Botswana teachers’ experiences of formative assessment in Standard 4 Mathematics

by

Sello Editor Moyo
16168136

Submitted in fulfilment of the requirements for the degree

PHILOSOPHIAE DOCTOR in

Assessment and Quality Assurance

FACULTY OF EDUCATION
UNIVERSITY OF PRETORIA

SUPERVISOR:

Prof Surette van Staden (UP)

CO-SUPERVISORS:

Dr Celeste Combrinck (UP)

Prof Billy Fraser

March 2021

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

DECLARATION

I declare that the thesis, which I hereby submit for the degree PhD in Assessment and Quality Assurance 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.

…………………………………

Sello E. Moyo

18 June 2021

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

ETHICAL CLEARANCE CERTIFICATE

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

RESEARCH ETHICS COMMITTEE

CLEARANCE CERTIFICATE	CLEARANCE NUMBER: **SM 18/10/02**
DEGREE AND PROJECT	PhD Botswana teachers' experiences of formative assessment in Standard 4 Mathematics
INVESTIGATOR	Mr Sello Editor Moyo
DEPARTMENT	Science, Mathematics and Technology Education
APPROVAL TO COMMENCE STUDY	29 November 2018
DATE OF CLEARANCE CERTIFICATE	08 July 2020

CHAIRPERSON OF ETHICS COMMITTEE: Prof Funke Omidire

CC

Ms Bronwynne Swarts
Dr Surette van Staden
Dr Celeste Combrinck

This Ethics Clearance Certificate should be read in conjunction with the Integrated Declaration Form (D08) which specifies details regarding:

- Compliance with approved research protocol,
- No significant changes,
- Informed consent/assent,
- Adverse experience or undue risk,
- Registered title, and
- Data storage requirements.

ETHICS STATEMENT

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 he has observed the ethical standards required in terms of the University of Pretoria's *Code of Ethics for Research* and the *Policy and Procedures for Responsible Research*.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

DEDICATION

This thesis is dedicated to my wife, Mrs Goitsemodimo Moyo, and my children, “my boys”. I hope this achievement will inspire you as much as you inspired me with your words of encouragement and support.

To my late mother, Mrs Kathiku Monica Moyo, I continue to derive inspiration from how hard you toiled to raise me as your only son.

ACKNOWLEDGEMENTS

The success of this study was dependent upon the guidance, support and inspiration of the people whom God has placed in my life, as outlined below.

- Let me start by expressing the greatest acknowledgement to my research participants, teachers and pupils, who have dedicated time and contributed enthusiastically to completing this research. Without them, this research would never have been presented.
- I owe my deepest gratitude to my supervisors, Prof Surette van Staden, Prof Bill Fraser, and Dr Celeste Combrinck, for the scholarly support and constructive supervision towards attaining this PhD. Prof Van Staden, with her expertise in educational research, gave me a comprehensive learning opportunity, timely feedback, guidance and the right direction throughout the study period. Thank you, and may God continue to bless you abundantly. Prof Bill Fraser, may your soul rest in peace; my co-supervisor, you remain a very special man in this project whose sense of humility, kindness and generosity during my initial phase of proposal development has greatly inspired me.
- Dr Celeste Combrinck joined the project team as my co-supervisor at the right time. I appreciate her contribution to the Rasch Measurement Model presentation and her statistical expertise in data management and analysis, which shaped my knowledge. She is inspirational and motivated me to know more about the Rasch Model and available software.
- Prof Gavin T. L. Brown joined this PhD thesis as an examiner. Indeed, Prof Brown is not only a critical examiner, instead he is a mentor and expert scholar in Assessment, Quantitative, Data Analysis and Research; thus, I would like to express my special appreciation and thanks to him. Professor Brown, you have been a tremendous mentor for me in this thesis. I would like to thank you for your unique, formative feedback and fair cross-examination of the initialled submitted report, which initially left me with mixed feelings. However, your critical eye on this project, your brilliant comments and suggestions have transformed my knowledge about assessment, Rasch modelling techniques

and thesis reporting. More especially, your contribution on the resequencing of the chapters had made this thesis indeed "a quite remarkable" and "coherent" writing creation; thank you.

- I would like to sincerely thank the University of Pretoria Postgraduate Scholarship and Selection Committee, which enabled me to fulfil my PhD dream.
- My sincere gratitude to my lovely wife Goitsemodimo, for continued support, understanding, pre-proofreading of the chapters and patience throughout the whole process of my study, which was so demanding physically and emotionally. Thank you for your love, and I love you.
- I want to say "thanks a bundle" to my boys of the house, you were so good and understood my sleepless nights in the study room. To my entire family, father, sisters, nephews, nieces, aunts, uncles and friends, thank you for the support, love and encouragement throughout my entire PhD journey. Thank you ALL!

Sello E. Moyo

June 2021

ABSTRACT

In 21st century education, assessment is considered a crucial role as it bridges the gap between teaching and learning. This study aimed to determine the possible effects of an intervention for Botswana Standard 4 mathematics use of formative assessment (FA) strategies in the classroom to enhance pupils' higher-order thinking skills (HOTS). There is a recurrent educational debate in Botswana centred on pupils' poor achievement in basic numeracy, literacy, and life skills. Specifically, Standard 4 pupils who progress to Standard 5 and 6 were perceived to be unable to handle mathematical HOTS tasks. They were considered to be weak when measured on the common achievement in large-scale assessments such as the Trends in International Mathematics and Science Study (TIMSS) 2015.

With this backdrop in mind, the current study took a sequential embedded mixed, single-group, pre-test, intervention (training) and post-test design as a methodological strategy to investigate assessment practices related to higher-order thinking skills. The findings from the baseline survey (Phase I) revealed that teachers had some challenges in FA practices, including integration of HOTS tasks. The lesson from this phase's research data is how to assist the teacher in implementing FA effectively to enhance mathematics teaching. Post-intervention classroom observation showed that teachers practised the integration and implementation of some FA strategies (Phase II). The findings in Phase III also revealed that the pupils' post-interventional results in HOTS tasks significantly improved compared to their pre-assessment results. Additionally, the teachers' experience and reflections were found to be favourably inclined to support formative assessment higher-order thinking skills as a strategy to enhance mathematics teaching. Throughout the phases of the study, Kivunja's assessment feedback loop (AFL) model was underpinned as a theoretical point of departure to determine the classroom FA practice and integration of teaching HOTS in mathematics. The current study provides evidence of strategies to improve achievement levels in mathematics higher-order thinking skills in primary education in Botswana.

Keywords: *formative assessment, higher-order thinking, Post-positivism, pragmatism, achievement, Kivunja's assessment feedback loop, pre- and post-assessment, the Partial Credit Model (Rasch Measurement Theory).*

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

LANGUAGE EDITOR

Member South African Translators' Institute
www.language-services.online

PO Box 3172
Lyttelton South
0176
17 APRIL 2021

TO WHOM IT MAY CONCERN

The thesis titled "Botswana teachers' experiences of formative assessment in Standard 4 Mathematics" by Sello Editor Moyo has been proofread and edited for language by me.

I verify that it is ready for publication or public viewing in respect of language and style, and it has been formatted as per the prescribed style of the relevant institution.

Please note that no view is expressed in respect of the subject-specific technical contents of the document or changes made after the date of this letter.

Kind regards

Anna M de Wet

BA (Afrikaans, English, Classical Languages) (Cum Laude), University of Pretoria.
BA Hons ((Latin) (Cum Laude), University of Pretoria.
BA Hons (Psychology), University of Pretoria.

LIST OF ABBREVIATIONS

FA	Formative Assessment
BOT	Republic of Botswana
BS4MT	Botswana Standard 4 Mathematics Teachers
HOTS	Higher Order Thinking Skills
PISA	Programme of International Student Assessment
OECD	Organisation for Economic Co-operation and Development
UNESCO	United Nations Educational, Scientific and Cultural Organization
TIMSS	Trends in International Mathematics and Science Study
ETSSP	Education and Training Sector Strategic Plan
MOBE	Ministry of Basic Education
RNPE	Revised National Policy on Education
BEC	Botswana Examination Council
MOESD	Ministry of Education and Skill Development
PSLE	Primary School Leaving Examination
NDP	National Development Plan
EFA	Education for All
MDGs	Millennium Development Goals
SDG	Sustainable Development Goals
ICT	Information Communication Technology

HIV	Human Immunodeficiency Virus
AIDS	Acquired Immune Deficiency Syndrome
ECCE	Early Childhood Care and Education
HE	Higher Education
DPE	Diploma in Primary Education
SBA	School-Based Assessment
SFAT	Standard Four Attainment Test
FAST	Formative Assessment for Students and Teachers
SCASS	State Collaborative on Assessment and Student Standards
CA	Continuous Assessment
AFL	Assessment for Learning
AfLA	Assessment for Learning in Africa
MFDP	Ministry of Finance and Development Planning
RNPP	Revised National Population Policy
GDP	Gross Domestic Product
TVET	Technical and Vocational Education and Training
JCE	Junior Certificate of Education
BGCSE	Botswana General Certificate of Secondary Education
BQA	Botswana Qualification Authority
ETP	Education and Training Providers

HRDC	Human Resource Development Council
OBE	Outcome-Based Education
INSET	In-Service Teacher Training
CAPA	Cultural Studies, Creative and Performing Arts
LOTS	Lower-Order Thinking Skills
SA	Summative Assessment
PARCC	Partnership for Assessment of Readiness for College and Careers
FAS	Formative Assessment Strategies
LG	Learning Goal
SC	Success Criteria
PASA	Peer and Self-Assessment
WAG	Welsh Assembly Government

Table of Contents

List of Figures

List of Tables

1 CHAPTER ONE: INTRODUCTION

The study aimed to find the possible effects of an intervention for Botswana Standard 4 mathematics teachers' (BS4MT) use of the Formative Assessment (FA) strategies in the classroom to enhance pupils' High Order Thinking Skills (HOTS). The study attempted to contribute to the shifting body of knowledge for pupils' learning in the 21st century, which strongly emphasises the important impact of assessment strategies, methods and delivery on the learning outcomes with particular attention to HOTS (Al-Wassia, Hamed, Al-Wassia, Alafari, & Jamjoom, 2015; Patel, Yoskowitz & Arocha, 2009, Pellegrino, DiBello & Goldman, 2016).

According to Balakrishnan, Nadarajah, Vellasamy and George (2016), education aims at enhancing individuals' problem-solving and critical thinking skills, which is known as *HOTS* to some scholars. Kruger (2013, cited in Chen 2016) states that HOTS involves concept formation, critical thinking, creativity/brainstorming, problem-solving, mental representation, rule use, reasoning, and logical thinking. HOTS is essential and pertinent to educate the pupils of the 21st century who face complex real-life problems, which often needs a complex solution (Rajendran, 2008). One of the teachers' roles in education entails the skills development of the pupils' HOTS based on the teaching methods and approaches that are essential with the transformation to enhance pupils' thinking skills accordingly (Serin, 2013). Scholars in education have suggested that teachers who implement the teaching approaches for HOTS are helpful for the pupils' academic growth' and work independently with minimum guidance from their teachers (Fahim & Masouleh, 2012; Goethals, 2013; Yang & Gamble, 2013).

In supporting the teaching approaches for HOTS, Thomas and Thorne (2009, as cited in Arase, Kamarudin & Hassan, 2016), had suggested that classroom lessons should be planned and designed to allow integration for HOTS components. For cognitive behaviour, the concept of HOTS is based on the Taxonomy of Educational Objectives popularly known as Bloom's Taxonomy (Bloom et al., 1956). This system involves a six-level hierarchical progression for the categorisation of human cognitive behaviour from a most basic to a higher order level of cognitive processing namely knowledge, comprehension, application, analysis, synthesis, and evaluation (Chen, 2016; Chidozie, Yusri, Muhammad-Sukri & Wilfredo, 2014; Nenty, Adedoyin, Odili & Major, 2007).

> *The first two levels of Bloom's Taxonomy involve accumulation and understanding of information only, while the other four levels which are often classified as higher-order thinking involve the application of such information for finding a solution to real-life problems, for creativity and critical thinking and judgment. These four levels of cognitive thinking are the more desirable ones for development and educators have been increasingly charged to develop these among the learners to enhance their ability to contribute to the development of society* (Nenty et al., 2007).

According to Nenty and Umoinyang (2004) application involves the transfer of what is learnt to concrete and new problem-solving situations. "After pupils learn and understand general ideas, rules and procedures, generalised methods, principles, theories, laws and formula in their area, at this level pupils should be able to put these into concrete applications in solving real-life problems, in new situations other than the examples involved in the acquisition of this knowledge" (Nenty, 2008).

Furthermore, analysis involves the skills necessary to break down a piece of communication (for example, analysis of a sentence or a paragraph) or works of art "into its constituent elements or parts such that the relative hierarchy of ideas is made clear and/or the relations between the ideas expressed are made explicit" (Bloom, 1956, p. 205 cited in Nenty & Umoinyang, 2004). In mathematics, for example, the analysis calls for pupils' skill to "tear apart" and detect the fault in a line of mathematical reasoning or proof, discover mathematical relationships and construct proof to substantiate such discovery (Nenty, 2008).

Synthesis is another cognitive level of the HOTS, which involves pupils' ability to create something unique or new by putting together related elements, parts, pieces, and components. According to Chen (2015), synthesis includes problem-solving in mathematics or creative skills like putting together concepts in a unique organisation to form a literal work in writing or in a speech that conveys to other novel ideas, feelings, understandings, and experiences, or create unique non-literal work or art (Nenty, 2008). Synthesis skills include divergent thinking which results in the generation of novel responses that are original, varied, and unusual. Such skills are scored in terms of the fluency of ideas, flexibility, and originality (Nenty & Umoinyang, 2004).

Chen (2016) explained evaluation skills as the pupil's ability to make a quantitative or qualitative value judgment about a piece of communication, an invention, a procedure or method, a proposal, and an idea based on criteria that may be given or derived; internal or external to what is being assessed. Evaluation as an educational objective is concerned with the pupil's ability to organise their thoughts and knowledge to reach a logical and rational judgment or decision which are dependable (Nenty, 2008).

These HOTS are activated when pupils of any age encounter unfamiliar problems, uncertainties, questions, or dilemmas. Successful application of these skills results in explanations, decisions, performances, and products that are valid within the context of available knowledge and experience, and promote continued growth in higher-order thinking, as well as other intellectual skills (Chen, 2016; Nenty, 2008).

According to Nenty et al. (2007):

> *High-level cognitive questions can be defined as questions that require pupils to use higher-order thinking or reasoning skills. By using these skills, pupils do not only remember factual knowledge, instead, but they also use such knowledge to solve problems, analyse, create and evaluate. It is believed that this type of question reveals the most about whether or not a pupil has truly learned that which is necessary for him or her to contribute to the development in society. But several research studies have indicated that student possesses limited abilities to think at higher levels of cognition.*

In Botswana, for instance, the main challenge facing the education sector is to improve the quality of education at all levels (UNICEF, 2019). This challenge is evidenced by the declining performance of primary school pupils in international and national assessments which are periodically administered, such as the Trends in International Mathematics and Science Study (TIMSS) in 2003, 2007, 2011 and 2015. TIMSS 2015 has described Botswana pupils who participated in the study as unable to apply HOTS such as critical thinking and problem-solving in mathematics (Masole, aa, Guga, Pharithi & BEC, 2016).

Notwithstanding the pupils' low mathematics achievement, education, along with development, remains a major priority in Botswana. The Botswana Education and Training Sector Strategic Plan (ETSSP) was recently envisioned and designed to attain a significant milestone in the collective effort of Botswana as a nation to bring about a more diversified knowledge-based economy (BOT, 2015). ETSSP aims to promote equitable, effective, efficient, and quality education and a rationalised education sector by restructuring the Ministry of Basic Education (MOBE) functions.

The ETSSP of 2015 seeks to refocus education and training towards the fulfilment of social and economic aspirations identified in the Revised National Policy on Education (RNPE) (BOT, 1994; BOT, 2015), to raise educational standards at all levels while emphasising mathematics, science, and technology. The central focus of the transformational shift through ETSSP is promoting pupils' learning outcomes and assessments. Thus, improving teaching and learning at all levels and undertaking intensive teacher development. This process develops appropriate assessment patterns by setting up a National Assessment Framework to organise school-based assessment and measuring skills better and linking with a national assessment. Activities will be focused on the qualitative and substantive diagnosis and effective targeted support of a new curriculum and assessment patterns that will also assess skills development, leadership, management and planning, capacity building, school monitoring, and foster partnerships (BOT, 2015; p. 41).

In an attempt to contribute a solution to the poor mathematics achievement in Botswana, this study proposes to selectively explore the Standard 4 teachers and their pupils' performance in mathematics within the specified education reform context, with emphasis on the application of FA strategies during instruction to provide feedback, to adjust ongoing teaching and learning in the classroom, and improve pupils' achievement in mathematics and HOTS development.

1.1 Problem Statement

Evidence of low mathematics achievement patterns has been discussed widely in a different educational setting by various stakeholders. The following section presents and analyses low pupils' achievement in mathematics, specifically in the context of Botswana.

1.1.1 Low mathematics achievement in both international and national assessment

Botswana's lower primary education as a means of ensuring that pupils acquire the relevant skills that the society intended them to develop as stipulated in the mathematics curriculum, is not achieving its intention. The pupils' performance in the TIMSS 2015 study has revealed that Botswana pupils have not acquired substantial mathematics skills when ranked along with their international cohort. It was evident through the TIMSS 2015 study in which Botswana's older pupils (Standard 6 instead of Standard 4 took part), were ranked third from the bottom of the 59 countries, and the 425 000 pupils who participated in TIMSS 2015.

TIMSS created a set of international benchmarks to provide countries with more meaningful descriptions of what pupils know. TIMSS defines four categories of benchmarks, namely:

- scores between 400 and 475 points are classified as achievement at a low international benchmark;
- scores between 475 and 550 points are classified as achievement at an intermediate international benchmark;
- scores from 550 to 625 points are classified as achievement at a high international benchmark; and
- scores above 625 points are classified as achievement at an advanced international benchmark.

Botswana pupils' mathematics mean performance was recorded at approximately 391 (*SE*= 2.0), below the international mean (500), as an indication of failing compared to their cohorts (Masole et al., 2016). The five lowest-performing countries were Botswana (391; *SE* = 2.0), Jordan (386; *SE* = 2.4), Morocco (384; *SE* = 2.3), South Africa (372; *SE* = 4.5) and Saudi Arabia (368; SE 4.6). The top-performing countries were from Asia, and the Northern Hemisphere and the lowest-performing countries were Middle-Eastern and African (Reddy et al., 2016).

Furthermore, it was observed that the Botswana pupils who took part in the TIMSS 2015 study were unable to apply HOTS, such as critical thinking and problem-solving in mathematics (Masole et al., 2016). According to Masole, et al. (2016), TIMSS studies have identified pedagogical issues such as availability of resources and teacher effectiveness as being among the background variables which are negative indicators of Botswana's achievement in mathematics. Despite these observations, there are other contributing factors tied to pupils' achievements. For instance, historically, an international report on achievement by Coleman (1966) gave credit to the pupils' family background as the main reason for pupils' success in school. His findings suggested that children from low-income families and homes, lacking the prime conditions or values to support education, could not learn, regardless of what the school did (Coleman, 1966). Edmonds (1979), and other scholars, refused to accept Coleman's report as conclusive, although they acknowledged that family socio-economic background does indeed make a difference. They set out to find schools where children from low-income families were highly successful thereby proving that schools can make a difference (Edmonds, 1979).

With this Coleman report (1966) in mind, pupils' achievement involves the interaction of at least six factors, which include the child, home, school, curricula, teacher and the approaches to teaching (Hattie, 2009). The meta-analysis by Hattie (2009), revealed that the interactions between these factors could influence the pupil's achievements. Most importantly, the teachers' conceptions of teaching, learning, assessment and the pupils are all related to teachers' views on whether all pupils can progress and whether achievement for all is changeable (or fixed) and whether progress is understood and articulated by teachers. The teacher remains the main factor tied to classroom teaching practice which can either enhance productive or unproductive learning among pupils.

Evidently, in the Botswana context, Raboijane (2005) pointed out that the "teaching to the test" strategy has resulted in a distortion of the curriculum for many pupils, narrowing it down to basic low-level skills. Such a teaching conception corroborates the findings from Fetogang's (2016) study, which showed a tendency among Botswana teachers to teach more to fulfil the alignment with the previous examination papers than those of the syllabus. In effect, teaching and assessment should all look at the curriculum for content and objectives and not at the content of past public examinations (Fetogang, 2015).

Also, according to Volante (2004, p15, cited in Fetogang. 2016), critics of the practice of teaching to the test argue that:

> *Pupils who are taught to the test lack a comprehensive and lasting understanding of the subject matter; even if it raises test scores - which it fails to do. Pupils may not truly grasp the key concepts of the domain, because of the teaching to the test centres on rote memorisation while excluding the strengthening of creative skills, and abstract thinking ability. Teachers who want to raise test scores, contrarily, must promote a deep conceptual understanding of the subject matter*.

Low achievement in mathematics results in Botswana was even observed in the pupils' national achievement examinations. For instance, a study by the Botswana Examination Council, (BEC, 2007), following the Standard 4 assessment, revealed that the mean performance for mathematics was averaged at 30%. Also, this less stable achievement in mathematics has further been seen upon the release of the yearly national examination results. Stakeholders are dissatisfied with the outcomes of the national examinations (Moyo & Nenty, 2017) and according to the Ministry of Education and Skills Development (MOESD), the pattern of undesirable achievement is still a recurrent phenomenon observed at the end of primary school leaving examination (PSLE) across the core subjects, including mathematics (MOESD, 2016). Evidently, in 2017, the mathematics pass rate at PSLE was among the lowest averaged at 65.94%, lower than in 2016 at 66.4%, for the percentages of candidates awarded Grade C[1] or better (BEC, 2017).

---

[1] Grade C means the candidate obtains a numeric aggregate of 60 - 69

The mathematics achievement problem also seems to cascade towards secondary schools. It was discernible upon the release of the 2017 Botswana Junior Certificate Examination (BJCE) results; where mathematics results were also observed to be the most failed core subject (BEC, 2017). Table 1.1 reflects on the more or less stable results of the secondary school candidates who have been awarded grade C or better in mathematics at BJCE conducted by BEC. The patterns of consistent low performance in mathematics were similar for the last seven years, as illustrated in

Table 1.1: BJCE Mathematics Percentages of Candidates Awarded Grade C or Better for Seven Years

Year	2012	2013	2014	2015	2016	2017	2018	2019
Percentages of C or Better	33.9%	28.7%	24.5%	24.4%	24.84%	26.20%	28.26%	25.57%
Total No of Candidates	37801	38791	40519	41893	41432	40886	41063	41048

The continued low achievement in mathematics by Botswana pupils' set back the government efforts to channel resources towards developing mathematics and science-related fields like mining and engineering (Tabulawa, 2014). Against this backdrop, the influence of teacher conceptions and beliefs in the daily school practice remains an important factor for pupils' academic success (Brown & Remesal, 2017). Teachers' understanding of the purpose, role, and effects of assessment have been published (Barnes, Fives, & Dacey, 2015). So far, the two predominant purposes of assessment (i.e., formative improvement vs summative evaluation) have created substantial tension among teachers (Bonner, 2016). Essentially, teacher's thinking about assessment orientation and beliefs tend to reflect the social, historical, and cultural priorities established within the jurisdiction in which teachers are employed (Fulmer, Lee, & Tan, 2015; Brown & Remesal, 2017). For instance, Botswana is a high-stake tests-oriented administration that may influence teachers' behaviours on classroom assessment, teaching, and learning strategies that are mostly about hastening the curriculum's completion. Therefore, their behaviour forces Botswana teachers to use the traditional technique of "teaching to the test" to prepare pupils for national examinations (Fetogang 2015; Lerner & Tetlock, 1999, Raboijane, 2005).

Nevertheless, Hattie (2009) emphasised that teachers and schools must make learning exciting, engaging and enduring (effective classroom engagement to enhance HOTS among pupils). What matters are conceptions of teaching, learning, assessment and teachers having expectations that all pupils can progress. Achieving this for all is not guaranteed (Rubie-Davies, 2006, 2007; Rubie-Davies, Hattie, & Hamilton, 2006; Weinstein, 2002). Common teaching approaches that are associated with pupils learning include making various strategies successful through effective planning. In such particular cases, the teachers should discuss teaching and planning, attention to learning intentions and success criteria with other teachers. Efforts should constantly be made to ensure that teachers are seeking feedback information on the successful impact of their teaching on their pupils learning (Black & Wiliam,1998; Hattie, 2009).

The following section will report some education reforms and implementation in Botswana as an attempt to provide the reader with an overview of the significant strides of the education and training background in which the department of primary and secondary education are currently operating as an effort to provide quality education. The section will consider the brief history of education and some major reforms, the teaching of mathematics in lower primary schools and pupils' assessment.

These foregoing issues are very integral to the current study since they are potential elements to give the direction and context to teaching and learning among Standard 4 pupils in the current educational set-up, most importantly on the practices of FA in the classroom and teaching of HOTS.

1.1.2 Botswana: Brief historical overview and education reforms

Upon the attainment of independence, Botswana mapped the direction of their education. The country introduced two major education policy reviews (Mareka, 2015: Monyaku & Mmereki, 2011). The initial National Education Policy (1977) was a significant milestone, as it provided a sound framework for education planning and provision of education (UNESCO, 2006). In 1994, the country introduced the Revised National Policy on Education (RNPE) (Republic of Botswana [BOT], 1994) and recognised the goal of education as setting up Botswana from a transition-based economy to an industrial economy to enable competition with other countries of the world.

The Botswana Government continues to place a great emphasis on the development of human resources for job readiness and preparing for a knowledge-based and globally competitive economy (BOT, 2016b). The attainment of Vison 2036 has been operationalised through the use of mathematics, science, technology, and innovation to propel economies to high levels of efficiency as well as the key to support socio-economic development (BOT, 2016bon, 2036). To that effect, the education priorities are strongly aligned with Botswana's Vision 2036 (BOT, 2016b), National Development Plan 11, the RNPE (BOT, 1994) and refocused in the recent Education and Training Sector Strategic Plan (ETSSP).

Thus, ETSSP pays great attention to inclusive and life-long learning policies and in doing so is aligned with international contexts and reflects long-standing commitments to Education for All (EFA), the Millennium Development Goals (MDGs) and Sustainable Development Goal (SDG). Cross-cutting issues, for example, gender, Information Communication, and Technology (ICT), Human Immunodeficiency Virus (HIV) and Acquired Immune Deficiency Syndrome (AIDS) are, whenever appropriately, mainstreamed across each sub-sector. The sector plan has a strong leaning towards ensuring a continuum from Early Childhood Care and Education (ECCE) to Higher Education (HE), non-formal and continuing education (BOT, 2015).

Figure 1.1 (next page) describes a commitment to inclusive, life-long learning, whilst paying due attention to issues of quality, relevance, efficiency and delivery. The current study has targeted the foundation phase of the first four (4) years of schooling, which the ETSSP has also considered as a crucial stage. According to the ETSSP (BOT, 2015), the foundation phase is aimed at "providing an excellent start in education so that they have a better foundation for future training" (BOT, 2015, p. 22) shown in Figure 1.1. Learning activities during these first four years are built around Literacy, Numeracy and Life Skills (BEC, 2007). In doing so, ETSSP gives serious consideration to the management and improved pupils assessment, monitoring and evaluation of the sector.

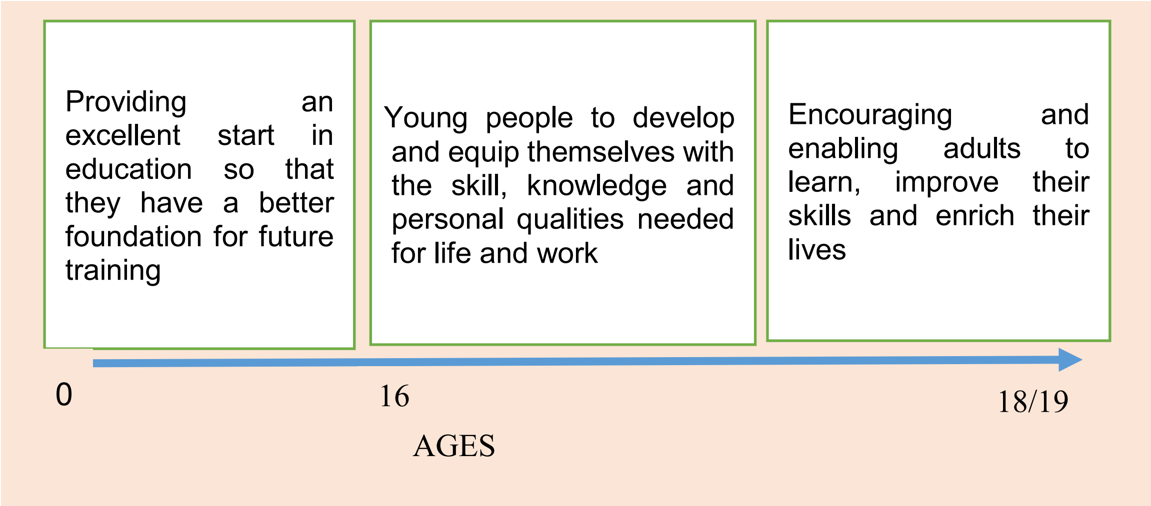


Figure 1.1: Continuum of the education system
Adapted from ETSSP (BOT, 2015)

A major setback in the ETSSP is the lack of implementation. Little is known in terms of the setting up of the National Assessment Programme which was proposed to customise and yield accurate information about the bottlenecks of the education system. Additionally, the intended planning for intensive teacher development to support a better organised school-based assessment and measuring skills must still be realised. In the light of these setbacks, the current study, however, is consistent with ETSSP reform to seek improvement in the context of mathematics teaching and learning at Standard 4 and specifically formative assessment.

1.1.3 Teaching mathematics at lower primary school

Statistics Botswana (2012) revealed that teachers in Botswana public schools are formally qualified. At the primary level, 98% of teachers had a Diploma in Primary Education (DPE) in 2012 (Statistics Botswana, 2012). Primary school teachers in Botswana are qualified by either holding a DPE from a primary college or a Bachelor's degree in primary education from the University of Botswana or universities abroad.

