch Measurement Model

The Rasch measurement model has the basic principles that need to be adhered to in the process of analysing the assessment instrument. These principles are applicable to this study; therefore, it is important to discuss each principle:

Unidimensionality. It is considered a good measurement practice because it allows the "estimation of one ability at a time and it will not intentionally or unintentionally confuse two or more attributes into a score" (Bond & Fox, 2007, p. 120). The assessment is considered as unidimensional, when it assesses one construct for example; one subject such mathematics or language. Multidimensionality will be found when there is another influence apart from mathematics proficiency evident, for example; language proficiency. Though this aspect is generally regarded as an important part of the Rasch measurement analysis it was not considered within the scope of this study.

Item Fit. Bond and Fox described Item Fit as a "quality control principle used to help decide whether the actual item and person performances are close enough to the Rasch measurement model's requirements to be counted as linear interval scale measures" (Bond & Fox, 2007, p. 48). "Rasch analysis provides fit statistics designed to aid the investigator" in making a number of interrelated decisions about the assessment instrument data (Smith, 2004, p. 9). That way, "the quantitative and qualitative aspects of investigation get a chance to work together to improve assessment instrument design" (Bond & Fox, 2007, p. 35).

It is important to note that, unlike classical test theory models and other IRT models, Rasch models are prescriptive rather than descriptive (Bond & Fox, 2007). This implies that an ideal item function is calculated and the actual item performance is then compared with this ideal model to estimate how effective the item is functioning at that particular level of difficulty. Item fit is an indication of the degree of correspondence between ideal item performance and actual item performance.

One way of estimating this fit is by looking at the magnitude of disagreement between predictions made by the model and the actual item scores for learners at specific ability levels. This disagreement may indicate an error and can "be referred to as residual fit measures in the context of the Rasch measurement model" (Bond & Fox, 2007, p. 116). Significant disagreement between the model and the data should be cause for concern, as this is an indication that the item is not performing as expected. These items are referred to as misfit items. Misfit items generally compromise test performance.

Two types of misfit are over-discrimination and under-discrimination.

Over-discrimination. Over discrimination occurs when learners with higher ability levels get the item right more often than the model predicts, but learners at lower ability levels get the item wrong more often than the model predicts. This occurrence is indicated by items that have large negative residual values and a relatively steep curve, indicating that the discrimination ability is too high. Though very high discrimination ability appears to be a desirable trait for an item to have, it could be an indication that the item may be disadvantaging low ability learners unnecessarily. In other words, ability level alone is not enough to account for the differential performance of the item.

A central in Rasch measurement theory is that "the model is based on the idea that useful measurement involves examination of only one human attribute at a time on a hierarchical "more than/ less than" line of inquiry" (Bond & Fox, 2007, p. 122). According to Masters (1988), over-discrimination in educational settings should be investigated as this could be an indication that subtle differences, such as learning opportunities or teaching content, are causing the differences between low and high ability learners.

Under-discrimination. Under-discrimination occurs when learners with higher ability levels get a question wrong more often than the model predicts, but learners at the lower ability levels get it right more often than the model predicts. Under-discrimination is problematic since it implies that, as the ability of the learner increases, the probability of gaining a higher score does not increase proportionally, as we would expect.

Items that exhibit under-discrimination indicate that an item might have been guessed or it might be testing a construct that does not fit well into the overall framework because learner ability only partially accounts for item performance. Under-discriminating items are usually indicated by large positive residual values and a relatively flat curve, indicating low discrimination ability.

Reliability. The reliability of the assessment instrument is illustrated using the Item Statistics, Person-item Location Distribution and the Person-Item Map (Bond & Fox, 2007). In Rasch measurement theory the Person Separation Index (PSI) provides a measure of reliability, indicating the robustness of the test. The PSI, specific to the Rasch model, contrasts the variance among the proficiency estimates of the learner cohort as a whole relative to the error variance within each person (Andrich, 1982). It provides a measure of internal consistency by providing an indicator of the separation of persons relative to the difficulty of the item. The equivalent in traditional test theory is the Kuder-Richardson 20, or Cronbach's Alpha, which provides a measure of the internal consistency of the items, but does not provide a measure of person consistency relative to items (Andrich, 1982).

Item Statistics. These statistics illustrate the location, residual; chi square and probability of items (refer to APPENDIX F, page 182).

The Person-Item Location Distribution. This distribution illustrates the relationship between grouped learner ability and item difficulties. The "Person-Item Location Distribution item mean is set at 0, and the item difficulties are calibrated with reference to the mean" (Bond & Fox, 2007, p. 116).

Person-Item Map. The Person-Item Map represents on the same scale both item difficulty and person proficiency (Bond & Fox, 2007). The data presented on the Person-Item Location Distribution is provided in a different form on the Person-Item Map. Again the item mean is set at 0, and the item difficulties are calibrated with reference to the mean are the same.

Item difficulty. The Rasch measurement model specifically addresses the conception of order, an idea fundamental to any account of developing human ability and basic to the idea of measuring more ability (Bond & Fox, 2007). In the Rasch measurement model, "performances are attributed relative importance in proportion to the position they hold on the measurement continuum" (Bond & Fox, 2007, p. 120). The learner ability location is defined as the point at which learners have a 0.5 probability (50% chance) of responding correctly to the item.

The Rasch measurement model provides "indices that help the investigator determine whether there are enough items spread along the continuum, as opposed to clumps of them, and enough spread of ability (more difficult and easier) among persons" (Bond and Fox, 2007, p. 40).

Smith (2004) described the Rasch measurement model purpose as "the model providing a direct estimate of the modelled variance for each estimate of person's ability and an item's difficulty providing a quantification of the precision of every person measure and item difficulty which can be used to describe the range within which item's true difficulty or person's ability falls, that is, both person reliability and item reliability" (Smith, 2004, p. 96). Person reliability can be described as the reproducibility of the person ordering whereas; item reliability is described as the true item variance divided by the observed item variance.

In conclusion, the Rasch principles were used to explore the test data in phase two of this study which focused on addressing the last three sub-research questions.

3.8 METHODOLOGICAL NORMS

In order to understand the methodological norms in this study, the validity that "deals with the fact, whether the content of the assessment is a fair reflection of the content and the aims of the subject" (Black, 2000, p. 412) is considered. As mentioned in Chapter 2, Anderson and Morgan (2008, p. 151) defined the two concepts as follows: "test validity refers to a broad concept that involves making appropriate interpretations and uses scores or test information while test reliability refers to the extent to which the evidence collected is sufficient for generalisations".

Here the researcher drew information in the conceptual framework for the concept validity in this study.

Validity and reliability checks will be applied to the 2012 ANA Grade 6 mathematics instrument. Checking the validity means making sure that the test is measuring what it is supposed to measure. Reliability can be seen as a pre-condition for validity, that is, in order to be valid a test should be reliable, and on the other hand reliability is no guarantee that the test will also be valid (Scheerens et al., 2003).

3.8.1 Validity

In the conceptual framework, validity centres on the extent to which meaningful and appropriate conclusions are made on the assessment instrument used in this study. Messick (1989, 1995) viewed the concept of validity as an "overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores" (p. 110). The interpretations of assessment scores are "valid to the extent that these interpretations are supported by appropriate evidence" (Scheerens et al., 2003, 97).

The "test development manager is responsible for coordinating with a nominated reference group of subject specialists, such as curriculum specialists, to ensure that the items represent an adequate sampling of a curriculum or construct" (Anderson & Morgan, 2009, p. 15). Furthermore "the expert group should determine if the test represents an adequate coverage of a specified subject (such as Grade 6 mathematics) and should consider if performance on the test provides adequate evidence of learner achievement in the subject area" (Anderson & Morgan, 2009, p. 15). One facet of validity is the "extent to which the content of the test is representative of the curriculum or construct that is being measured" (Anderson & Morgan, 2009, p. 16).

Whilst there are more than twenty different forms of validity (Cohen et al., 2008) in educational research, there are four commonly used types of validity that are of great importance. In this study, the focus will only be on two types considered to be important; a) content validity and b) construct validity, looking at the performance of a sample of learners on the ANA Grade 6 mathematics, reflecting learners' skills in numerical calculation. Content and construct validity have two aspects: content relevance and representativeness (Scheerens et al., 2003).

A judgement of content relevance and representativeness is based on "specification of boundaries and structure of the domain tested; the domain can be explored, using curriculum analysis and test analysis" (Scheerens et al., 2003, p. 100).

Content validity. In the conceptual framework, content validity is determined when considering the proportion of content, which ANA is representing, of the intended Grade 6 mathematics scope and relevance to the curriculum. This "validity is a theoretical concept which focuses on the extent to which the assessment instrument shows evidence of fair and comprehensive coverage of the domain of items that it is supposed to cover" (Scheerens et al., 2003, p. 100).

Messick explains that, "content validity is based on expert judgements about the relevance of the test content to the content of a particular behavioural domain of interest and about the representativeness with which item content covers that domain" (Messick, 1989, p. 8). The content validity of a national assessment test is checked by analysing whether the set of subject- matter components (that is, the test items) adequately represent all subject-matter elements that together constitute the subject-matter domain in question (Scheerens et al., 2003).

In this case the focus is on the analysis of the content validity of the ANA Grade 6 mathematics in particular.

Construct validity. In the conceptual framework, the 2012 ANA Grade 6 mathematics assessment instrument is analysed to check whether it measured what it was intended to measure. "Construct validity, is based on an integration of any evidence that bears on the interpretation or meaning of the test scores" (Messick, 1989, p. 8). The "construct validity of a national assessment test is checked by making sure that the test is measuring the construct it is supposed to measure, irrespective of the test format" (Scheerens et al., 2003, p. 20).

Messick further elaborated that; "construct validity also subsumes content relevance and representativeness as well as criterion-relatedness because such information about the range and limits of content coverage and about specific criterion behaviours predicted by, test scores clearly contributes to score interpretation" (Messick, 1989, p. 9).

In this circumstance the emphasis is on the analysis of the construct validity and the interpretation of the ANA of the Grade 6 mathematics instrument.

3.8.2 Reliability

In the conceptual framework, the aim of the reliability analysis is quantification of the consistency and inconsistency of the learner performance on the test. Reliability is the extent to which a test measures consistently, "a measure of reliability is an indicator of the consistency of the test results" (Anderson & Morgan, 2009, p. 74). Inconsistency stems from factors influencing the outcome of the measurement that are not part of the construct of interest; for instance, test length or the assessment procedure (Scheerens et al., 2003).

Reliability depends on the "quality of test items, the test itself, the way the tests were administered, the characteristics of the group of learners (such as the effort they make while taking the national assessment tests), and the quality of scoring of the test items" (Anderson & Morgan, 2009, p. 76). In addition "test reliability indicators range from 0 to 1, where 0 represents a test in which learners' responses are entirely inconsistent (for example, a test, where all learners guess randomly on all items), and 1 represents a test that measures a domain with perfect consistency" (Anderson & Morgan, 2009, p. 77).

Therefore "this information will provide a measure of the internal consistency of the test items, and that this approach assumes the selected items measure a single construct, such as mathematics ability" (Scheerens et al., 2003, p. 101).

Reliability can be defined as the degree to which an assessment instrument produces stable and consistent results (Meyer et. al, 2010). The Scottish Qualifications Authority (SQA, 1997) asserted that reliable assessment "gives consistent results on different occasions with different candidates and different assessors".

Cannon and Newble (2002) suggested that reliability can be improved by:

- "Ensuring that the questions are clear and suitable for the level of learners;
- Ensuring time limits are realistic;
- Ensuring that instructions are simple, clear and unambiguous;
- Developing high quality marking schemes" (Cannon & Newble, 2002, p. 22).

Reliability might be compromised in an assessment by the "instrument variables; poor domain sampling, errors in sampling tasks, the realism of the tasks and the relatedness to the experience of the assessor, poor question items, length of the assessment, mechanical errors, scoring errors and computer errors" (Cohen et al., 2008, p. 108). Similarly, the reliability of the assessment is affected by the "objectivity of the scoring of the test, the specificity of the items, the difficulty level and the test length" (Scheerens et al. 2003, p. 100).

To address reliability of the instrument in this study, the Rasch measurement model served the purpose of exploring the extent that items would retain the same ordering of item difficulties, given a different, but equivalent sample of the Gauteng North District.

The description of the major statistical concepts was investigated. The learner ability location was defined as the point at which learners have a 0.5 (50% chance) of responding correctly to the item at the same location. This statistic indicated that the item was experienced very differently by high proficiency learners to low proficiency learners.

The process of Rasch analysis aligned both the item difficulty and person proficiency on the same scale. This relationship was depicted in the Person-Item Map and Person-Item Location Distribution. The Person-Item Map (see Figure 4 in Chapter 4) represents, on the same scale, both item difficulty and person proficiency. As with the Person-Item Location Distribution (see Figure 5 page 124, in Chapter 4), the item mean was set at a 0, and the item difficulties are calibrated with reference to the mean.

Thus, the aim of this study was to gauge the reliability of the instrument and the validity of the interpretations based on its results, using those results to improve the quality, design of the assessment instrument.

3.9 ETHICAL CONSIDERATIONS

For this study, permission was requested from the University of Pretoria (UP) in the Faculty of Education to undertake this study. Ethical clearance was sought and approval granted by the University of Pretoria Ethics Committee.

Permission was also requested from the Gauteng Department of Education (Head Office Gauteng) to gain access to the Annual National Assessment Grade 6 mathematics 2012 and to use as sample of the Gauteng North District 2012 ANA Grade 6 mathematics results for this study (see Appendix A, page 163).

The research methodology and the procedures employed in this study were discussed in detail. The population and sample in the study of the Gauteng North District were highlighted. The research paradigm, research design and research methods employed in this study were discussed. Validity and reliability were explained. Methodological norms featured in this study were highlighted. The ethical clearance for conducting this study was specified. The next chapter 4 focuses on interpretation of data and presentation.

CHAPTER 4 – ANALYSIS AND PRESENTATION OF DATA

Introduction

The purpose of the secondary data analysis for this study is to describe, explore and report on the quality of the Annual National Assessment mathematics instrument, including its design with specific reference to the 2012 Grade 6 mathematics and the validity of using the ANA data for district level intervention. The overall results of the 2012 ANA Grade 6 mathematics learner performance in South Africa as described by the DBE earlier in Chapter 1, displayed the following averages attained by the learners writing the test; the national average was at 27.0 %, the Gauteng provincial average at 30.9 % and the Gauteng North District average at 27.2 % (see Table 2 page 25).

Primarily, two data analysis methods were used in order to address the research questions adequately. The first data analysis method required the analysis of the Annual National Assessment Grade 6 mathematics instrument using a Test Framework which had been informed by the conceptual framework and by the broader literature reviewed. The conceptual framework provided the structure according to which curriculum coverage, mathematics content and cognitive levels could be explored and analysed in a systematic way (see Figure 1 page 61). The second data analysis method required the statistical item analysis of the ANA Grade 6 mathematics instrument using the Rasch measurement model. These analytical methods utilised in this study are presented as follows:

4.1 ITEM CURRICULUM MATCH

The Department of Basic Education reports that the 2012 ANA Grade 6 mathematics assessment framework was used to arrange the items according to the content area specified in NCS. The Test Framework was developed by the DBE selected group of test developers who are also experienced subject experts, as indicated earlier.

The DBE test framework provided the parameters for the ANA test developers. The framework "was also given to schools and teachers to inform them regarding the scope of the ANA and the expected coverage of the curriculum" (DBE, 2012, p. 7). In this study the 2012 ANA Grade 6 mathematics items were mapped on the mentioned DBE test framework to enable the content analysis (see Table 11 page 96).

The 2012 ANA Grade 6 mathematics test had a total of 52 items as explained in Chapter 3. These items were mapped on the framework in order to answer phase one of the research (see Table 11). The mapping of the items was done to confirm whether the 2012 ANA Grade 6 mathematics instrument adhered to the DBE requirements on specified percentages in terms of the difficulty level, the cognitive levels and the content area.

Table 11[2]

2012 ANA Grade 6 Mathematics Items Mapped on DBE (2012) Test Framework (DBE, 2012)

Content Area	Assessment standard focus	Rasch Analysis Label	**Skills/competencies assessed** Testing whether the learner is able to…	Suggested Number of questions	Cognitive Level	Difficulty Level	Weighting of Content Area.
Numbers, Operations and Relationships	Recognising, classifying and representing numbers	1 9	Count forwards and backwards in decimals to at least 2 decimal places.	1	K	M	50%
		10	Recognise and represent whole numbers to at least 9-digits.	1	K	E	
			▪ Recognise and represent: 0 in terms of its additive property; 1 in terms of its multiplicative property. ▪ Recognise and use the commutative, associative and distributive properties with whole numbers.	1 (with sub-questions)	K	E	
		11	Recognise multiples of whole numbers.	1	K	M	
		12	Recognise factors of whole numbers.	1	K	M	
		13	Recognise prime numbers to at least 100.	1	K	M	
		15	Recognise the place value of whole numbers to at least 9-digit numbers.	1			

[2] The study used the original DBE Test Framework to map the ANA items back onto this.