During teacher training, student-teachers in primary colleges of education spend 50% of their course time on subject content, including mathematics content and 50% on pedagogy. As part of the teacher education training, in addition to academic and pedagogical studies, primary school student-teachers have to complete a six months course of teaching practice. During the teaching practice, student-teachers in schools' practice under the supervision of experienced teachers for three to four months. Similarly, the universities offer some further studies for qualified teachers, in which both subject content and pedagogy are also covered. In all cases, the lecturers would then visit student teachers who are doing teaching practice to assess their performance in schools, and their assessment scores contribute to their final grade (Masole et al., 2016).

Low education outcomes in schools, despite the presence of qualified teaching staff, is an undesirable phenomenon and may lead one to postulate that teachers may have limited knowledge to teach mathematics. Evidently, a study by Carnoy, Chisholm and Chilisa (2012) titled the *Low achievement trap; comparing schooling in Botswana and South Africa* revealed some association of teachers' knowledge and quality of teaching. According to Carnoy and his colleagues:

> *Teachers' mathematics knowledge (as measured by a test given to teachers), the quality of their teaching (as measured by assessing videotapes of lessons) and the amount of mathematics teaching they actually did during the year (as measured by an assessment of student notebooks) were all statistically related to each other. Teachers who did better on our mathematics test also, on average, taught mathematics more effectively and taught more lessons on the topics covered by our student test. This relationship was more evident in North West Province than in South-East, Botswana* (Carnoy et al., 2012).

Carnoy et al.'s (2012) study appear to corroborate Masole et al.'s (2016) study, in the TIMSS 2015 study, which identified that teacher pedagogy seems to be a problem for effective teaching of mathematics in primary schools in Botswana. There would, therefore, seem to be a definite need for in-service training of teachers in pedagogy to augment the training which they have acquired during pre-service training. In addition, some studies on the demographic of the teachers (qualification and experience), for instance, obtaining a graduate degree was found to be related to a higher level in teachers' assessment literacy (Hoover, 2009; King, 2010), and schooling experience also impacted teachers' assessment decisions (Campbell & Evans, 2000). However, Brown's (2008) study did not find a significant difference between teacher demographic and training experience in assessment.

Teachers in Botswana are employed as permanent and pensionable while others hold temporary contracts for two to three years. The primary school teachers are allocated for teaching at any level and all the subjects except in a few selected schools in which upper primary offer some specialised teaching. At the Standard 4-level, teachers are a jack of all trades and teach all the subjects, including mathematics. Interestingly, however, the RNPE (1994) had long recommended teacher development to take charge of major core subjects like mathematics as well as "from Standard 4 onwards pupils should gradually be introduced to teaching by specialist subjects" (RNPE, 1994, p 20). Lack of specialisation could be a barrier to achieving basic education excellency. According to Mokotedi (2013), one cannot be a teacher to efficiently and effectively teach all school subjects across the curriculum. Thus, lack of specialisation has an adverse effect on the thoroughness, mastery, skills; and the efficiency with which subjects are taught cannot be overemphasised.

An attempt has been made to put strategies in place for the improvement of HOTS. For example, the BEC had engaged a consultant from the Australian Council for Educational Research (ACER) in 2008, to train its staff and teachers on construction and implementing HOTS items that would allow pupils to develop critical and creative thinking, problem-solving, and performance skills (Masole et al., 2016). However, the training did not bear fruits which can be observed on the ground in terms of policy, and practice in the classroom settings.

The RNPE (1994) remains the only reference document based on the current implementation of the primary school mathematics curriculum. The curriculum emphasises mathematics for all pupils and suggests that in teaching mathematical concepts, teachers should first use children's prior knowledge of mathematical concepts before they begin teaching. This recommendation has been clearly articulated in mathematics curricula preambles (MEOSD, 2002). According to Mosothwane (2014), the current primary mathematics curriculum is based on the philosophy of group work and discussion. The teachers, therefore, ought to implement those teaching approaches which encourage pupils to construct their knowledge and skills. Mosothwane (2014) further claimed that the Botswana mathematics curriculum is in line with Vygotsky's 1978 theory of cooperative learning which asserts that “What children can do together today, they can do alone tomorrow”. Hence, the current study is consistent with the Standard 4 mathematics curriculum which seeks to improve the teaching and learning of pupils' HOTS using FA strategies.

1.1.4 Assessment at Standard 4 Level

Botswana is an administrative compliance examination country. According to Fetogang (2015), this statement means that Botswana has a history of high emphasis on the use of examination results to judge the quality of schools and that of pupils.

Standard 4 pupils are tested using the mandatory school-based assessment (SBA) and at the end of the Standard 4 period, they are also assessed using tests commonly known as the Standard 4 Attainment Test (SFAT). In the latter, assessment results are used by the schools to decide whether they can promote pupils to standard five (BEC, 2015). Passing the Attainment Test is a prerequisite for pupils to proceed to upper primary school (Monyaku & Mereki, 2011). The former, being the SBA, therefore, is supposed to be formative, and it is used to monitor the pupils’ learning (Masole et al., 2016). “Formative assessment in the case of SBA is not a test but a process” (Popham 2008, p 6). In this view, the process produces a score but rather than a qualitative insight into pupils understanding (Shepard 2008), when the results are actually used to adapt the teaching to meet pupils’ needs (Black & Wiliam 1998a, 140; Bennett, 2011). In this context, mathematics, as one of the core subjects at Standard 4 level, and pupils' learning progress, are also supposed to be monitored through SBA. Masole et al. (2016) have argued that SBA is supposed to be treated more like a formative assessment (FA), and it also should be related to the ongoing process of learning in the classroom.

In 2017, the Formative Assessment for Students and Teachers (FAST) State Collaborative on Assessment and Student Standards (SCASS), defined FA as a planned, ongoing process used by all pupils and teachers during teaching and learning to improve pupils' understanding of intended disciplinary learning outcomes and, supporting pupils in becoming more self-directed learners. The primary purpose of FA is considered as a means to:

> *Improve [pupils'] learning, as both [pupil] and teacher respond to the information that it provides. Information is needed about what knowledge, understanding, or skills students need. By finding out what [pupils] currently know, understand, and can do, any gap between the two can be made apparent. Assessment is the process of gaining information about the gap, and learning is about attempts to reduce the gap* (Wiliam, 2010).

According to Black and Wiliam (2009), FA has different names which include continuous assessment (CA), Authentic Assessment, SBA and Assessment for Learning (AfL) and these terms are used interchangeably (Firn, 2016; Kanjee, 2017; Wiliam & Thompson, 2014). For this study, FA will be used. In Botswana, teachers and schools use FA (either referred to as CA or SBA) as administrative compliance and record-keeping exercise of an individual pupil through quizzes, topic tests, end-of-term tests, and mock examinations.

According to Motswiri (2004), the reality at the classroom level in Botswana is that the use of SBA has been inclined towards summative purposes more than FA. The tension between formative purposes of assessment and summative purposes of assessment is striking. Teachers are often under pressure to prepare pupils for external examinations, which have to do with administrative compliance purposes. This use deflates the assumption that FA seeks to guide instruction and to assist pupils in learning. For instance, in the recent interview between the Botswana Guardian newspaper and the Regional Education Director, Kweneng District, the director reported that “every month the school heads must give monthly reports on the pupils, ensure monthly tests are written and the results are registered so that …can track the performance" (Botswana Guardian, 2018. p.8).

From the director's interpretation of monthly tests, FA is commonly erroneously labelled, FA is administrative compliance, and it is not used to adjust initial instruction. According to Strutchens and Petit (2017), an assessment used in the context of the director's statement cannot deliver the promised pupils' growth enabled by the FA process that should occur within the daily stream of planning and instructions.

Mills, Strutchens and Petit (2017) assert that assessment practitioners have general misconceptions across African states. Thus, the majority of educators in Sub-Sahara Africa, including teachers, principals, policymakers, and other leaders seem to have had difficulty distinguishing between summative assessments and the formative assessment process. Confounding these two processes is the word "assessment" that appears in both. For many, this conflated thinking about formative and summative assessment has meant that to adopt FA practices is to layer additional assessment events into the school calendar rather than incorporating a process of eliciting and using evidence of learning to adjust instruction into daily practice (Mills et al., 2017).

Fundamentally, the drive for FA is to improve pupils' understanding, which can mean that the pedagogical knowledge of the teacher is essential to know what the next step should be for the pupils (Baird, 2010; Baird, Andrich, Hopfenbeck & Stobart, 2017). FA literature affords extensive evidence that FA has the potential to improve pupil learning if well implemented and well perceived by pupils (Black & Wiliam, 1998, 2009; James & Pedder, 2006; Kyaruzi, 2017; Wiliam, 2011; Wiliam, Lee, Harrison, & Black, 2004;) (A detailed discussion of FA is done in Chapter 2). However, in the Botswana context, there is seemingly little integration of instructional teaching pedagogies and the FA, which can possibly enhance pupils HOTS.

1.2 The Rationale for the Study

Mathematics is considered as fundamental for learning in all aspects of life hence teaching and learning of mathematics should emphasise "the ability to develop and apply mathematical thinking to solve a range of problems in everyday situations which is important for all learners" (Department of Curriculum Management, 2014, p. 2). Through mathematics, pupils gain the foundation for competitiveness in the global economy (Mareka, 2015; Nkate, 2008; Venson-Moitoi, 2014). However, achievement in mathematics for Botswana's schools has been unsatisfactory and cannot help the country much to achieve its full economic potential (BEC, 2014) unless the current practices of teaching and learning mathematics are transformed and possibly towards FA approaches (Adedoyin & Chilisa, 2004). It is therefore important to embark on a research inquiry to learn about what is going on in the learning of mathematics in the classroom formative assessment to understand the discrepancy in pupils' achievement of HOTS and the magnitude of the contextual problem better. The rationale for this study is therefore twofold– (1) to explore FA practices in the Botswana context, and (2) to propose a possible effect of an intervention for BS4MT use of the FA strategies in the classroom to enhance pupils' HOTS.

According to Pellegrino et al., (2016), FA instructional settings help to guide the quality and value of the information about constructs that provide relative support for ongoing classroom teaching and learning. The benefits of the effective use of FA strategies possibly increase proficiency in mathematics. That is to say; pupils are provided with opportunities to deepen their mathematical knowledge and reasoning, to come more formally into contact with abstract and logical reasoning, and to appreciate and apply the communication possibilities that the mathematics medium offers better (Kanjee, 2017; Department of Curriculum Management, 2014). These opportunities will possibly result in pupils achieving good results in mathematics.

FA has the potential to contribute to the narrowing of the achievement gap between struggling and non-struggling pupils (Stiggins, 2002, 2005). Given the benefits of FA, this research seeks to explore the teacher's use of FA strategies to enhance the HOTS among the Botswana pupils who happen to perform poorly in mathematics both at the international and national level. Mathematics remains the foundational subject for many pupils who are vying for career ambitions in the various science fields. It is therefore imperative that relevant research in the area of FA to improve both pupils' thinking skills and academic records in mathematics is conducted.

Specifically, the rationale behind the current study claims that a properly managed and fully implemented FA may prepare pupils to develop HOTS at the Standard 4 level to perform better in the classroom and national examinations. The integration of the two domains (FA & HOTS) employs similar strategies to promote the quality of thinking and learning (Welsh Assembly Government, WAG, 2010).

1.3 Purpose and Research Questions of the Study

The anticipated potential significance of this study is the enhancement of mathematics teaching and learning through the generation of knowledge and teacher capacity development in the effective integration and use of FA to improve pupils HOTS. The involvement of BS4MT in this study would assist teachers in their specific classroom contexts to plan instruction and know when and how to elicit pupils' understanding of all aspects of HOTS relative to misconceptions and/or obstacles. They will also know how to take immediate steps to close the gap between where the pupils are and where they need to be.

This study intends to benefit all teachers, managers and improve the level of academic achievement of the pupils. The current study is positioned to contribute directly to the teachers' assessment literacy (specifically FA literacy) at the primary school level. Apart from enriching the researcher's practices as an educational measurement specialist, the results of the study will provide information to managers in education to support teachers. Pupils will benefit maximally as both teachers and managers would refocus to improve achievement in mathematics. Given that there was no detailed baseline data for teacher practices in FA and teaching HOTS in mathematics at the primary school level, this study is the first empirical study to establish and design intervention, and, hopefully, to set the ground rules and promote further research in Botswana relating to FA and HOTS.

Additionally, this study provides an important opportunity to advance the understanding of FA, not only for Botswana, but the study will also attempt to contribute to the body of knowledge on the project for Assessment for Learning in Africa (AfLA). AfLA is a research project that was ongoing in primary schools in Tanzania, East Africa and two sites in South Africa, during the development of this PhD. The project was aimed at improving teachers' numeracy skills and their understanding of how numeracy can be more effectively communicated to their young pupils and to improve children's outcomes in the form of test results through appropriate assessment strategies (Aga Khan University, 2017). Thus, this study will offer an insight into AfLA towards teacher development and the use of assessment for improving learning outcomes in the core curriculum area of numeracy in the context of Botswana school settings. Therefore, this thesis investigates Botswana teachers' experience of formative assessment in standard 4 mathematics.

1.3.1 Purpose

The purpose of the study was to determine the possible effects of an intervention for BS4MT use of the FA strategies in the classroom to enhance Standard 4 pupils' HOTS.

1.3.2 Research questions

The main research question that guided this study was:

To what extent do the intervention for teaching and learning within the FA strategies enhance Standard 4 pupils' HOTS in mathematics?

In an attempt to answer this question, one needs to ascertain the details of the teaching processes with teachers at the school level first, to understand the process and procedures which are involved in the particular practice for FA, the teaching of HOTS and pupils' level of achievement for HOTS. This evidence was achieved through three phases of this study directed by the seven research sub-questions classified.

Phase I

In **Phase I,** the research sub-question addressed was:

1. What are the current practices in FA and teaching of HOTS in Botswana primary schools?

To fully understand and appreciate the concerns and limitations imposed by the current practice in FA and teaching of HOTS, one needs to consult with the practitioners as well as observe them in classroom practice. For this study, a sufficient sample was used to ascertain the problem context; hence a mixed research methodology has been deemed appropriate in the form of baseline survey research. At this phase, a closed-ended questionnaire and observation tool were used to collect information from teachers on their practices in FA and teaching of HOTS.

According to Van Staden and Zimmerman (2017), data collection through a questionnaire alone is no longer considered effective due to the difficulty to report timely data about best practices. Zimmerman (2011, cited in Van Staden &

Zimmerman, 2017), criticised the use of a teacher questionnaire in relation to teaching practices in poor school performance contexts because it may be problematic, as teachers may feel vulnerable and defensive, resulting in unreliable or unrealistic answers. Thus, Van Staden and Zimmerman (2017) argued that teachers reported incorrect practices when they were asked about the priority given to identifying instructional practices that relate to high achievement. That is in the questionnaire data, they indicated effective instructional practice strategies, while in actuality, many did not implement strategies in the way envisioned to enhance learning.

For the reason of unreliable teacher questionnaire answers, the current study also considered using a questionnaire, followed by some classroom observations and reviewed the pupils' work during the baseline survey. The observation and checking of pupils' work helped, in addition, to ascertain the extent of the actual implementation of FA and the teaching of HOTS are done.

The use of classroom observation in this study is consistent with the findings of Molina et al. (2018), who, in their World Bank Group research, measured teaching practices through the Teach Classroom Observation tool. According to Molina and other authors: “The teach classroom observational tool was developed in response to concerns and to foster the measurement of teaching practices in low- and middle-income countries. The “teach tool measures the teacher practices at a primary school level and was intended to be used as a system diagnostic and monitoring tool and for professional development” (Molina et al., 2018). Classroom observation as a diagnostic and monitoring tool at the system level is therefore relevant to this current study as it helps identify bottlenecks in service delivery, monitor the effectiveness of the practice and focus efforts towards improving teacher practices.

2. How can FA be supported through intervention to enhance teaching and learning for Standard 4 pupils HOTS in mathematics?

Teaching and learning within FA has been discussed in different academic forums as well as widely covered in the literature on how FA can impact pupils learning (Black & Wiliam, 2003; Heritage, 2010; Kivunja, 2015). A similar assumption and avenue are explored in this Research Sub-Question 2, that FA Strategies should be embedded as a means to enhance pupils' HOTS. For any intervention to be acceptable to the users, it should be part of the developing team. Practitioners, in this case, teachers, should be able to recognise and implement FA materials for use to improve the acquisition of HOTS by the pupils.

Phase II

In **Phase II**, the research sub-questions were:

3. What are the current pre-intervention levels of mathematics achievement for Standard 4 pupils?

So far, the available evidence for Standard 4 specifically is the data from the 2007 project on mathematics achievement of Standard 4 pupils. The rest of the evidence was either achievement at the end of primary (PSLE) school or international studies like TIMSS. Research sub-question 3 specifically explores the level of pupils' mathematical proficiency in HOTS before the intervention to determine the level of deficiency.

4. To what extent does the FA strategies intervention enhance Standard 4 pupils' HOTS academic achievement in mathematics when comparing the pre-to post-intervention?

Following the intervention, the information was gathered by measuring the pupils' mathematics performance after the intervention was used to answer Research Sub-Question 4. The information gathered from pupils' tests was used to shed light on the impact of the possible effect of FA intervention when incorporated in the teaching of HOTS to improve mathematics achievement when compared to pre-achievement testing results.

Phase III

In **Phase III**, the research sub-questions were:

5. To what extent does the FA strategies intervention enhance Standard 4 teachers' teaching of HOTS in mathematics when comparing the pre- and post-observation?
6. What are teachers' experiences following the FA strategies intervention and mathematics teaching on the pupils learning outcomes?
7. What are teachers' reflections following the FA strategies intervention and mathematics teaching of pupils' HOTS?

Following the intervention, the teachers' observations and interviews were used to answer both Research Sub-Questions 5, 6, and 7 as an attempt to analyse teacher's gain, experiences and reflections on the implementation of the intervention for FA strategies to enhance HOTS for every mathematics lesson. Such information was considered imperative to determine the possible impact of the FA intervention when incorporated in the teaching of HOTS as well as pupils' improvement in learning mathematics.

1.4 Research Methodology

The ontological assumption which constituted this study involved multiple realities which encompassed determining frequencies of teachers' FA practice and teaching HOTS. It included observing the teacher's classroom practice, being involved in training and establishing how teachers' as individuals are affected by the entire intervention with which they were involved as well as determining pupils' achievement at the two critical instances of time.

The epistemological perspective of the current study was, therefore, underpinned by postpositivism. This study attempted to explore teachers' FA practice and after that engaged participating teachers in facilitating FA and teaching HOTS as an intervention in a more collaborative approach and democratising learning environment so that exposure and experiences of the participating teachers would benefit pupils' achievement in mathematics (Gray, 2014). In this way, the study's theoretical perspective was appropriately inclined to a pragmatist paradigm. According to Saunders and Thornhill (2012), pragmatism is a particular approach that constitutes a mixed methodology. This mixed-methods design for the current study was executed in three phases, as shown in Table 1.2.

Table 1.2: Summary of the Research Design and Methodology

Research Topic	Botswana teachers' experiences of formative assessment in Standard 4 Mathematics.					
Research Approach	Mixed Methods					
Research Design	**Sequential embedded mixed, single-group, pre-test, intervention (training) and post-test design (**Martin, 2016; Creswell 2014). The purpose of the study is to determine the possible effect of an intervention for BS4MT use of the FA strategies in the classroom to enhance Standard 4 pupils' HOTS.					
Main Question	To what extent does the FA intervention for teaching and learning to enhance Standard 4 pupils' HOTS in mathematics?					
Study Phases	PHASE I		PHASE II		PHASE III	
Research-Sub Questions	Question 1 What are the current practices in FA and teaching of HOTS	Question 2 How can FA be supported through intervention to enhance teaching and	Question 3 What are the current pre-intervention levels of mathematics achievement in HOTS	Question 4 To what extent does the FA strategies intervention enhance Standard 4 pupils'	Question 5 To what extent does the FA strategies intervention enhance Standard 4 teachers' teaching of	Question 6 & 7 What are teachers' experiences and reflections following the FA strategies

	in Botswana primary schools?	learning for Standard 4 pupils HOTS in Mathematics?	items for Standard 4 pupils?	HOTS achievement in mathematics when comparing the pre -and post-intervention?	HOTS in mathematics when comparing the pre- and post-intervention?	intervention and mathematics teaching of pupils' HOTS?
Data Collection Strategies	The baseline survey used 150 Standard 4 teachers from southern education region primary schools. 9 teachers who selected for FA intervention, were observed prior to the intervention.	Nine teachers in the nine (9) intervention schools will also document self-reflective journals after every mathematics lesson. The teachers were observed at least twice during the intervention implementation phase.	One pre-test for the Standard 4 pupils from nine sampled schools. That is a total of 272 pupils in the intervention schools.	One post-test for the Standard 4 pupils from the nine schools (272 pupils).	Teachers' pre- and post-observation findings and semi-interviewed on the possible effect of the intervention on teaching and learning HOTS in mathematics.	
Strategies for each sub-question	Baseline Survey Observation	Observation	Pre-test	Post-test	Pre- and post-observation and Semi-structured interviews	

Data Analysis	Rasch Measurement (RMT) Theory using Winsteps. For Quantitative data: SPSS 26.0 version statistical package descriptive and inferential statistics For observation data: computed using Microsoft Excel.	Microsoft Excel will be used to develop: coding, themes, detailed descriptions. For observation data: computed using Microsoft Excel.	RMT (The Partial Credit Model) SPSS 26.0 version statistical package descriptive and inferential statistics.	RMT (The Partial Credit Model) SPSS 26.0 version statistical package descriptive and inferential statistics.	Microsoft Excel will be used to develop: coding, themes, detailed descriptions Interrelated theme and member checking.

1.5 Structure of the Thesis

This thesis attempts to integrate the formative assessment model (Black & Wiliam 1998; Heritage 2010; Kivunja, 2015) in teaching with particular emphasis on Higher-Order Thinking Skills informed by Bloom's taxonomy (Bloom, Engelhart, Furst, Hill & Krathwohl, 1956) as a developmental guide of measures for cognition and achievement of educational objectives in many countries (Hattie & Brown, 2004). This PhD's position emanated from Chapter 1's exploration of assessment practices and teaching in the context of the Botswana education system. The existing evidence on assessment practice in continuous assessment (FA) and summative assessment and alignment to teaching pedagogy is scarce and seems to impinge on pupils' achievement in general across Botswana's education system. The literature's emphasis, also for other African countries that participated in the large international assessment such as TIMSS, is particularly on pupils' low achievement in higher-order thinking (Masole, et al., 2016; Reddy et al., 2016). Chapter 1 sets out the problem statement in the context of Botswana Primary School education at the Grade 4 level. The chapter also discussed the rationale for the study and listed the research questions of the study. Lastly, it introduced the research methodology utilised to address one general research objective:

- To determine the possible effects of an intervention for BS4MT use of the FA strategies in the classroom to enhance Standard 4 pupils' HOTS (Phase I reported in Chapter 4, Phase II reported in Chapter 5, and Phase III reported in Chapter 6.)

Chapter 2 of this thesis elaborates on the context of the educational elements mentioned in the introductory chapter. This chapter mainly focuses on the Botswana educational structure, policies and reforms that orientate the current education and training. Key aspects of this section are, in particular, the curriculum, teaching and learning, and assessment of mathematics in the lower primary schools in Botswana at the Standard 4 level.

Chapter 2, the literature review, also ushers in educational assessment information related to the study. The first section explores the nature of formative assessment, its effectiveness and briefly espouses Assessment for Learning (AfL) in Africa and the impact of teachers' use of FA and the need for teacher training in FA. The second section explores the empirical evidence on the FA strategies as an intervention, reviewing the teachers' use of FA strategies within the primary school mathematics context, its benefits to lower-achieving pupils, and the cause and effect of empirical studies espoused by different scholars. The chapter ends with the theoretical framework and Kivunja's (2015) Assessment Feedback model used as the conceptual framework to describe the FA framework in teaching mathematics which was used for addressing the main research question and sub-questions of the current study.

Chapter 3 focuses on the research design and methodology, which is a mixed-method approach. In this chapter, the paradigm orientation and epistemological assumptions from which the research design was distilled are discussed. A general methodology approach associated with the three Phases is briefly discussed, and a detailed account of the design for each phase is explained separately in the next three subsequent chapters of this thesis.

Chapter 4 elaborates on the Phase I research design, method and data analysis techniques employed. In Phase, I the baseline survey in a mixed-method approach comprised of quantitative data of 150 sampled teachers from the Southern Education region and follow-up data from nine observed teachers are discussed. Specifically, the sampling procedures employed to reach samples of Standard 4 teachers from the five Sub-regions (Kanye, Lobatse, Goodhope, Mabutsane-Jwaneng and Moshupa) of the Southern region are discussed. The baseline was mainly explored to measure the teachers' formative assessment practice used as an adapted quantitative survey instrument. Quantitative survey data was analysed using the Rasch Model in Winsteps, and descriptive content analysis was employed to analyse the classroom observation data. The survey findings were used to create an intervention. The baseline survey and intervention was investigated using two sub-research questions (SRQ):

- SRQ 1. What are the current practices in FA and teaching of HOTS in Botswana primary schools?
- SRQ 2. How can FA be supported through intervention to enhance teaching and learning for Standard 4 pupils HOTS in mathematics?

Chapter 5 covers teachers and pupils' data from Phase II of this thesis. The nine participating teachers who pre-observed during mathematics and their 272 pupils were recruited in this Phase. A single group was used for the pre- and post-test design. The researcher designed a formative assessment for the HOTS (FAHOTS) intervention for the enhancement of teaching of HOTS and to measure the Standard 4 pupils' achievement in mathematical HOTS items by comparing the pre- and post-intervention assessment. Rasch person logits analysis techniques were used in Winsteps software, while descriptive content analysis was employed to analyse the post-classroom observation data. The following two research questions were investigated:

- SRQ 2. What are the current pre-intervention levels of mathematics achievement in HOTS items for Standard 4 pupils?
- SRQ 5. To what extent does the FA strategies intervention enhance Standard 4 pupils' HOTS academic achievement in mathematics when comparing the pre- and post-intervention?

Chapter 6, Phase III also covers teachers and pupils' data as used in Phase II with the purpose in mind to evaluate the intervention. The Phase adopted a validated and equated scale to measure pupils' achievement in mathematical HOTS items during post-test to investigate the impact of the intervention. At the same time, teachers' FAHOTS intervention practice was also measured using the same scale used in pre-observation to compare with post-observation to assess any impact. A follow-up teacher interview was conducted to measure the experience and experience of using FAHOTS intervention. The teacher data for classroom observation was analysed using descriptive content analysis, while interview analysis was conducted based on thematic analysis. The following three research questions were investigated:

- SRQ 4. To what extent does the FA strategies intervention enhance Standard 4 teachers' teaching of HOTS in mathematics when comparing the pre- and post-observation?
- SRQ 6. What are teachers' experiences following the FA strategies intervention and mathematics teaching on the pupils learning outcomes?
- SRQ 7. What are teachers' reflections following the FA strategies intervention and mathematics teaching of pupils' HOTS?

Chapter 7 is a reminder of the three phases of empirical studies presented in detail to explore the possible impact of FA intervention and its contribution to the pupils' mathematics achievement and enhancement in learning HOTS. The findings from the literature are reviewed, and the findings from the theories that underlie the FA strategies and enhancement of HOTS are also considered. The research questions are addressed, and the limitations of the research are discussed. The chapter ends by pointing out the conclusions, recommendations, new knowledge innovation and insights, and closing thoughts for the study.

2 CHAPTER TWO: LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

This chapter articulates the literature and theoretical foundation upon which the present study was conducted. The first section provides an overview of Botswana with a brief description of the geographical location, demography and economic status. The second section explores the educational transition of Botswana from pre to post-independence with emphasis on major issues related to teaching and learning. Then an outline concerning transformational policies is provided as a guide to direct education for Botswana (third section). The fourth section on primary education in Botswana follows this to explore its composition and the expected competencies, mathematics curriculum in lower primary school, teaching mathematics in lower primary school, and teaching and assessment of HOTS in mathematics for the Standard 4 level.

The fifth section provides a review of summative and formative assessment (FA) with a close examination of its nature and purpose as well as the integration of FA in teaching and learning in the developed nations contexts. A recent research project in Africa, which is termed as an Assessment for Learning in Africa (AfLA) and FA practices in African contexts are also discussed. The fifth section also provides a review need for teacher training in FA and the problems associated with FA implementation. The sixth section is an empirical review of the relevant studies within the framework of FA strategies. The sixth section also includes the integration of FA and HOTS, which is how teachers could use FA and link it to teaching and learning of HOTS in mathematics. Following the review of FA and HOTS, there is a discussion about the intersection of the domains which form the theoretical framework of the current study. In the last section, the study's conceptual framework is presented and elaborated.

2.1 Botswana Context

Over the years, Botswana has evolved and made significant strides in different sectors of societal transformation. For the sake of the current study, the following section would selectively discuss some sectors which are relevant for this study. Among them are the geographical location of the country, its education transitions and the current education structure.

2.1.1 Landscape and climate

Botswana is a landlocked country in Southern Africa covering an area of 582 000 square kilometres, with a population of 2 024 904 people according to the 2011 Population and Housing Census (Ministry of Finance and Development Planning [MFDP], 2018). In terms of location, Botswana shares borders with Zimbabwe, South Africa, Namibia and Zambia (see Figure 2.1 below). The country has a dry, semi-arid climate with temperatures ranging from as low as -5^{o}C at night to as high as 43^{o}C during the day and an average rainfall of 450 mm. The population is concentrated in the eastern parts of the country where arable farming is feasible on account of better and more favourable climatic and soil conditions. The northern part of the country is a good tourist attraction because of its natural flora and fauna, in particular, the Okavango Delta (Masole, 2011, MFDP, 2018).

Figure 2.1: Location of Botswana in Southern Africa

2.1.1 Demography

> *Botswana's population has increased over the years, from 574 000 in 1971 to 2 025 000 in 2011. The rapid growth from 1971 to 1991 was the result of the country's high birth rate and declining mortality rates, which were due to the improved health of the population. However, since 1991 the rate of population growth has been declining, growing only by 2.4 % per annum over the decade to 2001, and 1.9 % to 2011* (MFDP, 2018).

According to the Botswana Revised National Population Policy (RNPP, March 2010), the contributory factors for the decline of the population growth rate include HIV/AIDS, declining fertility rates, increased female participation in economic activities, increased literacy rates, and access to better healthcare.

> *The life expectancy at birth, which had increased from 55.5 years in 1971 to 65.3 years in 1991, decreased to 55.6 years in 2001 as a result of HIV/AIDS. It then rose to 68.0 years in 2011 on account of the government's successful strategies of fighting the HIV/AIDS scourge, which included the universal provision of anti-retroviral drugs (ARVs) and 84 per cent of the population lived within 5 kilometres of a health facility in 2015* (MFDP, 2018, p. 11).

Botswana is a multi-ethnic and multilingual country. There are over 20 other languages which are mostly spoken by the non-Setswana speaking ethnic groups (Tlou & Campbell, cited in Masole, 2011). Setswana is the National Language, while English is the official language used in business in most government affairs. Christianity is the main religion, and there are several indigenous religions throughout the country (MFDP, 2016). “With literacy, the country's adult literacy rate increased from 81% in 1981 to 88.7% in 2014” (MFDP, 2018).

2.1.2 Economy

Botswana is considered an upper-middle-income country, having been one of the fastest-growing economies in Africa during the last decade. "Botswana is endowed with a variety of minerals such as diamond, nickel, salt, soda ash, coal, gold and Potash" (Ministry of Trade and Industry [MTI], 2007).

> *Diamond extraction is the main factor behind high growth rates in recent years. It accounts for more than one-third of GDP; almost all export earnings, and half of the government's revenues. Other key sectors are tourism, financial services, subsistence farming, and cattle raising. The GDP in Botswana expanded to 2.40% in the second quarter of 2018 over the previous quarter. The GDP growth rate in Botswana averaged 1.24 % from 1994 until 2018, reaching an all-time high of 13.40 % in the second quarter of 1997 and a record low of -13.80 % in the first quarter of 1998* (MFDP, 2018).

Historically, Botswana is an administrative compliance examination country (Fetogang, 2015). This examination compliance has brought some teaching practices and instigated tension between formative purposes of assessment and summative purposes of assessment (Motswiri, 2004). Teachers are often under pressure to prepare pupils for external examinations, which have to do with administrative compliance purposes and judge pupils' achievements. So, the mandatory school-based assessment (SBA) in the form of continuous assessment (CA) does not serve its purpose to guide teaching and learning (Masole et al., 2016).

Teaching and learning in a primary school in the Botswana context have pedagogical related problems (Masole et al., 2016). A significantly noted problem is teaching to the test, found in several studies over the decades (Fetogang, 2016; Moyo & Nenty, 2017; Raboijane, 2005). The negative impact of ineffective teaching practices affects pupils' comprehension and long-term understanding (Fetogang, 2015). Teaching to the test promotes rote memorisation and suppresses the pupils' strength for creating skills and growing abstract-thinking ability. More specifically, limited evidence on formative assessment and its alignment to teaching pedagogy in the Botswana education system is a major concern to the scholars and general stakeholders in education (Botswana Guardian, 2018. p.8; Masole et al., 2016; Moyo & Nenty, 2017). As outlined in Chapter 1, Botswana pupils were unable to handle mathematical HOTS tasks and generally low achieving when measured on the common scale of performance in large-scale international assessments such as TIMSS 2015 (Masole et al., 2016). The current study was motivated by pupils' poor mathematics achievements to seek possible assessment interventions to enhance learning and investigate the phenomenon under the title "Botswana teachers' experiences of formative assessment in Standard 4 mathematics with a focus on HOTS". FA strategies and the associated feedback loop in various contexts had provided a means of collecting data that constituted evidence of pupils' learning. These strategies help to identify gaps between what pupils know and they should know, and determine how to respond to pupils' learning tasks (HOTS) to close the gap (Black & Wiliam, 2003, 2006; Hattie & Timperley, 2007; Kanjee, 2017; Kivunja, 2015; Kyaruzi et al, 2020; Nicol & Macfarlane-Dick, 2006; Ramsey & Duffy, 2016; Weurlander et al., 2012; Wiliam, 2013; Wiliam & Thompson, 2014).

2.2 The transition of Education in Botswana

2.2.1 Traditional and Pre-Independence Education

In 1885, Botswana became a British Protectorate and later gained her independence on September 30, 1966, which marked the end of British colonial rule. Before the arrival of Western type of education, children in Botswana were socialised into adulthood through some kind of traditional initiation schools such as *Bogwera* for boys and *Bojale* for girls (Coles, 1985, p.1, cited in Ministry of Education [MOE], 2001). The *Bogwera* and the *Bojale* were done at the times when those approaching adulthood received general basic training. They learnt of their physiology, tribal history, games (incorporating riddles, puzzles and proverbs to indicate socially desirable attitudes) and skills of hunting and gathering. This was the time for those entering adulthood to understand that they had rights and responsibilities.

According to Coles (1985, as cited in MOE, 2001, pp. 2- 4) "Western type of education had its origins in the work of Moffat and the Kuruman Mission Station established by the London Missionary Society in 1821". Coles emphasised that though the educators were concerned with the spread of the Christian faith in Bechuanaland, they were also concerned with the broader aspects of education, such as the introduction of applied skills, with agriculture being the dominant area. The traditional and new educational systems were to exist side by side for some time.

In education, the change came with modern education with its formal teaching and assessment. According to Koloi- Keaikitse:

> *The community members also lost control of the children which became the responsibility of their close relatives and teachers at school. During this time, teacher training was not given much attention by the British colonialists, despite much responsibilities which were entrusted to teachers to take care of all aspects of educating children. The majority of the teachers were trained in South Africa and Zimbabwe. In the late 1930s, Botswana introduced its teacher education in Kanye and Serowe. The training programme was more focused on primary education, in which teachers were trained on how to teach all subjects in primary schools* (Koloi-Keaikitse, 2013, p10).

2.1.2 Post-Independence Education

At independence, the government inherited a western type of educational system that was poor in quality and catered for a tiny proportion of the population. The school curriculum did not address the needs of society, especially its children. In 1976, out of 3918 classes, there were only 2275 classrooms. Therefore, 1643 classes did not have classrooms. Untrained teachers made up 30% of the teaching cadre. Studies on learning achievement revealed that a majority of primary school leavers lacked minimal competency in reading and writing (Husen, 1977, cited in MOE, 2001). Thus, a radical approach was needed to make education a truly national entity.

The first President of the Republic of Botswana, Sir Seretse Khama, appointed a Commission on Education in December 1975, to chart the path of education. The Commission produced its first report in April 1977, after extensive consultations and studies undertaken within and outside the country (BOT, 1977; BOT, 1994). The commission envisioned an education system that provided access to all and supported social harmony after the realisation that the Botswana community was isolated according to different ethnic groups. The attempt was to unite the nation and maintain a peaceful co-existence. Observers can see this through the fact that Batswana can joke about each other's origin or ethnicity without causing any bloodshed or animosity. The thinking behind this principle led to the formulation of the National Policy on Education (NPE) of 1977 that addressed the four (4) national principles of democracy, development, self-reliance and unity popularly known as education for Kagisano which literally means social harmony (BOT, 1997; Makwinja, 2017).

As stated in the NPE (BOT, 1977), school fees were to be abolished, the development of infrastructure was to be accelerated, and teachers' training was to be intensified. Great strides were made following the implementation of the recommendations. This period saw a massive expansion of school places. The data indicates that 40% of the primary schools were built and opened between 1977 and 1987. Pupils' enrolment also doubled as a result of increased places as published in the Statistical Bulletin of 1998 (MOE, 2001). Currently, Botswana has 753 public primary schools and 283 public secondary schools. The overall enrolment in primary schools has significantly improved, rising to 97% for the 6 – 12 years cohort by 2014 (BOT, 2015).

In 1994, the RNPE identified the goal of education as preparing Botswana for the transition from a traditional agro-based economy to an industrial economy to compete with the other countries of the world. Additionally, the government considered access to basic education a fundamental human right.

The main objectives of the Ministry of Education were identified as follows:

- "to raise the educational standards at all levels;
- to emphasise science and technology in the education system;
- to make further education and training more relevant and available to more people;
- to improve partnerships between schools and communities in the development of education;
- to provide lifelong education to all sections of the population;
- to assume more effective control of the examination mechanism to ensure that the broad objectives of the curriculum are realised; and

- to achieve efficiency in educational development" (RNPE, 1994; UNESCO, 2010, p.2).

2.1.3 The Current Structure of the Education System

Botswana has made significant progress in creating broad opportunities in education. The current education structure is broadly classified into three levels, namely; basic education, tertiary education and technical and vocational education and training (TVET) as shown in Figure 2.2.

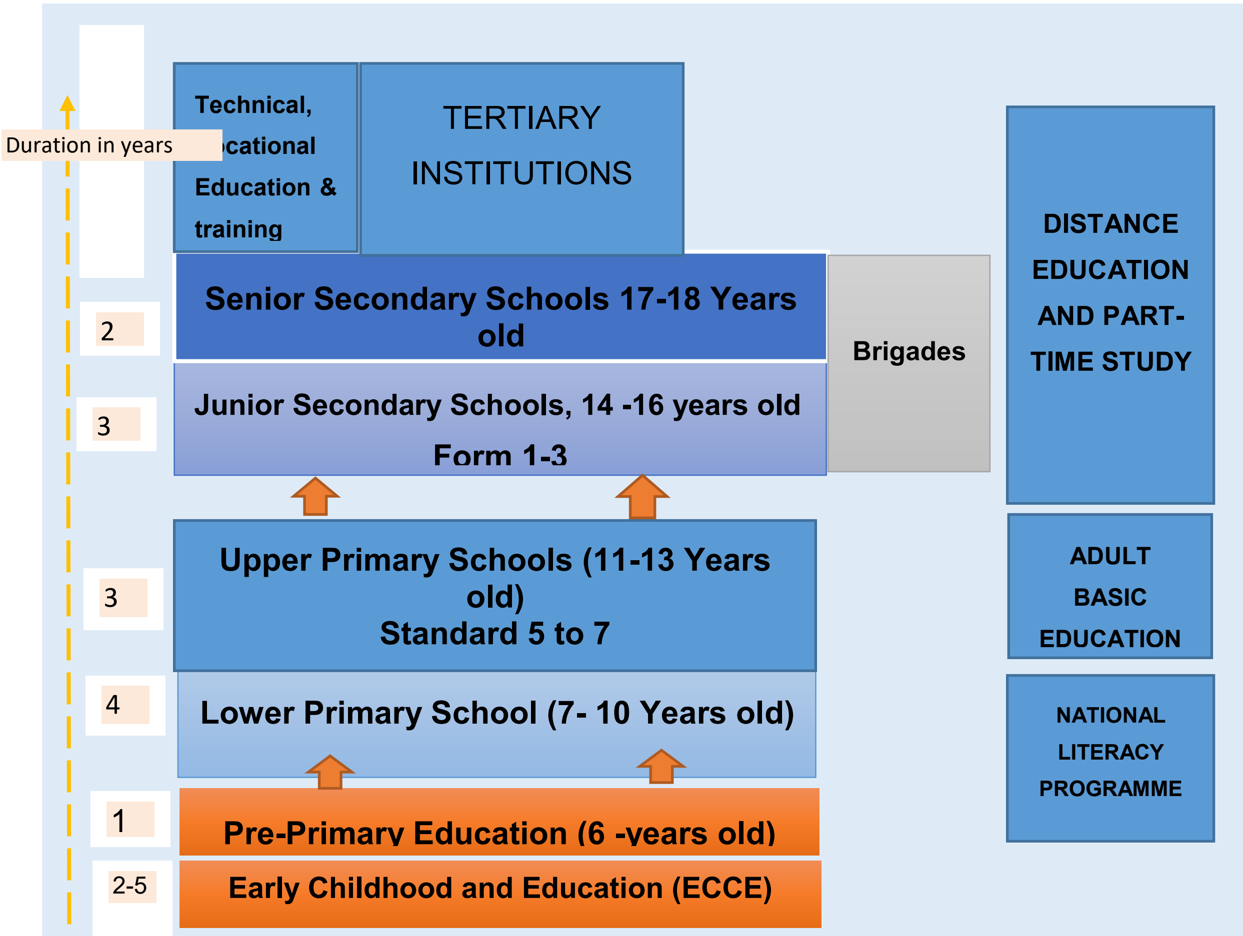


Figure 2.2: The structure of Botswana's education and training system
Source: (BOT, 2015, p. 21)

Basic Education

Basic education in Botswana comprises 12 years of schooling; seven years of primary school, three years of junior secondary school and two years of senior secondary school. The official school starting age is six years of age; however, the majority of the pupils start at seven years of age. Parents are charged a co-payment for education, but the children from lower-income families receive free education. The co-payment is very flexible and not fully implemented. Free schools' meals are provided to all school pupils (BOT, 2015).

The Botswana Government describe the provision of the Early Childhood Care and Education programme as follows:

> *The provision of the Early Childhood Care and Education (ECCE) programme for all children is essential as a foundation for primary education. Its benefits to a child among others include helping to boost cognitive and motor development as well as enhancing social skills. Pre-school education helps build a foundation of learning that is built upon throughout their school years. Pre-school also allows the child lots of interaction with peers which enhance their skills to share, negotiate and listening skills amongst others* (Statistics Botswana, 2015, p.4).

The RNPE emphasises the need for the provision of enabling environment for pre-primary education and increasing access to all children before enrolling in primary schools (Statistic Botswana, 2015).

ECCE is mostly limited to private schools with limited enrolment (BOT, 2015). Maunganidze and Tsamaase (2014) acknowledged the changes in pre-primary education, which have occurred since 1977 and the sector's phenomenal growth. The 2013 pre-primary data was collected from 570 pre-primary schools. Out of these, 391 (68.6%) schools are owned by private companies and individuals, 68 (11.9%) of the schools belong to the community, 52 (9.1%) to churches, 45 (7.9%) to Non-Governmental Organisations, 12 (92.1%) belong to councils while there was one institutional school (Statistic Botswana, 2015). From 2017, the MOBE also introduced and implemented a one-year reception programme education in all the 753 public primary schools.

According to Statistics Botswana (2015), there were 821 registered primary schools countrywide in 2013, of which 753 (91.1%) were government-owned schools and 68

(8.9%) were privately owned. This is an indication that the government is the main provider of primary education.

Progression is automatic from primary to junior secondary. Conversely, 12.5% of the members of each class can be allowed to repeat the year. The first ten years of education is available and compulsory to all children, while the last two years is not compulsory. There are four national examinations which are taken during the schooling period. These examinations include the SFAT at the end of Standard 4. The schools use the assessment results to decide whether they can promote pupils to standard 5 (BEC, 2015), then PSLE at the end of primary schooling. Those who pass the Junior Certificate of Education (JCE) at the end of secondary school can progress to senior secondary school. At the end of senior secondary schooling, students sit for the Botswana General Certificate of Secondary Education (BGCSE), (BOT, 2015).

Tertiary Education

Botswana tertiary education is provided by both public and private institutions. The institutions enrol the secondary school graduates and train them in a wide range of programmes to meet the aspirations of the country as well as to attain a globally competitive human capital. The students in higher education are mostly sponsored by the government through the Department of Tertiary Education Financing. Programmes leading to certificates, diplomas, and degrees are offered by universities, colleges of education and the Institute of Health Sciences (BOT, 2015).

Technical and Vocational Education and Training (TVET)

Historically the brigades were established by the communities to absorb children who would have failed to progress through secondary schools and were managed through a consultative process with the government. The Brigades have been transformed to TVET, which are currently delivered at different levels from certificate to diploma in different types of institutions (BOT, 2015, Masole, 2011).

2.3 The Transformational Policy and Legislative Education Framework

Botswana has articulated long-term national goals and values in Vision 2036 (previously revised vision, 2016 and the current vision of 2036), and in the series of National Development Plans (NDPs), and RNPE. The RNPE (BOT, 1994) and

National Development Plans have been the guiding basis for the implementation of providing quality education and training. These policies and plans set out the goals and strategies that Botswana will pursue to promote human, social, economic, and environmental development and choices that will define Botswana society over the next decades.

According to Vision 2036, Botswana society will be knowledgeable with a relevant quality education that is based on outcome-based learning, with an emphasis on technical and vocational skills as well as academic competencies (education with production). That education curriculum will be aligned to the need of the economy and business; hence science, mathematics and technology would be taught right from primary schools to tertiary education. In an attempt to achieve such aspirations, the government has made a strong alignment between priorities as outlined in the Botswana visions, NDPs and RNPE, which gave birth to Botswana Qualifications Authority Act (BQA) 2013 and the ETSSP in 2015 (BQA, 2013; BOT, 2015).

The following section will provide the background to the Botswana Education policy frameworks from the context of BQA, followed by the competency framework for National Credit Qualifications Framework (BOT, 2016a) and the recently ETSSP 2015-2020 aspirations for outcome-based education and its implementation as they are considered relevant to the current study (BOT, 2015).

2.3.1 National Credits and Qualifications Framework

The Botswana Qualifications Authority (BQA) is a parastatal organisation under the Ministry of Tertiary Education, Research, Science and Technology. BQA took over the quality assurance functions that were previously performed by the old Botswana Training Authority (BOTA) and the Tertiary Education Council (TEC). The organisation merged the two entities for quality assurance systems; the vocational training system that was under the remit of BOTA, on the one hand, and the higher education system that was the responsibility of TEC, on the different approaches to learning programmes accreditation (BQA Act, 2013). BQA comprises a board of 13 members appointed by the Minister of Tertiary Education, Research, Science & Technology and consists of a Chairperson, Vice-Chairperson, Chief Executive Officer and support staff (BQA Act, 2013).

The transition is a product of the BQA Regulations that were gazetted late in 2016. As required by the Act and the regulations, the Authority is mandated to register and accredit all education and training providers (ETPs), both private and public. Thus, BQA's main objectives are to provide for and maintain a National Credit and Qualifications Framework (NCQF) and to coordinate the education, training and skills development quality assurance system. BQA flows from the National Human Resource Development Strategy, which underscores the need to transform Botswana into a knowledge-based society that produces global workers. The framework in place, therefore, is seen as the mechanism to deliver that type of human resource (BOT, 2016).

This BQA framework is such that the national qualifications get to be registered in the system, and are defined in terms of what is called level descriptors, which state knowledge, competencies and skills expected to be imparted at each level. A level is defined in terms of standards required for the sub-system under which it falls, whether General Education, Technical and Vocational Education and Training or Higher Education. The standards are generally and technically described as outcomes; so we talk of an outcomes-based system. It is outcomes-based because the standards state what the learner should be able to do when he or she has completed a diploma or a degree (BOT, 2016).

In 2016, the BQA National Credit Qualifications Framework (NCQF) Regulations instructed all the education and training providers (ETPs) to have registered and accredited by the 31st December 2017 in line with the new requirements. The new requirements requested the ETPs also to review and align their existing qualifications and programmes into the new NCQF. The competency framework for NCQF Implementation is outlined below (BOT, 2016).

- Outcomes-based qualification and programme design and development
- Outcomes-based programme delivery strategies and methods
- Programme implementation and monitoring
- Outcomes-based assessment design and application
- Developing rubrics for outcomes-based assessment
- Inclusive education and training principles and practices

- Learner-centred approaches
- Training evaluation principles and processes
- Conducting self-studies or annual reviews
- Establishing and maintaining strategic partnerships for education and training

2.3.2 Botswana Education and Training Sectors Strategic Plan (2015-2020)

The ETSSP is an extensive reform and it represents a broad consensus derived from principles agreed with the key stakeholders prominent in the education sector. The goals, policy, objectives and activities are those derived through broad participation in discussions, workshops, and through a process of interactive reviews of the programmes and priorities. The MOBE led the development of the strategy at all levels with analytical work and advisory support from international and local consultants (Makwinja, 2017).

The central focus of the ETSSP is to improve quality education throughout the sector as an investment in improving pupils' outcomes. This will ensure high-quality education and provide a wide range and flexible learning opportunities with a lifelong learning framework. The cross-cutting issues and linkage involve the implementation of curriculum and assessment competency-based approach across all the sub-sectors of the educational system; in alignment with the standards established by the teaching council of Botswana, BQA and by the BEC (BOT, 2015).

ETSSP recognises the primary education to ensure that all children gain the basic knowledge and skills (literacy, numeracy, information and communication technology and life skills), and develop the attitude and values for continuing lifelong learning. Such a decision was reached after establishing the challenges faced by primary school education which among others were; (i) disparities between rural and urban schools in terms of quality education and pupils performance, particularly children from poorer and remote rural areas, and children with special needs; (ii) poor teaching and curriculum delivery in remote areas, (iii) in urban areas and major villages, there is a shortage of classrooms leading to congestion high teacher-pupil ratios and (iv) educational quality in core subjects such as English, Mathematics, Science and Agriculture have been failing and significant disparities in pupils' performance persist, again apparently correlated with geographic areas and socio-economic background of the pupils (Ministry of education and Skills Development; MoESD, 2014).

The current study also targets the primary school pupils, in particular, the Standard 4 pupils. Standard 4 level is a transitional foundation phase of pupils' schooling life in Botswana and pupils are expected to sit for attainment examinations in literacy, numeracy and other subjects to determine their progression to the upper primary schools (MoESD, 2014). Given the pupils' attainment challenges identified so far, the ETSSP then recommended some programmes and activities as an attempt to address the quality of education. One of the programmes was to improve achievement in reading, mathematics and science in primary education, as a means to enhance pupils' performance in literacy and numeracy. ETSSP proposed among other activities was that the schools be able to do the following:

- Develop and implement standards for school development and academic performance;
- Develop and implement continuous in-service teacher training and professional development in reading, mathematics and science;
- Develop and implemented tools and procedures for assessment in the classroom;
- Develop and implement school-based management assessment and monitoring tools;
- Develop a coaching model for professional support to teachers and school heads;
- Train School Heads and Pedagogical staff in each school to serve as a coach for reading, mathematics and science in line with the In-Service teacher training (INSET) policies, frameworks and guides.

The ETSSP tends to focus on the harmonisation of the implementation of outcome-based learning, by changing the teaching pedagogical approaches and assessment across all the sub-sectors of the education system (BOT, 2015). The current study is in line with the outcome learning-based about a shift in teaching pedagogy approach and assessment practice as an attempt to improve mathematics outcome at Standard 4 level.

It is clear from ETSSP proponents that Botswana's aspirations are for outcome-based education (OBE) practices. As purported in the Human Resource Development Council (HRDC (BOT, 2015) advocate for a shift from a resource-based economy to a knowledge-based economy aligning with the Botswana Education policy frameworks such as; ETSSP, (2015-2020) and National Curriculum and Assessment Framework (NCAF, 2015) whose mandate and priorities include improving the quality and relevance of education among others. Providing relevant classroom pedagogy and assessment that involves engaging pupils in independent learning as a way to address the OBE which ideally ought to be introduced in the Botswana Education system effectively in the year 2019 (Mannathoko, 2017).

OBE practice means focusing and organising everything in an educational system around what is essential for all pupils to be able to do successfully at the end of their learning experiences (Akir, Eng & Malie, 2012). OBE requires Government institutions to start with a clear picture of what is important for pupils to be able to do, then organising the curriculum, instruction and assessment to make sure this learning ultimately happens. Van Staden (2010) justified that OBE is characterised as a learner-centred approach, in which both teachers and pupils are focused on a predetermined result or outcome that is to be achieved by the end of each learning process. These outcomes are intended to relate to real-life needs and situations and integrate knowledge, competence and orientations which pupils need to become responsible, critically-thinking and competent adults.

The attainment of the OBE system in Botswana demands a new learning paradigm to pedagogy which posits that the purpose of education is to meet the increasingly dynamically complex societies truly. Learning, teaching, assessment and curriculum need to equip pupils with the skills that will enable them to contribute effectively to the productive capacities of the economy (Kivunja, 2015). The classroom pedagogy, therefore, must get pupils to engage, explore, explain, elaborate, and evaluate as they learn. These processes would enable them to maximise their participation in active learning and lead to deep than surface learning (Entwistle, 2000; Flewelling & Higginson, 2002).

In Botswana, the OBE seems not to take any significant stride and the most disturbing observation is the ETSSP's lack of implementation which is now only left with one year since its inception, and with limited evidence of INSET on the OBE and pupils are still not doing well across all levels. The recent study conducted by Makwinja (2017) confirmed that Botswana has innovations detailed in policies and they have not been implemented as expected. The findings also showed that policies were never monitored and for instance, if the RNPE: 1994 policy was implemented, the education system would be highly developed.

The participants in Makwinja's study also lamented about the lack of implementation capacity and that most policies were not used. When the author questioned participants, why new policies were drawn while the existing ones had not been fully implemented, and the respondents alleged that most of the decisions made were politically driven. The participants further commented on the latest policy (ETSSP) that it was merely the RNPE called by a different name. Additionally, about the ETSSP, the teachers had no clue about the existence of such a plan (Makwinja, 2017).

2.4 Primary School Education in Botswana

According to UNESCO (2010), the task of primary education schooling in Botswana is to provide the foundation of basic competencies that will prepare the child for continued in-school and out-school learning for social and economic life in modernising society as outlined in the RNPE of 1994. It is also important for this study to underscore the educational processes upon which the current study was conducted. Such educational processes include the composition of primary education and competencies expectations, the curriculum at lower primary level, teaching mathematics as well as the teaching of HOTS and its assessment at the lower primary education level.

2.4.1 The composition of primary school and competencies expectations

The primary school programme forms the lower level of the ten-year basic education. It comprises a seven-year level divided into two phases, being the lower primary (standards 1-4) and the upper primary (Standard 5-7). The programme is woven around among others are; the acquisition and application of foundation skills; vocational orientation of academic subjects and the acquisition of socially desirable skills and attributes (MoESD, 2008; Monyaku & Mereki, 2011; Spaull, 2011; UNESCO, 2010).

According to the MoESD (2008), on completion of seven years of primary education pupils should have:

- Acquired language skills to be able to express themselves appropriately in English and Setswana as tools of communication and also for learning;
- Developed awareness of the interrelationship between science, technology and society in everyday life;
- Developed desirable attitudes towards and appreciation for different types of work and the ability to assess personal capabilities/weakness and achievement;
- Acquired knowledge, skills in and appropriate attitudes towards food production and industrial arts;
- Acquired knowledge and understanding of their society through appreciation of their culture and tradition including languages, songs, ceremonies, customs, social norms and a sense of citizenship;
- Developed skills such as numeracy, literacy, communication, adaptability and problem-solving for further learning and vocational preparation;
- Acquired critical thinking, problem-solving and inquiry skills;
- Developed competence and confidence in the application of computational skills to solve day to day problems;
- Developed awareness and appreciation of the use of computers in everyday life;
- Developed awareness and appreciation of basic entrepreneurial skills in business and everyday commercial transactions;
- Developed the ability to recognise and appreciate the contribution of religion in the formation of values and behaviour patterns;
- Developed awareness of their rights and responsibilities related to health, gender, law, violence, identity, civic and other social and moral issues;
- Developed their special interests, talents and skills whether these be dexterity, physical strength, intellectual ability, and/or artistic gifts.

2.4.2 Mathematics curriculum in lower primary school level

The Botswana lower primary curriculum is designed for pupils from Standards 1 to 4 (UNESCO, 2002). In 2002, the ministry of education introduced the reviewed standard 1- 4 curriculum, and the revision took cognisance of the current societal, economic aspirations. The curriculum is now aimed at providing a platform that promotes the holistic growth of individuals who are relevant to the society in which they live (MoESD, 2008). The repackage of the curriculum include in addition to the traditional subjects of languages and sciences, some new learning area such as the cultural studies, creative and performing arts (CAPA), and environmental science.

According to UNESCO (2010), the subject packaging for standards 1 – 4 is broad with some subjects included together into broader areas to facilitate project teaching and integration. Music, physical education, design, art and craft are part of the area for CAPA. The set of agriculture, home economics and science are integrated into the Environmental Science area. Religious education, moral education and social studies are taught under the area of Cultural Studies. These broad areas are organised into modules. While Setswana, English and Mathematics are taught as a separate entity, (UNESCO, 2010).

The mathematics curriculum (syllabus) is also part of the programme for pupils in the lower primary level. The first two aims of the lower primary school mathematics programme include developing in all pupils (i) competence and confidence in the application of computational skills to solve day to day problems and (ii) inquiry skills, creativity, critical thinking and problem-solving ability (UNESCO, 2002). The specific objectives outline the breadth and depth of teaching required in a particular topic. Those specific objectives are entrenched from the six modules, namely the numbers and operations, geometry, measures, problem solving and statistics (Masole, Gabalebatse, Guga, Pharithi & BEC, 2016):

- Numbers and Operations -This module helps pupils understand the concept and use of numbers. Pupils practice counting, sorting, and classifying numbers, as well as matching objects and numbers. These concepts lay a foundation for addition and subtraction. By the end of Standard 4, pupils should be able to add and subtract three-digit numbers vertically and horizontally. They should be able to multiply using one-digit to three-digit numbers and do simple division. Pupils also should understand money, local currency denominations, and monetary units and be able to add, subtract, multiply, and divide with money.
- Geometry - Pupils are introduced to geometry with the study of shapes and solids. Pupils identify shapes such as rectangles and triangles and solids such as cubes and

cylinders. By the end of Standard 4, pupils should be able to describe shapes by the number of sides and number of angles and be able to name solids.

- Measures - This module intends to develop measuring skills. Pupils start by comparing lengths and weights in standard 1, and gradually begin using standard measuring instruments. By the end of Standard 4, pupils should be able to use formulas to calculate area and perimeter. They should be able to use instruments to measure volume and mass and to convert units of length, time, and mass.
- Problem Solving -This module introduces pupils to practical problem-solving skills. Skills are developed through mathematical games, simple puzzles, and simple investigations involving numbers and shapes. By the end of Standard 4, pupils should be able to conduct simple research projects.
- Statistics -This module introduces simple methods of data collection and simple statistical presentations, such as pictographs. By the end of Standard 4, pupils should be able to interpret information and draw simple statistical conclusions, such as finding the mode.