Content Area	Assessment standard focus	Rasch Analysis Label	Skills/competencies assessed Testing whether the learner is able to…	Suggested Number of questions	Cognitive Level	Difficulty Level	Weighting of Content Area.
					K	E	
		45	Recognise and use equivalent forms of numbers including: common fractions (1 digit or 2 digit denominators); decimal fractions and percentages.	1 (with sub-questions)	K	M	
	Different calculation types involving numbers	2 16	Round off to the nearest 5, 10, 100 or 1000.	1(with sub-questions)	A	M	
		40 41 28	Add and subtract whole numbers ○ in columns ○ by building up and breaking down numbers ○ by rounding off and compensating numbers.	1	A	M	
		29	▪ Add and subtract common fractions in which one denominator is a multiple of another. ▪ Add and subtract mixed numbers.	1	A	M	
		27 46	Multiply a 4-digit number by a 3-digit number (in columns).	1	A	M	
			Divide a 4-digit number by a 3-digit number (Including long division).	1	A	D	
		18 17	▪ Add and subtract positive decimal numbers with at least 2 decimal places. ▪ Multiply decimal fractions by 10 and 100.	1	A	M	

Content Area	Assessment standard focus	Rasch Analysis Label	Skills/competencies assessed Testing whether the learner is able to…	Suggested Number of questions	Cognitive Level	Difficulty Level	Weighting of Content Area.
	Properties of numbers	14	▪ Recognise and use divisibility rules for 2, 3, 4, 5, 10, 100, and 1000. ▪ Find the factors of 2-digit and 3-digit numbers. *	1	K * A	E M	
		4 6	Use multiple operations on whole numbers (with or without brackets).	2 (with sub-questions)	A	D	
	Problem solving involving whole numbers, fractions, and decimal fractions.	 47	▪ Compare 2 or more quantities of the same kind (ratio). ▪ Compare 2 quantities of different kinds (rate). ▪ Grouping and sharing with remainders. ▪ Solve problems in context involving common and decimal fractions. ▪ Find fractions of whole numbers. ▪ Find percentages of whole numbers. ▪ Solve problems in financial context (profit and loss, simple budgets, and discount).	2 (with sub-questions)	N	M	
Patterns, Functions and Algebra	Patterns	43	Investigate and extend geometric patterns looking for a relationship.	1	K	M / D	10%
		21	Investigate and extend numeric patterns not limited to sequences involving a constant difference or ratio and describe observed relationships or rules in own rules.	1	A	D	
	Equivalent representations	19 20	Determine equivalence of different descriptions of the same relationships or rule presented: ▪ in a flow diagram (input values, output values and rule). ▪ in a table. ▪ by a number sentence.	1	A	M	

Content Area	**Assessment standard focus**	**Rasch Analysis Label**	**Skills/competencies assessed** Testing whether the learner is able to…	**Suggested Number of questions**	**Cognitive Level**	**Difficulty Level**	**Weighting of Content Area.**
	Number sentences	3 22	Solve or complete number sentences (by trial and improvement, inspection) and check solutions by substitution.	1	N	D	
Shape and Space	2-D shapes and 3-D objects	7 48	▪ Recognise and name regular and irregular polygons including triangles, quadrilaterals, pentagons, hexagons, heptagons and octagons. ▪ Sort and compare 2-D shapes in terms of number of sides, lengths of sides and size of angles (right angles, smaller / greater than right angles). ▪ Recognise similarities and differences between rectangles and parallelograms. ***** ▪ Recognise and name 3-D objects including rectangular prisms, cubes, tetrahedrons and other pyramids. ▪ Recognise similarities and differences between tetrahedrons and other pyramids. ***** ▪ Compare 3-D objects in terms of the number and shape of faces, the number of vertices and the number of edges.	2 (with sub-questions)	K * A	E M	15%
	Symmetry	24	Recognise, draw and describe line(s) of symmetry in 2-D shapes.	1	K	E	
	Transformations	23	▪ Use the properties of rotations, reflections and translations to describe transformations of 2-D shapes and 3-D objects. ▪ Draw enlargements and reductions of 2-D shapes (at least triangles and quadrilaterals) on grid paper to compare their size and shape.	1 (with sub-questions)	K	M	
	Position	8	Interpret sketches of simple 3-D objects from different positions.	1	A	D	

Content Area	**Assessment standard focus**	**Rasch Analysis Label**	**Skills/competencies assessed** Testing whether the learner is able to…	**Suggested Number of questions**	**Cognitive Level**	**Difficulty Level**	**Weighting of Content Area.**
		28 29 30	Locate position on a coded grid, describe how to move between positions and recognise maps as grids.	1 (with sub-questions)	K	E	
Measurement	Time	42 44	Solve problems involving calculation and conversion between appropriate time units (including time zones and differences).	1 (with sub-questions)	N	D	15%
	Units and instruments	5 49 31 32	Solve problems involving calculating and converting between appropriate SI units including: • Mass using grams and kilograms. • Capacity using millilitres, litres and kilolitres. • Length using millimetres, centimetres, metres and kilometres. • Temperature using degree Celsius scale. • Conversions should include fraction form and decimal form (to 2 decimal places).	1(with sub-questions for each SI unit) (Temperature can be assessed as a separate question)	N	M	
	Perimeter, Area and Volume	25 26 27	▪ Determine the perimeter of 2-D shapes using rulers. ▪ Calculate the area of rectangles and squares using square grids. ▪ Determine the capacity / volume of 3-D objects (by packing or filling them) in order to develop rules for calculating volumes of rectangular prisms.	1	K	M	
Data Handling	Collecting and organising data	33 34	▪ Organise and record data using tally marks and tables. ▪ Distinguish between samples and populations.	1 (in the form of a table)	K	E	10%

Content Area	Assessment standard focus	Rasch Analysis Label	Skills/competencies assessed Testing whether the learner is able to…	Suggested Number of questions	Cognitive Level	Difficulty Level	Weighting of Content Area.
	Representing and interpreting data	39	Determine the mode and median of ungrouped numerical data.	1 (with sub-questions)	A	M	
		35 36 37	▪ Draw a variety of graphs including pictographs with a many-to-one correspondence, bar graphs and double bar graphs to display and interpret information. ▪ Interpret data represented in pie charts and bar graphs.	1(with sub-questions)	A	M	
	Chance		Predict the likelihood of events from daily life on a scale from 'impossible' to 'certain.	1	N	E	
		38	▪ List possible outcomes for simple experiments (including tossing a coin, rolling a die, and spinning a spinner). ▪ Count the frequency of actual outcomes for a series of trials.	1	N	D	
							100%

For the purpose of this investigation, the 52 items of the 2012 ANA Grade 6 mathematics were successfully mapped on the DBE framework, according to the allocated weights per content area (see column 3 of Table 11). At this point the specialist opinion of two external advisers was sought by the researcher. Firstly, the Senior Education Specialist for the subject mathematics in Gauteng North District was consulted; secondly, an experienced Grade 6 mathematics teacher in Gauteng North District was consulted to verify the researchers mapping of items on the DBE test framework. This action was undertaken to ensure validity and reliability within this investigation.

The results of the Test Framework, conceptual framework and literature reviewed are presented in this chapter per research question.

4.2 PRESENTATION OF PHASE ONES RESULTS

This section presents the results from the content analysis based on the research data.

4.2.1 Purpose of the Annual National Assessment Grade 6

In answer to the first research question, “What is the purpose of the Annual National Assessment Grade 6 instrument”? The broad and overarching purpose of the ANA “is to make a notable contribution towards better learning in schools, by serving broadly as a systemic tool and second, as a diagnostic tool in identifying areas of strength and weakness in teaching and learning” (DBE, 2013, p. 36).

In order to meet the broader purpose of the ANA, four specific objectives were set for this systemic assessment. This study does not set out to investigate the extent to which these objectives have been met; however for the sake of clarity these objectives are briefly discussed.

The first specific objective of the ANA outlined by the DBE is to expose teachers to best practices in assessment: Annual National Assessment for example caters for a range of cognitive skills levels and abilities of learners according to the prescribed cognitive levels derived from Bloom's Taxonomy and explained in the DBE documents, therefore, it aims to provide a valid measure of individual learners' attainment of knowledge in the subject, as explained in Chapter 1. The implication is that the ANA is supposed to assist teachers to teach and test according to these cognitive levels. Even the DBE framework for the ANA caters for the range of cognitive levels and can serve as an example of best practice to teachers. The range of skills provides "all teachers with a clearer idea of how to develop their own assessments at critical points in the school year" (DBE, 2013, p. 37).

In the present content analysis, it was calculated through the given formula, (refer to APPENDIX I, page 185) that the 2012 ANA Grade 6 resonates with this objective because it is evident through the mapped items that the test had catered for a range of cognitive skills levels and abilities of learners according to the prescribed cognitive levels (Blooms Taxonomy). The compliance with the prescribed cognitive levels is used here as an example of how the ANA test can serve to assist teachers in developing their own assessments.

The second specific objective "targeting interventions to the schools that need those most:" this objective is linked to the first objective of the ANA: ANA serves "as a diagnostic tool identifying areas of strength and weakness in teaching and learning" (DBE, 2012, p. 4). This outcome is dependent on the targeting and the feedback process. The third objective "giving schools the opportunity to pride themselves in their own improvement": also depends on the relationship to teaching and learning, the appropriate content and the quality of the instrument. As stated this objective will "allow schools to take pride in knowing that the efforts to improve the teaching and learning in the school is producing the desired outcomes" (DBE, 2012, p. 4). The average percentage of the 2012 ANA Grade 6 mathematics compared to the ANA 2011 had decreased (see Table 2 page 25 in Chapter 1). The fourth objective was "giving parents better information on the education of their children". It is potentially achieved after the administrations of the ANA results are disseminated to the relevant stakeholders, including parents. The quality of this information is dependent on the quality of the instrument.

The last three ANA objectives are mentioned here but they are not relevant for the focus of the study.

In terms of the broader purpose and the specific objectives of the 2012 ANA Grade 6 mathematics assessment instrument, a single aspect falls within the scope of this investigation, namely, whether the test served as an example of best practice to teachers. Therefore, the instrument was analysed only in terms of its content.

Referring to the conceptual framework, the purpose of assessment is explained, is, a) assessment promotes, assists and improves learning of mathematics; b) assessment informs programmes of teaching and learning of mathematics; and c) assessment provides information to relevant stakeholders about progress and achievement of learners (see Figure 1 page 61). The content analysis conducted in this study has the potential validity of using the ANA data for district level intervention to further inform programmes of teaching and learning of mathematics, especially including exposing teachers to best practices in assessment.

4.2.2 Criteria for Good Assessment Design

The next question asks "To what extent does the instrument fulfil the criteria for good assessment design". Conditions to ensure quality of the national assessments are outlined in Chapter 2. The four conditions are applicable to the 2012 ANA Grade 6 mathematics instrument because it is a national assessment:

Firstly, representation of knowledge and skills: According to Kellaghan et al. (2009, p. 42), "a test can measure only part of the knowledge and skills specified in a curriculum or construct (for example, mathematics), and it is important to ensure that it provides an adequate representation of that knowledge and those skills". This aspect was addressed through the mapping of the ANA items on the DBE framework (see Table 11 page 96).

Further, as reported in the literature review, Kellaghan et al., (2009, p. 43) explained that to "secure adequate representation of a domain, construct, objectives or subdomains (for example, content strands or skills in mathematics), a test should contain an adequate number of items, the small number of items in some national assessments might not provide adequate coverage". There is also the danger that too many items might tire the test respondents. From the content analysis it can be deduced that the 2012 ANA Grade 6 mathematics contained an adequate amount of items. This judgement, however, depends on a number of factors concluded in Chapter 5, that the assessment had too many items.

Secondly, the level of development of the learner: Kellaghan et al. (2009) explained that a "test should assess knowledge and skills at a level that is appropriate" for learners. This lack of appropriate level "is fairly common in developing countries, that is, the test will be too difficult for lower-achieving learners and therefore will fail to register their achievement" (Kellaghan et al., 2009, p. 44). The solution to this problem "lies in taking into account in the test development and the actual achievement" of learners in schools not just the standard of the intended curriculum (Kellaghan et al., 2009, p. 44).

As becomes clear through the Rasch analysis (see Figure 4), the 2012 ANA Grade 6 mathematics had items that were far too difficult for the level of development of this particular cohort of learners. In Table 19 the possible reasons for this difficulty level are set out, with specific reference to instructions, language and phrasing. This finding pertains to the next condition for designing an assessment.

Thirdly, language usage is an important condition in the designing of assessment. “A test designed to assess learners’ achievement in mathematics should not contain so much complex language that performance on it depends rather on the ability of learners’ to read rather than on their proficiency in … mathematics” (Kellaghan et al., 2009, p. 45).

This language problem occurs when learners responding to a test, possess different levels “of skill in reading, this will be the case when the language of the test differs from that which they normally use for some learners” (Kellaghan et al., 2009, p. 45).

Within the content analysis it had to be kept in mind that the sample in this investigation did not have English as home language, a factor that had influenced the performance. For this reason, the instructions, language and phrasing of the items were scrutinised for any potential to obstruct understanding applying professional judgement (see Table 13 page 114).

Lastly, assessment results to monitor change. In order to use the results, the assessment instruments must be comparable (Kellaghan et. al., 2009). “In order to achieve this, the monitoring of the results over time of the same test, which should be kept secure between administrations, should be used” (Anderson & Morgan, 2009, p. 58).

4.2.3 Construct Validity and Content Validity

The next question “To what extent does the instrument have construct validity and content validity”. In answering the question, the content analysis was used to analyse the construct validity and content validity of the ANA Grade 6 mathematics instrument. According to Babbie (2007, p. 23), “content validity is used as a theoretical concept which focuses on the extent to which an assessment instrument shows evidence of fair and comprehensive coverage of the domain of items that it is supposed to cover”. According to this explanation, it is important for the sake of content validity to look at: 1) relevance to the curriculum; 2) focus on what was taught; 3) comprehensive content coverage; 4) the proportion of the scope of learning; and 5) sampled potential content (Babbie, 2007). The analysis of the 2012 ANA Grade 6 mathematics was focused on these aspects, as follows:

- Relevance to the NCS mathematics policy document;
- Content areas as specified in the NCS;
- Item distribution according to the content areas;
- Proportions of items according to the allocated weightings in the NCS; and
- Representation of sampled items.

The structure of the 2012 ANA Grade 6 mathematics instrument was analysed to ascertain whether the scope of content area/topics was adhered to. According to the DBE framework, items for the ANA Grade 6 mathematics were distributed as set out as follows; 50% for ‘number operations and relationships’, 10% for ‘patterns, functions and algebra’, 15% for ‘shape and space’, 15% for ‘measurement’, and 10% for ‘data handling’ (see Table 11 page 96). This formula was used to calculate the items percentage in order to verify correspondence with the given content area:

$$\frac{\text{Item Number}}{\text{Total Number Items}} \times \frac{100\ \%}{}$$

$$\frac{22}{52} \times \frac{100}{} = 42,\ 3\%$$

After the calculations the result of this Test Framework presents the content areas of the 2012 ANA Grade 6 mathematics items as follows:

- Items falling under 'Number Operations and Relationships' comprised 42, 3% (DBE suggested 50 %)
- Items falling under 'Patterns Functions and Algebra' comprised 16 % (DBE suggested 10%)
- Items falling under 'Shape and Space' comprised 15, 3% (DBE suggested 15%)
- Items falling under 'Measurement' comprised 15, 3 % (DBE suggested 15%)
- Items falling under 'Data Handling comprised 13, 4 % (DBE suggested 10%)

The data upon which these finding are based is represented in a table format in APPENDIX I (page 185).

Although it is recognised that a systemic assessment has to make choices and cannot cover everything the results of the content analysis indicate that the ANA had partly managed to achieve evidence of fair and comprehensive coverage of the domain items that it is supposed to cover. The results of the content analysis confirm that the distribution of items on the Test Framework for the ANA Grade 6 mathematics content had discrepancies when compared with according to the DBE specifications.

Construct validity refers to "the degree to which a test or other measure assesses the underlying theoretical construct it is supposed to measure" (Pope, 2009, p. 1). According to Poe (2009), in order to demonstrate construct validity, evidence that the test measures what it intends to measure (in this case mathematics skill) as well as evidence that the test does not measure irrelevant attributes (reading ability) are both required. The construct validity was achieved to some extent, because some items on the ANA had language problems and some items had long instructions that might have confused the learners.

As explained in Chapter 1, this study was motivated by the perceived failure of the ANA to reach the broad purpose as set out above. This investigation is therefore focused on the analysis of the content of the mentioned test, in order to explore its strengths and weaknesses.

In the literature review Chapter 2, Scheerens et al., (2003, p. 23) explained that the "national assessments tests in a particular subject need not be administered each year; each subject may be tested every six years". He further explains that, "the information from the assessment programmes can lead to adaptations in the curriculum in the sense of goals (standards) or means (curriculum contents) and all conditions that have impact on the performance in a particular subject (e.g. teacher training in the particular subject matter area)" Scheerens et al., (2003, p. 23). Locally, the ANA is administered annually; it is argued that teachers are not given enough time to adapt to the curriculum and focus on strategies that are meant to remedy conditions that have contributed to performance.

The DBE reports that the "ANA results provide valuable information for the system to benchmark performance in both Literacy and Numeracy and have provided a measure so as to track progress over the years towards the achievement of the set targets" (DBE, 2012, p. 24). Over the past years the ANA has not yet managed to reach the target set by the DBE, even the results of the focus year were very low.

Since it was impossible, given the design of the ANA, which does not allow comparability across the years, this research focuses on a single year's test.

As set out in the previous paragraphs, there were discrepancies as far as the DBE and the 2012 ANA Grade 6 mathematics coverage of content areas.

Within the conceptual framework the matter of validity (construct and content) is classified under the structure of the assessment instrument. The other aspects pertaining to the validity of the instrument namely, designing assessment, comparability of quality standards, the purpose of assessment, and the outcome of assessment results were not addressed in this question (refer to Table 12 page 111 and Table 13 page 114). Therefore, as mentioned in the conceptual framework, instructions, written language and curriculum coverage were analysed (refer to Figure 1 page 61). This, in short, was used to determine whether the instrument assessed what it was supposed to assess.

Once the process of the mapping of assessment items to the Test Framework was completed, two problem categories emerged: the first was the mathematical correctness of items, and the second, the language clarity of the statements. The two problem categories form the part of the content analysis and are discussed below.

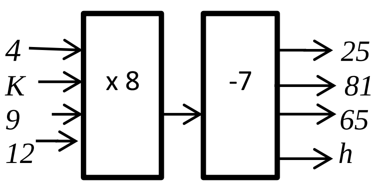
4.3 MATHEMATICAL CORRECTNESS OF THE 2012 ANA GRADE 6 MATHEMATICS

According to the observation, problems and challenges were identified with the 2012 ANA Grade 6. There were slight discrepancies found in terms of cognitive levels and content areas coverage and also learners might have had problems in attempting to respond to the mathematics instructions given. Some of the problems on items were perceived to be with the clarity of instructions. Furthermore, items where most learners underperformed were identified, that pointed to the quality of the questions as problematic.

In the discussion following, the Test Label has been assigned to the item number in the test. For the analysis the items were ordered differently. This ordering of items is labelled the Rasch Analysis Label (RA Label). Both the original test labels and the items numbered for Rasch analysis purpose will be used simultaneously for data presentation. The item numbers pertaining to the Rasch analysis will be put in brackets. The following Test Labels (TA) 18, 20, 2, 4.1, 7.6, 9.1, 9.2, 10, 24.1, 27, 7.4, 8.1 and 8.2, with corresponding Rasch analysis (RA) label (7, 8, 9, 11, 18, 19, 20, 21, 31, 38, 50, 51 & 52) exhibited problems. Table 12 presents the items with mathematical problems. In this table the quality of the instrument and of the items is analysed in terms of the reasons identified to pose problems for learners answering the questions.

Table 12

Items Posing Mathematical Concerns on the 2012 ANA Grade 6 Mathematics

RA Label	Description	Possible reasons for problem
Item (7)	**Test label 18** *Circle the letter of the net that cannot be folded into a cube.* *(Refer to Appendices C)*	It is most likely that the item was difficult because the learners could not recognise a cube (professional reason in conversation with the subject adviser).
Item (8)	**Test label 20** *Six cubes are used to build the 3-D figure shown below. The view from the right is given next to it. Circle the letter showing the front view.*	The learners might have had a challenge of recognising the properties of 3D shapes, making this item difficult.
Item (9)	**Test label 2** *Write down the next decimal number.* 0,79 ; 0,76 ; 0,73 ; 0,7 ; ______	The item was found to be difficult because the last number had only one decimal while preceding ones had two decimals.
Item (11)	**Test label 14.1** *From a given group of numbers, choose ONE number that fits a multiple of 11* __________ **36 120 19** **8 143**	On this item, the problem was with the factor of 13. At Grade 6 level, learners are conversant with the factors up to 12. The factor 13 (as in 11x13 = 143) was outside of this range and therefore unexpected.
Item (18)	**Test label 7.6** *Calculate the answer.* *24, 37 + 346, 83 =* __________. Item 18 required learners to add two numbers with two decimal places.	This error is a conceptual error on the part of the learners; many of them may have been confused with alignment of decimal places. Learners unable to add decimal numbers with 2 decimal places.
Item (19) and (20)	**Test label 9.1 & 9.2** *Find the value of h and k in the flow diagram.*  *h* = ___________________ *K* = ____________________	Item 19 and 20 were about a flow diagram. The instruction was fine to the level of the learners. The problem could be with mathematical content, especially, if the learner has difficulties in understanding flow diagrams and anything about the values.

RA Label	Description	Possible reasons for problem
Item (21)	**Test label 10** *Write down the next number in the sequence.* 1; 1; 2 ; 3 ; 5 ; 8 ; __________	The problem may have been that this is a Fibonacci sequence, which has an unexpected pattern. Learners might have previously only encountered sequences whose first differences were constant.
Item (31)	**Test label 24.1** *What is the mass indicated on the scale above? ______________ (Refer to Appendices C)*	It is more likely that learners were unable to recognise the mass in grams.
Item (38)	**Test label 27** *A bag contains black and white marbles. The "probability of taking a white marble out of the bag" above, without looking in the bag, is* _________________.	On this item, the problem was with the term "probability", it was not to the level of understanding of the learners especially those that started English in Grade 4 during NCS, and it might not have been understood. The bag contains black marbles and white marbles not black and white marbles. The statement regardless of the language requires that learners think carefully about the probability of taking out a particular number. Another possibility that needs to be taken into consideration is that the learners may have "guessed" the answer. Item 38 does not differentiate between learners of any ability level and even displays a trend of negative discrimination, where higher ability levels are associated with a decreased probability to respond correctly to the item.
Item (50)	**Test label 7.4** *Calculate the answer:* *2 067 x 189 =*______________.	Learners may have had a challenge of multiplication with 3 digits with carrying over answers. The nature of the question allows the learners with higher ability to be more succesful at answering the question correctly. The probability of learners answering a question like item 50 succesfully is much lower because of the level of skills needed. Learners at lower ability levels might have struggled with this particular item.
Item (51) and (52)	**Test label 8.1 & 8.2** *Calculate the answer and write your number as a mixed number:* $8\frac{3}{10} - 4\frac{1}{5} =$ _______________ *Calculate the answer and write your number as a mixed number:* $5\frac{1}{2} + 3\frac{1}{8} + 4\frac{3}{4} =$ ___________	On Item 51 and Item 52 the challenge might have been with subtracting mixed numbers.. Both Item 51 and Item 52 were problematic due to the fact that lower ability learners have a very low probability of responding successfully to the item, despite the fact that the item is only perceived to be moderately difficult.