The weakness of the lower primary mathematics curriculum in Botswana is an old-fashioned package of contents, that is the curriculum only outlined the content and specific objectives in a particular topic, without a provision for varied and appropriate learning opportunities (UNESCO, 2002). In the OBE system, the curriculum is expected to outline the learning opportunities. Thus, learning opportunities apparently help the teachers stay focused on the larger outcome of learning and guide teachers when implementing the nationally based curriculum innovations and customisations of integrated assessment and instruction (MoE: Singapore, 2012). This integration translate that the curriculum has to outline both; content, objective and learning experiences by level to provide appropriate learning opportunities.

Teaching and learning as undertaken by the teacher have a close relationship with the teacher’s proficiency and how each aspect of the curriculum is related to one another. According to Wilkins (2008), there is a significant association between the knowledge and proficiency of teachers in a curriculum and their classroom teaching. Wilkins also argued that a weak proficiency in educational content among teachers and lack of learning opportunities outlined in the curriculum may cause confidence problem in carrying out teaching activities for their pupils. Therefore, mathematics teachers as the target for this study should also be knowledgeable of their curriculum content to enable them to integrate FA and teaching of HOTS in mathematics in the classroom.

2.4.3 The teaching of mathematics in lower primary school

Botswana has two official languages which are: 1) Setswana, which is the most widely spoken language in the country, and 2) English, which is used for official purposes (Monyaku & Mereki, 2011; Spaull, 2011). In schools, both English and Setswana are the official media of communication and instruction in all educational and official communication. Despite that, however, the Botswana population constitutes of various ethnic groups who speak more than 20 different languages other than Setswana, which is the national languages.

Following the RNPE and introduction of the new syllabus for lower, the English language also became the medium of instruction from Standard 2 onwards (MoE, 2002). This was a change to early full English immersion from late English immersion in Standard 5 (Kasule & Mapolelo, 2005). Since then, the language situation in Botswana schools has not changed. A significant number of pupils whose language of teaching and learning is not the home language are subjected to language trauma in the early foundational phase of learning. The teacher's roles as the determiner of what understanding of the content goes on in the classroom and teaching approaches to apply are also strained to cater to bi-or multilingual pupils.

Teaching mathematics in such circumstances may likely to disadvantage the young pupils in any way, and it's a strong reason for not even considering the use of a second language. According to the study carried out by Kasule and Mapolelo (2005) regarding primary school mathematics teaching in bi/multilingual northern Botswana and the findings revealed that teachers had difficulties in devising teaching strategies to overcome bi/multilingual classroom. Little efforts have been made in Botswana towards the development of a language policy that seeks to encourage the recognition of other languages for instructional purposes. The use of multiple languages seems to be a pedagogical reality. The MoESD tried to undertake a study to identify the languages in the country and assess their level of lexicon development. Unfortunately, the findings of the study have neither been revealed nor published in the public domain while pupils are continuing to learn in such a challenging learning environment (UNESCO, 2010).

2.4.4 The teaching of HOTS in mathematics for standard 4 level

Thinking skills have always been part of human existence. This thinking skill is defined as the intellectual process that involves the formation of a concept, applying, analysis or evaluating information gathered through observation, experience, reflection, reasoning, or communication (Mohamed & Lebar, 2017). Thinking skills is the process of using the mind to make decisions and solve problems. There are two levels of thinking; lower-order thinking skills (LOTS) and HOTS. Other scholars classify these levels as surface or deep thinking, respectively (Hattie & Brown, 2004, Hattie & Purdie, 1998). According to Hattie and Brown (2004), education aims to strike a balance of surface and deeper learning to lead pupils to more conceptual understanding. Such thinking is modelled within three types of understanding: surface, deep and constructed or conceptual understanding, founded on the Bigg and Collis SOLO model (Hattie & Brown, 2004, Hattie & Purdie 1998; Hix, 2012).

Deep thinking is synonymous with HOTS; therefore, for this study, HOTS will be used. Thus, the higher-order thinking of an individual depends on the individual's ability to apply, develop and enhance knowledge in the context of thinking. These HOTS comprise cognitive levels of analysis, synthesis and evaluation and mastery in applying the routine things in new and different situations (McDavitt, 1994 cited in Mohamed & Lebar, 2017). HOTS has been considered an important component in pupils' learning which must be included in the curriculum and is integral to teaching and learning during lesson planning (Sengul & Ustundag, 2010). For instance, the education systems in some Asian countries like Hong Kong, Malaysia and Japan, have long revised their curricula to integrate HOTS as the main component to develop pupils who are critical and creative in thinking and are on par globally with other children in mathematics (Balakrishnan, Nadarajah, Vellasamy & George, 2016). The integration of HOTS in the classroom mainly "occurs when pupils take new information stored in memory, interrelates and or rearranges and extends this information to achieve a purpose or find possible answers in perplexing situation" (Lewis & Smith, 1993, p. 136, cited in Balakrishnan et al., 2016). Curriculum integration in this manner had seemingly made the Asian pupils superior when compared to their international cohort in performance in TIMSS large scale assessment studies.

Pupils from African countries do not traditionally do well in studies like TIMSS. This low achievement of pupils seems to indicate a real challenge for Africa. One may wonder what factors contribute to low achievement in large-scale assessment, including HOTS items, in developing nations. Some debates about the inadequacy of assessments and examinations in some African countries (for example, Uganda) to measure HOTS is not new (Allen, Elks, Outthred & Varly, 2016; Chapman & Snyder,2000; Mainali, 2012; Mitana, Muwagga, & Ssempala, 2018). Moreover, a study by Mitana et al. (2018) contends that Ugandan's assessment seems to measure superficial learning and can hardly enable the country to achieve its National Vision or Sustainable Development Goals (SDGs). Ideally, assessment supports teachers and pupils in the teaching and learning process.

Moreover, the evidence concerning the 2007 mathematics results for Botswana Standard 4 as discussed in Chapter 1 indicates that the pupils' performance was weak. The mathematics mean performance was 30%. Benchmark performance levels developed by teachers also exhibited that 50% of the pupils did not even reach the low-performance level (BEC, 2008; MoE, 2002). According to the criteria they developed, including for HOTS items, only 22.5% of the pupils who participated in the study qualify for progression in mathematics, while the rest of the pupils did not. Pedagogical practices were found to be associated with performance (BEC, 2008; MoE, 2002).

Consequently, if the tests teachers have at their disposal are inadequate and of poor quality, it is not surprising that the teacher cannot use test results or information to guide their daily classroom practices. Similarly, if the tests do not contain many HOTS items, perhaps the teacher practice of not teaching HOTS is understandable. If the tests don't help teachers gather diagnostic or formative information, they will likely be useless for HOTS or FA (Brown & Hattie, 2012). Most importantly, one has to note the balance between the LOTS (surface) approach to understanding both how and what they should learn (Brown, 2002a) and teachers' claim that the goal of their teaching is enhancing HOTS (deep) learning (Brown, 2002b). The teachers are supposed to know how to teach and assess LOTS and HOTS to improve pupils' preparation for either low or high-stakes assessment contexts (Brown, 2002b: Brown & Hattie, 2004).

There is also literature about the nature and quality of teachers' questioning of pupils learning, which indicates that some teachers' questioning may not elicit deep thinking from pupils, particularly when teachers ask more procedural questions. With this context in mind, pupils should understand that classroom practice is a learning environment for knowledge acquisition rather than the generation of the knowledge, and consider questioning how teachers lead and control the classroom activities (Torrance & Pryor, 1998; Wade & Moje, 2000). In other settings, teachers' questioning is considered an effective strategy for pupils' deep-thinking enhancement, and pupils see learning through teacher questioning as a way for them to generate knowledge. There is evidence of teachers implementing rich, divergent, higher-order thinking questioning based on an understanding of the LOTS (surface) knowledge as part of classroom repertoires to enhance pupils' learning and thinking (Black, Harrison, Lee, Marshall & Wiliam, 2003; Gall, 1984; Wood & Wood,1988).

The fundamental issue in addressing and guiding the teaching, assessment and learning of HOTS is dependent on the taxonomy of processing. The next section will review and discuss the learning taxonomy with specific reference to the Botswana context.

2.4.5 Taxonomy for teaching HOTS in mathematics for Standard 4 level

Bloom's taxonomy has been used for more than six decades as a developmental guide of measures for cognition and achievement of educational objectives in many countries (Hattie & Brown, 2004). This taxonomy refers to the type of thinking or processing required to complete tasks or answer questions aimed at understanding, comprehending, applying, analysing, synthesising and evaluating, as depicted in Figure 2.3 (Bloom et al., 1956).

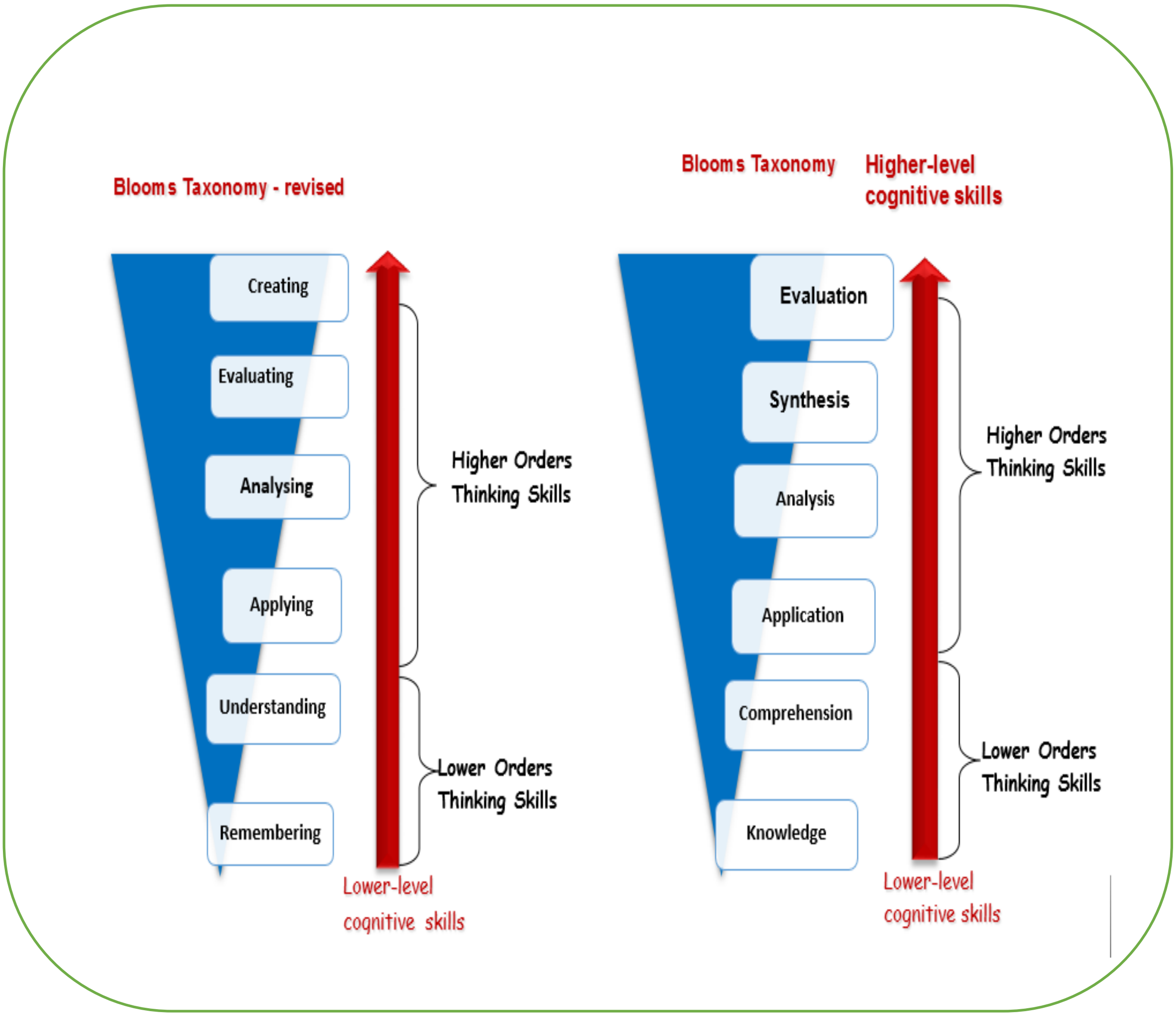


Figure 2.3: Levels of knowledge–Bloom's Taxonomy
Source: Anderson and Krathwohl (2001); Bloom (1956).

In fact, educators have typically used Bloom's taxonomy to inform or guide the development of assessments (tests and other evaluations of pupils' learning), the curriculum (units, lessons, projects, and other learning activities), and instructional methods such as questioning strategies (Forehand, 2010). The current study is interested in the integration of FA and HOTS in the mathematics classroom. The teaching of HOTS is rooted in Bloom's Taxonomy as the fundamental foundation for teaching and learning. It is, therefore, necessary to briefly reflect on the taxonomy

(Bloom taxonomy, original and the revised version, SOLO taxonomy) as well as the position of the current study on this taxonomy. The HOTS component is woven within Bloom's Taxonomy and expressed in the form of tasks that require pupils to analyse, infer, explain, synthesise and evaluate phenomena, as shown in Figure 2.3. As also reflected in Figure 2.3, the first two levels of Bloom's Taxonomy involve accumulation and understanding of information only, while the other four levels, which are often classified as HOTS, involve the application of such information for finding a solution to real-life problems, for creativity, critical thinking and judgment (Anderson & Krathwohl, 2001; Bloom, 1956; Nenty et al., 2007;).

The original taxonomy was developed by Bloom (1956) and was later revised by Anderson and Krathwohl (2001), who recognised the limitations of the earlier (1956) version. They changed the order of the cognitive processes' hierarchy (remember, understand, apply, analyse, evaluate, and create), and introduced a new dimension of knowledge type (i.e., factual, conceptual, procedural, and metacognitive) (Gichuhi, 2014; Nworgu, 2010).

According to Hattie and Brown (2004), the new version introduced a matrix of 24 process-object cells, wherein each cognitive process has four types of knowledge. Typically, the remembering process can include factual knowledge (i.e., knowledge of terminology or details), conceptual knowledge (i.e., knowledge of categories, principles, theories), procedural knowledge (i.e., knowledge of skills, methods, techniques), or metacognitive knowledge (i.e., knowledge of strategy, context, conditions).

Hattie and Brown (2004) criticised the revised taxonomy. They viewed it as serving the purposes of describing the cognitive process and the category of knowledge underpinning an educational objective, outcome, or task; arranged in a hierarchical order of increasing cognitive complexity. Metacognitive knowledge is more abstract than concrete factual knowledge and creative cognition is more complex than remembering. Hattie and Brown (2004) felt that whatever the merits of this reincarnation of Bloom's model (original and revised version of Bloom taxonomy), the most important addition should be the movement from a surface (LOTS) to a deep (HOTS) continuum. Thus, the original developers of Bloom's Taxonomy have abandoned their model primarily because of significant deficiencies in this simplistic and incorrect hierarchy of six steps. Hattie and Brown (2004) draw the attention of education stakeholders to such deficiencies and claim that they are best resolved using the Structure of Observed Learning Outcomes (SOLO) cognitive processing taxonomy (SOLO) model. However, for this study, the SOLO taxonomy will not be discussed in detail.

Despite the popularity of Bloom’s taxonomy, there is little support for its use in organising, instructing, curriculum design or assessment (Hattie & Brown, 2004). The major fundamental criticism of Bloom’s taxonomy is that there is little evidence to support the invariance and hierarchical nature of the six levels (McMillan, 2001; Hattie & Brown, 2004). Other problems of Bloom’s taxonomy include the necessary relationship between the questions asked and the responses and the separation of knowledge from the intellectual abilities or process that operate on this knowledge. The questions relate not only to complexity but also to the order of difficulty of such problems requiring behaviour at one level; and are not accompanied by criteria for judging the outcome of the activity (Ennis, 1985; Hattie & Brown, 2004; Hattie & Purdie, 1998; McMillan, 2001).

The current study sought to investigate the possible effects of a BS4MT intervention that uses the FA strategies in the classroom to enhance pupils HOTS. The original Bloom’s Taxonomy (1956) was used in the current system since Botswana’s education system still uses the framework as a foundation to guide curriculum development, teaching and learning, and setting off all educational assessments. Botswana’s education system still considers Bloom’s taxonomy applicable in characterising different levels of questions in assessments and the corresponding responses expected from pupils (MoE, 2002). Even the BEC’s national assessment items are developed based on Bloom’s taxonomy (BEC, 2008).

According to Thompson (2008), a view to creating a generation of pupils who cover all cognitive stages, mathematics teachers can use either the original or re-worked frameworks of Bloom’s taxonomy and integrate them into implementing HOTS in mathematics. For this reason, the study has chosen the original taxonomy, as it is still relevant in the Botswana context. Both the revised and original taxonomy still have some limitations to deal with how gradually teaching and learning shift between the levels (Hattie & Brown, 2004). However, the current study assumes that there are some slight shifts in learning accorded by Bloom's levels, mainly from comprehension to application and slightly towards analysis as reflected in the lower primary school mathematics curriculum (BEC, 2008, MoE, 2002; Nenty & Umoinyang, 2004).

The focus of education lies in seeking how to improve human thinking and cannot be attained by chance but must be diligently sought (Forehand, 2010). A teacher can accurately measure their pupils' ability only if they adopt and adequately use the classification of intellectual behaviour levels as identified by Bloom (1956). According to Nenty and Odili (2012), teachers must possess detailed conceptual knowledge and application of levels and sub-levels of the original Bloom's taxonomy of human cognitive behaviour and their skill demands. Therefore, teachers should be able to differentiate among the cognitive levels and develop items (tasks, questions, statements) that call on each level or sub-level of these domains.

Studies have revealed that a pupil who has acquired a higher level of thinking can do things such as analysing the facts, categorising them, manipulating them, putting them together and applying them in real-life situations (Balarishnan et al., 2016; Yee, Widad, Razali & Tee, 2011). HOTS involve the use of various learning processes and strategies in different complex learning situations. HOTS for an individual depends on the individual's ability to apply, develop, and enhance knowledge in thinking. These thinking contexts include cognitive levels of analysis, synthesis and evaluation and mastery in applying the routine things in new and different situations (Mohamed & Lebar, 2017).

HOTS are also learned in the problem-solving context, which is defined as one or more of a pupil's thought processes to achieve a goal or find a solution for a problem (Nitko & Brookhart, 2007). In all the HOTS development processes, the teachers' effort to select activities and assess a specific task is paramount. Teachers' ability to apply several modes of assessment in the classroom is very important. Besides the usual paper and pencil tests in their different forms, teachers can apply other modes, which include: oral questioning and presentation, group work or projects, portfolio, creative writing, reviews, coursework and assignments, short writing exercises, case study, research projects, learning journals, seminar presentation, peer assessment, demonstration, problem-based learning, laboratory work, work-based learning, role-playing and online assessment. The outlined methods, if well planned, is a rich means of provoking and assessing HOTS (Nenty & Odili, 2012).

Mathematics teachers have to be knowledgeable and skilful in mathematics to select an appropriate strategy to provoke and assess pupils' HOTS to prepare them for future success. It is, therefore, the responsibility of every education system to empower its teachers either during pre-service or through in-service training with skills and approaches on how to incorporate HOTS into their teaching and learning process (Abdullah et al., 2017). One of the apparent approaches is through professional empowerment of teachers on how to engage pupils in HOTS related learning activities (Rajendran, 2008).

The empowerment of teachers should influence a pupil-centred learning approach, which involves the use of hands-on activities, collaborative learning, and self and peer assessment learning opportunities in the lesson. Engaging pupils in the lesson instead of sitting quietly and listening may provide pupils with the necessary experiential learning opportunities (Balakrishnan et al., 2016). The use of appropriate questions and effective questioning techniques is also considered another means to enhance pupils' thinking skills (Supramani, 2006). In all the teaching and learning processes, the teacher is expected to play a role as a facilitator to allow discussion and encourage a free thought process to attain a pupil-centred approach to learning (Rajendran, 2008).

According to Murray (2011), the teacher should also be able to determine the pupils' level of cognitive development, use teaching techniques that are impactful to the pupils, and strategies that can be modified to ensure that all pupils can achieve the set learning goals. This notion is purported by Adams (2011), who believed that using varied strategies for teaching HOTS could improve the pupils' understanding of mathematical concepts. Pupils would know how to apply different strategies in solving problems and put efforts into proving that every solution has its justification (Abdullah et al., 2017).

In summary, the implementation of HOTS by engaging pupils in teaching and learning processes of mathematics is essential to change the stigma of the difficulty of mathematics (Abdullah et al., 2017; Serin, 2013). In this case, teaching HOTS should be attractive to pupils to foster their interest in mathematics. Tengku Zawawi et al. (2009) believe that mathematics will be appreciated as an easy and fun subject if pupils are allowed to build their understanding, attitude, and creativity. HOTS are consistent with this idea because one of HOTS's indicators is creating a continuous learning and instilling creativity among individuals.

Despite the obvious advantages of competent teachers who can elicit HOTS in their teaching and assessment of pupils, some studies show evidence of a limited understanding of using HOTS by teachers. A study carried out by Balakrishnan et al. (2016) revealed that teachers are still confused about elements of HOTS and how to incorporate these elements in teaching and learning. Such limitation on HOTS usage has a similar discrepancy observed in Botswana, in which pupils were reportedly unable to apply HOTS (Masole et al., 2016). The problem might probably have emanated from the traditional teaching technique in which teachers dominate the classroom activities, limited engagement of pupils, as well as teachers' inclination of teaching to the test (Fetogang, 2015, 2016; Hattie, 2009; Raboijane, 2005).

2.4.6 Assessment of HOTS in mathematics for Standard 4 Level

In Botswana, pupils' achievement in mathematics is assessed through continuous assessment (tests) and examinations. Criterion-referenced testing is utilised as the basis for assessing pupils for diagnosis and remediation purposes. The use of criterion reference testing implies that items for HOTS are included in the process of test development as outlined in the learning objectives (UNESCO, 2010). And HOTS items are assumed to be the items that assess cognitive skills for application, analysis, evaluation and creativity (Mohamed & Lebar, 2017).

The Standard 4 mathematics attainment test is developed in partnership with classroom teachers and Botswana Examination Council. However, the test is administered and marked locally by the teachers. The Standard 4 mathematics attainment test serves as a checkpoint to enable the teachers to diagnose the learning problems of the pupils to plan remedial measures, however, evidence of utility is very limited. Those tests cover the work in the four levels of the lower primary syllabus (Monyaku & Mereki, 2011; Spaull, 2011).

According to UNESCO (2010), at the classroom level, Botswana teachers are encouraged to keep performance records of all pupils and appraise their achievements against the skills that they should have achieved at each level as stipulated in the curriculum. Such a teacher-based assessment is assumed to be FA at the school level. As UNESCO further clarified that teachers are expected to develop pupils profile records to reflect what the pupils can do, what their strength and weakness. It is clear from the former statement that, FA is done for the sake of administrative compliance and record-keeping exercise of individual pupils through quizzes, topic tests, end-of-term tests, and mock examinations. There is little evidence regarding FA as a means to guide instruction in teaching and learning mathematics.

Such a discrepancy in the use of FA has influenced some stakeholders to accuse teachers of conducting both classroom and assessment practices that emphasise only lower levels of cognition for administrative compliance. As well as implementing classroom questioning and test items which are geared to encourage the development of cognitive skills (Nenty, Adedoyin, Odili, & Major, 2007). This is generally inadequate as a means of preparing the pupils to fit in and contribute to the development of self and society.

2.5 Nature of Assessment

The word “assessment” comes from the Latin verb “assidere” meaning “to sit with”. This word origin implies that in assessment, the teacher sits with the pupils and assessment is something teachers do with and for pupils rather than to pupils (Green, 1998 as cited in Heritage, 2010).

Several operational definitions have been advanced for the concept of assessment and learning in the literature (Baird, 2010; Baird et al., 2017). However, as the concept of assessment applies to learn and educate, it is anything done to find out what knowledge, skills, habits, attitudes, practices or generally what behaviour a pupil does or does not have, acquire or develop, before, during and at the end of an instruction, a period of instructions, or a course of study (Baird, 2010; Nenty, 2004; Nenty, 2008).

In the classroom, assessment is the means through which inputs into the teaching-learning process, the learning process itself, as well as the changes that accrue as a result of this process, are described and documented qualitatively or quantitatively, and are compared to what was expected or to what is determined to be desirable (Nenty, 2004). In general, assessment should serve these purposes; "measure learning and achievement, diagnose learning, promotes and markets institutions, provides feedback and feedforward for pupils and teachers, certification of learning, development of learning outcomes for the course and programme and development of knowledge, skills and dispositions for a long term and judgement assessment'' (Spiller, 2015, pp. 1-2).

In almost all cases, especially in a formal setting, there are various types of assessment, and the two main forms of assessment are summative and formative. Their results are used for making a different decision (Heritage, 2010; Nenty, 2004). For this study, formative assessment and summative assessment will be discussed further in the following next sections. After the discussion of the two forms of assessment separately, the section would end by describing the recent research project, which attempts to implement FA in African states and the need for teacher training in FA.

2.5.1 Summative Assessment

SA measures the achievement of tasks or goals, encapsulating all collected evidence up to a given point to yield either comparative or numerical ratings (Buchholtz, Krosanke, Orschulik & Vorholter, 2018; Taras 2005). SA is used for evaluation, in which there is limited or no feedback beyond the achievement report, and is usually a numerical or letter grade score. SA wants to find out whether the instructional goals of the unit have been achieved or not. SA is an activity, typically a written test given at the end of a term, chapter, semester, year, or the like, for grading, evaluation, or certification purposes. The goal of SA is to make a judgment of pupils' competency (Wiliam, 2013).

SA is delivered in different ways. Glazer (2014) outlined that SA is to include for example: interviewing; observing; measuring through a variety of classroom testing and examinations. SA is also termed assessment of learning, emphasising its nature as an assessment of an activity that has occurred (thus, after a period of learning). However, the term also emphasises a numeral aspect and is often associated with a number or letter grade. Where this grade gets a high weighting or has significant consequences for progress, it can be termed high stakes assessment (Bonner, 2016). In Botswana, SA seems to dominate the teaching and learning environment through the use of tests and national examinations. This SA is often a key driver for pupil learning ("if it's not assessed, it doesn't matter"). It can be associated with pupil and/or staff anxiety and workload and build pupils and staff confidence when the outcome is to their satisfaction (Evans, 2013). The pupils'/or staff's anxiety seems to negatively impact many high stakes' national examinations (including Botswana's examinations) since large amounts of homework, weekly testing and drill, and practice exercises characterise most classroom teaching and learning (Kennedy 2007). These activities are done to prepare pupils for public examinations, and teachers are more often reluctant to move away from these processes that usually have support from school leaders and parents (Kennedy, Chan, Fok & Yu, 2008).

2.5.2 Formative Assessment

FA often used interchangeably with the term "Assessment for Learning" (AfL) as it was explained in Chapter 1, Section 1.1.3, refers to any assessment activities undertaken by teachers and by pupils themselves that provide feedback, which is then used to adapt teaching methods to meet pupils' needs and improve learning outcomes (Black & Wiliam, 1998 cited in Education Endowment Foundation, 2018). FA is task-oriented that provides feedback to pupils on their learning achievements during the learning process (Glazer, 2014; Wiliam, 2013). The definition of FA has evolved over a while. For instance, during 2016-201ative Assessment for Students and Teachers (FAST) State Collaborative on Assessment and Student Standards (SCASS) members reviewed the original definition and attributes, identified areas that are emphasised in current FA research, theory, and practice that were not addressed adequately in the original definition phrased in the year 2006. The updated definition of FAST SCASS members developed as follows:

> *Formative assessment is a planned, ongoing process used by all [pupils] and teachers during learning and teaching to elicit and use evidence of [pupil]*

> *learning to improve [pupil] understanding of intended disciplinary learning outcomes and support [pupils] to become more self-directed [pupils] (FAST SCASS, 2016, p.1).*

The above definition means that FA practices in the classroom explore the pupils' present state of knowledge, where the pupils need to reach in their learning and how appropriately can the pupils be helped to reach a point, he/she supposed to be in their learning (Wiliam & Thompson, 2014; Wiliam, 2013). The FA concept is expected to both to provide support and improve teachers' instructional planning and pupils' progress (Heritage, 2010). As purported by the Partnership for Assessment of Readiness for College and Careers (PARCC) that teachers "need additional support to collect evidence of learning to inform instruction, second by second, minute by minute, hour by hour, day by day, and week by week" (PARCC, 2010, p. 56).

FA is all about informing pupils of their progress to empower them to take the necessary action to improve their performance. Teachers need to create learning opportunities where pupils can progress at their own pace and undertake consolidation activities where necessary (Foster, 2019). It includes; open-ended response questions, essays, problem-solving tasks and performance tasks, such as posters, presentations or projects. It may also include closed-ended questions, such as multiple-choice questions, when used for providing feedback to guide the pupil's growth. Race (2009) emphasised the importance of having qualified feedback by first restating an analogy he gave credits to John Cowan, "Assessment is the engine that drives learning" (Race, 2009, p. 47), and then extending it to add that, "feedback is the oil that lubricates the cogs of understanding" (Race, 2009, p. 47). Thus, the way feedback is produced are important for achieving maximum efficiency of the learning process (Black & Wiliam, 2003, 2006; Nicol & Macfarlane-Dick, 2006; Weurlander, Soderberg, Scheja, Hult, & Wernerson, 2012).

FA activities are ongoing and part of the learning process in the classroom; it features activities that provide feedback to the pupils and teachers during the learning process, rather than after a period of instruction (Education Endowment Foundation, 2018). FA as an Assessment for Learning has often been extended to include the idea of the empowerment of pupils to self-regulate and critically evaluate their learning and performance (Black & Wiliam, 2003, 2006; Carless, Salter, Yang, & Lam, 2011; Sadler, 2010). Furthermore, some authors separate this concept to highlight this important skill of self-assessment and refer to it as an assessment as learning (Earl & Katz, 2006). Pupils need support to develop this skill of self-assessment as a way of promoting the development of an independent individual through assessment as learning (Evans, 2013). The teacher is most responsible in FA, but in the assessment as learning, it is the pupils who become more empowered, are more responsible and can become the key decision-makers in setting goal their own goals, and monitoring their learning progress. However, both the facets of FA (Assessment FOR Learning and Assessment AS Learning) provide learning opportunities to integrate activities that encourage pupils to think critically and to practice lifelong skills, such as presentation, communication, analytical, and problem-solving skills, as well as to practice teamwork (Foster, 2019). The exposure to such lifelong skills could also help pupils who are not performing well on traditional assessment tasks to demonstrate their knowledge in alternative ways.