In view of the above analysis it can be concluded that the unfamiliarity with some of the concepts listed above, and unclear question formulation could have influenced learners' performance. A hypothesis to be further explored is that there was lack of curriculum coverage although this was out of scope of the study.

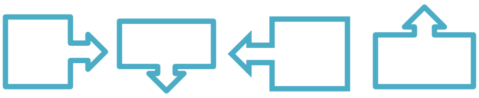
4.4 LANGUAGE CLARITY OF THE ITEMS OF THE 2012 ANA GRADE 6 MATHEMATICS

It has been argued that mother tongue "is the best medium of instruction at school because it is the language the child knows well, and in which they can express meanings" (Pattanayak, 2003, p. 6). The level of language used in the assessment instrument was investigated. The focus in this regard was firstly on the presentation of the instructions and the phrasing of questions. Secondly, it was on appropriateness of language in respect of the language of learners at Grade 6 level. Professional judgement was used to verify the readability of items on the ANA Grade 6 mathematics instrument. The following items showed language issues Test label; 1.3, 11, 16, 13, and 26.3 and with corresponding Rasch analysis label (3, 22, 23, 37, and 47) showed language problems (see Table 13).

Table 13

Items Identified as Having Language Problems on the ANA Grade 6 Mathematics

Rasch Analysis Label	**Description**	**Possible reasons**
Item (3)	**Test label 1.3** *Select a number sentence to match the following statement: Seven less than a certain number m is equal to twelve.*	Learners may be confused because they recognise *m* as a letter of the alphabet. The recognition of *m* as a number requires Algebraic knowledge.
Item (22)	**Test label 11** *Calculate the value of x if x – 41 = 12* ____________________. Item 22 is overdiscriminating (discussed in Chapter 3).	Algebraic expressions difficult terminology for learners who might have not being exposed to solving for *x* in algebraic expressions.

Rasch Analysis Label	Description	Possible reasons
Item (23)	**Test label 16** *Which kind of transformation is shown here?* 	It is more likely that the term "transformation" was difficult for the learners to comprehend. Item 23 needed learners to describe the transformation that they see. This item required learners to have a skill of recognising both movement and direction.
Item (37)	**Test label 26.3** *What did Suzi's part of the pizza cost?* ________ *(Refer to Appendices C for the Pizza)*	It could be possible that learners might not be able to interpret a problem stated in words, as a mathematical calculation.
Item (47)	**Test label 13** *Mr Msebenzi buys 480 sweets for R30, 00. He repacks the sweets into packets of 24 each. He sells the packets for R2, 50 each. How much profit will he make if he sells all the sweets?*	The instruction is very long, which raises reading complexity and it could be confusing for learners to attempt answering the item.

From the above it can be concluded that language issues on the 2012 ANA Grade 6 mathematics might have contributed to the low performance.

In order to establish the internal coherence of the 2012 ANA Grade 6 mathematics assessment instrument, a Rasch analysis was conducted. With this analysis, "the item difficulty and the person proficiency are located on the same scale" (Bond & Fox, 2007, p. 100). This alignment of person and item enabled an investigation of the instrument itself and each of the items.

4.5 PRESENTATION OF PHASE TWO RESULTS

The summary statistics from the Rasch analysis, including item and person means and standard deviations, point to the appropriateness of the test for this cohort. The results from the five schools, ranging from high performance within the one district to low performance, indicate that the test was too difficult to provide optimum information on this cohort. We note that the item mean is set at zero. The difficulty of each of the items is calibrated in relation to the mean. The person mean for this sample is -1.550 logits, which means that this test is too hard. The standard deviation of the items (1.752) shows the spread to be large. However the learner spread is closer together (0.976).

Table 14

The Person Separation Index (Adapted from Andrich, 1982)

	Item	**Person**
Mean	0.000	-1.550
Standard Deviation	1.752	0.976
Person Separation Index	0.857	

In Rasch measurement theory the Person Separation Index (PSI) provides a measure of reliability, indicating the robustness of the test. The PSI, specific to the Rasch model, contrasts the variance among the proficiency estimates of the learner cohort as a whole relative to the error variance within each person (Andrich, 1982). It provides a measure of internal consistency by providing an indicator of the separation of persons relative to the difficulty of the item. The equivalent in traditional test theory is the Kuder-Richardson 20, or Cronbach's Alpha, which provides a measure of the internal consistency of the items, but does not provide a measure of person consistency relative to items (Andrich, 1982).

The PSI (0.857) shows the test to have good spread and therefore to be providing meaningful information (see Table 14). However this claim is mitigated when considering other factors such as targeting.

The four basic principles for Rasch measurement model (Item Fit, Location Distribution and Person-Item Overview, and Item Difficulty) are used to answer the last three research questions of this study. The results of the Rasch analysis are presented per research question.

4.5.1 Fit to the Model

This section deals with the fourth research question, which is concerned with Item Fit. This question investigated whether the assessment instrument functioned as expected, specifically whether the items fit the requirements of the Rasch model. Item fit is an indication of the degree of correspondence between ideal item performance and actual item performance. One way of estimating this fit is by looking at the magnitude of disagreement between predictions made by the model and the actual item scores for learners at specific ability levels.

This disagreement can be thought of as a form of error and is referred to as residual fit measures in the context of the Rasch analysis. Significant disagreement between the model and the data should be cause for concern, as this is an indication that the item is not performing as expected (Bond & Fox, 2007). The results of item fit in this study for this specific research question is presented by explaining the two types of misfit: over-discrimination and under- discrimination. As revealed by the analysis, Table 15 presents five items listed as over- discrimination on the 2012 ANA Grade 6 mathematics for the following Test labels; 6, 9.2, 9.1, 21.2 and 7.4 and with corresponding Rasch analysis label (16, 20, 19, 29 and 50).

Table 15

Over- Discrimination Items for Gauteng North District

Seq	Item	Type	Location	SE	**Residual**	DF	ChiSq	DF	Prob
16	I0016	Poly	-0.343	0.104	**-4.053**	524.63	32.708	8	0.000069
20	I0020	Poly	1.065	0.151	**-3.283**	530.49	24.425	8	0.001946
19	I0019	Poly	0.38	0.122	**-3.25**	531.47	20.803	8	0.00769
29	I0029	Poly	-0.221	0.106	**-3.004**	529.51	29.013	8	0.000315
50	I0050	Poly	0.379	0.07	**-2.982**	530.49	16.713	8	0.033243

Figure 2 presents Test label 1.6 (Rasch analysis label 16) in a graph as one example out of five of over-discrimination item.

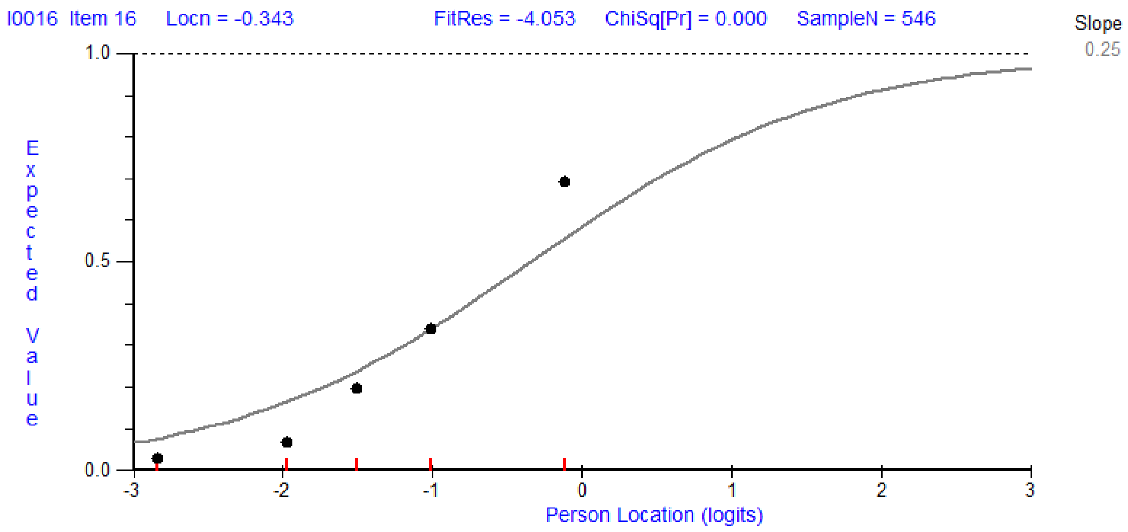


Figure 2. Test label 1.6 (Rasch analysis label16): Over- discrimination for Gauteng North District sample.

Test label 1.6 (Rasch analysis label 16) is about rounding off. Rounding off is an important skill that requires understanding of place value. This item differentiates excessively between learners of different ability levels. Rounding off by 5 may not have been the focus in the South African classroom, rounding off by 10 is more common. Learners at ability levels < -1.0 logits appear to be struggling with Item 16.

Higher ability learners may conceptualise understanding of rounding quicker and therefore are more able to respond correctly to this item.

For Test label 1.6 (Rasch analysis label 16), learner at ability levels < -1.0 logits are performing significantly poorer than the model predicts and learners with ability levels > 0 logits are also perfroming significantly higher than the model predicts. The over-discrimination trend in test label 1.6 (Rasch analysis label 16) is problematic due to the fact that lower ability learners (> -1 logits) have a lower than expected probability of responding successfully to the item.

Another type of misfit is identified as under- discriminating. Table 16 present five items listed as under- discriminating on the 2012 Annual National Assessment Grade 6 mathematics, Test label; 14, 21.1, 17, 1.6 and 18 with corresponding Rasch analysis label (43, 28, 24, 6 and 7).

Table 16

Under -Discrimination Items for Gauteng North District

Seq	Item	Type	Location	SE	**Residual**	DF	ChiSq	DF	Prob
-----	-----	------	-------------	-----	**-----------**	-------	--------	----	----------
43	I0043	Poly	-1.469	0.059	**6.595**	508.02	82.415	8	0
28	I0028	Poly	-1.606	0.094	**5.113**	530.49	45.147	8	0
24	I0024	Poly	-0.976	0.097	**4.369**	517.79	60.671	8	0
6	I0006	MC	-1.073	0.098	**4.24**	497.27	36.493	8	0.000014
7	I0007	MC	0.114	0.117	**3.601**	493.36	55.244	8	0

Figure 3 presents Test label 17 (Rasch analysis label 24) as an example of an under-discrimination item.

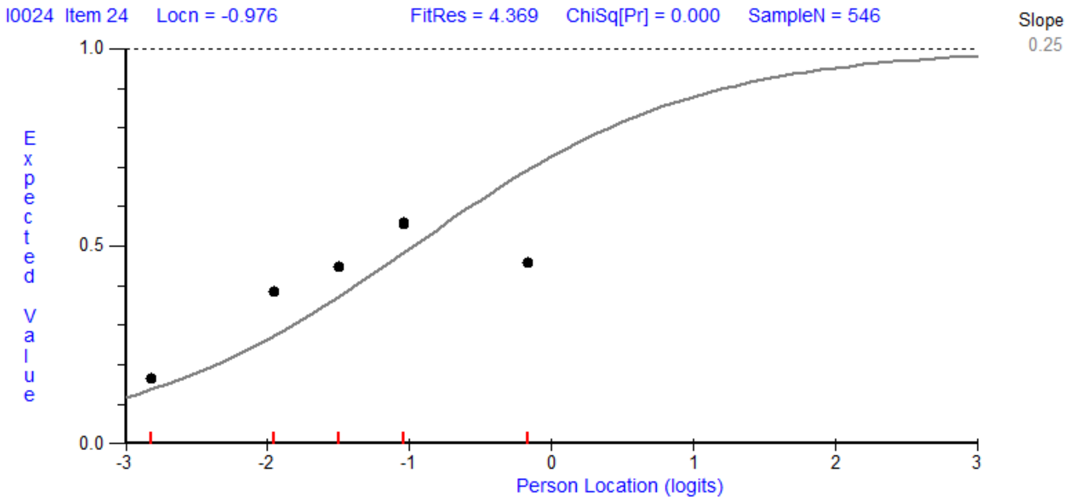


Figure 3. Test label 17 (Rasch analysis label 24): Under discrimination for Gauteng North District sample.

Test label 17 (Rasch analysis label 24) is about lines of symmetry. It does not fit the model well and indicates under-discrimination. This item requires the knowledge and skill of identifying the lines of symmetry. This item might not be contributing in a valuable way to the test due to lower ability level learners who have a higher probability of responding correctly to the item than the model predicts, while learners of higher ability have a lower probability than the model predicts. Test label 17 (Rasch analysis label 24) does not differentiate between learners across all ability levels and is thus not contributing to the assessment in a meaningful way.

Lastly, Test label 8.2 (Rasch analysis label 52) and Test label 23 (Rasch analysis label 49) were identified as having significant disagreement between the model and the data, the items did not perform as expected. These items were deleted for the purposes of the analysis, refer to Table 17.

Table 17

Deleted Items on the ANA Grade 6 Mathematics of Gauteng North District Sample

Rasch Analysis label	Item on ANA Grade 6 mathematics	Misfit	Possible reasons
Item (52)	**Test label 8.2** *Calculate the answer and write your number as a mixed number:* $5\frac{1}{2} + 3\frac{1}{8} + 4\frac{3}{4} =$ ____________	Deleted	Instruction not clearly presented is confusing. Learners not familiar with calculating mixed numbers or they do not know the term mixed numbers. Learners may not have covered the fraction $\frac{1}{8}$ in class. The relationship between the fractions might have been a challenge to learners. No learner got this item correct.
Item (49)	**Test label 23** *Tamara invited 37 friends to her party. Each friend drinks 2 glasses of cool drink. If each glass holds 200ml, "how many 2-litre bottles of cool drink should her mother buy"*	Deleted	Only one learner got this item correct. The phrasing of the question was to the level of learners however, the reading comprehension of learners may not have been on the appropriate level. It is possible that learners are not sufficiently exposed to problem solving in this format. The question might have been totally confusing to learners. Learners may have been confused on how to approach it. The instruction is also excessively long for grade level. The learning of mathematics in everyday life context. The item itself, including its mathematical requirements seems reasonable, therefore the reason for low performance in this item should be sought in other factors, for example; the opportunity to learners was not adequate.

The two items were deleted after a first analysis as they explicitly showed misfit on inspection. For possible reasons identified, (see Table 17).

4.5.2 Item Distribution

This section deals with the fifth research question, which is concerned with the reliability of the 2012 ANA Grade 6 mathematics instrument as far as the distribution of items along the continuum of the variable is concerned. This question deals with the outputs Person-Item Map and Person- Item Location Distribution. These two Rasch outputs depict same information in different format.

The Rasch model incorporated a method for ordering items, for example, (from the sample of school learners in the Gauteng North District) “according to their ability (see Figure 4), and ordering items according to their difficulty” (Bond & Fox, 2007, p. 99).

The Person-Item Map. The Person-Item map represents “both item difficulty and person proficiency on the same scale” (Bond & Fox, 2007, p. 89). The locations of both the items and the persons in terms of logits, or units of mathematical proficiency are displayed on Figure 4. All units are of the same size, the highest values are located at the top of the map and the lowest values are located at the bottom. The spread of items and the spread of learners are depicted. The relation of item to learners is such that deductions can be made on abilities presented on performance and the difficulty level through the learners’ response.

As with the Person-Item Map the item mean is set at 0, and the “item difficulties are calibrated with reference to the mean” (Bond & Fox, 2007, p. 89). The items “at the top end of the scale have greater difficulty, for example Test label 29 (Rasch analysis label 45), while the items at the lower end of the scale, for example Test label 25.1 (Rasch analysis label 33), are of lesser difficulty” (Bond & Fox, 2007, p. 87).

Likewise on the left, learners’ ability is estimated in relation to item difficulty, with learners of greater ability or proficiency at the top of the scale and learners with less proficiency at the bottom of the scale.

The item difficulty estimates are expressed in logits on Figure 4, in which a logit of 0 is arbitrarily set as the average, or mean, of the item difficulty estimates. Items 51, 7 and 8 are close to 0. Test labels 25.2, 25.1, and 1.2 (Rasch analysis label 34, 33 & 2) have negative logit scores; a negative logit indicates that the item was relatively easy. However, Test labels 29, 27 and 26.2 (Rasch analysis label 45, 38 & 36) have positive logit estimates, “a positive logit means that the item was progressively more difficult” (Bond & Fox, 2007).

Person ability is estimated in relation to the item difficulty estimates; for example, the more negative the value, the lower the learner’s ability, and the more positive the value, the higher the ability of learners (Bond & Fox, 2007). Figure 4 indicates that there are no learners with ability levels high enough to justify an item as difficult as Test label 29 (Item 45) (located above 4 logits). Conversely, Test labels 25.1 and 25.2 (Rasch analysis label 33 and 34) appear to be very easy, since the majority of learners have ability estimates above -3 logits, while these two items are both located below -3 logits. However it is important for there to be items targeted at very weak learners.

```
LOCATION        PERSONS     ITEMS [locations]
---------------------------------------------------------------------------------------
 5.0                          |
                              | I0045
                              |
                              |
                              |              A
 4.0                          |
                              |
                              | I0038
                              |
                              |
 3.0                          | I0036
                              |
                              |
                              | I0037
                              | I0017  I0042
 2.0                          |
                              |
                              | I0032  I0046  I0039
                              | I0044
                              | I0047
 1.0                          | I0020
                            o | I0023  I0003  I0021
                           oo |
                            o | I0022  I0011  I0009
                           oo | I0031  I0050  I0019
 0.0                    ooooo | I0051  I0007  I0008
                      ooooooo |
                        ooooo | I0016  I0035  I0013  I0029
                       oooooo | I0001  I0014
                 oooooooooooo | I0004
-1.0             oooooooooooo | I0024  I0005  I0027  I0048
                 oooooooooooo | I0012  I0040  I0018  I0006  I0041
            ooooooooooooooooo |
             oooooooooooooooo | I0043
         oooooooooooooooooooo | I0010  I0028
-2.0               oooooooooo | I0026  I0025  I0015  I0030
             oooooooooooooooo | I0002
                      ooooooo |
                       oooooo |
                     oooooooo |            B
-3.0                    ooooo |
                           oo |
                           oo | I0034
                          ooo |
                              | I0033
-4.0                       oo |
                              |
                              |
                              |
                              |
-5.0                          |
                              |
                              |
                              |
                              |
-6.0                          |
---------------------------------------------------------------------------------------
     o = 3 Persons
---------------------------------------------------------------------------------------
```

Figure 4. The Grade 6 Person-item map for Gauteng North District sample.

Person-Item Location Distribution. The Person-Item Location Distribution in Figure 5 illustrates the relationship between grouped learner ability (red, and above the horizontal axis) and item difficulties (blue, and below the horizontal axis) for the 2012 ANA Grade 6 mathematics cohort (N = 547) of GND sample.

The item difficulty scale ranges from -4 logits (very easy) to 5 logits (very difficult). The logit is an “interval in which the unit intervals between the locations on the person-item map have a consistent value or meaning” (Bond & Fox, 2007, p. 35). The GND sample shows a learner ability scale ranging from -5.5 logits (very low ability) to +1.5 logits (moderately high ability), refer to Figure 5. The mean of the item difficulty is set at zero. The mean person ability estimates are at -1.550 logits.