For this PhD thesis, FA is discussed within the context of Assessment for Learning primarily due to the feedback it provides on teaching and pupils learning. As the current study is concerned with pedagogical and classroom contexts, FA is appropriate to enhance pupils' motivation and commitment to learning. When teachers commit to learning as the focus of assessment, they change the classroom culture to one of pupil success (Torrance and Pryor 2001; Black et al. 2003 a, b). Most importantly Botswana has little empirical evidence on FA, hence this study will attempt to contribute towards the similar main concerns of Black and Wiliam (1998).

2.5.3 Integration of Formative Assessment in Teaching and Learning from Developed Nations contexts

FA has been an integral and important tool in the teacher's hands through which the country's quality of education could be assured in developed contexts. According to Leahy, Lyon, Thompson and Wiliam (2005), FA has been implemented for several years by the elementary, middle and high school teachers in some American states like Arizona, Delaware, Maryland, Massachusetts, New Jersey, New Mexico, Pennsylvania, and some countries like the United Kingdom, Canada, Singapore, Malaysia and Australia respectively and have demonstrated desirable results in teaching and learning. However, according to Brown, Gebril and Michaelides (2019), the implementation contexts for assessment in these developed nations are grouped as to whether each jurisdiction is defined as being a relatively low-stakes assessment environment (e. g. New Zealand, Queensland, Cyprus, and Catalonia) or examination dominated (e.g. Hong Kong).

In low-stakes assessment environments, like New Zealand, schools use assessment to improve pupils' learning outcomes and provide guiding information to managers, parents, and governments about the status of pupil’s learning (Ministry of Education, 1994). Additionally, the Ministry of Education, New Zealand, provided professional development programmes that focused on teachers’ use of formative assessment for learning (New Zealand Ministry of Education, 1994). Queensland in Australia, similar to New Zealand, had an outcome-based curriculum framework, had limited use of mandatory national testing, and a highly-skilled teaching force. Primary school assessment policies (years 1–7) usually differed from those of secondary schooling (years 8–12) (Brown et al., 2019). Spain’s assessment policy prioritises low-stakes, school-based, continuous, formative, and holistic practices (Brown et al., 2019). At the same time, Finland is currently implementing phenomenon-based learning, a purely formative assessment related model (Lonka, 2018).

As for the high stakes’ assessment environments, teachers’ implementation context of assessment is inclined towards the role of summative examination, and a paradigm shift from these notions could pose substantial challenges to their teaching practice. According to Brown et al. (2019), some high stakes assessments like those in Hong Kong are still influenced by the UK Assessment Reform Group. Thus, Hong Kong has discussed the use of assessment for learning extensively (Berry, 2011), while at the same time maintaining a strong examination system and culture (Chee-Cheong, 1999). FA seems to have been firmly accepted, and while summative testing is inevitable, calls have been made for the formative and diagnostic use of summative testing (Brown et al., 2019; Carless, 2011). Similar sentiments are observed with South American schools, particularly in Ecuador, where schooling is generally characterised by traditional solid conventions of examination and pedagogical practices (Brown et al., 2019). The next section will discuss FA practices with a specific emphasis on the African context.

2.5.4 Formative Assessment in Africa

The attempts to pursue FA in the classroom environment for African states is long overdue when compared to other developed countries. Some African countries have been claiming to have a system of assessment in place that utilise both informal and formal methods to measure pupils' proficiency and provides teachers with data to inform instruction (Perry, 2013). Such countries among others are Botswana (Ottevanger, Akker, and Feiter, 2007); Ethiopia (Smith, Stone, & Coming 2012); Ghana (Pryor & Akwesi, 1998); Lesotho (Sebatane, 1994); Malawi (Mchazime, 2003); Namibia (Ottevanger et al., 2007) Nigeria (Dibu-Ojerinde, 2005); South Africa, (Kuze & Shumba, 2011; Lubisi & Murphy, 2002) and Zambia (Kapambwe, 2010). However, more observational evidence is needed to understand how teachers utilise formative assessment data (Perry, 2013).

Despite this unclear picture, however, an FA trend is beginning to shift towards the African states to join the developed world in pursuing FA in the classroom environment. As noted earlier that FA is interchangeably used with AFL. Thus, FA strategies intervention is currently implemented in other parts of Africa which include Tanzania, East Africa and two sites in South Africa where it is called Assessment for learning in Africa (AFLA). The project research aimed to improve teachers' numeracy skills and their understanding of how numeracy can be more effectively communicated to their young pupils and to improve children's outcomes in the form of test results through appropriate assessment strategies. Their approach involves the development of classroom materials for primary school numeracy development and uses these as the basis for workshops and the development of teacher learning communities in each research site (Aga Khan University, 2017).

According to Aga Khan University, (2017), the research team for AFLA includes researchers with extensive experience in implementing formative assessment internationally, numeracy specialists and academics based in South Africa and Tanzania who are familiar with their contexts and local conditions. The FA (AFLA) project focuses on the use of assessment for improving learning outcomes in the core curricular area of numeracy in particular in challenging educational settings described as schools with harsh realities with limited resources such as large classes and few resources (Aga Khan University, 2017; Kanjee, 2017).

Following this research project, McGrane, Hopfenbeck, Halai, Katunzi and Sarungi (2018) also carried out a follow-up study on:

- FA -primarily teacher-focused;
- The psychometric challenges of assessing the learning of students in schools located within the urban slums of Dar es Salaam Tanzania,
    - Developing an appropriate mathematics assessment for the student in this context, and
    - Developing an appropriate Socio-economic status indicator for those students

This was a mixed-method study which involved teacher workshops; classroom observation; interviews with teachers, students, head-teachers; and assessment of mathematics learning across the school year. The data was analysed using the Rasch Model, as it was deemed as the simplest Item Response Theory model. Simple parameterisation made the estimates more robust with samples sizes, and performance on different tests was represented on the same scale using common-item linking across the tests (McGrane et al., 2018).

The findings of the study revealed no significant difference between treatment schools and control schools in the average end line performance. The results also revealed that differences in socio-economic and cultural circumstances do not significantly predict end line performance. However, the preliminary findings led to propose some area of further discussion as outlined below;

- Close and ongoing collaboration with local expertise is key to the psychometric research in this challenging African context in terms of assessment design, translation, administration, piloting and analysis.
- Proper targeting of the assessments to the students in these contexts is crucial.
- The impact of AFLA may be more visible in the other assessment, particularly those focused on the teachers.

The socio-economic circumstances of predicting learning outcomes were not significant. However, these factors can further be explored in terms of teacher perception in FA and enhancement of teaching and learning. For instance, some other recent studies on teachers' perceptions and formative assessment practices in the Tanzanian context include an empirical study conducted by Kyaruzi (2017) as part of a PhD dissertation. This study addressed two general research aims: (i) to investigate Tanzanian secondary school mathematics teachers' perceptions and practices of formative assessment and pupils' perceptions of their teachers' practices, (ii) to investigate the impact of an intervention on feedback processes during mathematics education in secondary schools in Tanzania, i.e., feedback provided by the teacher; how this is perceived by pupils, and whether pupils incorporated feedback. Surveys, interviews, focus group discussions, and video observations were used to obtain data. The findings seemed to encourage teachers and pupils to use pupils' errors formatively to improve the instructional process (Kyaruzi, 2017). The methodological approaches were found relevant to this study.

Another similar study FA was carried out by Kanjee (2017) to investigate the teachers' understanding of FA in the classroom in South Africa. The results of the interview on using questioning technique revealed the following

- No evidence of planning, one of the participants said:

"Because I am teaching for several years, I thought I need not prepare questions beforehand. I now plan deliberately, and learner enjoy answering it"

- Across all quintiles, teachers asked lower-level questions

"I give my learners a range of answers, and they must do the mathematics and explain to me how they obtained the answer."

While the observation results on the knowledge and use of FA were ranked from low to average. The following recommendations were made by the study:

- Increase the time to develop skills in writing and using learning intentions (LI) and success criteria's (SC);
- Develop a template to assist teachers to plan and prepare lessons using available material;
- Introduce and model new FA techniques in every workshop and;

Allow teachers to reflect on current practices and use new knowledge to make relevant changes (Kanjee, 2017).

The few empirical studies mentioned above on teacher perceptions and practices of FA provide an overview of the preliminary findings of the nature and impact of FA in schools in African countries. Given those findings, Botswana teachers are almost in a similar situation and are faced with the challenge of ensuring that pupils perform in mathematics, and hence FA (AFLA) interventions are the potential area for benchmarking for the current study. Thus, the AFLA approaches are in line with this study which seeks to find a solution to improve teachers' numeracy skills and their understanding of how numeracy can be more effectively communicated to their pupils and to improve pupil's outcomes in the form of test results through appropriate assessment strategies.

2.5.5 Need for teacher training in Formative Assessment

Several studies have been conducted about teachers' support and effective implementation of formative assessment strategies towards the improvement of pupil's achievement through professional development (Ramsey & Duffy, 2016, Kyaruzi, Strijbos and Ufer, 2020; Black & Wiliam, 2003; 2006). Kyaruzi et al. (2020), purport training as a short-term professional development opportunity. In the case of FA, such training would require support for capacity building, including time for reflection, in-school modelling and coaching, access to materials and ongoing targeted feedback. In mathematics education, professional development has focused on assessing professional training effectiveness in pre-post-test forms (Kyaruzi et al., 2020).

Gaining insights and understanding of teachers' perceptions and use of FA as professional development has been explored in previous studies (Accado, 2017; Melina, 2017). As an example, Melani's (2017) findings for an in-depth case revealed that, when provided with specific information about FA through staff development, teachers became more positive toward such assessment, and their implementation skills were greatly improved. When teachers are trained, they can support the FA as a method for monitoring pupils. Accado (2017) conducted a project for mathematics teachers in which they received professional development (PD) in FA provided by a university faculty with expertise in mathematics education and instructional methods for students. Specifically, the findings revealed that regular use of specific questions during lessons enhances mathematics teaching (Accado, 2017).

According to Accado, (2017) teachers who received support for PD on formative assessment methods, including training on breaking problems into steps for error analysis, using data collection charts to identify student response patterns, providing multiple probes to assess pupils understanding, and embedding one key question into a formative assessment for analysis, were able to implement the FA strategies and improvement in pupils' performance.

In African countries, as already captured in the previous reviews, Perry 's (2013) study for formative assessment use and training in Africa has revealed that there was support for teachers in understanding how to use FA data. Thus, additional teacher training was being advocated for in this area due to the lack of attention it has previously received and because insufficient teacher training is frequently identified as a barrier to implementation and change (Broun & Kanjee, 2006).

The following sections will pay particular attention to the evidence and the effects of providing training on how to use FA specifically in African countries. Recent studies in Tanzania include Kyaruzi et al. (2020), who revealed that mathematics teachers who received short-term professional development training appeared more error-friendly and positively utilised errors in teaching. Moreover, three other related empirical studies reviewed dealt with teachers' perceptions of FA and feedback practices in mathematics education in the Tanzanian context. The findings of those studies support FA practices. Teachers and students were encouraged to use student errors formatively as learning opportunities and to improve the instructional process. Teachers should improve their teaching strategies while students must improve their learning strategies (Kyaruzi, 2017; Kyaruzi et al., 2018, 2019, Kyaruzi et al., 2020).

Akom (2010) in Cameroon conducted a one-day assessment workshop on effective questioning, how to analyse students work, and the formative assessment cycle in which pre and post classroom observation data was collected. The findings revealed that teachers were assessing students throughout instruction after receiving the training. The teachers' level of questioning and quality of feedback was also improved. However, the teachers had difficulty in making real-time instructional changes based on assessment information.

In another study, Miske (2003) studied the Malawian teachers who received four weeks of intensive training throughout the school year on assessing students, providing remediation, and giving feedback. Teachers also received classroom visits, feedback, and support from coaches. Findings showed that teachers were able to assess student learning effectively and use the data, modify their teaching and quantitative data which suggested that the training impacted students' Mathematics and English performance.

In South Africa, Kanjee's (2009) study involved teachers who attended three workshops on how to use assessment resource banks, monitor student progress and develop interventions. While in Zambia, Kapambwe (2010) teachers received intensive training on the use of continuous assessment materials. The findings of the two studies found that teachers were effectively utilizing the formative assessment tools in their classroom and planning of instruction.

As much as possible, the experts support training teachers to impart knowledge that must play a critical role in shaping their assessment practices and a bearing on their ability to convert espoused theories about assessment into actual classroom practice. Notably, there are externally imposed barriers that constrained the use of formative assessment which included expectations, habits, and dispositions of students; the pressure that teachers felt to "cover all of the curricula to prepare students for the end-of-year, high-stakes exam; and an instructive rather than a constructivist approach to teaching and learning" (Box, 2015).

For instance, a study by Chemeli (2019a) investigated teachers' support for the effective implementation of the five key formative assessment strategies in mathematics instruction in secondary schools in Kenya. Findings revealed that there was inadequate teachers' support, lack of training offered to teachers on Formative Assessment Strategies (FAS), adequate resources and materials, big class sizes which did not favour the good use of FAS, lack of time to plan for FAS, FAS was not included in the curriculum, inadequate head teacher support and lack of understanding of learners' context by the teachers. The study recommends that teachers support is needed in terms of time, curriculum modifications, resources/materials, class sizes and periods. Intensive training of head teachers and all stakeholders should also be done on the support needed for the successful implementation of FAS. While on another similar study by Chemeli (2019b) findings were significantly reflecting the positive impact of formative assessment on students' achievement in mathematics.

In summary, the evidence presented in these FA studies in the African context is dominated by East African countries, with two studies in Southern Africa. Botswana teachers' perspectives and FA practices were not reflected in any studies, and in general, the evidence does not converge on the most effective methods of training teachers in Africa. Additionally, the evidence of effective training programmes should be examined from both the developing and developed world (Perry, 2013). It is also worthy to further explore FA in the Botswana context and identify its acceptance, hence this study.

2.5.6 Problems with Formative Assessment

Even though there is a growing volume of literature reporting the positive effects of FA on teaching practice and pupils' learning outcomes, there is also an increasing volume of literature on the difficulties of introducing FA in regular classroom settings (Antoniou & James, 2014; Dunn & Mulvenon, 2009). There are several reasons why the successful implementation of FA is still problematic.

Firstly, the effectiveness issues identified by Black and Wiliam's (1998) meta-analysis of more than 250 studies reported effect sizes of between .4 and .7 in favour of pupils taught in classrooms where FA was employed (Stiggins, 2006). However, Bennett (2011) noted the inconsistency of the original meta-analysis of Black and Wiliam (1998) about the approach used to explore research studies and draw conclusions on FA. Black and Wiliam (1998) reviewed unrelated studies which included feedback, pupil goal orientation, self-perception, peer assessment, self-assessment, teacher choice of the assessment task, teacher questioning behaviour, teacher use of tests, and mastery of learning systems (Bennett, 2011). Such a collection of themes is simply too diverse to be sensibly combined and summarised by a single, mean effect-size statistic (or range of mean statistics), (Bennett, 2011; Dunn & Mulvenon, 2009). With FA it is also difficult to achieve the required learning outcome because empirically derived models of learning are not generally available and the shift in teacher practices required is large and may also involve changing teacher beliefs and values related to effective teaching and learning (Brown, 2019; Dunn & Mulvenon, 2009; Webb & Jones 2009). Therefore, the standard of rigour being advocated is a scientific one, similar to that required for effectiveness claims behind any education intervention (Bennett, 2011).

Secondly, the domain dependency is still a significant implication of teachers' weak cognitive domain understanding in terms of what questions to ask of pupils, what to look for in their performance, what inferences to make from that performance about pupils' knowledge, and what actions to take to adjust instruction. The other implication is that the intellectual tools and instrumentation that ought to be given to teachers may differ significantly from one domain to the next because they should be tuned specifically for the domain in question (Hodgen & Marshall, 2005). Nevertheless, Bennett (2011) suggests that FA should be essentially curriculum embedded, a position that Shepard (2006, 2008) has espoused, and Shavelson (2008) has illustrated. It may be workable, for instance, to provide FA materials for the key ideas or core understandings in a domain, which should be common across curricula (Bennett, 2011).

The third consideration deals with measurement issues. According to Dunn and Mulvenon (2009), FA, like all educational measurement, is an inferential process because we cannot know with certainty what understanding exists inside a pupil's head. Teachers can only make conjectures based on what they observe from class participation, classwork, homework, and test performance. Backing for such conjectures' validity is stronger when teachers observe reasonable consistency in pupil behaviour across multiple sources, occasions, and contexts (Bennett & Gitomer 2009; Shepard, 2006). Bennett (2011) supports the centrality of inference in FA which becomes quite clear when teachers consider the distinctions among errors, slips, misconceptions, and lack of understanding.

According to the Organisation for Economic Co-operation and Development (OECD), many countries developed commonalities of understanding and practice about FA (Sebba 2006). However, the difficulties in effective implementation need to be identified and tackled by researchers and policymakers if FA is to fulfil its promise (Baird 2010). FA seemed to be supported in rhetoric terms rather than any real understanding of the process involved (Antoniou & James, 2014; Brown et al., 2019; Darling-Hammond 2004, Carless, 2011; Gattullo, 2000). Thus, as Antoniou and James (2014) argue, without a systematic analysis of FA, based on empirical research in classrooms, research evidence can only provide a reliable and valid understanding of the nature and process of formative assessment (Torrance & Pryor, 2001).

The fourth criticism with FA is in line with professional development issues. Much of the literature on FA conceptualise it as pedagogical knowledge rooted in teaching and learning (Black & Wiliam, 2009; Harrison, 2005; Wiliam & Thompson, 2008). Bennett (2011) expands on the professional development issue by considering a deep understanding of domain and measurement fundamentals as a way to enhance pedagogical knowledge of the teachers. According to the literature, professional development has a central role in developing a firmer foundation upon which in-service programmes can subsequently be built (Bennett, 2011).

Lastly, as already mentioned earlier, the systems issue of having various definitions and an understanding of the FA concept has created confusion about what FA implies in terms of classroom practices (Brown, 2019; Klenowski, 2009; Shavelson,2008; Swaffield, 2011). The jurisdiction in which the teacher operates influences the assessment practice. For instance, in high stakes examination environments, most teachers are familiar with summative assessment, and only a few implement FA effectively in their classrooms (Black and Wiliam 1998). Several studies (Morgan 1996; Preece and Skinner; 1999; Shen 2002) have shown how summative assessment requirements dominate many teachers' assessment practice. The challenge with the summative assessment about its ability to improve the teaching and learning process as observed by Black and Wiliam (2009); and educational accountability becomes synonymous with pupil’s achievement outcome testing and the sanctions that accompany the results (Darling-Hammond 2004).

Teachers usually do not distinguish between formative and summative purposes (Antoniou & James, 2014; Bachor & Anderson, 1994; Shepard, 2000). This assumption presents a challenge for the system's assessment components that can either work against one another or larger societal goals (Bennet, 2011). Moreover, effective implementation of FA requires developing new tools and changing classroom practices in terms of practical issues (Black & Wiliam, 2003; Brookhart, 2010). FA's effectiveness will be limited by the nature of the larger system in which it is embedded and, particularly, by the content, format, and design of the accountability test (Bennet, 2011; Bennett & Gitomer, 2009). However, changing the system means remaking the accountability tests to maximise the impact on learning and instruction through FA, which may be a big challenge to other countries, particularly the developing countries.

From this perspective, this study further explored empirical literature within the five formative assessment strategies to search for a guide to developing an effective training programme conceptually, based on a framework of actions and strategies encompassing the relationship between teaching, assessment and learning.

2.6 Empirical Literature in The Formative Assessment Framework

Many scholars have written articles about FA. Black and Wiliam (2006, 2009), Wiliam (2007, 2011), Heritage (2010) and Kivunja (2015) provide a theoretical grounding for FA. The framework is centred around the common FA strategies, which are (i) sharing learning goals, (ii) engineering effective classroom, (iii) providing feedback to pupils and (iv) pupils self and peer assessment. The following section outlines the findings of several related studies within the FA framework from both developed and developing countries. Given this outline of already existing literature, it will attempt to fill the gaps and add to the body of literature on the application of FA and teaching HOTS to enhance pupils' achievement, specifically in the Botswana context with pupils with low achievement in mathematics backgrounds.

FA is a systematic process to continuously gather evidence about learning for both pupils and teachers. The data is used by teachers to identify a pupil's current level of learning and to adapt lessons to help the pupil reaching the desired learning goal. Implementing classroom FA improves student achievement (Black & Wiliam, 2003, Heritage, 2010; Stiggins, 2009). Recognition of the effectiveness of FA comes at a time when the Botswana nation is choosing to focus and change the traditional modes of assessment for high-stakes administrative summative tests as the sole instrument needed to evaluate all educational stakeholders (BOT, 2015) and not on building teacher capacity to capitalise on the research-proven effective instructional strategies.

As for FA strategies implementation, teachers are required to attempt to address all five broad formative assessment strategies in their classroom, but the specific techniques that they used within each strategy are up to the individual teacher (Education Endowment Foundation, 2018). For this study, each FA strategy would be explored and limited to the empirical review towards the articles relevant to FA in mathematics primary school's classroom setting as well as focused on the assessment of the HOTS. The empirical review of some relevant articles within the theoretical framework of the FA strategies would provide guidance and direction for the current study to develop its conceptual framework to answer the research questions of this study.

2.6.1 Sharing Learning Goals and Success Criteria

Classrooms in which teachers and pupils engage in FA processes are ones in which teachers are explicit about expectations for learning. Teacher and pupils are to monitor pupils' work in terms of progress towards those expectations. The first process in FA is clear about what is it that one would like to know. As supported by Tobey and Goldsmith (2013) that the shared element of reflecting on pupils' understanding in the context of clearly identified learning goal (LG) to help pupils learn and monitor their progress, will receive feedback which is intended to promote further learning, and incorporate the feedback into subsequent work.

According to Anderson and Palm (2017), emphasis on the clarification of learning intentions may also involve the teacher setting up activities with which pupils' understanding of the goals may be facilitated through discussion and negotiation. Teachers must explain LG and check that pupils' understanding of what they have to do during the task, what they have to learn from doing it and why they have to learn it.

Teachers often do not communicate the LG either by writing it on the chalkboard or in manila charts (Black & Wiliam, 2006, 2009). This practice poses challenges to teaching and learning because pupils do not understand why they have to undertake a task and how it fits into the curriculum. If they did, they would be more inclined to do it (Black & Wiliam, 2006). Jones, Alexander and Estell (2010) suggest learning outcomes, in terms of the work that pupils have to produce, have to be clearly communicated to pupils and involve them in the learning process. Learning cannot be done to individuals; it has to be done with them and by them. It is therefore evident, that effective two-way communication is the key to unlocking pupils' full potential to learn and ultimately achieve the desired outcomes (Heritage, 2010; Wiliam, 2011). Teachers should encourage pupils to observe how others are responding to a task so that they will begin to apply the assessment criteria to their work. Jones urges teachers the need to explain:

- the learning objectives and why pupils have to achieve them (and check pupils' understanding);
- the assessment criteria and how to use them; and
- what pupils have done well and what they need to do to improve.

There is a body of research that indicates that when pupils are given learning goals that describe the intended learning, they perform significantly better than pupils who are given performance goals that focus on task completion. The teacher's effort of teaching focuses their attention on learning by helping pupils understand that the assignment is the means and the learning is the end (Hattie & Timperley, 2007; Tobey & Goldsmith, 2013; Wiliam, 2011).

Specifically, success criteria (SC) has been defined and linked to the learning outcome and that it merely tells pupils what criteria are being assessed to measure whether the learning outcome has been achieved (Wiliam, 2011). Additionally, the success criteria are linked to the learning outcome and indicate achievement. According to Tobey and Goldsmith (2013), the SC must be shared with the pupils and must relate directly to the learning outcome and the task. It has to be explained in a language the pupils can understand. In many instances, the SC are not "owned" by the teacher but can be generated with the pupils. Sharing the SC with pupils allows them to self-edit and self-assess. Thus, take greater ownership of their learning (Heritage, 2010).

The SC helps pupils to strengthen their learning, assess themselves independently and know what is expected of them. It takes away the mystery pupils do not always know what the teacher's criteria for assessing their work are, but by sharing the criteria, it becomes transparent. It also means that one is not assessing pupils on what they 'don't know' but on what they have been learning/doing. SC can be presented in various ways and be it for the whole class or tailored to individual pupils needs (Hattie & Timperley, 2007; Wiliam, 2011).

A few empirical studies have revealed that identifying and sharing LG and SC are not typically a part of most teachers' classroom assessment practices, and when implemented a critical feature of the strategy was routinely absent. For instance, Wylie and Lyon (2015) specifically explored the extent and quality of teacher' FA practices used to identify and share learning goals and success criteria through a two-year professional development. The findings revealed that teachers frequently implemented only one FA practice associated with the learning goal and the teachers scored significantly lower on the quality of implementation scale for this FA strategy as compared to the other FA practices implemented.

Another similar study by Suurtamm (2004) explored five mathematics teachers' authentic practices in Ontario. Through interviews, classroom observations, and focus groups, Suurtamm found that teachers in the study discussed success criteria with the class before an activity. However, these five teachers were the exceptions, as teachers in this study noted that other teachers in their schools continued to use traditional forms of assessment.

Rahman (2018) explored science teachers' perception of the classroom assessment in secondary schools of Bangladesh. On teaching-learning activities, the findings revealed that teachers strongly agreed with both explaining learning objectives and discussing topic according to objectives as imperative. Nevertheless, during the interviews, most of the teachers asserted that they only mention learning objectives before starting a topic, but they would not always discuss all the prescribed topics, instead, they opt to progress the lesson as they preferred. Even though the reviewed studies have demonstrated some evidence about LGSC, findings remained limited given the research designs (mainly qualitative), and sample sizes employed.

2.6.2 Engineering effective classroom discussions, questions, and tasks that elicit evidence of learning

The second operational strategy of FA is the monitoring of pupils' learning. This strategy of engineering effective classroom discussions, questions, and learning tasks focus on the teacher's ability to diagnose the state of pupils learning on an ongoing basis. This strategy involves teachers engineering and coordinating classroom activities based on the learning goals to facilitate the collection of accurate learning evidence from all pupils and providing a valid interpretation of pupils' understandings (Bennet, 2011; Campbell, 2013). The key point for this strategy of engineering effective classroom discussions, questions, and learning tasks is that these tools are used for the intentional purpose of monitoring pupils' learning so that the learning gap may be closed (Pinchok, Brandt & Learning Point, 2009) and this strategy is considered as the heart of FA (Campbell, 2013).

The common techniques of engineering effective classroom learning are questioning and collaboration. Effective questioning is one technique which can help elicit information, probe thoughts and ideas, tapping into different types of knowledge, and investigating deeper levels of understanding (Ruiz-Primo, 2011). According to Wiliam (2011), teachers can encourage pupils to express themselves using effective questioning and active listening. Some of those techniques and strategies may include:

- Increased wait time
- No hands up
- Follow up on answers (correct and incorrect)
    - "Can you explain your answer?"
    - "How did you find that out?"
    - "Why do you think that?"
    - "What made you decide to do it that way?"
- Convert some closed questions to open questions.
- Ask questions that require pupils to evaluate their work/ or reason or generalise at their current level of mathematical development. These types of questions can challenge all pupils, but they have a particularly significant role in challenging our more mathematically able pupils.

- Explain why?
- Can you explain why this might not always or will always be the case?
- Try to find other possibilities?
- Explain to me why you think this is the most efficient solution/ method?

- Asking pupils to pose questions and/or design their problems for themselves and their peers.

(Wiliam, 2011.p 1-2)

These techniques are effective for questioning, which a skilled teacher will use to elicit evidence of learning from pupils as well as to assess pupils' progress towards the instructional goal (Michael & Susan Dell Foundation, 2016). Unfortunately, for some teachers, questioning in the classroom is often done too shallowly, narrowly, or ineffectively (Leahy et al., 2005; Pinchok et al., 2009). Some teachers' questioning only focuses on whether the pupils have provided a correct or wrong response (Wiliam, 2011). This observation might also apply to the Botswana context, where teachers tend to ask procedural questions.

However, when questioning is employed formatively, as Wiliam (2010) put it, it may result in focused questioning, which prompts thinking. Wiliam (2010) found that teachers who took part in their intervention study have become aware of the need to plan the questions they use in class carefully. According to Wiliam (2018), teachers who participated in the study considered spending more time planning for the instruction than grading pupils' work, a practice that emphasises the shift from quality control to quality assurance. By thinking more carefully about the questions they ask in class, teachers can check on pupils' understanding while the pupils are still in the class rather than after they have left, as is the case with grading.

Collaboration is another important technique under *engineering effective classroom learning* strategy. Ramsey and Duffy (2016) consider collaboration should be used as an instructional strategy to engineer effective classroom discussions to assess pupils thinking and generation of pupils' ownership of their thinking and work. A rich and purposeful classroom dialogue is an essential component of engaging pupils with their learning. The teacher can promote pupils' mathematical thinking and discussion using carefully planned and designed lessons (Wiliam, 2011). According to Wiliam, some of the useful strategies or activities to support purposeful classroom dialogue may include: Same and Different; Agree, Disagree, Depends; Always, Sometimes, Never; Card Sorts; Card Matching; and prediction.

An empirical study was carried out by Harrison (2005) in the United Kingdom which included 24 teachers in science and mathematics involved in an 18-month study designed to develop their FA practices. Teachers were trained on the professional development sessions and in groups developed an action plan of FA practices they sought to implement in their classrooms. Through lesson observations, teacher meetings and interviews, teachers' reflections, and student focus groups, researchers revealed that teachers' FA practices changed as a result of professional development. For example, teachers noted devoting more preparation time to planning questions and instructional activities that encouraged the visibility of student thinking, and to anticipating students' misconceptions so potential remedial instructional activities were readily available.