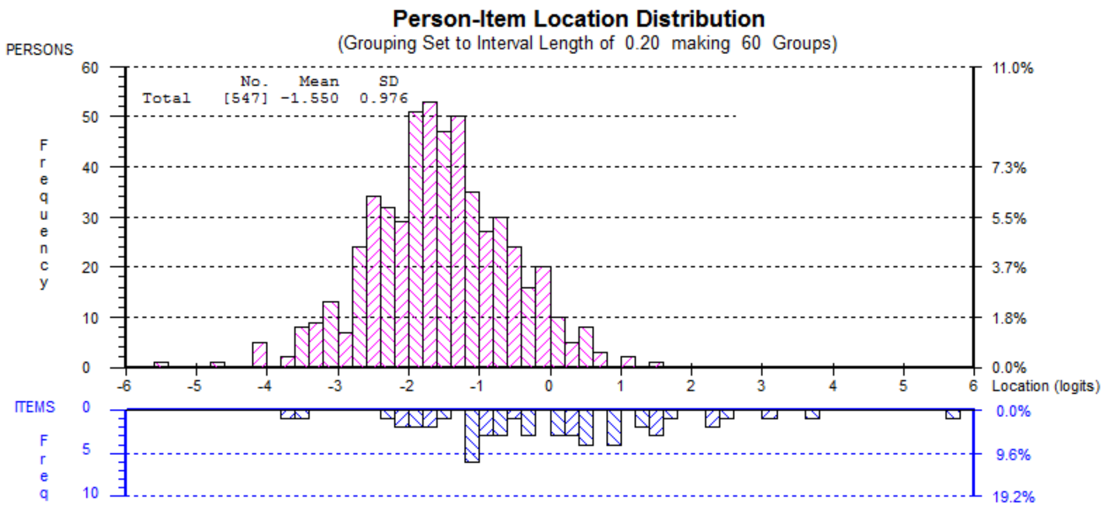


Figure 5. Grade 6 Person-item location distribution for Gauteng North District sample.

4.5.3 Targeting of the Instrument

This section deals with the sixth research question, which is concerned with the extent to which the assessment instrument was targeted to the abilities of the learners. As discussed earlier in Chapter 3 Item Difficulty addresses learner abilities in this section. The following ten items, Test labels; 29, 27, 26.2, 26.3, 12, 7.5, 28, 7.3, 24.2, and 22 with corresponding Rasch analysis label (45, 38, 36, 37, 42, 17, 39, 46, 32 and 44) on the 2012 ANA Grade 6 mathematics were identified as the most difficult items, see Table 18.

Table 18

Ten Most Difficult on the ANA Grade 6 Mathematics

Seq	Item	Type	**Location**	SE	Residual	DF	ChiSq	DF	Prob
-----	-----	------	------------	-----	-----------	-------	--------	----	----------
45	I0045	Poly	**4.843**	0.47	0.006	471.87	4.373	8	0.82203
38	I0038	Poly	**3.77**	0.483	-1.205	528.53	6.826	8	0.555502
36	I0036	Poly	**3.174**	0.363	0.051	531.47	9.143	8	0.330353
37	I0037	Poly	**2.595**	0.279	-0.62	529.51	5.336	8	0.721145
42	I0042	Poly	**2.372**	0.334	-0.082	515.83	6.762	8	0.562508
17	I0017	Poly	**2.298**	0.244	-0.821	528.53	21.376	8	0.006214
39	I0039	Poly	**1.77**	0.197	-1.434	523.65	21.323	8	0.006339
46	I0046	Poly	**1.668**	0.219	-0.586	530.49	1.309	8	0.995446
32	I0032	Poly	**1.642**	0.188	0.389	515.83	11.911	8	0.155243
44	I0044	Poly	**1.418**	0.162	-0.503	521.7	5.777	8	0.672149
Display: 10 MOST DIFFICULT ITEMS									
--									

In Table 18 it can be seen that Test label 29 and Rasch analysis label 45 (4.843 logits), Test label 27 and Rasch analysis label 38 (3.77 logits) and Test label 26.2 and Rasch analysis label 36 (3.174 logits) were the extremely difficult items in the ANA test. Items at the extreme >3 logits, will only be answered correctly by learners with very high mathematical abilities.

The extreme difficulty of Test labels 29, 27 and 26.2 (Rasch analysis label labelled 45, 38 and 36) were investigated earlier (see Tables 14 page 116, and 15 page 117) for Test label 29 and 27 (Rasch analysis label 45 and 38); see the explanation in the summary section and the recommendation for the items, Test label 26.2 Rasch analysis label 36 required only an understanding of a pie chart.

A summary of the 10 most difficult items identified for the Gauteng North District sample are presented in order of difficulty as they appear on Table 19 and possible reasons for each item are stated.

Table 19

Rasch Analysis Items Identified as Most Difficult on the ANA Grade 6 Mathematics

Rasch Analysis Label	**Item on ANA Grade 6 mathematics**	**Possible reasons**
Item (45)	**Test label 29** *Refer to Table 12 for explanation, possible reasons and recommendations.*	
Item (38)	**Test label 27** *Refer to Table11 for explanation, possible reasons and recommendations.*	
Item (36)	**Test label 26.2** *What percentage of the pizza did Peter eat?* ______________ *(Refer to Appendices C for the Pizza)*	It is more likely that the item was difficult because learners could have challenges to calculate percentages of whole numbers.
Item (37)	**Test label 26.3** *What did Suzi's part of the pizza cost?* ________ *(Refer to Appendices C for the Pizza)*	It could be possible that learners might not be able to interpret a problem stated in words, a mathematical calculation.
Item (42)	**Test label 12** *A car travels at 100 "km per hour". "How far will it travel in 45 minutes"*	It is more likely that learners could not be acquainted with calculation and conversion of Time units.
Item (17)	**Test label 7.5** *Calculate the answer.* 5 – 3,64 ________________	Instruction is confusing as to what is expected of learners to do with the 5, 3 and 64. The standard writing procedure for decimal numbers alternate between the use of a full stop and a comma. It might be possible that learners encountered problems by subtracting a Decimal number with 2 Decimal places. This item has a formatting problem.
Item (39)	**Test label 28** *Jacob listed the marks for his mathematics class tests. Jacob's marks:* 20 16 10 3 12 10 11 14 5 19 4 *What is Jacob's median mark?* _______	The learners had difficulty understanding the concept of Median when the data is not ordered.
Item (46)	**Test label 7.3** *Calculate the answer.* 6 960 ÷ 145 ________________	It is possible that learners were not conversant with dividing 4-digit by 3-digit.
Item (32)	**Test label 24.2** *Covert your answer in Question 24.1to kilograms.* __________	The problem might be that learners lack the knowledge of the metric system conversion and application.
LItem (44)	**Test label 22** *Miriam left Durban at 21:45 and arrived in Johannesburg at 04:30 the next day. How long did her journey take?*	It is more likely that learners were not acquainted with calculation and conversion of Time units.

Test labels 25.1, 25.2, 1.2, 19.2, 19.1, 5, 21.3, 3, 21.1 and 14 on the Annual National Assessment were identified as the easiest items; see Table 20 with corresponding Rasch analysis label used in the analysis (33, 34, 2, 26, 25, 15, 30, 10, 28 and 43).

Table 20

Ten Easiest on the ANA Grade 6 Mathematics

33	I0033	Poly	**-3.682**	0.134	-0.55	529.51	15.405	8	0.051736
34	I0034	Poly	**-3.385**	0.123	-0.228	531.47	33.133	8	0.000058
2	I0002	MC	**-2.166**	0.098	2.397	525.6	17.826	8	0.022569
26	I0026	Poly	**-1.954**	0.095	-0.561	532.44	8.22	8	0.412236
25	I0025	Poly	**-1.952**	0.095	-1.819	532.44	16.692	8	0.033478
15	I0015	Poly	**-1.93**	0.095	1.245	532.44	23.914	8	0.002371
30	I0030	Poly	**-1.926**	0.095	-2.605	530.49	20.124	8	0.009876
10	I0010	Poly	**-1.669**	0.094	-2.695	531.47	21.265	8	0.006477
28	I0028	Poly	**-1.606**	0.094	5.113	530.49	45.147	8	0
43	I0043	Poly	**-1.469**	0.059	6.595	508.02	82.415	8	0
-----	-----	------	------------	-----	-----------	-------	--------	----	----------
Seq	Item	Type	**Location**	SE	Residual	DF	ChiSq	DF	Prob
Display: 10 MOST EASIEST ITEMS									

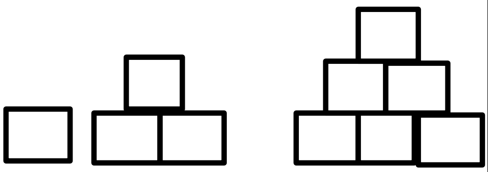
Table 20 indicates that Test label 25.1 Rasch analysis label 33 (-3.682 logits) and Test label 25.2 Rasch analysis label 34 (-3.385) were the easiest item in the test, meaning these items were located at the extreme (< -3 logits). The next easiest item in the ANA test was Test label1.2 Rasch analysis label 2 (-2.166 logits), with item difficulty close to 3 logits. This table indicates that even learners with relatively low ability (< -3.5 logits) levels have a 0.5 probability (50% chance) of getting the easiest items correct.

A summary of the ten easiest items identified in the Gauteng North District sample are presented on Table 21 in order of easiness.

Table 21

Items Identified as Easiest on ANA Grade 6 Mathematics in Order of Easiness

Rasch Analysis Label	Item on ANA Grade 6 mathematics	Possible reasons
Item (33)	**Test label 25.1** *Complete your answers in the table above.* *How many lions did the learners see?* __________ *(Refer to Appendices C for the Table)*	Instruction was clearly laid out. Learners competent to Record Data using Tally marks and tables. The item represents daily live activities of learners.
Item (34)	**Test label 25.2** *Complete your answers in the table above.* *Fill in the correct number of tally marks for the number of giraffes seen.* *(Refer to Appendices C for the Table)*	Phrasing clearly presented. Learners can distinguish between Samples and Populations.
Item (2)	**Test label 1.2** 39 569 *was rounded off to* 40 000. *To which of the following numbers was it rounded off?* *A 5* *B 10* *C 100* *D 1 000*	Instruction clearly presented. Learners competent to Round off numbers.
Item (26)	**Test label 19.2** *On the grid below, each block represents 1 cm by 1 cm. ABCD is a rectangle.* *The length of side AD = _______ cm.* *(Refer to Appendices C for the Grid)*	Phrasing clearly presented. Learners knowledgeable to calculate the Area using grids. Learners familiar with grids.
Item (25)	**Test label 19.1** *On the grid below, each block represents 1 cm by 1 cm. ABCD is a rectangle.* *The length of side AB = _______ cm.* *(Refer to Appendices C for the Grid)*	Same as above (Item 26)

Rasch Analysis Label	Item on ANA Grade 6 mathematics	Possible reasons
Item (15)	**Test label 15** *"What is the value of the underlined digit in the following number"* 367 049 215______________	Language appropriate to the level of learners. Learners skilled to distinguish the value of numbers.
Item (30)	**Test label 21.3** *"Study the grid below and answer the questions that follow".* *In which block will you find a rectangle?* ______________ *(Refer to Appendices C for the Grid)*	Phrasing clearly presented. Learners are familiar with the shape rectangle and can recognise it from the grid. Learners can identify items from the grid without difficulties.
Item (10)	**Test label 3** *Write the number in digits.* *Four million two hundred and eighty three thousand one hundred and sixty-four.* ________________	Learners are knowledgeable on number digits. Instruction clearly laid out. Straightforward not complicated.
Item (28)	**Test label 21.1** *"Study the grid below and answer the questions that follow".* *How many pentagons are there in this grid?* ______________ *(Refer to Appendices C for the Grid)*	Learners familiar with the shape "pentagon". The shape pentagon has five sides and not as complex as the octagon. Language appropriate to level of learners.
Item (43)	**Test label 14** *Ingrid has stacked some blocks according to a certain pattern. Use the same pattern to draw Stack 5.*  *Stack 1 Stack 2 Stack 3 …….* *Stack 5* _________________	Instruction was clear and the language used was to the level of understanding for learners. Learners skilled to draw geometric patterns.

4.6 VALIDITY AND RELIABILITY OF THE 2012 ANA GRADE 6 MATHEMATICS

Validity and reliability were discussed in the literature review in Chapter 2 with the definitions being adopted for this study defined in the literature review. Content analysis and item analysis applying the Rasch measurement model were used to evaluate the validity and reliability of the ANA Grade 6 mathematics instrument in this study.

Possible factors that could have affected the reliability and validity of the 2012 ANA Grade 6 mathematics instrument that contributed to low performance were identified and discussed in the presentation of the results of each research question. The identification of possible factors was attempted to evaluate each item by looking at the content and by studying the item statistics. The reasons for low performance on items were reasonably speculated, however, were not confirmed by a measure such as learner interview during and after the test.

Content analysis and item analysis were used to check the validity and reliability on the 2012 ANA Grade 6 mathematics instrument, mainly for mathematical errors and language problems. The short summary following here serves to remind the reader of the broader discussion previously and to link the results to the issues of validity and reliability.

4.6.1 Validity

The results of the data analysis of the 2012 ANA Grade 6 mathematics serve as evidence to support interpretation of assessment scores and to make meaningful conclusions on this study.

Content validity. The mapped items were used to ascertain construct validity and content validity of the Annual National Assessment Grade 6 mathematics. The results of the content analysis show that the content of the 2012 ANA Grade 6 mathematics was representative of the intended mathematics scope and relevant to the curriculum. The results further showed that there was disagreement in terms of the specified percentages of the domain of items as recommended by the DBE on the ANA instrument.

Construct validity. As far as construct validity is concerned, the results revealed that the 2012 ANA Grade 6 mathematics instrument had a number of items that had language problems and long instructions, meaning the ANA had questions that did not measure the construct "mathematics ability".

4.6.2 Reliability

The Rasch measurement model aims "to provide social scientists with the means to produce genuine interval measures and to monitor the adherence of those scales to scientific measurement principles, so that the Rasch estimates of ability/attitude/difficulty become the data for statistical analysis" (Bond & Fox, 2007, p. 117). The Rasch measurement was particularly suited to this study on the grounds that:

- It is sensitive to the ordered acquisition of skills or abilities under investigations (i.e., it aims at uncovering the order of development or acquisition);
- It is capable of estimating the developmental distances between the ordered skills or persons (i.e., it tells us how much Person T is more developed, more capable, or more rehabilitated than is Person S); and
- Allows us to determine whether the general developmental pattern shown among items and persons is sufficient to account for the pattern of development shown by every item and every person (Bond & Fox, 2007, p. 118).

The summary statistics from the Rasch analysis, including item and person means and standard deviations, point to the appropriateness of the test. The results from the five schools, ranging from high performance within the one district to low performance, indicate that the test was too difficult to provide optimum information on this cohort. We note that the item mean is set at zero. The difficulty of each of the items is calibrated in relation to the mean. The person mean for this sample is -1.550 logits, which means that this test is too hard. The standard deviation of the items (1.752) shows the spread to be large. However the learner spread is closer together (0.976).

To sum up, the analysis of the Annual National Assessment involved the mapping of the content on the Test Framework to match the mathematics curriculum, and the application of the Rasch measurement model to establish authenticity of the data.

Therefore, the results can be used to establish the development of a theoretically based, empirically tested instrument to measure mathematical knowledge and skills for the Grade 6 learners through its submission to the Rasch model. Successfully the data gathered from the instrument can be used to provide meaningful information to refine the ANA Grade 6 mathematics instrument in terms of achieving its validity and high reliability to inform district-level interventions.

In addition Table 22 a summary of the qualitative observation of the instrument is presented. Language, mathematical errors, clarity of instructions, format and phrasing of the items were used as the main categories of the qualitative observation. Verbal descriptions based on the content analysis information are provided in this table. Additionally, possible actions to be taken on the 2012 ANA Grade 6 mathematics instrument are suggested.

Table 22

Summary of the Qualitative Observation of the Instrument

Test label	Item	Qualitative Observation	Content Analysis Information	Follow up Action required
1.1	1	Instruction clearly laid out		None
1.2	2	Instruction clearly laid out		None
1.3	3	Instruction not clearly laid out	Language clarity	Rephrase
1.4	4	Instruction clearly laid out		None
1.5	5	Language appropriate to level		None
1.6	6		Overfit	Rephrase
2	9		Mathematical concern	Curriculum intervention
3	10	Instruction clearly laid out		None
4.1	11		Mathematical concern	Curriculum intervention
4.2	12	Language appropriate to level		None
4.3	13	Language appropriate to level		None
4.4	14	Language appropriate to level		None
5	15	Instruction not clearly presented		None
6	16	Language not appropriate to level	Overfit	Rephrase
7.1	40	Instruction not clearly presented		None
7.2	41	Instruction not clearly presented		None
7.3	46	Instruction not clearly presented		None
7.4	50		Mathematical concern	Curriculum intervention
7.5	17	Instruction not clearly presented		None
7.6	18		Mathematical concern	Curriculum intervention
8.1	51		Mathematical concern	Curriculum intervention
8.2	52		Mathematical concern	Curriculum intervention
9.1	19		Mathematical concern	Curriculum intervention
9.2	20		Mathematical concern	Curriculum intervention
10	21		Mathematical concern	Curriculum intervention
11	22	Language not appropriate to level	Language clarity	Rephrase
12	42	Language appropriate to level		None
13	47	Language appropriate to level	Language clarity	Rephrase
14	43		Overfit	Rephrase
15	48	Language appropriate to level		None
16	23	Language not appropriate to level	Language clarity	Rephrase
17	24		Overfit	Rephrase
18	7		Mathematical concern	Curriculum intervention
19.1	25	Instruction clearly laid out		None
19.2	26	Instruction clearly laid out		None
19.3	27	Instruction clearly laid out		None
20	8		Mathematical concern	Curriculum intervention
21.1	28	Language appropriate to level	Overfit	Rephrase
21.2	29	Phrasing not clear	Overfit	Rephrase
21.3	30			None
22	44	Language appropriate to level		None

Test label	Item	Qualitative Observation	Content Analysis Information	Follow up Action required
23	**49**	Phrasing not clear		None
24.1	**31**		Mathematical concern	Curriculum intervention
24.2	**32**	Instruction clearly laid out		None
25.1	**33**	Instruction clearly laid out		None
25.2	**34**	Instruction clearly laid out		None
26.1	**35**	Language appropriate to level		None
26.2	**36**	Language appropriate to level		None
26.3	**37**	Language not appropriate to level		None
27	**38**		Mathematical concern	Curriculum intervention
28	**39**	Language appropriate to level		None
29	**45**	Phrasing not clear	Language clarity	Rephrase

In summary Chapter 4 presented the results of content analysis and of item analysis, in addition the summary of the qualitative observation of the instrument was provided. Chapter 5 presents conclusions and recommendations.

CHAPTER 5 – CONCLUSIONS AND RECOMMENDATIONS

Introduction

The aim of this study was to investigate the quality of the 2012 Annual National Assessment Grade 6 mathematics instrument, including its design with specific reference to the Gauteng North District. Factors related to the validity and reliability on the quality including design on the ANA mathematics instrument was explored.

Although learner attainment could be influenced by a variety of factors including factors internal to the learner as well as factors originating from characteristics of the assessment environment, the focus area of this study was on investigating the instrument itself. The assessment instrument, the preparation for it, and conditions under which the assessment takes place are recognised as external factors influencing learner attainment.

In this chapter a summary of the research is firstly provided, followed by a summary of results. A reflection on conceptual framework for this study is presented; this is followed by a reflection on the methodology followed in the investigation, the conclusions about the main research questions directed by sub-questions, leading to the conclusions pertaining to that. The limitations of this study and recommendations for further research are subsequently presented.