In another similar study, Panizzon and Pegg (2008) explored teachers' questioning practices. The authors conducted a two-year study with 25 secondary mathematics and science teachers in Australia, and their report focused on how teachers' assessment practices changed over time. Teachers participated in professional development workshops and collaborated to develop school action plans outlining formative assessment practices they sought to implement. Interviews and workshops revealed how substantial changes in teacher questioning practices, such as their perceived value of using open-ended, problem-solving questions, contributed to supporting teachers in understanding students' levels of understanding and identifying misconceptions. Making students' thinking visible facilitates teachers implementing activities that build new knowledge onto students' prior knowledge. The professional development workshops supported teachers with the use of a cognitive structural model-Structure of the Observed Learning Outcome (SOLO) taxonomy. This model provided teachers with a cognitive-developmental framework that helped them reflect on the quality of students' responses and assess their levels of understanding. Also, the framework helped teachers recognise the types of questions they routinely asked students; for example, several mathematics teachers noted that they often asked questions focusing on procedural understanding over conceptual understanding.

Wylie and Lyon (2015) specifically explored the extent and quality of teachers' formative assessment practices used to elicit evidence of learning. Of the five formative assessment strategies, Wylie and Lyon found that teachers most commonly implemented questions or tasks to elicit evidence of learning and these questions or tasks often focused on the collection of data from students. Further, even in the presence of relevant professional development, Wylie and Lyon found no significant difference in the process of "how teachers developed questions" (Wylie & Lyon, p. 150) but found a significant difference in "how they approached asking questions" (p. 151). For example, teachers more often selected students randomly to answer questions and used practices that allowed all students to respond, e.g., "student polling devices and individual student whiteboards" (Wylie & Lyon, 2015, p. 151).

The review of empirical studies illustrates that teachers use a strategy (*engineering effective classroom discussions, questions*, and *tasks that elicit evidence of learning*) frequently after being professionally exposed to training, particularly the questioning practices to assess pupils' levels of understanding.

2.6.3 Providing Feedback Process

Feedback is an integral part of FA processes for both teacher and pupils. Kollar and Fischer (2010) and Kyaruzi (2017) argue that the feedback process involves various activities such as feedback provision by a teacher (or peer), feedback reception by pupils and acceptance by pupils to apply such feedback to improve the quality of their work. Effective feedback should promote self-regulated learning and allow the pupils to interact with the feedback to confirm, add, overwrite, tune or restructure their previous knowledge (Hattie & Timperley, 2007; Jonsson, 2013; Nicol & Macfarlane-Dick, 2006, Kyaruzi, 2017).

FA practices within the classroom are usually carried out in two dimensions; assigning tasks to the pupils and providing feedback on the assigned tasks. Feedback, in this case, deems as a helpful strategy in describing information generated about the gap between current and desired performance (Wiliam, 2011). According to Rycroft-Smith,

(2017) feedback is widely regarded as a key element of formative assessment and therefore plays a crucial role in learning. And it involves three main functions which include changing achievement, changing interest and changing self-regulation (Harks, Rakoczy, Hattie, Besser & Klieme, 2014). Making sense of pupils' work and giving good feedback as to what pupils need to do next involves significant "fundamental" subject knowledge (Hodgen & Wiliam, 2006). This view for feedback substantiates Hattie and Timperley's (2007) consideration of the purpose of feedback to reduce discrepancies between current understanding or performance and the desired goal. For instance, when providing feedback to pupils' in mathematics, it is not helpful to tell them that they need to improve their work, even if this is true. It is, however, more helpful to point out what kind of errors they are making, and what they need to do to improve.

Building on the work of Hodgen and Wiliam (2006) and Hattie and Timperley (2007) about the purpose of feedback, teachers are responsible for reducing the discrepancies by changing the difficulty or the specificity of the goals or by providing more support to the pupils. Hattie and Timperley (2007) classified classroom feedback as either evaluative or descriptive. According to Hattie and Timperley (2007) evaluative feedback, involves the use of percentage marks, letter grades and even praise, when focused on characteristics of the pupils rather than on the characteristics of the work, can have the opposite of the intended effect. While descriptive feedback entails helping pupils to learn by providing information about their current achievement (Where am I now?) concerning a goal (Where am I going?) and identifying appropriate next steps (How can I close the gap?). Descriptive feedback provides specific information in the form of written comments or conversations, and help pupils to understand what he or she needs to do to improve (Hattie &Timperley, 2007; Hodgen & Wiliam, 2006; Wiliam, 2006; Rakoczy et al., 2017, Narciss, 2008).

Poulos and Mahony (2008) found that perceptions of feedback were related to the meaning pupils assigned to it, dependent on how the feedback was delivered, and the degree to which feedback was related to criteria, marks and grades. Specifically, Kyaruzi (2017) found that exposing teachers to training on how to monitor, scaffolding, and feedback delivery practices from pupils to teachers improved significantly from pre-test to post-test measures. Some previous studies on feedback training intervention improved those teacher' assessment practices and ensured that pupils could identify changes in their teacher' practices (Lizzio & Wilson 2008; Van de Pol, Volman, Oort & Beishuizen, 2014). Pupils' perceptions of teachers' feedback (or peers) play an essential role in their learning process (Harks et al., 2014; Kyaruzi, 2017; Poulos & Mahony, 2008; Strijbos, Narciss & Dunnebier, 2010; Rakoczy et al., 2013). In another similar study, Rakoczy et al. (2017), the results indicated that feedback was perceived as more useful in the FA condition, self-efficacy was more significant, and interest tended to increase; however, learning progress did not differ between the experimental and control groups. The assumed indirect effects were partly confirmed; FA showed an indirect effect on interest via its perceived usefulness.

Shah (2017) also specifically explored feedback practices of primary school teachers on students' homework in a public school in Pakistan. The study revealed that the teachers are used to assigning homework to their pupils and providing feedback on the assigned work but in a narrower scope. Teachers' feedback comprised mostly of a direct form of corrective feedback and also in an inconsistent way. However, the study also explored the challenges which constrained the teachers in carrying out effective feedback practices. These challenges and obstacles were: (a) inadequate professional training of teachers, (b) ineffective supervision of school administration, and (c) absence of clear institutional policy regarding homework and feedback. Recent findings by Rakoczy et al. (2019) indicate that feedback was perceived as more useful in the formative assessment strategy; self-efficacy was greater, and interest tended to increase, but learning progress did not differ between the experimental and control groups. The studies suggest detailed and individualised feedback on students written drafts help to increase learning as well as scaffolding assistance for students in lower grades.

Concerning the type of feedback given by teachers, the findings from Phase I of this thesis showed that the teachers' oral feedback provided evaluative rather than descriptive feedback. Concerning the written feedback, it was also observed that teachers were writing evaluative phrases in pupils' exercise books denoting brief descriptions such as "good" or "poor". It is acceptable to use both evaluative and descriptive feedback, but evaluative feedback is not of much help to the pupils because it is focused on the characteristics of the pupils rather than the characteristics of the work. Therefore, it can have the opposite of the intended effect (Hattie & Timperley, 2007). According to Hattie and Timperley (2007), pupils need descriptive feedback, which they can use to improve their learning and classwork. Thus, descriptive feedback helps pupils to learn by providing information about their current achievement (Where am I now?) with respect to a goal (Where am I going?) and identifying appropriate next steps (How can I close the gap?) (Hodgen & Wiliam, 2006).

Prior research studies have noted that when pupils receive a grade and a comment simultaneously, they would ignore the comment (Hattie &Timperley, 2007; Hodgen & Wiliam, 2006; Wiliam, 2004). The first thing they look at is the grade, and the second thing they look at is their neighbours' grade. Feedback needs to encourage thinking and engagement to be effective. Grades or comments like "good job" would not make pupils think. What encourages cognitive engagement is comments that address what the pupils need to improve (Hodgen and Wiliam, 2006; Wiliam, 2004).

The evaluative feedback findings for the current study agreed with Shah's (2017) study in a Pakistani public school. These findings showed that the teachers provided inconsistent feedback on the assigned work in a narrow scope, comprised mainly of the direct form of evaluative/corrective feedback. Interestingly, Shah's (2017) findings further elaborated on the reasons for ineffective feedback practices and found it mostly emanated from: (a) inadequate professional training of teachers, (b) ineffective supervision of school administration, and (c) absence of clear institutional policy regarding feedback. These challenges that may contribute to ineffective feedback could also be prominent in the Botswanan context, which may require more teacher professional development on the feedback process.

2.6.4 Pupils' Self and Peer Assessment

Pupils' self-and-peer assessment is another strategy for FA. Evidence for Policy and Practice Information and Co-ordinating Centre by Higgins (2007) described self-assessment as means of making pupils judge about their achievement and learning process and take part in decisions about action for further progress in learning. While peer assessment involves pupils in assessing others' work, again through reflection on the goals and what it means to achieve them. In both cases, the teachers have to possess the specific skills which enable them to execute pupils' self-and-peer assessment (Heritage, 2007).

Such skills involve helping pupils to set goals and criteria for success to reflect on their own and other's understanding, and evaluate learning according to the criteria. For example, Strijbos, Pat-El and Narciss (2010) found that pupils' perception of peer feedback adequacy (fairness, usefulness and acceptance) predicted their willingness to improve and effect learning changes. Feedback needs to be perceived well by pupils because pupils' positive perceptions of feedback were related to their learning outcome (De Kleijn, Mainhard, Meijer, Brekelmans, & Pilot, 2013).

Heritage (2007) reported strategies to involve pupils in self-assessment can be as simple as asking pupils to reflect on their performance through such questions as "Do you think that your response demonstrated understanding? If so why do you think this? If not, why do think you did not demonstrate understanding?" (Heritage, p. 144). These kinds of questions can help pupils to learn to be more independent and can recognise when they do not understand when they need to do something about it, and what they can do to improve (Heritage, 2007).

Teacher skills also include assisting pupils in learning to give constructive feedback to their peers that can provide for future growth. Heritage (2007) suggested that one can start with a simple statement like, "It wasn't clear to me when …" or "I didn't understand your point about …," (Heritage, 2007, p. 144). Pupils can process a detailed analysis of their peer performance against specific criteria. All these skills are important to teachers in modelling the classroom so that pupils see that they are collaborators with their teachers and peers in developing a shared understanding of their current learning status and what they need to do to move forward.

Studies have shown that processes of planning, monitoring reflecting on and modifying one's learning play an important role in the effectiveness of self-regulated learning (Kostons, Tamara & Fred, 2012; Pieger, Mengelkamp, & Bannert, 2016). The pupils' role in the FA practice may also be used to support each other's learning through peer-assisted learning, involving peer-assessment and subsequent peer feedback through explanations and suggestions to peers on how they can act to reach their learning goals (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010).

In exploring the practice, Harris and Brown (2013) investigated peer- and self-assessment (PASA) in which student-led assessment practices with the potential to positively affect achievement. Three case study in New Zealand classrooms explored the teacher and students' perspectives of and purposes for PASA based on four themes (i.e., improvement, accountability, social interaction, and accuracy). The result of the study indicated that teachers and students needed a deeper understanding of how to use PASA for improvement and self-regulation purposes. This suggests teachers must provide concrete instruction in PASA and carefully manage interpersonal issues for successful implementation. Teachers have cited mainly improvement purposes for PASA, with students primarily focused on accuracy and social interaction concerns.

Similar studies explored the nature and frequency of peer - and self-assessment use in schools. Thus, Noonan and Duncan (2005) surveyed 118 Canadian high school teachers and found that fewer mathematics or science teachers used peer-and self-assessments compared to teachers in other subjects. Another related study was carried out by Suurtamm (2004), who found secondary mathematics teachers in Ontario (*N* = 5) used peer- and self-assessments, but noted that the use of peer- and self-assessments were not common across teachers in their schools and that their colleagues did not understand the purpose of these strategies.

Similarly, Wylie and Lyon (2015) found that teachers implemented peer- and self-assessments with low frequency, as well as with a low to moderate quality of implementation. Specifically, teachers displayed a "lack of planning or structure" (Wylie & Lyon, 2015, p. 155) when using peer-assessments and sometimes underestimated the amount of support to those students. These empirical reviews revealed that none of the studies was carried out in the context of African states either in primary or secondary schools. Also, studies have illustrated that even in developed countries like Canada and New Zealand, peer- and self-assessment are not commonly used across classrooms, and when implemented critical features to support pupils with these strategies are absent. In summary, pupils' perception of feedback appears to be a key determinant for feedback usefulness or application (Gibbs & Simpson, 2004; King, Schrodt, & Weisel, 2009).

Despite these findings, peer- and self-assessment remain the key concept for FA, which requires teachers to possess specific skills to execute these activities effectively and efficiently. Such skills help pupils set goals and criteria for success to reflect on their own and others' understanding and evaluate learning according to the criteria. The observed teachers' challenges in implementing peer- and self-assessment deflate the feedback drive requiring a more profound understanding from the pupils of the given phenomenon. The current study supported Harris and Brown's (2014) findings that pupil-led assessment practices can improve achievement. Their findings had also indicated that teachers and pupils needed a deeper understanding of how to use teachers peer- and self-assessment (PASA) for improvement and self-regulation purposes. In this context, the participating teachers may require training to assist them in understanding PASA.

Based on reviewed FA strategies, there were limited studies in which the effect of a sole strategy was considered. Many literature reviews emphasise that the implementation of the FA strategies must be concurrently done through embedding it and adjusting instruction to use the five strategies.

2.6.5 Formative Assessment for HOTS in mathematics

The FA practices underscore the central role of evidence in effective teaching and learning; similarly, policy-makers and practitioners need evidence to build effective formative practices. Fundamentally, FA involves using assessment to "form" subsequent instruction (Black & Wiliam, 2003; Herman, Osmundson, Dai, Ringstaff & Timms, 2015).

More explicitly, instruction generally carries the connotation of teacher responsibility. The broader term "teaching and learning" more clearly acknowledges both teacher and pupils' responsibility and role in this process (McManus, 2008). For this reason, FA is described as a process used by teachers and pupils during instruction that provides feedback to adjust ongoing teaching and learning to improve pupils' achievement of intended instructional outcomes (McManus, 2008). The effective use of the FA processes requires pupils and teachers to integrate the following strategies as much as possible:

- clarifying learning goals within a broader progression of learning;
- eliciting and analysing evidence of pupil thinking;
- engaging in self-assessment and peer feedback;
- providing actionable feedback; and
- using evidence and feedback to move learning forward by adjusting learning strategies, goals or next instructional steps (McManus, 2008).

FA, to a large extent, is seen as teachers' content and pedagogical knowledge in action. The teachers' knowledge of the subject (in this case mathematics) is needed in specific classroom contexts for instructional planning. The teachers' knowledge is also imperative to know when and how to elicit pupils' understanding, analyse and interpret pupils work relative to the likely misconceptions and/or obstacles and; how to take immediate steps to close the gap between where pupils are and where they need to be (Black & Wiliam, 2009; Hattie & Timperley, 2007; Heritage, 2010; Herman et al., 2015). This translates that FA focuses on classroom practices and interactions, which can produce quantitative or qualitative information.

The notion behind this FA is to improve pupils' understanding, which is dependent on the practice knowledge of the teacher is essential to know what the next step should be for the pupils (Baird, 2010; Baird, et al., 2017). The FA should take place continuously throughout the learning period and requires pupils to collaborate with their peer and teachers (Mohamed & Lebar, 2017).

In support of formative assessment, Wiliam (2011) embarked on a meta-analysis of the 681 publications and concluded: "attention to formative assessment can lead to significant learning gains". The following will pay particular attention to empirical studies related to formative assessment intervention and mathematics achievement, as some background building up towards the current study.

Some researchers have made various efforts to improve pupils' HOTS in some countries. Foong (2000) researched open-ended problems for higher-order thinking in mathematics in Singapore. In Georgia in the USA, Murray (2011) examined the implementation of higher-order thinking in middle school mathematics classrooms. Ghasempour, Kashefi, Bakar, and Miri (2012) conducted a similar study in Malaysia on higher-order thinking via mathematical problem posing tasks among engineering students, while Tajudin (2015) studied mathematical knowledge and higher-order thinking skills for teaching algebraic problem-solving in Turkey. In the African context, the Ugandan study by Ssempala et al. (2018) revealed an overall mean value of 86.8% of LOTS questions and a mean value of 13.2% of HOTS questions. It recommended that more HOTS questions be included in the primary school leaving examination for that specific context.

Some other studies investigated the relationship between HOTS and the academic performance of pupils. Yee et al., (2011) conclude that there is a very low positive relationship between the level of understanding HOTS and gender, academic achievement and socioeconomic status. On the other hand, Ramos, Dolipas, and Vilamor (2013) examined the relationship between HOTS and academic performance in physics for college students. They concluded that the HOTS level on analysis, comparison, and evaluation significantly influenced male students' physics performance. On the other hand, the HOTS level on analysis, inference, and evaluation significantly influenced female students' physics performance. Yoshida's (2015) study with students of mathematics education from University of Papua, Manokwari West Papua Province, Indonesia, concluded that task-specific coaching rubrics enhance pupils' knowledge and understanding of curriculum development for higher-order thinking.

Studies on FA intervention were also examined for this study's literature review. These included a rethinking primary school mathematics teaching study that was conducted in Kenya (Kiplagat's, 2016). Through experimental exposure of the sampled participants of 140 pupils of grade 6, explored the effects of formative assessment classroom teaching strategy. The findings for Kiplagat's study were significant ($f = 131.14$, $p < 0.05$) and concluded that formative assessment as a classroom strategy improved achievement in primary school. Another similar study from Balan and Metcalfe (2012) sampled 47 participants and employed a mixed-method quasi-experimental intervention study on assessment for learning (formative assessment) in mathematics education. The finding revealed improvement in problem-solving for students in the intervention group. These positive findings formed the basis of the current study, however, the two studies used small sample sizes, and a questionnaire alone could not be used to gain deeper insight.

Moyosore (2015) conducted an experimental study on the effect of formative assessment on the students' mathematics achievement in Nigerian secondary schools. The findings of the 120 sampled students showed strong significance in the means of the difference of mathematics scores for exposed students in formative assessment. The most recent study by Chemeli (2019b), employed a pre- and post-test, control quasi-experimental method intervention and investigated the impact of the five key formative assessment strategies on learner's achievement in mathematics instruction in secondary schools. The sample for the study was 534 students and 33 teachers from Nandi County, Kanye. Chameli's (2019a) findings revealed a positive impact on the learner' achievement. The study provides reasons for the positive impact which emanated from formative assessment strategies including:

- they eased teachers' workload;
- they raised learners' attitudes and interest;
- they improved learners' critical thinking; and
- teachers and learners enjoyed using formative assessment strategies.

Most importantly, Chemeli (2019b) pointed out that the use of formative assessment strategies improved learners' acquisition of problem-solving skills. Although this study explored using data from mixed methods research, the shortfall of Chemeli study included:

- Limited evidence of how the pupils measured in the acquisition of problem-solving (HOTS). Thus, the study relied on the classroom observation and did not complement with measuring items which were specifically for HOTS with valid and reliable evidence of measurement instrument to gain a deeper insight into the skills; and

- Statistical analysis without, rack and stacking analysis render challenges to determine the level of significant learning gains. Chemeli's study did not provide enough information to calculate an effect size, which describes the magnitude of the effect of the intervention.

The previous studies have pointed out some gaps which the current study aimed to address. The question arose whether FA can enhance the teaching of HOTS and pupils' academic performance in mathematics instruction. This study attempted to explore and investigate Botswana teachers' understanding of the alignment of FA and teaching of HOTS in mathematics. Alignment in this context refers to the "degree of correspondence" between teachers' educational objectives, instruction methods, and forms of assessment (Anderson & Krathwohl, 2001). These domains of FA and HOTS are complementary to each other. The alignment between learning objectives, methods of instruction, and assessment towards higher levels of cognition is essential in creating a culture of thinking in teacher preparation (Anderson & Krathwohl, 2001). The current study also focused on the integration of FA strategies and teaching HOTS in the mathematics classroom.

Evidence suggests that more often teachers do not model the culture of thinking towards FA and teaching HOTS in the mathematics classroom. A recent study by Abdullah et al. (2017), on mathematics teachers' level of knowledge and practice on the implementation of HOTS, revealed that the level of knowledge and practice of the assessment facet was weak. Another related study was also carried out by Nenty et al., (2007) to determine the extent to which primary school teachers in Botswana and Nigeria perceived the six levels of Bloom's cognitive behaviour as being different in the extent to which they enhanced quality in basic education and the level to which their current classroom assessment practices involved items that measure each of these levels of cognitive behaviour.

The main finding revealed a significant discrepancy between the level to which teachers perceived each of Bloom's level of cognitive behaviour to enhance the quality of education and the level to which their classroom assessment practices can provide for the development of such behaviour among pupils (Nenty et al., 2007).

It is worthwhile to note from the studies (Abdullah et al., 2017; Nenty et al., 2007), that they confirm the observations which Heritage et al. (2009) identified, namely that teachers have more difficulty in determining the next instructional steps aligned to identifying mathematics principles and analysing pupils’ understandings. Despite this misgiving, the alignment of FA and teaching HOTS in mathematics are inseparably linked so that the development of one should influence the other. The integration of the two domains also means that similar strategies may be employed to promote the quality of thinking and learning. This was outlined by the Welsh Assembly Government (WAG, 2010) with a Venn diagram which summarises the characteristics of HOTS and FA in the classroom and how they overlap (See Figure 2.4). The characteristics of lessons that develop thinking have a considerable overlap with the FA practices in the formal classroom settings.

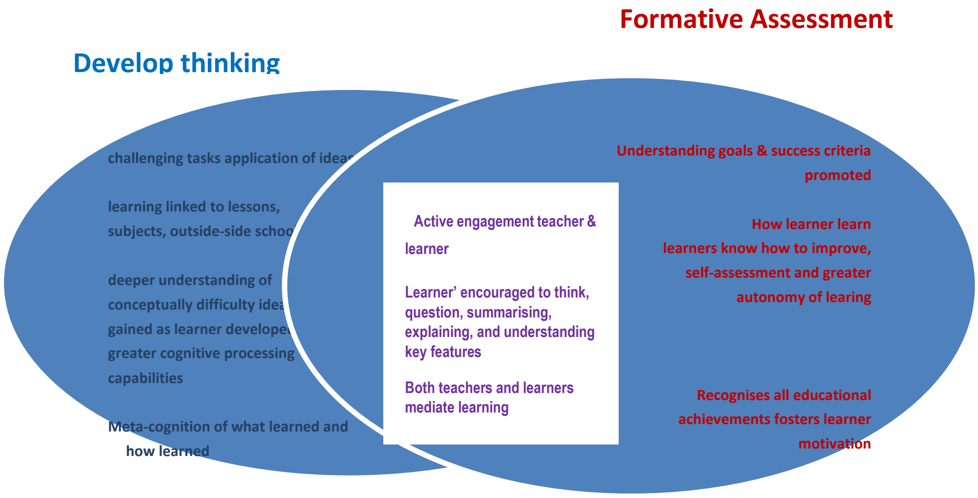


Figure 2.4: Developing Thinking and Formative Assessment in the classroom
Welsh Assembly Government (2010, p.10)

Figure 2.4 attempts to explain that the integration of the two domains can possibly promote the quality of HOTS thinking and learning. These common characteristics of FA and HOTS are:

- Pupils are actively engaged in lessons from the very start.
- Teachers and pupils explore and take account of:

- what pupils already know (subject knowledge and thinking strategies)
- what pupils can do
- what strategies may be useful to tackle the problem
- pupils' misconceptions.

- Pupils are encouraged to think, question and talk.
- Teachers and pupils need to actively listen, ask questions, summarise and explain their understanding.
- Group talk and collaboration are encouraged. Through articulation, using appropriate vocabulary, learners clarify their learning. Focused talk in lessons allows pupils to evaluate their understanding and add to that of their peers.
- Teachers and pupils play a key role in mediating learning experiences, through active listening, asking appropriate questions, summarising and explaining understanding.
- The environment is sensitive, constructive, and reflective so that pupils feel safe to make mistakes.

These common characteristics reveal how FA and teaching HOTS are related, which can possibly be implemented in teaching mathematics concomitantly. These common characteristics suggest that the features of FA converge with the features of HOTS. Due to this effect, demands are placed upon the teachers in terms of required pedagogical content knowledge to implement both HOTS and FA simultaneously in a mathematics lesson. The link suggests that the teacher who properly implements the FA strategies can also effectively influence pupils' acquisition of HOTS thinking and learning. Hence the current study sought to determine the extent to which teachers implement the FA in the mathematics classroom to enhance HOTS.

Teachers' roles in education include skills development of the pupils' Higher-Order Thinking Skills (HOTS) based on the teaching methods and approaches essential for the transformation to enhance pupils' thinking skills accordingly (Serin, 2013). HOTS are essential and pertinent to educate the pupils of the 21st century to handle complex real-life problems and provision of solutions (Rajendran, 2008). Scholars in education have suggested if teachers implement some teaching approaches which address HOTS acquisition, it would be very helpful for pupils' academic growth and working independently with minimum guidance from their teachers (Fahim & Masouleh, 2012; Goethals, 2013; Yang & Gamble, 2013).

Fundamentally Formative Assessment (FA) has to empower the teacher's pedagogical knowledge and improve pupils' understanding and learning. These actions are essential to know the next steps to improve teaching and pupils' learning (Baird, 2010; Baird, Andrich, Hopfenbeck & Stobart, 2017). FA literature affords extensive evidence that if FA is well implemented and well perceived by pupils, it has the potential to improve pupils' learning (Black & Wiliam, 1998, 2009; James & Pedder, 2006; Kyaruzi, 2017; Wiliam, 2011; Wiliam et al., 2004).

The teaching of HOTS and assessment are informed by taxonomy as the foundation for teaching and learning. Taxonomy such as Bloom's taxonomy, the original (Bloom et al., 1956) and the revised version (Anderson & Krathwohl, 2001), and the structure of observed learning outcomes (SOLO) taxonomy (Hattie & Brown, 2004, 2012), are all important in guiding and teaching mathematics in primary school, despite the criticism of the utility of the original and revised version of Bloom's taxonomy (Hattie & Brown, 2004, 2012). Specifically, Bloom's taxonomy is a classification system used to define and distinguish different levels of human cognition, namely thinking, learning, and understanding (Forehand, 2010).

So far, teaching evidence is assessed mainly with emphasis on two predominant purposes of assessment (i.e., formative improvement vs summative evaluation) have created substantial tension among teachers (Bonner, 2016). Primarily, teachers' thinking about assessment orientation and beliefs reflects the social, historical, and cultural priorities established in each jurisdiction in which teachers are employed (Brown & Remesal, 2017; Fulmer et al., 2015).

2.7 Conceptual Framework

This study claims that FA strategies intervention can contribute to pupils' HOTS achievement in mathematics. The claim for this study was engrained from the leaders in the field of FA (Black and Wiliam, 2003) who proposed that improving pupils' understanding of how they can learn better to improve their educational outcomes is the most important purpose of assessment.

The conceptual framework for this study aimed to guide the whole process of the data collection, analysis and interpretation of the results. The conceptual framework of the current study was closely linked and embedded into FA's potential to provide frequent information about learning and teaching which can be used to improve teaching effectiveness and pupils' learning outcomes for HOTS simultaneously. For the purpose of assessment, this study was inclined to the Kivunja (2015) and Heritage (2010) conceptual framework, which referred to as the power of the Assessment Feedback loop (AFL) model. As shown in Figure 2.5, framed as a cycle, illustrating that FA is a continuous process, integrated into instruction to close the gap.

The idea of closing the "gap" comes from D. Royce Sadler in 1989, who stressed feedback as the centrepiece of FA (Heritage, 2010). This implies that the feedback generated from FA must be used to make changes in the pupils learning status and help them close the gap between their current status and the intended learning goals (Heritage, 2010: p 10; Kivunja, 2015). It is the interest of the current study to use the AFL model, as a way of improving teaching and learning HOTS in mathematics.

2.7.1 Kivunja's conceptual framework of the assessment feedback loop

According to Kivunja (2015), the power of AFL was adapted from the Kivunja (2015a, p. 239) and Heritage (2010, p11) and it comprises of nine processes which are; (1) Defining learning goals, learning progression and criteria for success (2) Elicit evidence of learning (3) Interpret the evidence (4) Identify the gaps in knowledge and skills (5) Gather and provide feedback (6) Respond to learning needs (7) Scaffold learning in the zone of proximal development (8) Close the gap and (9) Define learning goals in a new iteration of these processes.

Kivunja (2015) devised the presentation version as a series of clockwise arrows which starts at process 1 "Defining learning goals, progression and criteria", and progresses through the other seven processes illustrated in Figure 2.5, to finish at "Closing the gap" in number 8, which allows the formative nature of the process to start all over again as a new cycle at number 9. Kivunja described every process in the AFL.

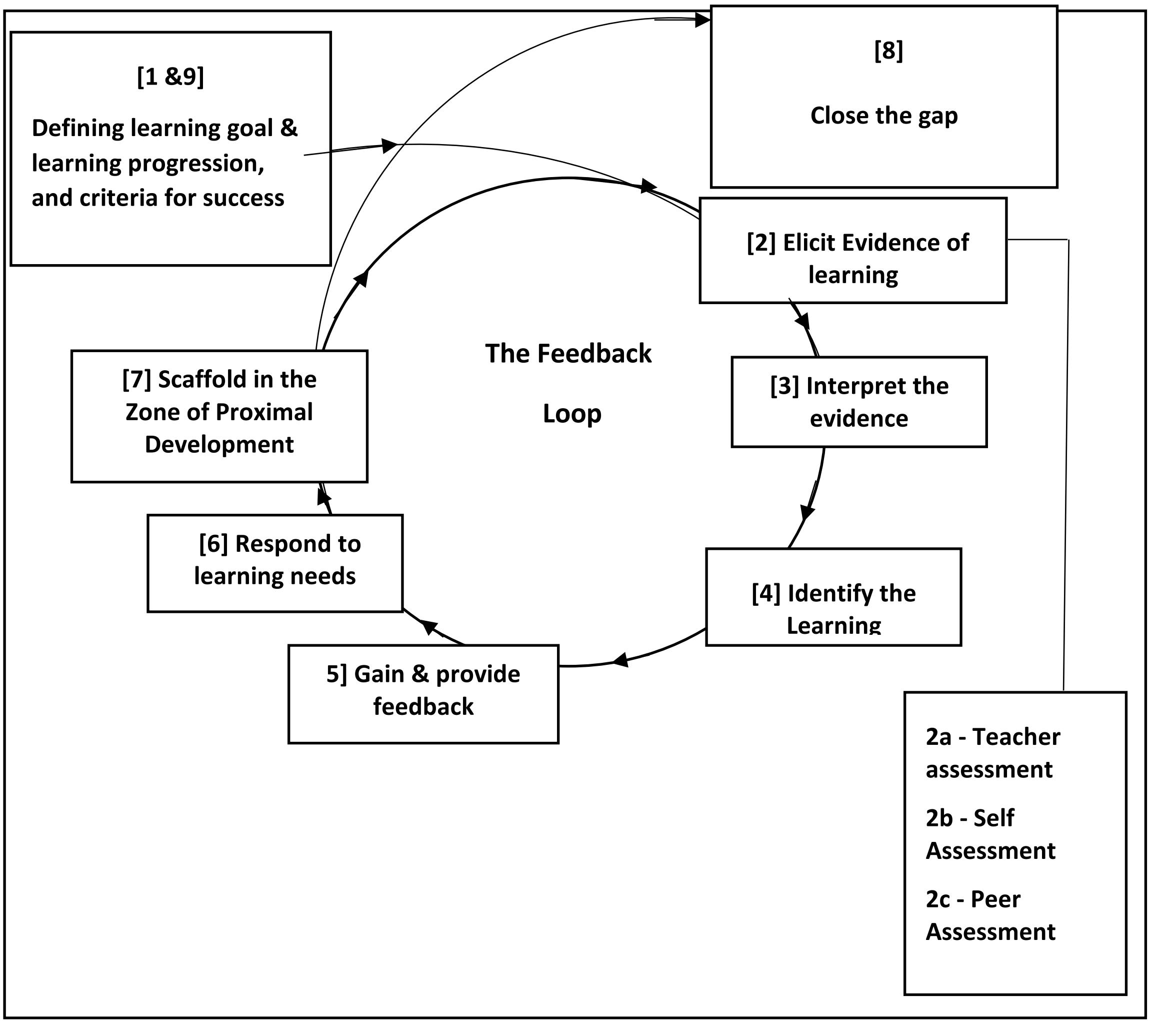


Figure 2.5: The assessment feedback loop
Source (Kivunja, 2015a, p. 239) and Heritage (2010, p. 11)

Determining learning goals and defining criteria for success

The AFL begins with (1) defining learning goals, learning progression and criteria for success. Process 1 is linked to process 8 to determine the gap closure, and selection of the new learning goal to another gap is created to revisit the previous learning goal if the pupils have not attained it. That as a teacher, it is important to articulate the learning goals for the lesson and an initial stage as derived from the pupils' learning progression on the current topic or sub-topic. It is upon the teacher also to determine the learning outcomes, which the pupils need to meet to demonstrate knowledge and achievement of the defined learning goals and make them very clear to the pupils.

Before the lesson(s) begin, the LG and SC are shared with the pupils. Success criteria are the guide to learning while the pupil is engaged in the learning tasks (Heritage, 2010).

Elicit evidence of learning

In process 2, Kivunja (2015) supports Wiliam and Thompson (2014). Thus, teachers are required to elicit evidence of learning, as the pupils engage with the learning activities that include the use of a variety of strategies to gather evidence on how pupils are learning and moving towards the defined goal. Elicit evidence learning should involve a wide range of assessment tasks so that every pupil has an opportunity to demonstrate what she/he have learnt. In other words, the enlarged perspective promotes the integration of FA with each instructional activity and therefore allows for more diversity in how learning is assessed (Heritage, 2010). For example, these diverse approaches could include questioning, observation, representations, performance, demonstration, exit cards, notes from the pupils to a teacher, worksheets completed, or more structured, curriculum-embedded assessments.

Interpreting the evidence

Process 3 of the model is about interpreting the evidence. According to Kivunja (2015) and Heritage (2010), Process 3 deals with examining the assessment data in light of the SC defined in Process 1 to determine how pupil's learning is occurring. At this level teachers are expected to know, for instance, what knowledge the pupils have, what they understand, what their misconceptions are, and what skills are or are not being acquired. This previous process leads to Process 4 which identify the gaps in knowledge and skills, the gap is defined as "the difference between your pupils' current understanding and intended learning goals" (Kivunja, 2015a: p. 240) as defined in process 1. When pupils are involved in peer assessment, they also use SC to interpret the evidence and provide each other with feedback about how learning can be improved. To do this, pupils must understand what success criteria mean (Heritage, 2010).

Identifying the Gap

This gap process is so crucial to maximising the effectiveness of the assessment feedback loop because it is only when the teacher has an accurate understanding of the gap that can lead to taking steps to bridge it (Kivunja, 2015). Bridging this gap is the primary purpose of FA. As explained by Sadler (1989, p. 121), assessment data is useful "only when the feedback is used to alter the gap" (Sadler, 1989, p. 121) between what the pupils know and what they need to know according to the criteria defined in Process 1. Interpreting the evidence from FA is critical to identify the gap between pupils' current learning status and the goal of the instruction. Closing the gap is achieved by responding to the evidence through feedback, which results in adaptations to instruction and learning (Heritage, 2010).

Feedback

According to Kivunja (2015) to gather and provide feedback (Process 5) is the central focus in which the assessment feedback loop is maximised only when teachers use feedback to help pupils take steps to improve learning towards the learning goals. Hattie and Timperley (2007) suggested that whatever the source of the feedback, it must answer three questions asked by a teacher and or by a pupil:

- Where am I going? (What are the goals?)
- How am I going? (What progress is being made toward the goal?)
- Where to next? (What activities need to be undertaken to make better progress?) (Hattie & Timperley, 2007, p.86)

Both the teacher and pupils are to provide feedback about teaching and learning. As noted earlier, peer and self-assessment can provide feedback that helps their classmates and feedback about their learning to improve learning, respectively. This is important because when pupils are monitoring their learning, they are engaged in metacognition (Heritage, 2010).

Adapting and Responding to Learning Needs

In response to learning needs in process 6, teachers use the feedback from Process 5 to plan for the instructional strategies that will enable one to put appropriate, additional demands on their pupils in place; to help them take responsibility for their learning as they pursue the fulfilment of their learning outcomes articulated in Process

1 (see Figure 2.5). Process 6 is focused on closing the gap between where the pupils currently are in learning, and where they need to go (Heritage, 2010; Kivunja, 2015).

Scaffolding New Learning

The term "scaffolding" characterises the support that teachers (or peers) give to pupils to move them from what they already know to what they can do next and close the gap between their learning status and learning goal (Heritage, 2010). In Process 7, scaffold learning is sometimes known as the zone of proximal development as identified by Vygotsky (1978). The teacher provides the necessary instructional support so that learning is incrementally internalised, ultimately becoming part of the pupils' independent achievement and by so doing closing the gap. According to Kivunja (2015), working along the scaffold, and using the feedback and pupils' learning strategies, can incrementally internalise learning as they gain a deeper understanding and learn new skills needed to close the gap (see Figure 2.5).

Closing the Gap

The final step in the process of FA is to close the gap (Process 8), for the realisation of that, it forms the beginning of the next set of learning goals that you define for your class (see Figure 2.5). Thus, starts a new FA cycle shown in Process 9 in Figure 2.5 (Kivunja, 2015). As the gap closes, the teacher selects new learning goals and another gap is created, renewing the FA cycle (Heritage, 2010).

Classroom Culture

According to Heritage (2010) and Kivunja (2015), the whole process of FA depends on a classroom culture where pupils feel safe to say they do not understand something and give and receive constructive feedback from peers. It is the responsibility of the teacher to establish a classroom culture characterised by cooperative learning, recognition, and appreciation of individual differences. The classroom should be a conducive learning environment in which pupils listen respectfully to each other, respond positively and constructively, and appreciate the different skills levels among peers; all pupils will feel safe to learn with and from each other (Heritage, 2010).

2.7.2 The conceptual framework for this study

Kivunja (2015) and Heritage (2010) confirmed that the AFL model is an effective pedagogy of teaching, learning and assessment which needs step by step consideration of nine (9) essential processes in which FA is infused to provide information on pupils' progress to pupils and the teacher, to improve teaching and learning. Similarly, in this study, the AFL approach and WAG (2010) were used. The common characteristics of FA and HOTS, as identified by WAG (2010), are linked to and can be implemented within the AFL framework. The AFL model was adapted for the study to address the research questions with a view of expressing the impact of FA to enhance pupils HOTS in mathematics. This AFL model is useful as it explains the link between the research questions in all the phases and it visually assists the researcher and the reader in understanding the possible relationships among various issues that are essential in ensuring the integration of FA into the instruction for teaching of HOTS to collect evidence about pupils and how pupils are progressing towards learning goals as well as the use of assessment results to identify gaps in learning (see Figure 2.5). The main question that serves as a basis for this study is centred on the integration of FA and the teaching of HOTS. Hence, this study claims that assisting teachers following the view of Kivunja (2015) and Heritage (2010), that effective FA occurs:

- when teachers adjust teaching and learning of HOTS in response to FA information;
- when pupils receive feedback about their learning, with advice on what they can do to improve;
- and when pupils are involved in the process through peer and self-assessment, can result in improved pupils HOTS and achievement in mathematics.

2.8 A Botswanan Assessment Feedback Loop to Enhance Hots in Mathematics

Figure 2.6 shows the adaptations of Kivunja's Model to serve as an assessment feedback loop to enhance teaching HOTS in the mathematics classroom using variables from the three phases of the current study.

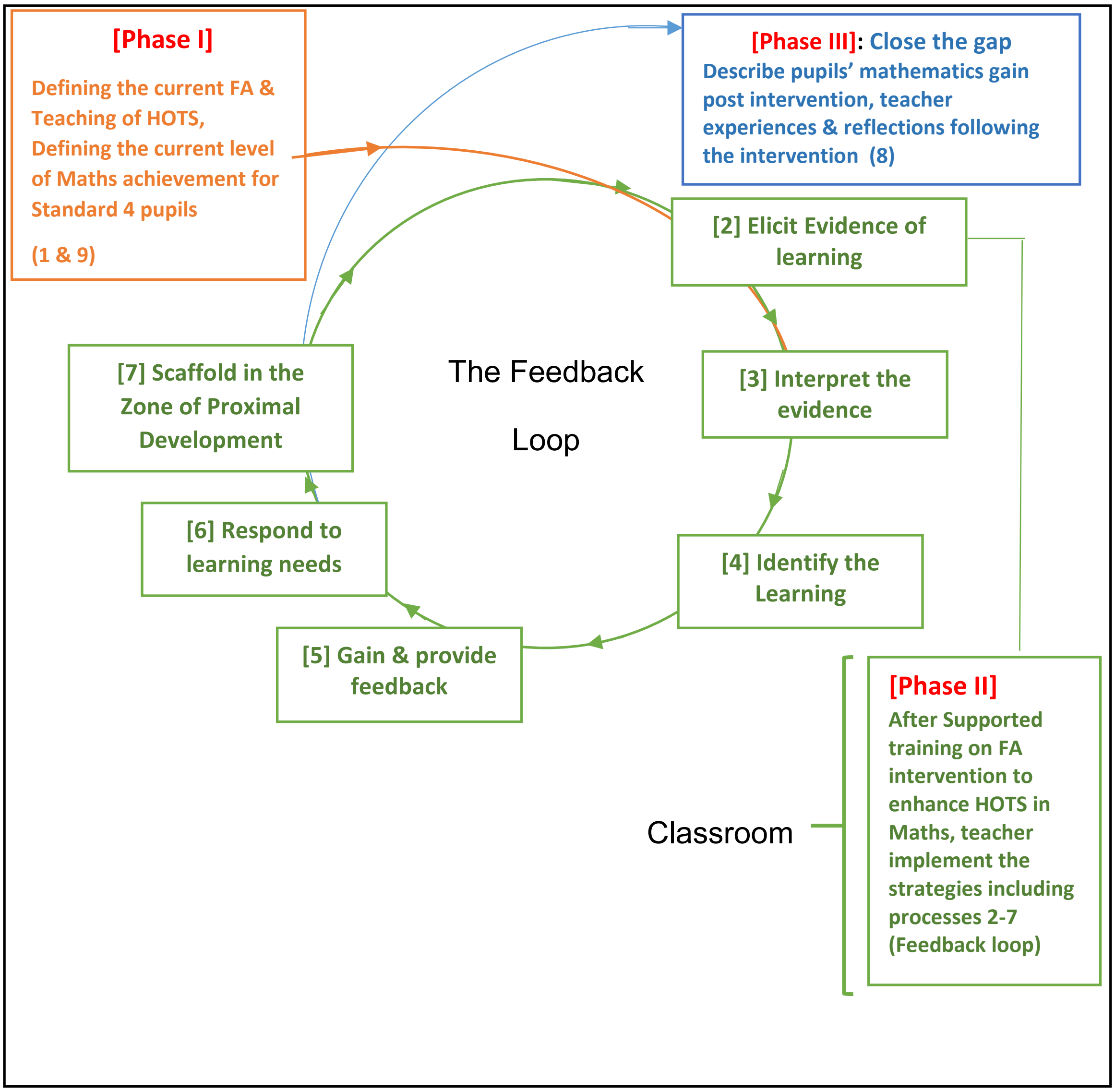


Figure 2.6: Assessment feedback loop to enhance HOTS in mathematics
As adapted from Kivunja's Assessment Feedback Loop (Kivunja, 2015c)

This study will employ the exploratory approach, as shown in Figure 2.6 to establish the impact of FA to enhance HOTS in Mathematics and data collection. This exploratory research is defined as research used to investigate a problem which is not clearly defined. In the case of the current study, it is conducted to have a better understanding of the existing problem of poor achievement of HOTS in mathematics (Phase I), and then embed the different FA variables as an intervention (Phase II) and determine the extent to which the gap would be closed (Phase III). For this research, a researcher starts with a general idea and uses this research as a medium to identify issues, which can be the focus of the intervention to close the gap.

A research study carried out in the current context was found appropriate to be guided by the use of a survey and pre-experiment. The reader is therefore asked to be aware that, for the purposes of this current study, a sequential embedded design with a single group was employed in three phases within the Kivunja's Feed Loop. All the phases are informed and taken from the purpose of this study to determine the data analysis of the related identified variables in the adapted assessment feedback loop framework (see Figure 2.6).

3 CHAPTER THREE: RESEARCH DESIGN AND METHODOLOGY

This chapter focuses on the research design and methodology of the overall study that was used to understand and explore the possible effects of FA intervention for Botswana Standard 4 mathematics teachers to use in the classroom to enhance Standard 4 pupils' HOTS academic achievement. The overall design for the study took place in several phases: firstly, conducting a survey (Phase I), followed by developing an appropriate intervention (Phase II) and then the pupils' gain, interventional experience and reflection of participants (Phase III). For this study, all three phases were derived from the main research question of the study;

- To what extent does the formative assessment intervention for teaching and learning enhanced Standard 4 pupils' HOTS in mathematics?

The research design and methods identified from the three phases are discussed broadly in this chapter. Section 3.1 comprises Section 3.1.1, which deals with the epistemological issues, and Section 3.1.2, which covers the theoretical perspectives of the study. Section 3.2 outlines the general approach of the study. Specifically, Section 3.2.1 outlines the phases of this study, while Section 3.2.2 discusses the overview of the study's research design. Section 3.4 covers ethical issues and considerations.

3.1 Epistemology, Ontology and Theoretical Perspective of the Study

3.1.1 Epistemology and ontology

The study is concerned with the teachers' possible exposure to FA intervention to address pupils' low achievement of HOTS in mathematics. It is within the effort of the researcher to explore and select an epistemology and ontology, which is appropriate for the current study. The epistemological assumption intends to provide a philosophical background for deciding "what kinds of knowledge is legitimate and adequate" (Gray, 2014, p. p19). Epistemology assists in clarifying the issues of design, the kind of evidence that is being gathered, where, and how it is going to be interpreted as well as help the researcher to recognise which designs will work (for a given set of objectives) and which will not (Gray, 2014). While ontology is all about the nature of reality (all that is or exists), and the different entities and categories within reality existence. These two perspectives are however considered as unitary and holistic (Gray, 2014; Denzin, 2012; Creswell,2014; Maree, 2008).

According to Gray (2014), there are three positions of epistemology which include objectivism (Positivism/or Postpositivism), subjectivism and constructivism. For this study, a postpositivist approach was deemed appropriate as it represents the challenge towards "absolute truth". Postpositivism tends to be deterministic and reductionist in its philosophy to reduce concepts and ideas to small and discrete sets (Ong, 2020; Popper, 1966). From the epistemological perspective, objectivity, impartiality and distance are maintained between the researcher and what is being researched (Creswell, 2014; Willis, 2012). In general, postpositivism postulates that research is defined by examining relationships between variables and associated meanings.

According to Romm (2020), this paradigmatic position is the least likely to admit that social research itself is a future-forming enterprise because the quest is, Popperian style, to concentrate on trying to get "nearer to the truth" (Popper, 1966, p. 377), and it is possible through the mixed methods community (Hunter & Brewer, 2015a, 2015b). This research approach is based on empirical observation, measurement of variables, and using surveys as a method that has its basis in postpositivism. It considers the numeric measures of observations through frequency counts, statistical analysis and other measures or statements made about meanings and views accorded to research participants, all arising in the research contexts wherein observations become imperative for the postpositivist (Phillips, 2000; Willis, 2012).

Like positivism, postpositivism depends on observation and measurement to develop strong causal understandings of the world (Willis, 2012; Young & Ryan, 2020). Postpositivism retains the assumption that there is an objective truth but concedes that it is unlikely ever to find such truth; instead, as researchers, we build our understanding of the world within the limitations of our times, techniques, and currently available knowledge (Hunter & Brewer, 2015a; Young & Ryan, 2020). This stance recognises that scientists (as humans) are fallible and subject to many influences, and bias (while undesirable) is inevitable. Observation and measurement are considered imperfect. Understanding will never be complete (Young & Ryan, 2020).

Young and Ryan (2020) postulate that for postpositivists, scientific investigation is a slow and iterative approach which places the researcher in the position of a lifelong learner–one that values problem-setting in addition to problem-solving. The postpositivist researcher progressively works toward a more complete understanding of our external and multiple realities, including shaping better questions to address the problem at hand. With this perspective in mind, the teachers' perceptions as participants, classroom observations and pupils' assessments in the current study were anticipated to investigate multiple realities concerning the phenomenon, in this case, application of FA strategies in the classroom to enhance teaching and learning (Young & Ryan, 2020).

Given the epistemological underpinning of postpositivism, which was adopted for this study, the next section outlines the theoretical perspective of the current study as an attempt to contribute a solution to the identified problem.

3.1.2 The theoretical perspective of the study

The theoretical foundation of this study was rooted in the dialectic stance, commonly known as pragmatism. The pragmatism paradigm posture allows a researcher to use more than one paradigmatic tradition in the research project. The paradigm also assumes to use multiple paradigms which contribute to a better understanding of the problem under study (Venkatesh, Brown & Sullivan, 2016). Pragmatism may not reflect reality; it assumes that reality is acknowledged but not presumed to be known directly (Mertens, 2010; Morgan, 2007; Teddlie & Tashakkori, 200). It is for this purpose that pragmatists hold absolute knowledge as a worthy but probably unreachable goal, and accentuate theories of the meaning of what works were adopted.

Fundamentally underpinning Peirce's notion of truth is the exclusive pragmatic idea that if one knows all the conceivable effects of an object one has grasped its conceptual meaning, and if this occurs, (Murgolis, 2006) then “the only effect which real things have is to cause belief” (Peirce, 1878, p. 53). Thus, the truth of a truth bearer flows necessarily from its practical effects on human experience; a proposition is true if it engenders a justified and useful belief that works out in such a way as one would expect it to work in all of its conceivable ways (Biesta & Burbules, 2003, Longo, 2013; Murphy, 2003). This assumption sounds important for the teaching and learning philosophy as far as the problem-solving task is concerned.

In addition, Saunders and Thornhill, (2012), advance the pragmatists’ school of thought, which deals with the tool for prediction, action, and problem-solving and not with describing, representing, or mirroring reality. Considering this conjunction of pragmatism, the current study employed a mixed-methods approach (as outlined in Section 3.2) to improve low pupil achievement in mathematics by applying a different approach to the real-world problem (Feilzer, 2010), rather than making assumptions about the nature of knowledge. Pragmatic researchers link their approach to the purpose and nature of the research questions (Creswell, 2003) and often follow a “what-works” strategy (Armitage, 2007; Ornstein & Levine, 2008).

Saunders and Thornhill (2012) continue to recognise the peculiar way in which pragmatists interpret the world and commission research, assuming that no single point of view can ever explain the entire picture. For this reason, it was found suitable for the current study, which intended to involve more than one approach in an attempt to address low pupil achievement of HOTS in mathematics and hence pragmatism provided theoretical justification for mixing methods.

This study adopted the classroom learning environment for the pragmatic approach, which dictates giving pupils a great deal of freedom of choice in seeking out the experiential situations that would have the most meaningful impact (Venkatesh et al., 2016). The challenge is that pupils must vigorously interact in the teaching-learning process. In this regard, the teacher prepares and presents the experiential learning activities. Simultaneously, the pupils are given learning opportunities to work on materials provided that they are personally experiencing the outcome of their actions (Aggarwal, Nagpurnanand & Manju 2002; Nwafor, 2013). Figure 3.1 outlines the relationship between epistemology, theoretical perspectives, and the methodology and research methods as applied in the current study.

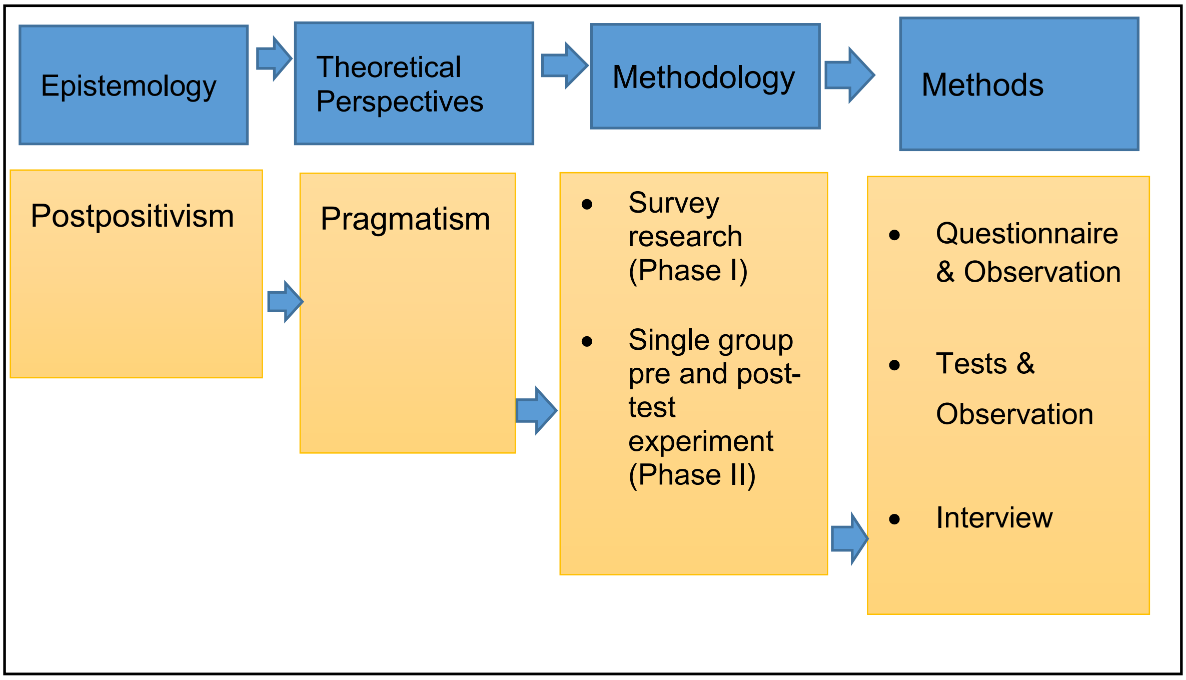


Figure 3.1: Relationship between epistemology, theoretical perspectives, methodology and research methods

Based on Figure 3.1, using the postpositivism from mixed community context with a pragmatism learning perspectives, the researcher for this study would interact with mathematics teachers to explore their classroom practice, and encourage them to implement a variety of pupils' centred approaches by using FA strategies. The teachers were also projected to employ FA strategies and focused on what works in practice. This FA approach would accord the possibility that pupils would be given a chance for experiential learning that enhance their learning of HOTS in mathematics. In a nutshell, the postpositivist perspective emphasises using a hybrid design, a combination of both quantitative and qualitative approaches as methodological drive, while a pragmatic research approach fits most comfortably within a postpositivist epistemology (Panhwar, Ansari, & Shah, 2017), hence its use in this study as a theoretical lens for investigation. A discussion on the research approach is provided in Section 3.2.

3.2 Research Approach

The most primary factors for one to consider after rendering the appropriateness of mixed-methods research for the phenomena under study include the decisions associated with the phases of the research, the priority of the methodological approach, design-investigation strategies, mixing strategies, and orientation (Venkatesh et al., 2016). The current study also follows all the strategies for mixed-methods research designs as outlined following this section.

3.2.1 Phases of the research

In this study, mixed methods were applied in the context of multistrand designs which contained three phases. According to Teddlie and Tashakkori (2009), multistrand designs involve mixing the quantitative and qualitative components in or across all stages. This study was also constituted in multiple faces in a broader research programme, with each phase encompassing all of the stages from conceptualisation through inference (Venkatesh et al., 2016).

Phase I of the study involved the collection of data by a questionnaire over a wide population of the Standard 4 teachers and followed by a carefully planned observation of teacher practice of FA in the classroom. The results of Phase I of this study informed Phase II while Phase III of the study involved data collection with the pupils' assessment and semi-structured interviews from teachers to get their gained achievement, experience and reflection following up on the FA interventions respectively.

All the three phases enfolded together into an embedded approach as an attempt to answer the main research question, which determined the extent to which the FA intervention for teaching and learning to enhance Standard 4 pupils' HOTS in mathematics. Table 3.1 below provides a summary of all aspects of the research methodology of this study into three phases.

Table 3.1: Summary of the Research Design and Methodology

Research Topic	Botswana teachers' experiences of formative assessment in Standard 4 Mathematics.					
Research Approach	Mixed Methods					
Research Design	**Sequential embedded mixed, single-group, pre-test, intervention (training) and post-test design** (Martin, 2016; Creswell 2014). The purpose of the study is to determine the possible effect of an intervention for BS4MT use of the FA strategies in the classroom to enhance Standard 4 pupils' HOTS.					
Main Question	To what extent does the FA intervention for teaching and learning to enhance Standard 4 pupils' HOTS in mathematics?					
Study Phases	PHASE I		PHASE II		PHASE III	
Research-Sub Questions	Question 1 What are the current practices in FA and teaching of HOTS	Question 2 How can FA be supported through intervention to enhance teaching and	Question 3 What are the current pre-intervention levels of mathematics achievement in HOTS	Question 4 To what extent does the FA strategies intervention enhance Standard 4 pupils'	Question 5 To what extent does the FA strategies intervention enhance Standard 4 teachers' teaching of	Question 6 & 7 What are teachers' experiences and reflections following the FA strategies

	in Botswana primary schools?	learning for Standard 4 pupils HOTS in Mathematics?	items for Standard 4 pupils?	HOTS achievement in mathematics when comparing the pre -and post-intervention?	HOTS in mathematics when comparing the pre- and post-intervention?	intervention and mathematics teaching of pupils' HOTS?
Data Collection Strategies	The baseline survey used 150 Standard 4 teachers from southern education region primary schools. 9 teachers who selected for FA intervention, were observed prior to the intervention.	Nine teachers in the nine (9) intervention schools will also document self-reflective journals after every mathematics lesson. The teachers were observed at least twice during the intervention implementation phase.	One pre-test for the Standard 4 pupils from nine sampled schools. That is a total of 272 pupils in the intervention schools.	One post-test for the Standard 4 pupils from the nine schools (272 pupils).	Teachers' pre- and post-observation findings and semi-interviewed on the possible effect of the intervention on teaching and learning HOTS in mathematics.	
Strategies for each sub-question	Baseline Survey Observation	Observation	Pre-test	Post-test	Pre- and post-observation and Semi-structured interviews	

Data Analysis	Rasch Measurement (RMT) Theory using Winsteps. For Quantitative data: SPSS 26.0 version statistical package descriptive and inferential statistics For observation data: computed using Microsoft Excel.	Microsoft Excel will be used to develop: coding, themes, detailed descriptions. For observation data: computed using Microsoft Excel.	RMT (The Partial Credit Model) SPSS 26.0 version statistical package descriptive and inferential statistics.	RMT (The Partial Credit Model) SPSS 26.0 version statistical package descriptive and inferential statistics.	Microsoft Excel will be used to develop: coding, themes, detailed descriptions Interrelated theme and member checking.

3.2.2 Overview of the research design and methodology

The primary focus of the current study was to find out the possible effect of an intervention for Botswana Standard 4 mathematics teachers' (BS4MT) use of FA strategies in the classroom to enhance pupils' HOTS academic achievement. A mixed-methods design was found appropriate for the current study. The study involved Standard 4 teachers and pupils for the intervention. This design was in line with pragmatism theoretical assumption, which is concerned with the multiple realities and moves away from the purely objective stance adopted by the positivist (Waismann, 2011; Creswell, 2014).

According to Baird et al., (2017), pragmatism is regarded as practical experimentation or intervention as an essential part of studying human practices using a mixed approach. Denzin (2012) refers to mixed research as a means of triangulation. The current study, therefore, triangulated a single group pre/post-test design which was a sequential, embedded mixed method-QUAN (qual) approach. The researcher opted for the sequential mixed methods research because the quantitative and qualitative data was collected in different phases of the study (Creswell, 2014; Creswell & Plano Clark, 2017).