5.1 SUMMARY OF THE RESEARCH

The quality of the Annual National Assessment Grade 6 mathematics instrument written in the year 2012 was investigated. The ANA results of the Gauteng North District were used for this research. The sample represents a spread of achievement from low to high across 5 schools from learners who wrote the ANA test in English and not in Afrikaans. A secondary analysis design was followed applying mixed methods which followed the analysis of content and the analysis of items. My investigation was divided into two phases: the first phase of the research focused on content analysis of the instrument for which the original DBE ANA Grade 6 mathematics test framework was used to analyse the ANA items. The second phase focused on the application of the Rasch Measurement model which involved the statistical analysis of all the items in the instrument. Learners' scripts were collected from the district office for the purpose of capturing and collecting the data for this study. The learners' response from each question was captured using codes from each script. The data was captured using the RUMM 2020 programme and converted to Microsoft excel to allow statistical analysis through the application of the Rasch measurement model. The results revealed that the ANA Grade 6 mathematics instrument appeared to have some content validity considering that the items were aligned to the Grade 6 mathematics curriculum in an appropriate way. As far as reliability is concerned the investigation revealed that the instrument had a reasonable person separation index, a measure of both item and person reliability.

5.2 SUMMARY OF RESULTS

The summary of results is done here according to the six secondary research questions, as follows:

Question 1: What is the purpose of the Annual National Assessment Grade 6 instrument? The first sub-research question dealt with the exploration of the purpose of the ANA. This question explored the purpose of the original implementation of the DBE with regard to the Grade 6 mathematics instrument, and what it intended to achieve.

Question 2: To what extent does the instrument fulfil the criteria for good assessment design? This question investigated the design of the ANA instrument. The structure of the ANA was analysed in terms of its quality as a national assessment instrument. The ANA instrument was further evaluated in order to determine whether the instrument met the conditions of a good national assessment. To this end, the criteria applied for good assessment design were, representation of knowledge and skills, the level of development of learners, the language use, and assessment results as means of monitoring change (refer to Chapter 2), the conceptual framework in Chapter 3.

Question 3: To what Extent does the Instrument have Construct Validity and Content Validity? This research question intended to examine the ANA instrument in terms of achieving the construct validity and content validity. Further, to ascertain the extent to which the instrument provided valid and reliable information for district use.

Question 4: Are the Items on the Mathematics Assessment Instrument Functioning as Expected? Specifically do the Items Fit the Requirements of the Rasch Measurement Model? This question proposed to investigate the degree of correspondence between ideal item performance and actual item performance. The Rasch measurement model intended to find out whether the items on the instrument were functioning as expected in terms of assessing the difficulty level and to confirm whether the items complied with the requirements of the model or were the items misfitting.

Question 5: How well are the Items Distributed Along the Continuum of the Variable? This sub-research question compared the relationship between learner abilities and item difficulties on the same scale. The Rasch measurement model analysed the results through an interactive process including quantitative method verification, specifically for the district concerned. This was done for the purpose of checking assessment construct validity and reliability and for aligning item difficulty and person proficiency on the same scale. The reliability of instrument was examined to ascertain the extent to which one might expect to render the same results if the instrument were re-administered on the same population and in the same conditions. The person separation index provided a measure here.

Question 6: How well is the Assessment Instrument Targeted to the Abilities of Learners? This question investigated the extent to which the ANA instrument had targeted the abilities of the learners. The Rasch measurement model was applied to investigate in more detail the problem conditions, evident in the ranges of the item difficulty and person ability.

5.3 REFLECTIONS ON THE CONCEPTUAL FRAMEWORK

The conceptual framework, as presented previously in Chapter 3 (Figure 1), was brought forward to this Chapter for the purpose of reflection and discussion.

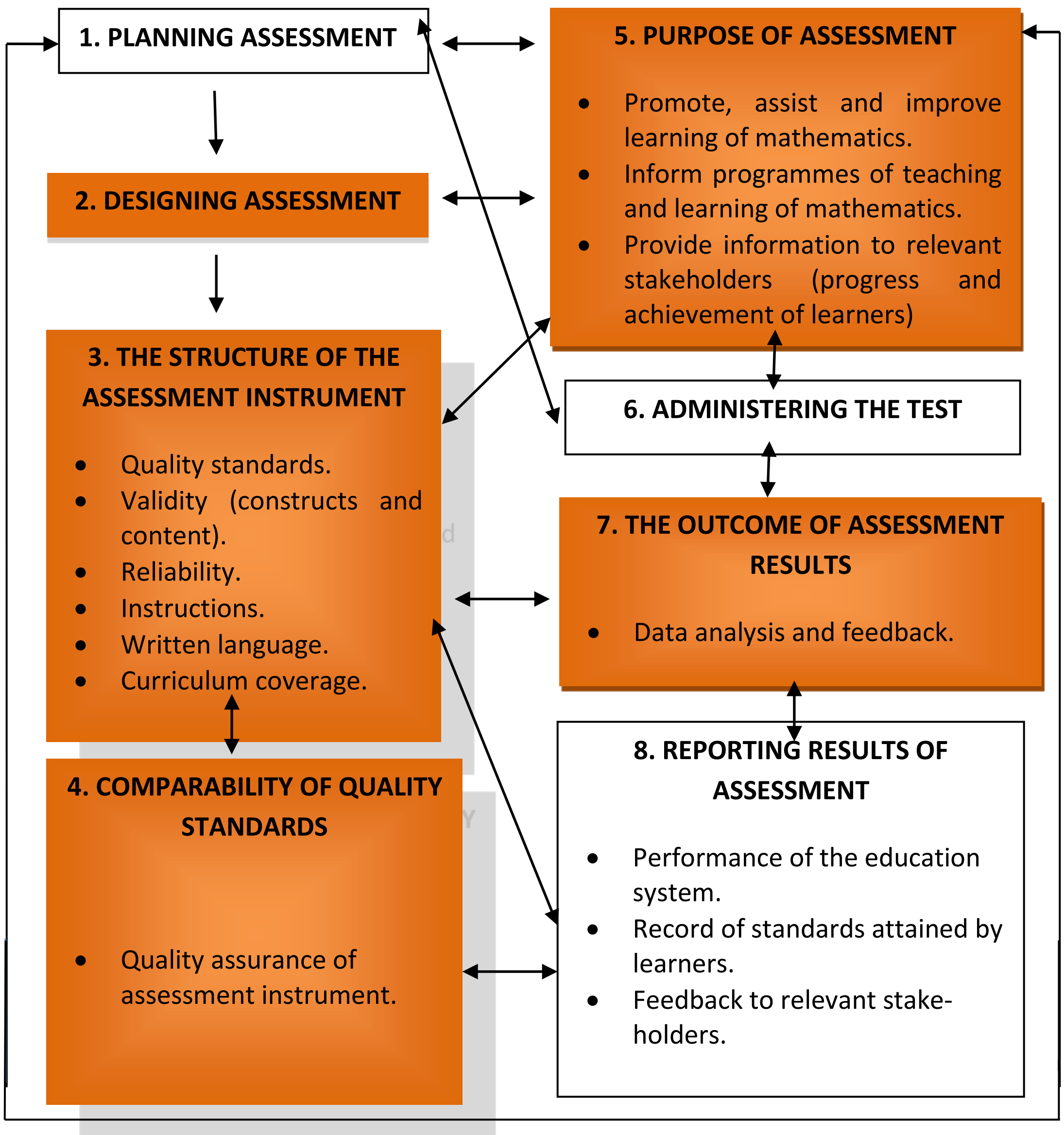


Figure 1. A Model for quality systemic assessment

This study focused on the five highlighted components on the conceptual framework because the scope of the research was not broad it focused specifically on the assessment instrument and not the administration of the ANA programme. The framework was set up from the literature review to provide a road map in order to investigate the quality of the 2012 ANA Grade 6.

The data gathered in the study was tested against the indicators in the framework. The criteria for a good assessment design are properly stipulated on the conceptual framework. Based on the research results the ANA Grade 6 mathematics instrument did not fully meet the criteria for a good assessment design. When reflecting on the indicators of the conceptual framework I infer that a systemic assessment that is not designed properly will have a negative influence. It is necessary to eliminate the negative variables when designing the assessment instrument. The instrument should adhere to standards of assessment design.

Based on the research results the construct validity was achieved to some extent; content validity was fully achieved; and reliability was low on the ANA Grade 6 mathematics instrument. The conceptual framework informs that in order to ensure a quality standard assessment instrument, the structure should be balanced, comprehensive and varied.

Achievement and progress on assessment are recognised through the comparability of quality standards. The assessment instruments must be comparable, in order to monitor learners' performance against the national indicators of learner performance. The design of the existing ANA programme over successive years, though comparable in terms of content tested, on a similar model to the South African matric exam, does not have common items across years. For this reason it is claimed that there is no comparability.

Assessment is administered for different purposes and in different situations. Assessment is seen as an integral part of teaching and learning. Based on the research results the purposes of the ANA Grade 6 mathematics instrument were met to some extent.

The outcomes of the assessment involve the analysis of results and giving feedback. Based on the research results, the ANA Grade 6 mathematics instrument was analysed at national level through the diagnostic report and feedback was provided to all the stakeholders.

Within the model of quality systemic assessment which is the conceptual framework in this study, it is clearly indicated that the link between the stages of assessment must be taken into consideration in order to produce a quality assessment instrument that could be administered for different purposes.

5.4 REFLECTIONS ON THE METHODOLOGY

The research was conducted through a secondary analysis design applying mixed methods. The strong point concerning this design is that, it allowed individual items analysis of the 2012 ANA Grade 6 mathematics instrument, and an in-depth investigation of the quality of the ANA instrument. Feedback on each item was made possible by this design. The research design was not time consuming and was user friendly. The design focused on the core principles provided by the Rasch measurement model to explore the quality of the instrument. I was advised to follow the pragmatic paradigm for this study. I then realised that the pragmatism theory was the most relevant for the study. The emphasis on the pragmatic paradigm is based on engagement with teaching, learning and critical thinking. The contribution made by this study is in the assessment component.

5.5 CONCLUSIONS

The main question for this study was: "To what extent does the 2012 Annual National Assessment Grade 6 mathematics instrument provide meaningful information for making appropriate interpretations at district level?"

This exploration was done with a view to investigate the extent to which the ANA instrument impacts on performance of learners in Grade 6 mathematics in the Gauteng North District.

5.5.1 The Purpose of the Annual National Assessment Grade 6 Instrument

The ANA instrument appeared to have met its desired purposes to some extent. The investigation did not entirely confirm that the ANA Grade 6 mathematics instrument met the purpose of assessment in terms of achieving its desired purpose as expected (see Figure 1 page 61). The information gathered on the Test Framework, the Classification Table codes for 2012 ANA Grade 6 mathematics, the conceptual framework, and the literature review provided findings on this research question, from which the following conclusions can be drawn. The literature review explained in detail the purpose of the ANA reported by the Department of Basic Education. The literature review in Chapter 2 summarised the purposes of the Annual National Assessment, but the question remains, whether these purposes were achievable within districts, for example the Gauteng North District.

The first purpose of the ANA test was to establish whether learners attained the desired level of understanding in mathematics in Grade 6. As far as this purpose is concerned, a matter of concern regarding the attainment of this purpose resulted from the events of 2012, where the Department of Basic Education provided exemplar papers as examples to prepare learners for the Annual National Assessment. This practice inevitably led to a "teaching-to-the-test" situation which affects the validity and reliability of the results. These results became questionable in two respects. Firstly, this practice could result in score inflation, because learners were coached to answer prepared questions. Secondly, the practice undermined the learners' confidence to solve problems and might have inculcated routines not directly matched to their expectations or requirement.

Furthermore, the implication is that valuable teaching time was spent on coaching learners to answer ANA test questions, rather than teaching the content of the subject to enable learners to respond competently to a variety of questions based on mathematics. When teachers teach to the test, the focus maybe is shifted away from good assessment practice and curriculum coverage.

The second purpose of the ANA was to use the results to improve learning outcomes. Since the Department of Basic Education (DBE), in collaboration with the provinces implemented a number of curriculum interventions in mathematics and languages with the aim of improving the Annual National Assessment results in South African Schools. This purpose was not achieved merely got strategies moving.

There seems to be an overload of plans to intervene and to remedy the situation. But the volume of information given and the accompanying paperwork within a short period of time overwhelm teachers and affect their health in most cases. This does not assist them at all but creates more challenges. The interventions implemented showed no improvement or only slight improvement. It ends up posing this question: Have the strategies for improving the ANA has been effectively implemented?

The third purpose was achieved to some extent. Is the ANA still a systemic evaluation or has it become a performance indicator for the schools and individuals? The schools are labelled as 'underperforming schools' by the Department of Basic Education.

This identification is brought about by the failure of some schools to use the ANA results to identify areas of weakness and, the inability to develop school based – interventions.

With regard to the final purpose, it remains the responsibility of the DBE to inform all the stakeholders involved about education issues. In conclusion, the purposes of the ANA were partially achieved in the Gauteng North District based on the results of the 2012 mathematics. These findings are supported by the discussion about the purposes of systemic assessment in the literature review Chapter 2 and the results of the purposes of the ANA are found in Chapter 4.

5.5.2 The Criteria for Good Assessment instrument Design

The ANA instrument seemed to have met the criteria for a good assessment instrument to some extent. On review, from the perspective of the literature, and the analysis, involving both professional judgement and statistical analysis, the 2012 ANA Grade 6 mathematics instrument did not satisfactorily meet the four conditions of a quality national assessment as required. The four conditions on the design of a quality national assessment are: representation of knowledge and skills; level of development; language use; and assessment results to monitor change. The literature review in Chapter 2 provided criteria for evaluating the national assessment instrument and enabled the answering of this research question.

As far as the presentation of knowledge and skills and the level of development of learners are concerned, these two conditions were partly met because it can be argued that in some cases the level of knowledge and skills that was required from the items was very high for the grade level and it could have distorted the outcomes.

As far as the requirement of language use is concerned, the investigation confirmed that the ANA Grade 6 mathematics instrument slightly met the criteria for good assessment design in terms of the written language used and the level of instructions. The instructions on some items were not clearly presented, the written language on some items was not to the level of development of learners, and these affected the quality, design of the ANA instrument. The written language was in the LOLT, as required. These findings are to some extent contradicted to some extent by the literature review in Chapter 2 and the conceptual framework in Chapter 3.

5.5.3 Construct Validity and Content Validity in an Assessment Instrument

The mathematics instrument appeared to have some content validity. The content analysis revealed that the items were aligned to the Grade 6 mathematics curriculum in an appropriate way. The results of the content analysis can be found in Chapter 4 and Table 11. In terms of construct validity the ANA instrument seemed to have achieved it to some extent. The results of construct validity are found in Chapter 3.

This research question was addressed by the literature review as well as the mapped items on the test Framework . The mapped items were used to ascertained construct validity and content validity of the Annual National Assessment Grade 6 mathematics (see Table 11 page 96).

As far as content validity is concerned, the findings revealed that there was a slight discrepancy with regard to the percentage requirements specified on the classification table used for this analysis (refer to APPENDIX I, page 185) for the mathematics content area, the difficulty level and mathematics cognitive levels of items for 2012 ANA Grade 6 mathematics with the required percentage given by the DBE column two on the Test Framework (Refer to APPENDIX I, page 185). The findings revealed that there was a slight discrepancy with regard to the percentage requirements specified on the classification table for the difficulty level of items for 2012 ANA Grade 6 mathematics with the required percentage given by the DBE test framework and might have contributed slightly on the low performance of Gauteng North District learners.

It can therefore be concluded that the classification of the items on this instrument as analysed for this did not concur with the DBE test framework specification of learner attainment at various cognitive levels according to the easy and difficult questions.

Not only were discrepancies found in terms of cognitive levels, difficulty levels and content areas coverage but also learners might have had problems in attempting to respond to the mathematics instructions given. Some of the problems on items were judged to be with the lack of clarity of the instructions. Furthermore, items where most learners underperformed were identified, pointing to a possibility that the quality of the questions could have been problematic (Refer to Table 12 and 13 pages 111 and 114).

The overall findings for Phase One revealed the following features: the time allocation was short for the ANA test, which led to some of the items being left unanswered by learners; 52 items were allocated 90 minutes translating to less than two minutes per item (refer to Appendix B, page 164). In contrast to the NAPLAN which had 32 items allocated 40 minutes. The duration of the test was one and a half hours long and comprised 13 pages excluding the cover page, which was very long for the learners, not fair and also affected performance. The disadvantage was that learners were not used to writing long papers as compared to School Based Assessment of 30 minutes.

The way that the ANA was designed, therefore imposed difficulty on the items themselves. The length of the ANA influenced learner performance, the concentration span from start to finish. A total of 52 items were too much for the Grade 6 level and the scope was very broad for the level of learners.

The following features of the ANA were acceptable, the cover page of the test had all the necessary information expected, the sequence of the items on the test ranged in questions from simple to complex as expected. The numbering of items on the ANA Grade 6 mathematics test was correct in sequence and not confusing to learners (see APPENDIX B, page 164).

As far as content validity is concerned, an exploration of the mathematics Grade 6 curriculum and that of the assessment instrument was undertaken. The Test Framework was successfully used to map the 52 items of the 2012 ANA Grade 6 mathematics test items; there was no item left unplaced (see Table 11, page 96).

Furthermore, the findings revealed that the 2012 ANA Grade 6 mathematics items for content area on ‘number operations and relationships’, ‘patterns, functions and algebra’, and ‘data handling’ did not adhere precisely to the specified percentage set by the DBE test framework (see Table 10 page 76) and contributed to the low performance of Gauteng North District learners, but was within the test specification. Curriculum coverage in terms of the scope was adhered to on the 2012 ANA Grade 6 mathematics. Phase Two of the study sought to address the last three sub-research questions, mainly pertaining to the issue of reliability of the ANA test. An exploration of the learners’ abilities in mathematics was undertaken, using the Rasch measurement model.

5.5.4 The Functioning of the Items on the Mathematics Assessment Instrument and the Requirements of the Rasch Measurement Model

The ANA Grade 6 mathematics instrument seemed to have some items which were consistent with the responses to other items on the assessment. The findings revealed that for this sample of learners the ANA was not well targeted, the mean item and person measures were not approximately equivalent. Item Fit statistics confirmed that over discrimination in the 2012 ANA Grade 6 mathematics was indicated by a total of 16 items that had large negative residual values and a relatively steep curve, indicating that the discrimination ability was too high. This high discrimination ability disadvantaged the low ability learners (see APPENDIX F page 182 and Figure 2 page 117). The overwhelming majority of learners got these items wrong; therefore, these particular items were considered to be insensitive to learner ability, something that is obviously undesirable in a test, designed for diagnostic purposes or could be a result of a biased, small sample.

Furthermore, a total of 10 items exhibited under-discrimination in the 2012 ANA Grade 6 mathematics. Under-discriminating items are usually indicated by large positive residual values and a relatively flat curve, indicating low discrimination ability (see APPENDIX F page 182 and Figure 3 page 119). The findings revealed that these items were biased and tested a construct that did not fit well into the overall framework because learner ability only partially accounted for item performance. Therefore, the items were undesirable in the test, because they were too difficult and had compromised the test's ability to differentiate between low, moderate and high ability learners due to its poor discrimination properties.

Test labels 23 and 7.4 (Rasch analysis label 49 and 50) were considered to misfit (see Table 17 page 120). The two misfit items compromised the test's ability to assess learners.

As far as the expectations for functioning of the items is concerned it can be concluded that the ANA items partially functioned as the Rasch measurement model expected. These findings are supported by the analysis of results in Chapter 4.

5.5.5 Are the Items Distributed Well Along the Continuum of the Variable

. The ANA instrument seemed to have a reasonable person separation index, a measure of both item and person reliability as it appeared not to have items that are well distributed along the continuum of the variable. The Item Analysis of Using the Rasch Model is discussed in Chapter 3. The Rasch analysis results on the reliability of the ANA test are found in Chapter 4, see also Figures 4 and 5 (pages 123 and 124), and APPENDIX F. These findings are supported by the definition and explanation of reliability on the assessment instrument in Chapter 3.

The ANA Grade 6 mathematics test was very difficult for the GND learners hence the unduly low performance and therefore it could not contribute to good diagnostic ability. This conclusion casts doubt about the reliability of the 2012 ANA Grade 6 mathematics as a systemic assessment instrument. These findings are supported by the analysis of results in Chapter 3.

5.5.6 The Abilities of the Learners Targeted by the Assessment Instrument

The Item Difficulty confirmed that the GND sample had a number of learners with an ability measure that did not correspond to any item's difficulty measure. In the 2012 ANA Grade 6 mathematics for the Gauteng North District sample, the findings revealed that there were items that had fallen significantly beyond learners' abilities in terms of difficulty as well as those which are well above the learners' expected abilities (refer to Table 19 page 126 difficult items and Table 20 page 127 easiest items).

The listed items on both Tables 19 and 20 were identified to have compromised the test's ability to assess learners. Therefore the 2012 ANA Grade 6 instrument was considered to be limited in its ability to accurately measure those learners' abilities.

In conclusion, the investigation indicated that the ANA Grade 6 mathematics was not entirely appropriate but had room for improvement, including the overall reliability (see Chapter 4). The identification of anomalies was not restricted to guessing, but addressed any potential measurement disturbance on the ANA Grade 6 mathematics instrument that had influenced negatively on the abilities of the learners. The Rasch analysis proved to be useful in identifying problematic features of the assessment instrument that the Department of Basic Education will be able to incorporate.

5.6 LIMITATIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

The weak point for this design was that it did not make room for the investigation of external factors that might have contributed to the low performance of the ANA Grade 6 mathematics of the year 2012. Because of this limitation, it is recommended that a future study be undertaken where both internal and external factors impacting on the quality of the assessment are investigated.

`The body of research based knowledge was very limited due to the fact that there were not enough research studies done related to this topic. The recommendation is therefore, that the body of knowledge be extended through multiple academic studies in order to improve the quality of the systemic assessment in South Africa.

The study focused on the quality including design of the ANA instrument. The recommendation is therefore, that the whole process involved in the administration of the ANA programme be extensively researched.

The focus was on GND with a population of 2 976, which is regarded as a small size and not representative of the total number of 148 367 learners registered for the ANA in the province and a total number of 944 397 learners registered nationally in the year 2012. The recommendation is therefore, that a large cohort be investigated country wide by a team of investigators, because this study was undertaken by a single investigator. The comparison of two districts in a province can be undertaken, in order to get a larger population to generalise the findings.

The research was directed to the investigation of the ANA on the subject mathematics. Since the subject English was also tested it should in future be investigated as well.

The investigation focused on Grade 6 level. In future Grade 3 and 9 should be included in similar investigations.

The study focused on the mathematics test written in English and excluded the learners writing the test in Afrikaans. The comparison of performance in terms of LOLT was not included; the ANAs were administered in the Language of Teaching and Learning of the school, in this case English and Afrikaans were used in Grade 6. The recommendation is therefore, that Afrikaans be investigated in a comparative study which could reveal various dynamics not taken into account within the study.

Investigating bias of the ANA on gender performance was not integrated into this study, though was explored as a factor. A focus on this aspect in future research could reveal dynamic factors that have an influence on performance. The comparison of gender performance on the ANA comparing the test scores with the actual knowledge levels and checking for systematic differences in these comparisons for boys and girls can be undertaken.

The comparison of the performance in the ANA administered between 2012 and 2013 to monitor progress could not be featured in this study due to lack of comparability. Therefore, longitudinal studies are needed to compare results, quality of tests and attainment across years.

Interviewing of teachers about their attitude towards the ANA was not incorporated in this study, this is important to find out the challenges and experiences teacher have towards the ANA. Therefore, there is room for a case study to investigate this phenomenon.

Furthermore, future research based on the study can take the following form:

Research can be undertaken on the comparison of ANA performance looking at the language used especially in the Foundation Phase where the ANA is offered in Home Languages (HL).

5.7 RECOMMENDATIONS TO IMPROVE ANA

Responding to the main research question that was investigated in this study in conclusion, the results of the Annual National Assessment inform the Gauteng Department of Education, the teacher, the district, the parents and the system. Some contribution to and explanation of the poor performance in the 2012 ANA Grade 6 mathematics could be attributed to poor quality of some of the items. The following are additional recommendations, not directly related to the findings of the study, are suggested based on my professional experience and require further research.

It is suggested that the test include items that are more evenly spaced in terms of difficulty, especially at the lower and higher difficulty levels; the assessment instrument should include items of a high quality that provide reliable information on the learners.

From the overview of the instrument and taking into account the many factors that impact on performance it is proposed that inferences be made from test results should be seen in the light of both the assessment environment and other learner factors such as curriculum coverage.

Therefore, interventions should be implemented in the form of ongoing professional development for all in the system as part of the diagnosis.

In Grade 6 mathematics learners are unfamiliar with mathematical terminology and how to use it effectively. Basic computational skills are not mastered and learners do not know how to solve problems. In geometry learners cannot cope with spatial manipulations.

The implication is that the foundations for key elements are not laid and it becomes impossible for senior grade teachers to build on prior knowledge when those required skills were not captured. Thus, there is great concern regarding the opportunity to learn, limited time for repetition and reinforcement of learnt skills, especially in the lower grades, the Foundation Phase and the Intermediate Phase.

The Rasch model analysis allowed for improvement of the instrument, even though the sample size of learners was small and limited to one district. The information on district performance by schools and subgroups with the school was possible and the results can be used by districts.

The consequences for these findings are that the DBE may be informed by these findings that the ANA test should be revised. Secondly, the investigated ANA test was written in 2012, which is four years ago, and there may have been progress recognised on performance in the subsequent years.

The 2012 ANA Grade 6 mathematics items identified not to fit the Rasch measurement model requirement should be revised for future use.

Results in the Intermediate Phase and in particular Grade 6 remain problematic. Mathematics learning and teaching needs to be strengthened by means of research and focused intervention into achievable results. The results largely confirm the concerns regarding the quality of the assessment instrument. The content gap between Grades 4 to 6 is too wide and accomplishing the set targets is a challenge. It seems that the system as a whole cannot cope with the amount of work that needs to be covered each year. Conceptual understanding does not take place. The teachers struggle to cover the curriculum and build on prior knowledge. Therefore, the scope of work to be covered in each phase should be revisited in order to find solutions towards the problem.

Among other factors the poor subject knowledge of many teachers in mathematics needs to be investigated; many teachers have not specialised in mathematics. The learners' inability to think critically; interpret information, to solve problems and struggle with a limited vocabulary seem to reflect the problems many teachers face themselves.

The ANA is administered annually which has a disadvantage of not providing enough time for remedial interventions. Therefore, the ANA should not be administered annually but rather be written in a three year - cycle and should be limited to exit Grades 3, 6 and 9 in order to achieve its purpose successfully.

Recommendations are made with a view of addressing the issue of the quality assessment instrument. Some suggestions about the instrument itself are reported in order to inform future testing. Suggestions are made about the interpretation that can be made from the systemic testing.

Finally, the main question for this study was: "To what extent does the 2012 Annual National Assessment Grade 6 mathematics instrument provide meaningful information for making appropriate interpretations at district level?"

The data gathered from the instrument and the literature review in this study make it possible for the of the 2012 ANA Grade 6 mathematics instrument to provide meaningful information that allow interpretations at district level. The data of this study may possibly inform the DBE to apply the Rasch model to determine the extent of the validity and reliability of the ANA Grade 6 mathematics instrument more generally.

The data gathered from the Rasch model might further be of assistance in the development of a theoretically based, empirically tested instrument to measure mathematical knowledge and skills for the Grade 6 learners.

In addition, the data gathered from the instrument may perhaps be used to provide meaningful information to refine the ANA Grade 6 mathematics instrument in terms of achieving its validity and high reliability and to inform district-level interventions.

REFERENCES

ACARA. (2008). *National Assessment Program Literacy and Numeracy.* Sydney; Australia: Queensland government.

Airasian, P.W. (2005). *Classroom Assessment: Concepts and Applications (5th Ed).* New York: McGraw-Hill.

Anderson, P., & Morgan, G. (2009). *Developing Tests and Questionnaires for National Assessments of Educational Achievement*. Vol. 2. Washington, DC: World Bank.

Andrich, D. (1982). *An index of person separation in Latent Trait Theory, the traditional KR-20 index, and the Guttman scale response pattern. Educational Research and Perspectives.* Vol. 9 (1), pp. 95–104.

Bell, M., Bell, J., &Flander, J. (1998). *Everyday Mathematics k-3 teachers Reference Manual: The university of Chicago school mathematics project.* Everyday Learning Corporation. Chicago: Everyday Learning Corporation.

Babbie, E. (2007). *The Practice of Social Research.* (11th Ed). Belmont, USA: Thomson High education.

Beaton, A.E., &Postlethwaite, T.N.,Ross, K.N., Spearritt, D & Wolf, R.M., 1999. *The Benefits and Limitations of International Educational Achievement Studies. Paris: UNESCO/International Institute of Educational Planning: International Academy of Learning.*

Biesta, G. (2009). *Good Education: What it is and why we need it. Inaugural Lecture*. *Stirling, the Sterling Institute of Education.*

Black, P. (1998). *Testing: Friend or Foe? Theory and Practice of Assessment and Testing.* Washington DC: Falmer Press.

Black, P. (2000). *Research and the Development of Educational Assessment.* Oxford review of education. V 26 (3 + 4), pp. 407-419.

Bond, T., & Fox, R. (2007). *Applying Rasch Model: Fundamental Measurement in the Human Sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Canon, R., & Newble, D. (2002). *Assessment to Promote Learning*. http//www.ucd.ie

Chambers, P. (2009). *Mathematics is the Study of Patterns Abstracted from the World Around Us*. California: Thousand Oaks.

Chinapah, V. (1997). Handbook of Monitoring Learning Achievement. Towards Capacity Building. Paris: UNESCO

Clement, D.H., Sarama, J.H., & Liu, X.H. (2015). *Development of Measure of Early Mathematics Achievement Using the Rasch Model: The Research-Based Early Maths Assessment*. http://dx.dol.org/10.1080/01443410701777272

Cohen, L., Manion, L., & Morrison, K. (2008). *Research Methods in Education.* London: Routledge.

Cresswell, J.W. (2003). *Research Design. Qualitative, Quantitative, and Mixed Methods Approaches (2nd Ed).* Thousand Oaks, CA: Sage Publications.

Croty, M. (1998). *The Foundations of Social Research: Meaning and Perspective in the Research Process.* Thousand Oaks, CA: Sage

Department of Education. (1998). *Assessment Policy in the General and Education and Training band.* Grades R-9 and ABET. Pretoria. Retrieved from http://www.education.gov.za

Department of Education. (2000). *Assessment Standards of the National Curriculum Statement*: *NCS*. Pretoria. Retrieved from http://www.education.gov.za

Department of Education. (2001). *C2005 Revised National Curriculum Statement Grades 4 – 6 (Schools) Intermediate Phase.* Pretoria. South Africa: Author.

Department of Education, (2003). *Qualification and Assessment Policy Framework Grades 10-12.* Pretoria. Retrieved from http://www.education.gov.za

Department of Education. (2005). *Lifelong Learning for the 21stCentury*. Pretoria, South Africa: Author

Department of Basic Education. (2009). *Report on the Task Team for the Review of the Implementation of the National Curriculum Statement (NCS).* (2009). Pretoria. Retrieved from http://www.education.gov.za

Department of Basic Education. (2011). *Annual National Assessment Report of 2011*. Pretoria. Retrieved from http://www.education.gov.za.

Department of Basic Education, (2011). *Curriculum and Assessment Policy statement Grades R- 12*. Pretoria, South Africa: Author.

Department of Basic Education. (2012). *Diagnostic Report of the Annual National Assessment.* (2012). Retrieved from http://www.education.gov.za

Department of Basic Education. (2013). *Report on the Annual National Assessment of 2013*. (2013). Retrieved from http://www.education.gov.za

Dreyer, J.M. (2008). *The Assessment Journey*. In: Dreyer. J.M (Ed). *The Educator as Assessor.* Pretoria: Van Schaick.

Du Plessis, E. (2013). *Introduction to Curriculum and Assessment Policy Statement (CAPS).* UNISA. Pretoria.

Du Plessis, P., Conley, L. & Du Plessis, E. (2007). *Teaching and Learning in South African Schools.* Pretoria: Van Schaick.

Fischer, F. (1988). *Beyond Criticism: Policy Enquiry in Post-Positivist Perspective*. Policy Studies Journal, Vol. 1 (26), pp. 129-146. Flick, U. (1988). *An Introduction to Qualitative Research.* London: Sage.

Greaney, V., & Kellaghan T. (1996). *Monitoring the Learning Outcomes of Education Systems.* Washington, DC: World Bank

Greaney, V., & Kellaghan T. (2008). *Assessing National Achievement Levels in Education*, Vol. 5. Washington, DC: World Bank.

Greaney, V., & Kellaghan T. (1992). *Using Examinations to Improve Education. A Study in Fourteen African Countries.* Washington, DC: World Bank.

Greaney, V., & Kellaghan T. (2001). *The Globalisation of Assessment in the 20th Century. In: Assessment in Education*, Vol. 8, pp. 87-102.

Graven, M.H. (2014). *Poverty, Inequity and Mathematics Performance: the Case of South Africa's Post-Apartheid Context.* V 46, p. 1039-1049. South Africa: Springer.

Hasen, T., & Postlethwaite, T.N., (1996). *A Brief History of the International Association for the Evaluation of Educational Achievement (IEA). In: Assessment in Education,* Vol 3, pp. 129-141.

Howie, S. (1997). *Mathematics and Science Performance in the Middle School Years in South Africa: A Summary Report on the Performance of the South African Students in the Third International Mathematics and Science Study*. Pretoria: HRSC.

Howie, S. (2002). *English Language Proficiency and Contextual Factors Influencing Mathematics Achievement of Secondary School Pupils in South Africa.* Published Thesis of Twente. PrintPartners Ipskamp, Enschede.

Howie, S. (2012).*High Stakes Testing in South Africa: Friend or Foe? Assessment in Education Principles: Policy and Practice* (V 9) pp. 81-89. Routledge: England.

Jansen, J. D. (1997). *Why Education Policies will Fail?* Indicator SA (V12) pp. 58-58.

Kellaghan, T., & Greaney, V. (2001). *Using Assessment to Improve the Quality of Education*: International Institute for Educational Planning. Paris: UNESCO.

Kellaghan, T., Greaney, V., & Murray, T.S (2009). *Using the Results of a National Assessment of Educational Achievement:* V5. Washington DC: The World Bank.

Kelly , M.J., & Kanyika, J. (2000). *Learning Achievement at the Middle Basic Level. Summary Report on Zambia's National Assessment Project* (1999). Lusaka: Examinations Council of Zambia.

Kurdek, L.A., & Sinclair, R.J. (2001). *Predicting Reading and Mathematics Achievement in Fourth Grade Children from Kindergarten Readiness Scores. Journal of Educational Psychology.* Vol 93 (3), pp. 451-455.

Long, C. (2015). *Learning Pathways within the Multiplicative Conceptual Field: Insights reflected through a Rasch Measurement Framework.* Munster: Waxman.

Maree, J.G., & Fraser, W.J. (2008). *Outcome-based Assessment: Facilitating best practice in classrooms. (2nd ed.):* Heinemann. Johannesburg South Africa.

Mason, T. (2005). *Senior Phase – Grade 7Revised National Curriculum Statement – Generic.* GDE: The collage of Education, University of the Witwatersrand.

Masters, G.N (1988). *A Rasch Model for Partial Credit Scoring*. Psychometrika. V 47, pp. 149-174.

Matters, G. (2009) *A Problematic Leap in the Use of Test Data: From Performance to Inference*. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational Assessment in the 21st Century* (pp. 43-62). Dordecht: Springer.

McMillan, J.H. (2007). *Classroom Assessment: Principles and practice for effective standards-based instruction (4th ed.)*: Boston: Allyn and Bacon.

Messick, S. (1989). *Meaning and values in test validation: The science and ethics of assessment. Educational Researcher*, 18(2): pp. 5-11.

Messick, S. (1988). *Validity. In R.L. Linn (Ed), Educational Measurement* (3rd ed). pp.13-103. New York: MacMillan.

Messick, S. (1995). *Standards of Validity and the Validity of Standards in Performance Assessment. Educational Measurement: Issues and Practice*, 14, Vol 4. pp.5-8.

Meyer, L., Lombard, K., Warnich, P., and Wolhuter, C. (2010). *Outcome-Based Assessment for South African Teachers*. Van Schaik Publishers. Pretoria.

Miles, M.B., Huberman, A.C. (1994). *Qualitative Data Analysis* (2nd Ed). SAGE Publications: Thousand Oaks. London. New Delhi.

OFSTED. (2003). *Good Assessment in Secondary Schools.* London: Office of Standards in Education.

Olson, J.F. (2007). *Trends in International Mathematics and Science Study.* International study centre. Boston college

O'Neill, T. (2000). *Boys' Problems Don't Matter*. Report: National Edition. Vol27 (15) pp. 54-56.

Ornstein, A.C., & Hunkins, F.P. (2011). *Curriculum: Foundations, Principles, and Issues.* (5th Ed). United States of America: Pearson.

Pattanayak, D.P. (2003). *Mother Tongue: The Problem of Definitions and the Educational Challenge.* In A. Ouane (Ed). *Towards a Multilingual Culture of Education.* pp. 23-28. Hamburg: UNESCO Institute for Education.

Pope, G. (2009). *Understanding Validity: What is Construct-Related Validity?* Retrieved from Copyright © 1995-2017 Questionmark.

Popham, W.J. (2003). *Test Better, Teach Better: The Instructional Role of Assessment.* Alexandria: Association of Supervision and Curriculum Development.

Queensland Studies Authorities Assessment Policy document. (2009). ACARA

Rasch, G. (1960/1980). *Probabilistic Models for Some Intelligence and Attainment Tests.* Chicago: University of Chicago Press.

Reddy, C. (2004). *Assessment Principles and Approaches.* In: Maree, J.G., & Fraser, W.J., eds. *Outcomes-Based Assessment.* Sandown: Heinemann. pp. 30-44.

Reddy, V. (2006). Mathematics and Science Achievement at South African Schools in TIMSS 2003. Cape Town: HRSC Press.

Ross, K.N., & Genevois, I.J. (2006). *Cross-National Studies of the Quality of Education: Planning their Design and Managing their Impact*. Paris: UNESCO.

SACMEQ. (1995-2013). Retrieved from http://www.sacmeq.org/about.htm. 23/11/2012

SACMEQ. (2010-2017). Retrieved from http://www.sacmeq.org/about.htm.

Scheerens, J., Glas, C., and Thomas, S.M. (2003). *Educational Evaluation, Assessment, and Monitoring:* A Systemic Approach. The Netherlands: Swets & Zeitlinger Publishers.

Schoenfeld, A.H. (2007). *What is Mathematical Proficiency and How Can It Be Assessed?* Assessing Mathematical Proficiency. MSRI Publications. Vol. 53 pp. 59-73.

Scottish Qualifications Authority (SQA). (1997). *Guide to Assessment: Guide to Assessment:* Glasgow: SCOTVEG.

Smith, E.V. (2004). *Evidence for Reliability of measures and Validity of measure Interpretation: A Rasch Measurement Perspective.* In E.V. Smith and R.M. Smith (Eds). *Introduction to Rasch Measurement.* pp. 93-122. Maple Grove, MN: JAM Press.

Teacher Union Queries differing Results in School Tests and ANAs. (2015, October 17, p.2). Pretoria News.

Tshabalala, P.M. (2008).*Numeracy Performance of Grade 3 Learners in Rural and Urban Primary Schools.* Faculty of Education: University of Pretoria.

UNESCO (United Nations Educational Scientific and Cultural Organisation). (2007). *Global Monitoring Report*. Paris: UNESCO.

Van den Akker, J. (2003). *Curriculum Perspectives: An Introduction*. In van den Akker, J., Kuiper, W. & Hameyer, U. (Eds). *Curriculum Landscapes and Trends.* Dordrecht: Kluwer Academic Publishers.

Van Rooyen, M., & Prinsloo, F. (2003). *Outcomes -Based Assessment Facilitated: A Comprehensive Handbook for South Africans*. Cape Town: Cambridge University Press.

Webb, N.L. (1992). *Assessment of Students Knowledge of Mathematics: Steps Toward a Theory*. In D. A. Grouws (Ed.), *NCTM Handbook of Research on Mathematics Teaching and Learning.* pp.661-683. New York: Macmillan Publishing Company.

Webb, L., & Coxford, A.F. (1993). *Assessment in the Mathematics in the Classroom.* University of Wisconsin: Madison.

Wiliam, D. (2009). *Assessment For Learning: Why, What and How?* University of London: Institute of Education.

APPENDIX A

Letter from Head Office of the Department of Basic Education

GAUTENG PROVINCE
Department: Education
REPUBLIC OF SOUTH AFRICA

For administrative use:
Reference no. D2014/084

GDE RESEARCH APPROVAL LETTER

Date:	29 May 2013
Validity of Research Approval:	29 May 2013 to 20 September 2013
Name of Researcher:	Modzuka C.
Address of Researcher:	P.O. Box 17514
	Pretoria-North
	0116
Telephone Number:	012 542 4880 / 072 945 4749
Fax Number:	086 608 0832
Email address:	Dzukis.c@gmail.com charlotte.modzuka@gauteng.gov.za
Research Topic:	Investigating the Annual National Assessment (ANA) programme with reference to the Grade 6 Mathematics in the Gauteng North District (GND)
Number and type of schools:	NONE
District/s/HO	Gauteng North District

Re: Approval in Respect of Request to Conduct Research

This letter serves to indicate that approval is hereby granted to the above-mentioned researcher to proceed with research in respect of the study indicated above. The onus rests with the researcher to negotiate appropriate and relevant time schedules with the school/s and/or offices involved to conduct the research. A separate copy of this letter must be presented to both the School (both Principal and SGB) and the District/Head Office Senior Manager confirming that permission has been granted for the research to be conducted.

The following conditions apply to GDE research. The researcher may proceed with the

Hdlido
2013/05/30

1

Making education a societal priority

Office of the Director: Knowledge Management and Research
9th Floor, 111 Commissioner Street, Johannesburg, 2001
P.O. Box 7710, Johannesburg, 2000 Tel: (011) 355 0506
Email: David.Makhado@gauteng.gov.za

APPENDIX B

The 2012 ANA Grade 6 Mathematics Instrument

basic education

Department:
Basic Education
REPUBLIC OF SOUTH AFRICA

MARKS	

ANNUAL NATIONAL ASSESSMENT 2012
GRADE 6 MATHEMATICS
TEST

MARKS: 75

TIME: 1½ hours

PROVINCE ______________________________

REGION ______________________________

DISTRICT ______________________________

SCHOOL NAME ______________________________

NATIONAL EMIS NUMBER (9 digits)

CLASS (e.g. 6A) ______________________________

SURNAME ______________________________

NAME ______________________________

GENDER (✓)

BOY		GIRL	

DATE OF BIRTH

C	C	Y	Y	M	M	D	D

This test consists of 13 pages, excluding the cover page.

Instructions to learners
1. Read all the instructions carefully.
2. Question 1 consists of 6 multiple-choice questions. Circle the letter of the correct answer.
3. Answer Questions 2 to 29 in the spaces or frames provided.
4. Show all working on the question paper.
5. The test duration is 90 minutes.
6. The teacher will lead you through the practice exercise before you start the test.
7. You may not use a calculator.

Practice exercise

Circle the letter of the correct answer.

8 x 6 =

(A) 48
B 84
C 72
D 60

You have answered the question correctly if you have circled A.

NOTE:
• You will answer more questions in the test like the one you have just completed.
• Do your best to answer each question even if you are not sure of the answer.
• Look only at your own work.

The test starts on the next page.

1. Circle the letter of the correct answer.

1.1 Which decimal number is the biggest?

A 1,01
B 1,1
C 0,11
D 1,001 (1)

1.2 39 569 was rounded off to 40 000.
To which of the following numbers was it rounded off?

A 5
B 10
C 100
D 1 000 (1)

1.3 Select a number sentence to match the following statement:
Seven less than a certain number *m* is equal to twelve.

A $7 - m = 12$
B $12 - m = 7$
C $m + 7 = 12$
D $m - 7 = 12$ (1)

1.4 Calculate: 4 x (12 – 8) ÷ (4 + 0) x 1

A 8
B 16
C 0
D 4 (1)

1.5 How much water will a full kitchen kettle hold?

A 1,7 mℓ
B 1,7 ℓ
C 107 ℓ
D 1,7 kℓ (1)

1.6 With which operation sign must you replace the Δ so that the number sentence 4 x 3 Δ 6 ÷ 2 = 15 is correct?

A +
B –
C x
D ÷ (1)

2. Write down the next decimal number.
0,79 ; 0,76 ; 0,73 ; 0,7 ; ______ (1)

3. Write the number in digits.
Four million two hundred and eighty three thousand one hundred and sixty-four.

______________________ (1)

4. From the given group of numbers, choose ONE number that fits each of the descriptions below.

36 120
19 8 143 87

4.1 A multiple of 11 __________ (1)
4.2 1 , 2 , 3 , 4 , 6 , 9 , 12 , 18 are factors of _________ (1)
4.3 A prime number __________ (1)
4.4 A number divisible by 5 __________ (1)

Grade 6 Mathematics English 3

5. What is the value of the underlined digit in the following number?

   367 049 215 ___________ (1)

6. Round off 29 702 to the nearest 5. __________ (1)

7. Calculate the answers in Questions 7.1 to 7.6.

   7.1 654 + 235 583 + 32 912 (2)

   7.2 394 067 – 63 625 (2)

   7.3 6 960 ÷ 145 (3)

7.4 2 067 x 189 (4)

7.5 5 – 3,64 (1)

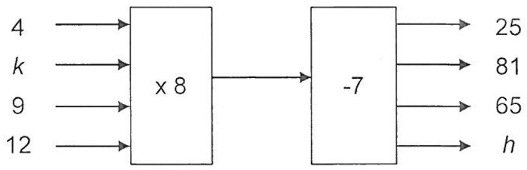
7.6 24,37 + 346,83 (1)

8. Calculate the answer in Questions 8.1 and 8.2 and write your answer as a mixed number.

8.1 $8\frac{3}{10} - 4\frac{1}{5}$ (4)

8.2 $5\frac{1}{2} + 3\frac{1}{8} + \frac{3}{4}$ (4)

9. Find the value of *h* and *k* in the flow diagram below.



9.1 h = __________ (1)

9.2 k = __________ (1)

10. Write down the next number in the sequence.

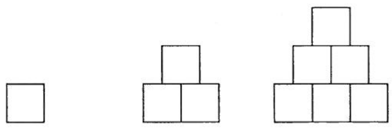
1 ; 1 ; 2 ; 3 ; 5 ; 8 ; __________ (1)

11. Calculate the value of x if $x - 41 = 12$. (1)

12. A car travels at 100 km per hour. How far will it travel in 45 minutes? (2)

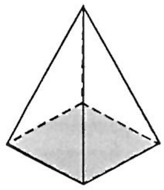
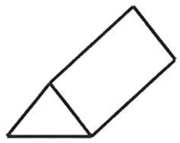
13. Mr Msebenzi buys 480 sweets for R30,00. He repacks the sweets into packets of 24 each. He sells the packets for R2,50 each.
How much profit will he make if he sells all the sweets? (3)

14. Ingrid has stacked some blocks according to a certain pattern.
Use the same pattern to draw Stack 5.



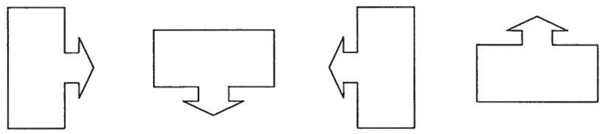
Stack 1 Stack 2 Stack 3 ... Stack 5 (2)

15. Examine the two objects below and complete the table.

		
Name of object	Square-based pyramid	
Number of vertices	5	
Number of edges		9

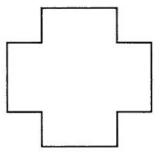
(3)

16. Which kind of transformation is shown here?



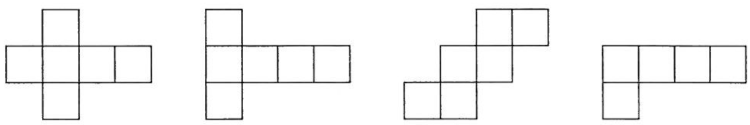
____ (1)

17. How many lines of symmetry does the diagram below have?



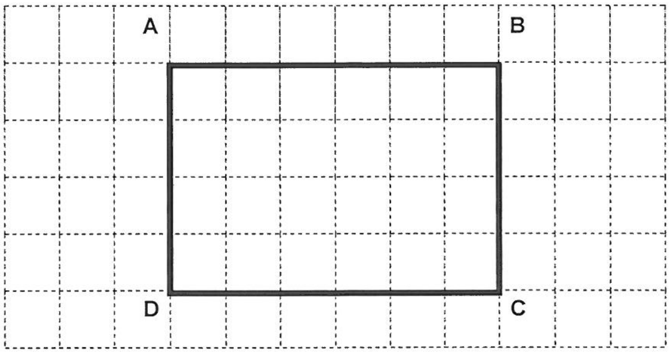
____ (1)

18. Circle the letter of the net that cannot be folded into a cube.



A B C D (1)

19. On the grid below, each block represents 1 cm by 1 cm. ABCD is a rectangle.

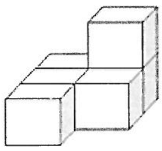


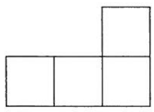
19.1 The length of side AB = ___________________ cm (1)

19.2 The length of side AD = ___________________ cm (1)

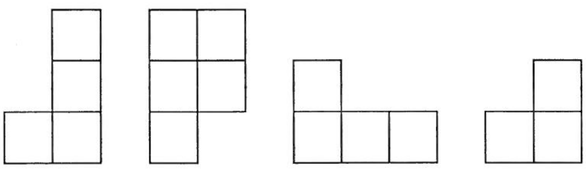
19.3 How many square centimetres are in ABCD? __________ cm^2 (1)

20. Six cubes are used to build the 3-D figure shown below. The view from the right is given next to it. Circle the letter showing the front view.



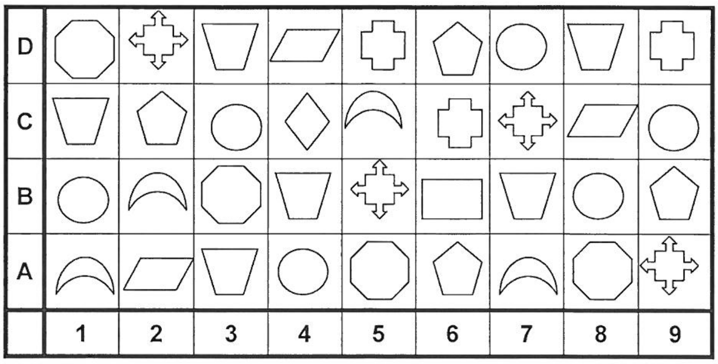


View from the right



A B C D (1)

21. Study the grid below and then answer the questions that follow.



21.1 How many pentagons are there in this grid? _______________ (1)

21.2 Which shape is in block 8A? ____________________________ (1)

21.3 In which block will you find a rectangle? __________________ (1)

22. Miriam left Durban at 21:45 and arrived in Johannesburg at 04:30 the next day. How long did her journey take?

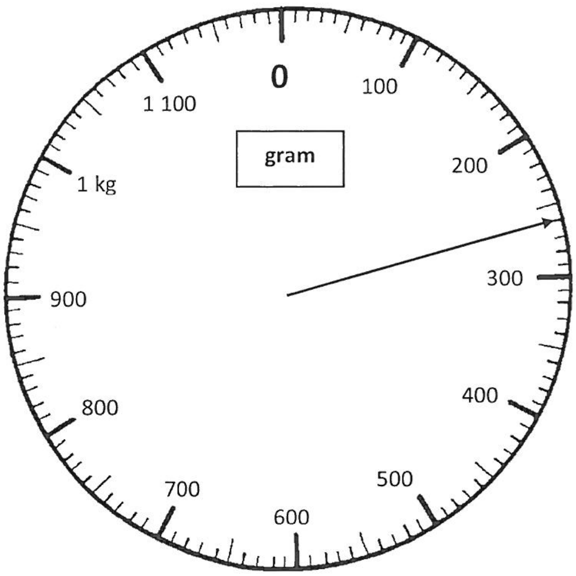
(2)

23. Tamara invited 37 friends to her party. Each friend may drink 2 glasses of cool drink. If each glass holds 200 mℓ, how many 2-litre bottles of cool drink should her mother buy?

(3)

Grade 6 Mathematics English 10

24. Refer to the scale below to answer the questions.



24.1 What is the mass indicated on the scale above? __________ (1)

24.2 Convert your answer in Question 24.1 to kilograms. ________ kg (1)

25. The tally chart below shows the number of lions and giraffes which learners saw in the zoo.

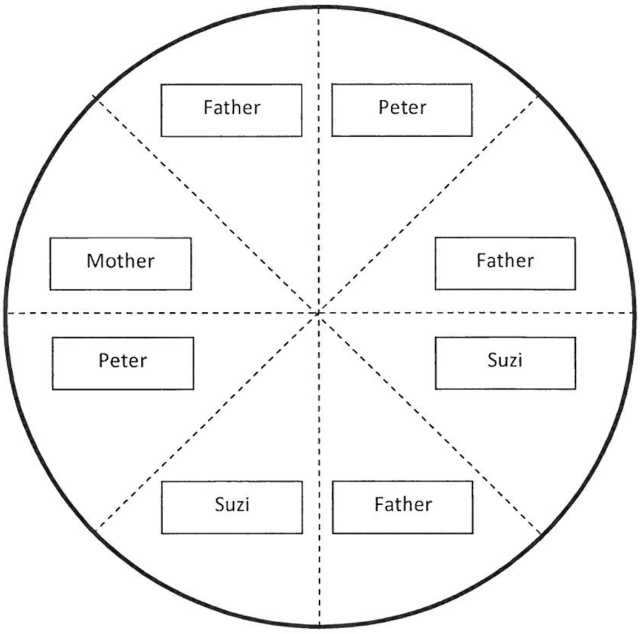
Animal	Tally marks	Total
Lion	~~IIII~~ ~~IIII~~ ~~IIII~~ I	
Giraffe		12

Complete your answers in the table above.

25.1 How many lions did the learners see? (1)

25.2 Fill in the correct number of tally marks for the number of giraffes seen. (1)

26. Mother bought the family a pizza for R120,00. She divided it amongst the family members as indicated below.

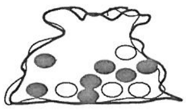


26.1 What fraction of the pizza did Father eat? ________________ (1)

26.2 What percentage of the pizza did Peter eat? _______________ (1)

26.3 What did Suzi's part of the pizza cost? __________________ (1)

27. Study the picture below.



A bag contains black and white marbles. The probability of taking a white marble out of the bag above, without looking in the bag, is _____________. (1)

Grade 6 Mathematics English 12

28. Jacob listed the marks for his Mathematics class tests.

Jacob's marks: 20 16 10 3 12 10 11 14 5 19 4

What is Jacob's median mark? ______________________________ (1)

29. Phiti has five numbered cards. How many different two-digit numbers can she make with these five cards?

6	2	0	7	3

____ (2)

TOTAL: 75

APPENDIX C

The Marking Tool for the 2012 ANA Grade 6 Mathematics

basic education
Department:
Basic Education
REPUBLIC OF SOUTH AFRICA

ANNUAL NATIONAL ASSESSMENT 2012
GRADE 6 MATHEMATICS
TEST
MEMORANDUM

MARKS: 75

This memorandum consists of 3 pages.

Important information:

1. **Give full marks if only correct answers are given, unless stated otherwise.**
2. **Accept any mathematically correct solution that is not included in the memorandum.**

QUESTION		EXPECTED ANSWERS	CLARIFICATION	MARKS	TOTAL
1.	1.1	B✓		1	6
	1.2	D✓		1	
	1.3	D✓		1	
	1.4	D✓		1	
	1.5	B✓		1	
	1.6	A✓		1	
2.		0,67✓		1	1
3.		4 283 164✓		1	1
4.	4.1	143✓		1	
	4.2	36✓		1	
	4.3	19✓		1	
	4.4	120✓		1	4
5.		40 000✓		1	1
6.		29 700✓		1	1
7.	7.1	235 583 + 32 912 654 269 149✓✓	All digits of the answer must be correct.	2	
	7.2	394 067 - 63 625 330 442✓✓	As above.	2	
	7.3	4 8 ✓ 1 4 5 ⟌ 6 9 6 0 5 8 0 ✓ 1 1 6 0 1 1 6 0 ✓ 0	Answer ✓ Step 1 ✓ Step 2 ✓	3	
	7.4	2 067 x 189 18 603✓ 165 360✓ + 206 700✓ 390 663✓	1 mark for each product. 1 mark for the answer.	4	

Grade 6 Mathematics English 1

	7.5	5 – 3,64 = 1,36 ✓ OR 5,00 - 3,64 1,36 ✓		1	
	7.6	24,37 + 346,83 371,20✓		1	13
8.	8.1	$8\frac{3}{10} - 4\frac{1}{5}$ OR $= \frac{83}{10} - \frac{21}{5}$ ✓ $= 4$✓ $+ \frac{3}{10} - \frac{2}{10}$ ✓ $= \frac{83}{10} - \frac{42}{10}$ ✓ $= 4 + \frac{1}{10}$ ✓ $= \frac{41}{10}$ ✓ $= 4\frac{1}{10}$ ✓ $= 4\frac{1}{10}$ ✓	Convert $\frac{1}{5}$ to $\frac{2}{10}$ ✓	4	
	8.2	$8 + \frac{1}{2} + \frac{1}{8} + \frac{3}{4}$ $= 8 + \frac{4}{8} + \frac{1}{8} + \frac{6}{8}$ ✓ $= 8 + \frac{11}{8}$ ✓ $= 8 + 1 + \frac{3}{8}$ ✓ $= 9\frac{3}{8}$ ✓ OR $= \frac{11}{2} + \frac{25}{8} + \frac{3}{4}$ $= \frac{44}{8}$ ✓ $+ \frac{25}{8} + \frac{6}{8}$ ✓ $= \frac{75}{8}$ ✓ $= 9\frac{3}{8}$ ✓	Convert $\frac{3}{4}$ to $\frac{6}{8}$ ✓ Add fractions to $\frac{11}{8}$ ✓ Convert $\frac{11}{8}$ to $1 + \frac{3}{8}$ ✓ Answer ✓	4	8
9.	9.1	h = 89 ✓		1	
	9.2	k = 11✓		1	2
10.		13✓		1	1
11.		53✓		1	1
12.		$\frac{45}{60}$ x 100 ✓ = 75 km✓		2	2
13.		Number of packets = 480 ÷ 24 = 20✓ Selling price = R2,50 x 20 = R50✓ Profit = R50 – R30 =R20✓		3	3

Grade 6 Mathematics English 2

APPENDIX D

Item Number and Test Label of 2012 ANA Gauteng North District Grade 6 Mathematics

Item Number	Test Label
1	1.1
2	1.2
3	1.3
4	1.4
5	1.5
5	1.6
7	18
8	20
9	2
10	3
11	4.1
12	4.2
13	4.3
14	4.4
15	5
16	6
17	7.5
18	7.6
19	9.1
20	9.2
21	10
22	11
23	16
24	17
25	19.1
26	19.2
27	19.3
28	21.1
29	21.2
30	21.3
31	24.1
32	24.2
33	25.1
34	25.2
35	26.1
36	26.2
37	26.3
38	27
39	28
40	7.1
41	7.2
42	12
43	14
44	22
45	29
46	7.3
47	13
48	15
49	23
50	7.4
51	8.1
52	8.2

APPENDIX E

Codes for RASCH Analysis 2012 ANA Gauteng North District Grade 6 Mathematics Assessment Data

DESCRIPTOR	CODE
LEARNER ID (Person ID):	
▪ Learners of School one	1001 – 1127
▪ Learners of School two	2001 – 2038
▪ Learners of School three	3001 – 3175
▪ Learners of School four	4001 – 4185
▪ Learners of School five	5001 – 5019
SCHOOL ID:	
▪ School one	01
▪ School two	02
▪ School three	03
▪ School four	04
▪ School five	05
CLASS	
▪ 6A	1
▪ 6B	2
▪ 6C	3
▪ 6D	4
▪ Missing	99
GENDER	
▪ Boy	A
▪ Girl	B
MULTIPLE CHOICE ANSWERS	
▪ A	1
▪ B	2
▪ C	3
▪ D	4
▪ Missing	9
SCORES PER QUESTIONS	
▪ Captured according to learner mark on scripts.	0 – 4
▪ Missing answers	9

APPENDIX F

Item Statistics of 2012 ANA Gauteng North District Grade 6 Mathematics Assessment Data

Display: ITEM FIT STATISTICS									
Seq	Item	Type	Location	SE	Residual	DF	ChiSq	DF	Prob
1	I0001	MC	-0.59	0.1	2.082	525.6	22.393	8	0.004238
2	I0002	MC	-2.166	0.098	2.397	525.6	17.826	8	0.022569
3	I0003	MC	0.892	0.147	1.345	496.3	27.174	8	0.00066
4	I0004	MC	-0.61	0.102	1.134	506.06	11.896	8	0.155926
5	I0005	MC	-0.836	0.098	1.812	524.63	15.296	8	0.053631
6	I0006	MC	-1.073	0.098	4.24	497.27	36.493	8	0.000014
7	I0007	MC	0.114	0.117	3.601	493.36	55.244	8	0
8	I0008	MC	0.114	0.119	2.422	470.89	37.232	8	0.00001
9	I0009	Poly	0.528	0.127	-2.804	531.47	20.499	8	0.008606
10	I0010	Poly	-1.669	0.094	-2.695	531.47	21.265	8	0.006477
11	I0011	Poly	0.522	0.127	-2.828	527.56	25.739	8	0.001164
12	I0012	Poly	-1.125	0.095	0.795	530.49	6.223	8	0.622296
13	I0013	Poly	-0.232	0.106	-1.122	529.51	15.369	8	0.052348
14	I0014	Poly	-0.564	0.1	0.063	529.51	8.996	8	0.342658
15	I0015	Poly	-1.93	0.095	1.245	532.44	23.914	8	0.002371
16	I0016	Poly	-0.343	0.104	-4.053	524.63	32.708	8	0.000069
17	I0017	Poly	2.298	0.244	-0.821	528.53	21.376	8	0.006214
18	I0018	Poly	-1.079	0.095	-2.799	529.51	15.587	8	0.048686
19	I0019	Poly	0.38	0.122	-3.25	531.47	20.803	8	0.00769
20	I0020	Poly	1.065	0.151	-3.283	530.49	24.425	8	0.001946
21	I0021	Poly	0.923	0.144	-1.807	529.51	7.277	8	0.507071
22	I0022	Poly	0.492	0.131	-2.079	474.8	18.239	8	0.019505
23	I0023	Poly	0.877	0.147	-2.415	472.85	14.485	8	0.069972
24	I0024	Poly	-0.976	0.097	4.369	517.79	60.671	8	0
25	I0025	Poly	-1.952	0.095	-1.819	532.44	16.692	8	0.033478
26	I0026	Poly	-1.954	0.095	-0.561	532.44	8.22	8	0.412236
27	I0027	Poly	-0.832	0.097	-0.185	530.49	12.823	8	0.11807
28	I0028	Poly	-1.606	0.094	5.113	530.49	45.147	8	0
29	I0029	Poly	-0.221	0.106	-3.004	529.51	29.013	8	0.000315
30	I0030	Poly	-1.926	0.095	-2.605	530.49	20.124	8	0.009876
31	I0031	Poly	0.337	0.122	-0.807	518.77	4.629	8	0.796359
32	I0032	Poly	1.642	0.188	0.389	515.83	11.911	8	0.155243
33	I0033	Poly	-3.682	0.134	-0.55	529.51	15.405	8	0.051736
34	I0034	Poly	-3.385	0.123	-0.228	531.47	33.133	8	0.000058
35	I0035	Poly	-0.327	0.104	-0.366	531.47	3.881	8	0.867677
36	I0036	Poly	3.174	0.363	0.051	531.47	9.143	8	0.330353
37	I0037	Poly	2.595	0.279	-0.62	529.51	5.336	8	0.721145
38	I0038	Poly	3.77	0.483	-1.205	528.53	6.826	8	0.555502
39	I0039	Poly	1.77	0.197	-1.434	523.65	21.323	8	0.006339
40	I0040	Poly	-1.124	0.055	-1.08	533.42	14.53	8	0.068955
41	I0041	Poly	-1.051	0.056	-0.337	532.44	16.411	8	0.036858
42	I0042	Poly	2.372	0.334	-0.082	515.83	6.762	8	0.562508
43	I0043	Poly	-1.469	0.059	6.595	508.02	82.415	8	0
44	I0044	Poly	1.418	0.162	-0.503	521.7	5.777	8	0.672149
45	I0045	Poly	4.843	0.47	0.006	471.87	4.373	8	0.82203
46	I0046	Poly	1.668	0.219	-0.586	530.49	1.309	8	0.995446
47	I0047	Poly	1.267	0.137	1.001	517.79	11.342	8	0.18308
48	I0048	Poly	-0.829	0.055	1.003	529.51	22.932	8	0.003454
49	**10049**				**REJECTED ITEM**				
50	I0050	Poly	0.379	0.07	-2.982	530.49	16.713	8	0.033243
51	I0051	Poly	0.11	0.061	0.852	524.63	16.173	8	0.039976
52	**10052**				**DELETED ITEM**				

APPENDIX G

Item Fit of 2012 ANA Gauteng North District Grade 6 Mathematics Assessment Data

(from Over Discrimination to Under Discrimination)

Seq	Item	Type	Location	SE	**Residual**	DF	ChiSq	DF	Prob
16	I0016	Poly	-0.343	0.104	**-4.053**	524.63	32.708	8	0.000069
20	I0020	Poly	1.065	0.151	**-3.283**	530.49	24.425	8	0.001946
19	I0019	Poly	0.38	0.122	**-3.25**	531.47	20.803	8	0.00769
29	I0029	Poly	-0.221	0.106	**-3.004**	529.51	29.013	8	0.000315
50	I0050	Poly	0.379	0.07	**-2.982**	530.49	16.713	8	0.033243
11	I0011	Poly	0.522	0.127	**-2.828**	527.56	25.739	8	0.001164
9	I0009	Poly	0.528	0.127	**-2.804**	531.47	20.499	8	0.008606
18	I0018	Poly	-1.079	0.095	**-2.799**	529.51	15.587	8	0.048686
10	I0010	Poly	-1.669	0.094	**-2.695**	531.47	21.265	8	0.006477
30	I0030	Poly	-1.926	0.095	**-2.605**	530.49	20.124	8	0.009876
23	I0023	Poly	0.877	0.147	**-2.415**	472.85	14.485	8	0.069972
22	I0022	Poly	0.492	0.131	**-2.079**	474.8	18.239	8	0.019505
25	I0025	Poly	-1.952	0.095	**-1.819**	532.44	16.692	8	0.033478
21	I0021	Poly	0.923	0.144	**-1.807**	529.51	7.277	8	0.507071
39	I0039	Poly	1.77	0.197	**-1.434**	523.65	21.323	8	0.006339
38	I0038	Poly	3.77	0.483	**-1.205**	528.53	6.826	8	0.555502
13	I0013	Poly	-0.232	0.106	**-1.122**	529.51	15.369	8	0.052348
40	I0040	Poly	-1.124	0.055	**-1.08**	533.42	14.53	8	0.068955
17	I0017	Poly	2.298	0.244	**-0.821**	528.53	21.376	8	0.006214
31	I0031	Poly	0.337	0.122	**-0.807**	518.77	4.629	8	0.796359
37	I0037	Poly	2.595	0.279	**-0.62**	529.51	5.336	8	0.721145
46	I0046	Poly	1.668	0.219	**-0.586**	530.49	1.309	8	0.995446
26	I0026	Poly	-1.954	0.095	**-0.561**	532.44	8.22	8	0.412236
33	I0033	Poly	-3.682	0.134	**-0.55**	529.51	15.405	8	0.051736
44	I0044	Poly	1.418	0.162	**-0.503**	521.7	5.777	8	0.672149
35	I0035	Poly	-0.327	0.104	**-0.366**	531.47	3.881	8	0.867677
41	I0041	Poly	-1.051	0.056	**-0.337**	532.44	16.411	8	0.036858
34	I0034	Poly	-3.385	0.123	**-0.228**	531.47	33.133	8	0.000058
27	I0027	Poly	-0.832	0.097	**-0.185**	530.49	12.823	8	0.11807
42	I0042	Poly	2.372	0.334	**-0.082**	515.83	6.762	8	0.562508
45	I0045	Poly	4.843	0.47	**0.006**	471.87	4.373	8	0.82203
36	I0036	Poly	3.174	0.363	**0.051**	531.47	9.143	8	0.330353
14	I0014	Poly	-0.564	0.1	**0.063**	529.51	8.996	8	0.342658
32	I0032	Poly	1.642	0.188	**0.389**	515.83	11.911	8	0.155243
12	I0012	Poly	-1.125	0.095	**0.795**	530.49	6.223	8	0.622296
51	I0051	Poly	0.11	0.061	**0.852**	524.63	16.173	8	0.039976
47	I0047	Poly	1.267	0.137	**1.001**	517.79	11.342	8	0.18308
48	I0048	Poly	-0.829	0.055	**1.003**	529.51	22.932	8	0.003454
4	I0004	MC	-0.61	0.102	**1.134**	506.06	11.896	8	0.155926
15	I0015	Poly	-1.93	0.095	**1.245**	532.44	23.914	8	0.002371
3	I0003	MC	0.892	0.147	**1.345**	496.3	27.174	8	0.00066
5	I0005	MC	-0.836	0.098	**1.812**	524.63	15.296	8	0.053631
1	I0001	MC	-0.59	0.1	**2.082**	525.6	22.393	8	0.004238
2	I0002	MC	-2.166	0.098	**2.397**	525.6	17.826	8	0.022569
8	I0008	MC	0.114	0.119	**2.422**	470.89	37.232	8	0.00001
7	I0007	MC	0.114	0.117	**3.601**	493.36	55.244	8	0
6	I0006	MC	-1.073	0.098	**4.24**	497.27	36.493	8	0.000014
24	I0024	Poly	-0.976	0.097	**4.369**	517.79	60.671	8	0
28	I0028	Poly	-1.606	0.094	**5.113**	530.49	45.147	8	0
43	I0043	Poly	-1.469	0.059	**6.595**	508.02	82.415	8	0
-----	-----	------	------------	-----	-----------	-------	--------	----	----------

APPENDIX H

Item Difficulty of 2012 ANA Gauteng North District Grade 6 Mathematics Assessment Data *(from Difficult to Easiest order)*

Seq	Item	Type	**Location**	SE	Residual	DF	ChiSq	DF	Prob
-----	-----	------	**------------**	-----	-----------	--------	--------	----	----------
45	I0045	Poly	**4.843**	0.47	0.006	471.87	4.373	8	0.82203
38	I0038	Poly	**3.77**	0.483	-1.205	528.53	6.826	8	0.555502
36	I0036	Poly	**3.174**	0.363	0.051	531.47	9.143	8	0.330353
37	I0037	Poly	**2.595**	0.279	-0.62	529.51	5.336	8	0.721145
42	I0042	Poly	**2.372**	0.334	-0.082	515.83	6.762	8	0.562508
17	I0017	Poly	**2.298**	0.244	-0.821	528.53	21.376	8	0.006214
39	I0039	Poly	**1.77**	0.197	-1.434	523.65	21.323	8	0.006339
46	I0046	Poly	**1.668**	0.219	-0.586	530.49	1.309	8	0.995446
32	I0032	Poly	**1.642**	0.188	0.389	515.83	11.911	8	0.155243
44	I0044	Poly	**1.418**	0.162	-0.503	521.7	5.777	8	0.672149
47	I0047	Poly	**1.267**	0.137	1.001	517.79	11.342	8	0.18308
20	I0020	Poly	**1.065**	0.151	-3.283	530.49	24.425	8	0.001946
21	I0021	Poly	**0.923**	0.144	-1.807	529.51	7.277	8	0.507071
3	I0003	MC	**0.892**	0.147	1.345	496.3	27.174	8	0.00066
23	I0023	Poly	**0.877**	0.147	-2.415	472.85	14.485	8	0.069972
9	I0009	Poly	**0.528**	0.127	-2.804	531.47	20.499	8	0.008606
11	I0011	Poly	**0.522**	0.127	-2.828	527.56	25.739	8	0.001164
22	I0022	Poly	**0.492**	0.131	-2.079	474.8	18.239	8	0.019505
19	I0019	Poly	**0.38**	0.122	-3.25	531.47	20.803	8	0.00769
50	I0050	Poly	**0.379**	0.07	-2.982	530.49	16.713	8	0.033243
31	I0031	Poly	**0.337**	0.122	-0.807	518.77	4.629	8	0.796359
7	I0007	MC	**0.114**	0.117	3.601	493.36	55.244	8	0
8	I0008	MC	**0.114**	0.119	2.422	470.89	37.232	8	0.00001
51	I0051	Poly	**0.11**	0.061	0.852	524.63	16.173	8	0.039976
29	I0029	Poly	**-0.221**	0.106	-3.004	529.51	29.013	8	0.000315
13	I0013	Poly	**-0.232**	0.106	-1.122	529.51	15.369	8	0.052348
35	I0035	Poly	**-0.327**	0.104	-0.366	531.47	3.881	8	0.867677
16	I0016	Poly	**-0.343**	0.104	-4.053	524.63	32.708	8	0.000069
14	I0014	Poly	**-0.564**	0.1	0.063	529.51	8.996	8	0.342658
1	I0001	MC	**-0.59**	0.1	2.082	525.6	22.393	8	0.004238
4	I0004	MC	**-0.61**	0.102	1.134	506.06	11.896	8	0.155926
48	I0048	Poly	**-0.829**	0.055	1.003	529.51	22.932	8	0.003454
27	I0027	Poly	**-0.832**	0.097	-0.185	530.49	12.823	8	0.11807
5	I0005	MC	**-0.836**	0.098	1.812	524.63	15.296	8	0.053631
24	I0024	Poly	**-0.976**	0.097	4.369	517.79	60.671	8	0
41	I0041	Poly	**-1.051**	0.056	-0.337	532.44	16.411	8	0.036858
6	I0006	MC	**-1.073**	0.098	4.24	497.27	36.493	8	0.000014
18	I0018	Poly	**-1.079**	0.095	-2.799	529.51	15.587	8	0.048686
40	I0040	Poly	**-1.124**	0.055	-1.08	533.42	14.53	8	0.068955
12	I0012	Poly	**-1.125**	0.095	0.795	530.49	6.223	8	0.622296
43	I0043	Poly	**-1.469**	0.059	6.595	508.02	82.415	8	0
28	I0028	Poly	**-1.606**	0.094	5.113	530.49	45.147	8	0
10	I0010	Poly	**-1.669**	0.094	-2.695	531.47	21.265	8	0.006477
30	I0030	Poly	**-1.926**	0.095	-2.605	530.49	20.124	8	0.009876
15	I0015	Poly	**-1.93**	0.095	1.245	532.44	23.914	8	0.002371
25	I0025	Poly	**-1.952**	0.095	-1.819	532.44	16.692	8	0.033478
26	I0026	Poly	**-1.954**	0.095	-0.561	532.44	8.22	8	0.412236
2	I0002	MC	**-2.166**	0.098	2.397	525.6	17.826	8	0.022569
34	I0034	Poly	**-3.385**	0.123	-0.228	531.47	33.133	8	0.000058
33	I0033	Poly	**-3.682**	0.134	-0.55	529.51	15.405	8	0.051736

APPENDIX I:

Calculated Percentages of Weightings on Test Framework 2012 ANA Grade 6 Mathematics Assessment Instrument (DBE, 2012)

1.

GRADE 6 MATHEMATICS CONTENT AREA	ANA GRADE 6 MATHEMATICS GDE WEIGHTING	MAPPED 2012 ANA ITEMS (52)	RESULTS OF MAPPED 2012 ANA ITEMS
Number "operations and Relationships"	50%	22	**42,3%**
"Patterns, Functions and Algebra"	10%	6	**16%**
"Shape and Space"	15%	8	**15,3%**
"Measurement"	15%	9	**15,3%**
"Data Handling"	10%	7	**13,4%**

2.

GRADE 6 MATHEMATICS DIFFICULTY LEVEL	ANA GRADE 6 MATHEMATICS GDE WEIGHTING	MAPPED 2012 ANA ITEMS (52)	RESULTS OF MAPPED 2012 ANA ITEMS
Easy	25%	10	**19,2%**
Moderate	60%	31	**60%**
Difficult	15%	11	**21,1%**

3.

GRADE 6 MATHEMATICS COGNITIVE LEVELS	ANA GRADE 6 MATHEMATICS GDE WEIGHTING	MAPPED 2012 ANA ITEMS (52)	RESULTS OF MAPPED 2012 ANA ITEMS
Knowledge of Basic Concepts	25%	21	**40,3%**
Application of Concepts	60%	21	**40,3%**
Non-Routine Problem Solving	15%	10	**19,2%**

APPENDIX J

ANA Question Analysis Template

ANNUAL NATIONAL ASSESSMENT QUESTION ANALYSIS: MATHEMATICS GRADE 6

EDUCATOR: ____ SCHOOL: ____

CLASS: ___________________

	NUMBER SEQUENCE, PLACE VALUES, 2 SHAPE, MULTIPLE, 3 D OJECT, DITSNCE, CAPACITY, **MASS, 2 D SHAPE(VIEW)**	NUMBER SYMBOL	PRIME NUMBER	VALUE OF A DIGIT NUMBER	ADDITION, SUBTRACTION, MULTIPLICATION & DIVISION	ORDER OF OPERATION	ORDER OF OPERATION	PROBLEM SOLVING/WORD SUMS	MISSING SUMBER IN A NUMBER SENTENCE	MULTIPLES	PROBLEM SOLVING/WORD SUMS	PERCENTAGES, DECIMAL,COMMON FRACTION	DECIMAL NUMBERS	INPUT & OUTPUT IN A TABLE FORM	FLOW DIAGRAM	2 D SHAPE(SQUARE)	ANGLES	INPUT & OUTPUT IN A TABLE FORM	3 D OBJECTS(VERTICES, SDGES, FACES)	LINE OF SYMMETRY	REDUCTION OF A OBECT ON THE GRID	DATA HANDLING	CONVERSION: PHYSICAL QUANTITIES: CAPACITY, DITANCE	MONEY(CALCULATION)	TEMPARATURE	DISTANCE	BAR GRAPH	STATS(MODE)	2 D SHAPE	PICTOGRAPH	TOTAL

THE QUALITY OF THE ANA INSTRUMENT

QUESTIONS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
MARK/S ALLOCATED	10	1	1	1	18	1	1	2	1	1	3	3	1	1	2	2	2	2	3	1	2	2	2	2	1	26	2	1	2	2	75
1.																															
2.																															
3.																															
4.																															
5.																															
6.																															
7.																															
8.																															
9.																															
10.																															
11.																															
12.																															
13.																															
14.																															
15.																															
16.																															