In a mixed-methods design, the researcher has several options regarding the priority of the research. According to Creswell and Plano Clark (2017), priority refers to the relative importance or weighting of the quantitative and qualitative methods for answering the study's questions. That is a priority can be given to both methods or can emphasise one method over the other (Byrne & Humble, 2007). This study, embedded a fully mixed methods approach to integrating one data set and provide a supportive, secondary role in a study based primarily on the other data set (Creswell, 2014). Leech and Onwuegbuzie (2009) illustrate that the design of a fully mixed-method involves using both quantitative and qualitative research across all components, for instance, objectives, type of data and operations, type analysis, type of inference.

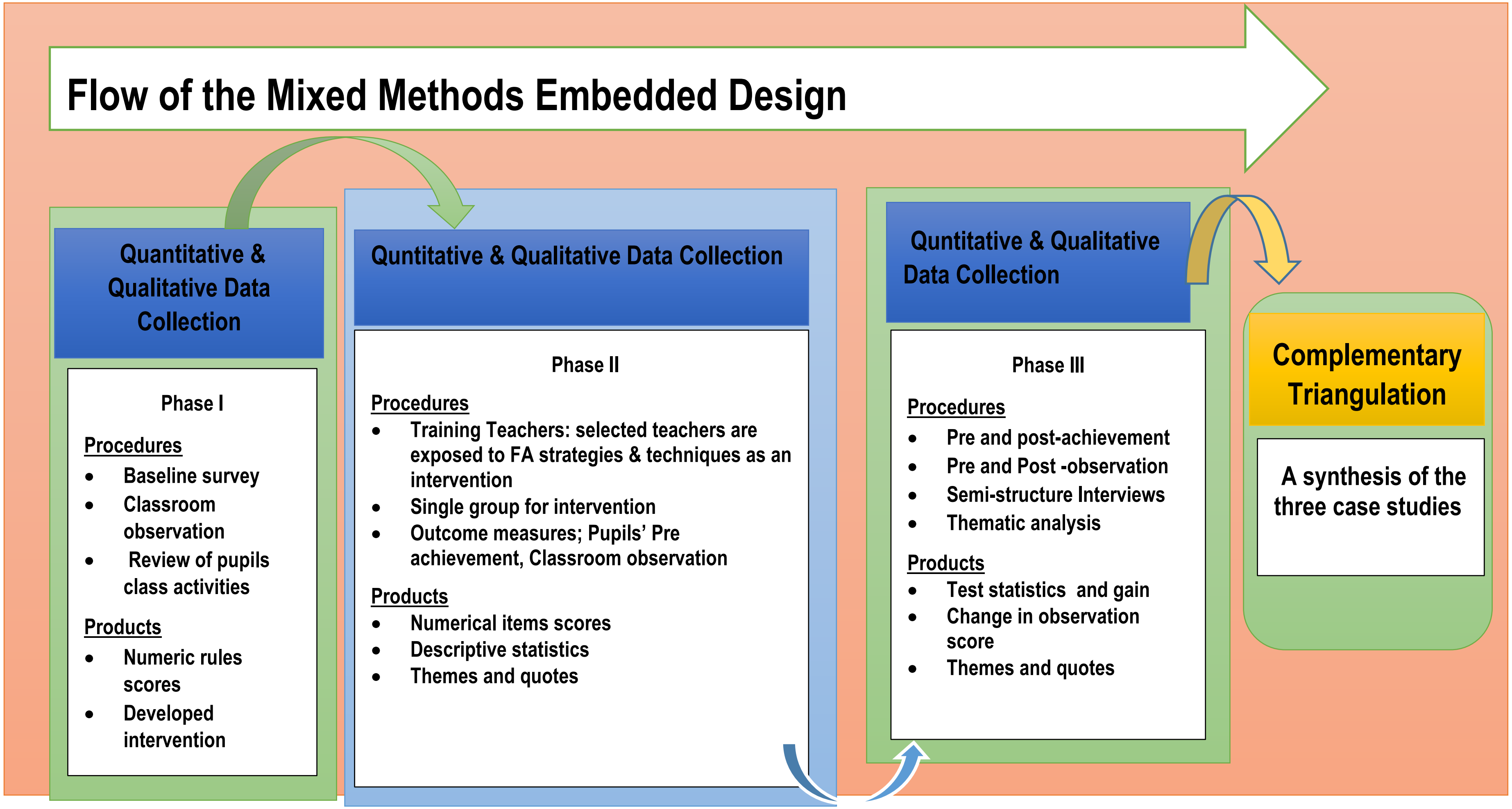


Figure 3.2: The complementary research design of this study

Figure 3.2 reflects the flow of the activities which were embedded in the design. As can be seen in Figure 3.2 (above), the current study is also deemed a sequential mixed method as far as time orientation was concerned. The flow of the research activities was typically conducted in one phase of the study at a given time and then followed by the other strand of the study. The sequence of phases was guided by research objectives and questions as used in the study (Creswell, 2014).

This research suggests that by mixing the methods, the data from the quantitative measurements and analyses could be fruitfully compared with those obtained from the classroom observation and interviews. This combined data would lead the researcher to appreciate better the realities being investigated (or, in Popperian terms, allow a closer approach to the truth) (Romm, 2020). The teacher would therefore obtain a better understanding of FA and teaching HOTS in mathematics.

In the mixed methods approach, a type of triangulation across the different methods have to be clearly defined, otherwise, it gives rise to practical problems in aligning data. The challenges that could be experienced in aligning data include differences in data collection procedures, the complexity and instability of the construct being investigated, difficulties in making data comparable, lack of variability in participant responses, greater sensitivity to context and seemingly emotive responses (such as interviews) and possible misinterpretation and greater control of content exposure in the questionnaire (Harris & Brown, 2010; Molina et al., 2018, Van Staden & Zimmerman, 2017).

Triangulation in the current study was much more towards complementary findings than confirmatory findings from different phases. The triangulation application in this study corroborates Harris and Brown's (2010) findings suggesting that complementary research findings are good to avert the cost of a separate confirmatory study approach. Strong alignment of data is required to avert "the loss of rich complementary data obtained through allowing each method to be analysed in its own right" (Harris & Brown, p.1). Firstly, the current study explored the teacher's understanding of how FA was used to enhance HOTS in mathematics. A baseline survey was carried out, and it involved Standard 4 teachers through questionnaires and pre-observation (Phase I). After that, the post-observation in some selected schools followed.

Based on the baseline findings for Phase I, the teachers were then workshopped in Phase II in an FA intervention and were expected to practice the learned materials in the classrooms. A single-group design for pupils' pre- and post-achievement of HOTS in mathematics was used to compare and determine the possible effect of the schools' FA intervention workshops. The findings for Phase II were used to show if the FA intervention for teaching HOTS in mathematics can help pupils better understand mathematics concepts and enhance their achievement. Teachers' experiences and reflections on their perception regarding FA after attending the intervention workshops and FA implementation in the classroom were paramount in this study (Phase III). This last phase was consistent with MacDonald's (2012) suggestion that intervention research offers a radical alternative to knowledge development in its mandate to remain a collective, self-reflective inquiry to improve a situation (Figure 3.2).

A detailed description of the phases involved in the current study would be discussed in the next subsequent chapters. Each chapter (Chapter 4, 5 and 6) describes a phase, and the discussion is focused on the design and methods, interpretation of the findings and discussion.

3.3 Quality Assurance Issues of the Study

This section describes the reliability and validity which were established in and across the two sets of data. Section 3.3.1 deals with how reliability and validity were applied in this study. As qualitative data, the term "trustworthiness" (Nieuwenhuis, 2016) is used to ascertain reliability and validity and how they were applied in this study, as discussed in Section 3.3.2.

3.3.1 Quantitative data

The instruments' content validity was supported through the literature (Thompson & Wiliam, 2008) and experts' judgment concerning the FA strategies to enhance HOTS academic achievement. The research supervisor, lecturers from the university and a candidate PhD student, were considered for content expert review for the FA strategies intervention and the questionnaires. All the instruments were piloted to check that instructions were understood and that the content of the items and response format yielded usable data.

Similarly, the content validity for the tests was examined by subject specialists who compared the adapted items for BEC instruments with the current use and assessed the instruments for coverage and measure for HOTS. Quality criteria for the instruments are discussed in Chapter 4 (see Section 4.4.3 and Section 5.4.3), and these instruments were subjected to the Rasch statistics measurements.

3.3.2 Qualitative data

The criteria to ensure trustworthiness for qualitative data involved measuring credibility, transferability, dependability, and confirmability, as Guba (1981, as cited in Gerryts, 2018) suggested. Credibility implies the extent to which the findings are believable and congruent (Nieuwenhuis, 2016). As for this study, the credibility of FA to enhance teaching and learning has been successfully done in other contexts (Kanjee 2017; Kivunja 2015; Wiliam 2011). The researcher has also been teaching and has been in an assessment environment as a teacher from primary schools to higher education institutions and holds a Master’s of Education in Research and Evaluation. This background means he is familiar with formative assessment in the classroom.

According to Nieuwenhuis (2016), the findings' transferability is accomplished through a rich description of the context, participants, and research design. Similar to the current study, formative assessment and teaching of HOTS (FAHOTS) was explained in the context of classroom assessment practice in Botswana and how the lack thereof could negatively impact pupils’ achievement. The criteria for selecting the participants were explained at each phase of the study.

Dependability criteria for this study were attained through the use of observation (its quality criteria as described in Section 4.4.4) and an interview guide (see the quality criteria discussed in Section 6.3.2). Additionally, during data collection, it was ensured that a research journal was kept and captured the process, perceptions, and ideas through the data collection process. Dependability was also ensured by keeping documentation and notes on the categories, labels and other data during the data analysis process as well as using an external coder who assisted and externally commented on the coding and interpretation of data (Nieuwenhuis, 2016).

By using triangulation, this study opted to use quantitative and qualitative data to complement the findings in more than one way. Additionally, for qualitative data, the external decoder who assisted in data coding and discussion of the findings with the research supervisors also enhanced the study's neutrality. The application of confirmability at this level of the current study was consistent with the notion suggested by Cook (2016), that confirmability is the degree of neutrality of the data.

3.4 Ethical Issues and Considerations

As far as ethics are concerned, special care was taken to ensure valid outcomes for the study. The researcher applied for ethics approval at the University of Pretoria’s (UP) ethics committee, detailing how the study would be conducted without violation of the participants' rights and privacy. Upon approval by the UP-ethics committee, permission to research in Botswanan schools was sought from the Southern Education Region, the Ministry of Basic Education (see *Appendix O*).

(i) Voluntary participation and Informed Consent

The prospective participants were voluntarily requested to sign the consent form if they consented to participate in this study. To achieve this consent, the researcher provided letters of informed consent to the Standard 4 mathematics teachers, the principals and teachers to obtain permission to research in their schools, which paved the way for finally seeking permission from individual participants. The study’s participants were professional teachers and pupils in the schools. To this effect, these regional officials, principals and participating teachers were assured that the study's findings would be reported in a manner that protects the anonymity of the participants.

Before requesting signed consent forms and the data collection process started, the researcher clarified to the participants what their commitments would be if they participated in the study. The study's purpose, potential risks and benefits were explained to obtain buy-in from participants. The Standard 4 mathematics teachers were informed that they would participate in a survey questionnaire completion, training workshop, implementation of the intervention, administration of the two tests for the pupils, and an individual post-intervention interview. The participants were informed that participation was voluntary and not compulsory. Informed consent documents were written in non-technical language so that all participants would understand them. The participants were assured that information about their personal details would be secured and their privacy would be respected at all times. The researcher obtained the participants' approval on using an audio recorder before the commencement of the research.

The Standard 4 pupils were given the consent form to be signed by parents or guardians to participate in the study. Pupils who returned the signed form participated in the study, and those without signed forms were not used in the study. The pupils also received information about the research from their teachers, explaining the researcher's presence in the classroom during observation visits in the form of a letter.

(ii) Safety

Because they were part of this study, the participants were exposed to possible effective formative assessment practices that could enhance methods and improve pupils' learning outcomes and HOTS development among pupils. The study used an intervention; therefore, participants' safety was more important than the advancement of the researcher's career. The researcher avoided information that compromised the relationship between the researcher and the participants, which took time and effort to establish, at all cost.

Unintended harm may have occurred when the researcher was unable to listen with complete attention to the responses provided by the interviewees. The participants (teachers) might also have felt uncomfortable during classroom observations and interviews, but any intentional discomfort was avoided as far as possible. Since this study was an intervention, the researcher provided support to the participants when there was a risk of undue stress, for instance, where teachers might fear the discovery of incompetence regarding teaching mathematics and poor classroom assessment.

Risks that were anticipated include lack of resources from schools, poor time management and unwillingness to adapt to the new classroom assessment strategies.

(iii) Privacy, Confidentiality and Anonymity

The researcher ensured that a relationship based on mutual respect, honesty and trust existed during the sample selection phases so that the selected participants could share information without the risk of harm. The selected schools and teachers were provided with pseudonyms in the study. Only the researcher knew the names of the schools and participating teachers. Such a sample list of participants' details was drawn up and kept safe. The schools' headteachers were fully aware of the teachers' involvement in the research, as observations were scheduled during class time. Similarly, participating teachers were made aware that the principal approved their involvement in the study.

Before the commencement of the data collection at every phase of the study, instructions on the data collection instruments, such as the questionnaire, were made clear for the participants to make sure they understood what was expected of them. Similarly, during the interviews, the researcher read out the confidential issues in the informed consent form to reassure participants of their anonymity in the study. Any information that can link the participants directly to the research were kept separately from the main findings.

All the data obtained from participants were used to write this dissertation. No information would be provided to identify the participants as pseudonyms were used for both the schools and the teachers and pupils to ensure anonymity and confidentiality. To enhance the data's credibility, participants were informed that all the interview data would be audio recorded for verification and inference before it was disseminated. Data collected through an audio recording were coded, kept confidential and only transcribed for this study's analysis.

4 CHAPTER FOUR: BASELINE DESCRIPTION (TEACHER SURVEY AND OBSERVATIONS) LEADING TO THE DESIGN OF FAHOTS INTERVENTION

This chapter discusses the research design and methodology of the baseline survey employed in Phase I of the study. Section 4.1 is devoted to detailing the research design and methods for the baseline study, while Section 4.2 is dedicated to the research site and Section 4.3 presents the sampling and participants involved in the study. This chapter discusses instrument development (Section 4.4) and procedure (Section 4.5) as used in Phase I of this study. Section 4.6 and 4.7 present the data analysis and findings respectively. Section 4.8 presents the discussion. The last section of this chapter is the intervention (Section 4.9).

4.1 Research Design

In an attempt to address Phase I, a cross-sectional survey was considered fitting this phase of the study. A survey was used to collect quantitative data about teachers' practice of FA to enhance pupils' HOTS in mathematics. The survey was also chosen based on the purpose that it can guarantee one will do a follow-up analysis of the problem at hand. In the case of this study, it was recognised in chapter 1, that Botswana Standard 4 pupils have been identified as mathematically weak in HOTS as compared to their cohort in international large-scale assessment. It was, therefore, imperative to have commissioned a survey as a follow-up study to ascertain the nature of the problem in the context of teacher classroom practices (Avedian, 2014). Thus, the baseline survey was aimed to further explore the participants' conduct and understanding of the FA practices about strategies and techniques as a way of enhancing pupils HOTS in mathematics.

Specifically, this baseline survey is enclosed within FA as a yardstick to check the classroom assessment in the Botswana context, since the literature affirmed the possibility of FA strategies to improve teaching results as far as the pupil's performance is concerned (Heritage, 2010; Kivunja & Kanjee 2017). Thus, survey the survey instruments were aligned with Kivunja (2015)’s assessment feedback loop Model (AFL) which was employed as a theoretical departure point to enhance the teaching of HOTS in the mathematics classroom by determining teachers' FA practice based on exploratory data.

According to Cohen, Manion, and Morrison (2011), a survey is a systematic method for gathering information from entities (in this case a sample of teachers) to construct quantitative descriptors of the attributes of the larger population of which the entities are members. The current study considered a survey to be an efficient way to collect information about a large group of teachers; a flexible medium that measured attitudes, knowledge, preferences for FA practices; using standardised data collection methods and so is less susceptible to error; easy to administer and tailored exactly to the phenomena under-study (Avedian, 2014; Babbie, 2009). Keeping in mind this purpose, using a wider population of teachers through a survey, ultimately led to the provision of data analysis quantitatively.

The success of survey research closely depends on how the participants respond to the survey questions match, how people think and act in reality (Avedian, 2014). In the current study, the researcher also gathered observational information as an attempt to augment the teachers' reflection of reality act into classroom assessment practice which cannot be observed directly with a questionnaire (Van Staden & Zimmerman, 2017).

The effort for classroom observation came as a result of Van Staden and Zimmerman's (2017) findings which cautioned against just using classroom process data collected through questionnaires. According to Van Staden and Zimmerman, the questionnaire alone does not always provide useful information, due to the difficulty to report timely data about best practices, especially when a teacher questionnaire is used concerning teaching practices in low school performance contexts. Using the questionnaire was discovered problematic as teachers may feel vulnerable and defensive, resulting in unreliable or unrealistic answers (Van Staden & Zimmerman, 2017).

Basically, for this phase, a mixed-methods approach was chosen as a way of assuring validity through the use of a variety of methods that involved different types of samples as well as methods of data collection (Martin, 2016). As for the data collection, mixed methods for this phase of the study was essential to seek complementary of the results from different methods (triangulation) and was used to compensate for the weakness of one approach by using the other (Venkatesh, Brown & Sullivan, 2014; Harris & Brown, 2010). The baseline survey and intervention was investigated using two sub-research questions:

- SRQ 1. What are the current practices in FA and teaching of HOTS in Botswana primary schools?
- SRQ 2. How can FA be supported through intervention to enhance teaching and learning for Standard 4 pupils HOTS in mathematics?

The following section describes the methods used for data collection and analysis to describe the FA practices by the teachers. This section provides a discussion focusing on how participants were selected and recruited for Phase I of the study. Firstly, a description of the site for the study is provided. Secondly, the description of the sample participant characteristics and the inclusion and exclusion are provided.

4.2 Research Site

This study was conducted in one of the ten education regions in Botswana. The regional offices represent the department in the Ministry of Basic Education, which implements educational policies and ensures sound management and supervision of schools in a region. These schools are the custodians of the regional academic affairs and enforce the common curriculum for lower primary mathematics as well as for the Standard 4 Attainment Test and Primary School Leaving Examination respectively. The selection of the research site was informed by the public schools that participated in the common curriculum implementation.

As observed in the problem statement section in Chapter 1 of this study, the poor mathematics performance in primary schools is a problem across the country, and the pattern is cascading from lower primary to upper primary and beyond regarding the pupils' pass rate at PSLE and JCE. For these reasons of a national academic performance crisis in schools and across education regions, the Southern Education Region was the setting of this study as it was convenient to the researcher. According to Lodico, Spaulding and Voegtle (2010), the convenience sampling procedure is the best of the non-probability sampling methods when limited resources are available. Employing the convenience sampling technique reduced expenses and time associated with travelling and accommodation. Teachers are also deployed centrally by the Ministry of Basic Education headquarters based on the available posts in a particular region, and they have no choice of schools they are posted to; hence selection bias in the study area was minimised.

4.3 Sampling and Participants

In the Southern Education region, this study proposed a multistage random sampling strategy. Nenty (2013) described multistage random sampling as a successive stage by stage random sampling from more inclusive to less inclusive sampling units until the population elements that constitute the desired sample mix has been put together. When one is involved with large-scale survey studies (that cut across states, school education regions, local government areas, schools, and classes), one seldom uses simple or stratified random sampling methods because of the enormous expenses involved. For these reasons, the sample design for the current study is precisely considered as a four-stage stratified cluster sample that was executed in phases in totality.

Foy and Joncas (2003) and Van Staden (2010:104) greatly purport the stratified cluster sample in a large-scale study due to the following reasons:

- It produces reliable estimates, using different sampling designs, for sub-national domains, for instance, education regions, education sub-regions;
- It improves the sampling efficiency. Thus, improving the reliability of national estimates without necessarily increasing sample sizes; and
- It guarantees that different parts of the population are appropriately represented in the sample.

Specifically, the current study was consistent with Polit and Beck (2018) who emphasised an appropriate sample as very necessary for any research as either too small or too large sample sizes which are not a good representation of the population. The sample for this study was ensured to be adequate as a strategy to allow the investigator to be confident in the findings of the study (Majid, Ennis & Bhola, 2017). In this way, computing the sample size became an important step during the baseline survey of the current study. There were 119 schools in the Southern Education region, and each school had two (2) Standard 4 teachers on average; hence the estimated teachers' population at Standard 4 levels were 238. The formula below was used to compute the sample size.

$$n = \frac{(N)(p)(1-P)}{(N-1)\left(\frac{B}{C}\right)2+(p)(1-P)}$$

From the formula, the researcher adopted the Beiemer (2003)' identification components, where ***n*** is the computed sample size of the desired level precision; ***N*** is the population size; ***P*** is the proportion of population expected to choose; ***B*** is set precision at .05 or 5%, and lastly ***C*** represents the ***Z*** statistics associated with the confidence level which is 1.96 that corresponds to the 95% level. The confidence level of 1.96 corresponds to the formula below:

Where *N* = 238, *p* = 0.5, *B* = 0.05, *C* = 1.96.

$$n = \frac{(238)(0.5)(1-0.5)}{(238-1)\left(\frac{0.05}{1.96}\right)2+(0.5)(1-0.5)}$$

$$n = \frac{59.5}{0.406}$$

$$n = 146.5$$

$$n = 147$$

In this study, the researcher identified five existing educational sub-regions that form the Southern Education Region of Botswana. According to Polit and Becker (2018), stratified random sampling is obtained by separating the population into mutually exclusive sets or strata, then drawing simple random samples from each stratum. The sub-regions all have a different number of schools with varying population sizes from which samples were drawn. Twenty-five per cent (25%) of schools were randomly drawn from the proportion of sub-region areas with more schools, and at least seventyfive per cent (75%) were drawn from sub-region areas with fewer schools as shown in Table 4.1. Moreover, the sample size was also calculated online with Rao software (roasoft, 2004) and revealed a total sample of 148 teachers, at a 5% margin of error, which imply a weak power (beta = .67 for f = .25 medium size) sample distribution for the 5 sub-regions identified (http://www.raosoft.com/samplesize.html).

For this study, 162 questionnaires instead of 147 were distributed randomly among Standard 4 teachers, using proportional distribution, compensating for this power weakness of the sample distribution as shown in Table 4.1. Specifically, oversampling was also done to make up for the possible loss due to non-cooperative participants, assuming that the estimated number of Standard 4 teachers may not be available in some schools (Bambale, 2014). The 81 sampled schools would enable the researcher to reach 162 teachers in schools that offer mathematics for Standard 4 pupils (see Table 4.1). Table 4.1 displays the response rate of the teacher participants. From the expected total of 162 teachers, a 95% response rate (150 teachers) was attained.

Table 4.1: Sampling Frame of Primary Schools in the Southern Education Region by Sub-Regions

Sub-Regions	Expected number of participants	Attained number of Participants	Cumulative percentage of Participants
Kanye	40	61	40.7
Lobatse	16	32	21.3
Mabutsane-Jwaneng	32	8	5.3
Moshupa	30	34	22.7
Goodhope	44	15	10
Total	162	150	100

A total of 61 (40.7%) of the teachers were from the Kanye sub-region, while the lowest response rate was observed from Goodhope (*n*= 34, 10%) and Mabutsane-Jwaneng (*n*=8, 5.3%), respectively. The sample distribution difference between expected and attained numbers by regions was analysed using over Chi-square testing, and the result was significant (χ^2 *=38.20, df=4, p=.0000001*), suggesting that there are large deviations for at least four sub-regions of the five total regions used for this part of the study (https://www.icalcu.com/stat/chisqtest.html). Such deviations are a potential threat to this study's results; therefore, the overall results of the baseline survey cannot be compared by sub-regions. The low response rates in these two regions might have been influenced by how the questionnaire was received, that is, it was post-mailed to each school, and some participants did not return the questionnaires. In other regions (Kanye, Lobatse, and Moshupa), the researcher delivered the questionnaires and made frequent visits to those schools to collect them upon completion, and the response rates were higher (See Table 4.1).

Small sample sizes (n= 150) may pose some potential threat towards multivariate data analysis. However, some literature supports small sizes and yield satisfactory findings to some extent (Combrinck, 2018; Linacre 2004, 2005). Thus, even n< 200 would allow the true score to vary from observed scores by +/-1 logit (Bond & Fox, 2015; Boone, Staver & Yale, 2014; Linacre, 2005). Contrary to this support for small samples, Houts, Edwards, Wirth and Deal (2016) criticised such a claim and found small sample sizes may lead to unreliable findings. He and Wheadon (2013) further criticised this notion by pointing out:

> *Nevertheless, the results obtained from this study demonstrate that both the sample size and the item type have an important influence on model parameter estimation. The Root Mean Square Error (RMSEs) for sample category measures decrease with increasing sample size. When the sample size is fixed, the RMSEs generally increase with an increasing number of categories in an item. Model parameter estimation is also affected by the score distribution between categories of the items* (p. 312).

Variation of sample parameter estimation shows that increases in sample size (e.g. n=300), can reduce the differences varying between -1.03 logit and 1.22 logit. Small sample sizes of 150 lead to differences between estimates and models' actual values to vary between -1.08 logits and .81 logits (He & Wheadon, 2013). Such variation is also observed with the RMSE of category measures, with a small sample size (n=150), category measure is as high as .90, while the increased sample size is observed for a majority of categories within .20 logits. This variance was also similar to analysing the impact of sample size on model parameter estimates for items with different numbers of categories (He & Wheadon, 2013).

Despite this criticism of sample size, there is also evidence on Rasch analysis that support small sample size studies. For example, an exploratory study about the accuracy of estimation can use available sample size; however, the result analysis of such a study has to be approached with caution (Brown & Abdulnabi, 2017; Revicki et al., 2014). Revicki et al.'s (2014) recommendation to use Rasch analysis based on small samples should only be used for exploratory purposes with extreme caution due to related items that may incorrectly order the parameters. Similarly, the current study was exploratory with a limited number of Standard 4 teachers per sub-region, and hence the analysis had to be interpreted with caution.

With item ratio, Costello and Osborne (2005) found that the majority of the studies (62.9%) which they reviewed revealed that analysis with "subject to item ratios of 10:1 or less which is earlier and still-prevalent rule of thumb is used by many researchers for determining a priori sample size" (p.4). In the current study, the sample size of 150, FA items n=48 (ratio 3:1) and HOTS items n=15 (ratio 10:1), are considered consistent with this rule of thumb, to allow an analysis of the item communalities (low to moderate communalities of .04 to .07), and loading of an item (.32 or .50 and better) (Costello and Osborne, 2005; Tabachnick & Fidell, 2001).

4.3.1 Inclusion and exclusion criteria

The inclusion and exclusion criteria for selection in this study were determined before sampling. The inclusion criteria for the study were as follows:

(i) The primary schools had to be in the Southern Education Sub-Regions of Botswana, and those schools who sit for the Standard 4 Attainment tests and Primary School Leaving Examination, respectively;
(ii) Only the public primary schools, which were accessible, were included;
(iii) Teachers and pupils who were teaching and learning in primary schools in the Southern Education Sub-Region, respectively;

(iv) Teachers are either currently teaching Standard 4 or have taught Standard 4 levels for the last 2-3 years;

(v) Pupils must be doing Standard 4 and taking mathematics as a core subject;

(vi) Within the classroom, intellectually able pupils who could follow the general instructions of the test were included. This criterion did not discriminate and exclude those pupils with poor academic performance, but only those who had learning barriers.

As for exclusion, all teachers and pupils not meeting eligibility criteria would be excluded.

4.3.2 Participant characteristics

Table 4.2 presents the demographic characteristics of the teachers who participated in the study. As can be seen from Table 4.2 (below), 105 of the participants (70%) were females, while only 45 participants (30%) were males. Male teachers were observed to be either few in numbers or absent in some primary schools who were teaching Standard 4 during the study period or have taught Standard 4 in the previous two years. The limited number of male teachers at the lower level is to a large extent, the situation at many primary schools in Botswana and globally (Statistics Botswana, 2015).

Table 4.2:Demographic data of the teacher participants

Variable		**Frequency of Participants**	**No of Participant in Percentage (%)**
Gender	Females	105	70.0
	Males	45	30.0
Level of education	Certificate	6	4.0
	Diploma	107	72.0
	BSc/Ba degree	30	20.0
	Master's Degree	4	2.7
	Other	2	1.3
Missing	System	1	.7
Teaching Experience	1 – 5	20	13.3
	6 – 10	23	15.3
	11 – 15	13	8.7
	16 – 20	24	16.0

Variable		Frequency of Participants	No of Participant in Percentage (%)
	21 – 25	31	20.7
	26 – 30	21	14.0
	31 – 35	15	10.0
	36 – 40	2	1.3
Missing	System	1	.7

Of the 150 participants who responded to the questionnaire, 107 (70.1%) reported having attained a diploma, followed by 31 (20.7%) of participants with a degree qualification. On the other hand, their teaching experience ranges from 1 to 40 years. From this teacher experience data, it can be seen that 31 (20.7%) of the teachers were the most experienced at primary school teaching, ranging between 21 to 25 years, then followed by 24 (16%) teachers with 16-20 years of experience. For the first 10 years (1-10) of teaching experiences, the participants accounted for 43 (28.3%), while and last ten years (31-40) accounted for 15 (10%) of the total participants who participated in the study. These results revealed that teaching experience at the primary schools was relatively broad, and many teachers had stayed longer in the teaching fraternity. The crosstabulation analysis revealed a significantly higher proportion of experienced teachers with diploma (n=74,70% of 107) and degree (n=10, 66.7% of 30) qualifications than the less experienced teachers (n= 66), χ^2 *(8, n =149) =68.04, p< .001, phi=.67*. Such association of teaching experience and qualification may have some implications on the classroom instructional strategies, assessment skills and practices.

In addition, more than half of the participants (n= 100, 66.7%) were teaching Standard 4, while 50 (33.3%) of the participating teacher have previously taught Standard 4 within the last two years. A chi-square goodness of fit test indicates no statistically significant difference in proportion between current teachers (n= 100) and those who previously taught Standard 4, (n =50), χ^2 *(1,), (n= 150) =.000, p< .993*.

4.3.3 Teacher level of training in educational assessment

Figure 4.1 illustrates the level of training in education assessment as perceived by primary school teachers who participated in the study.

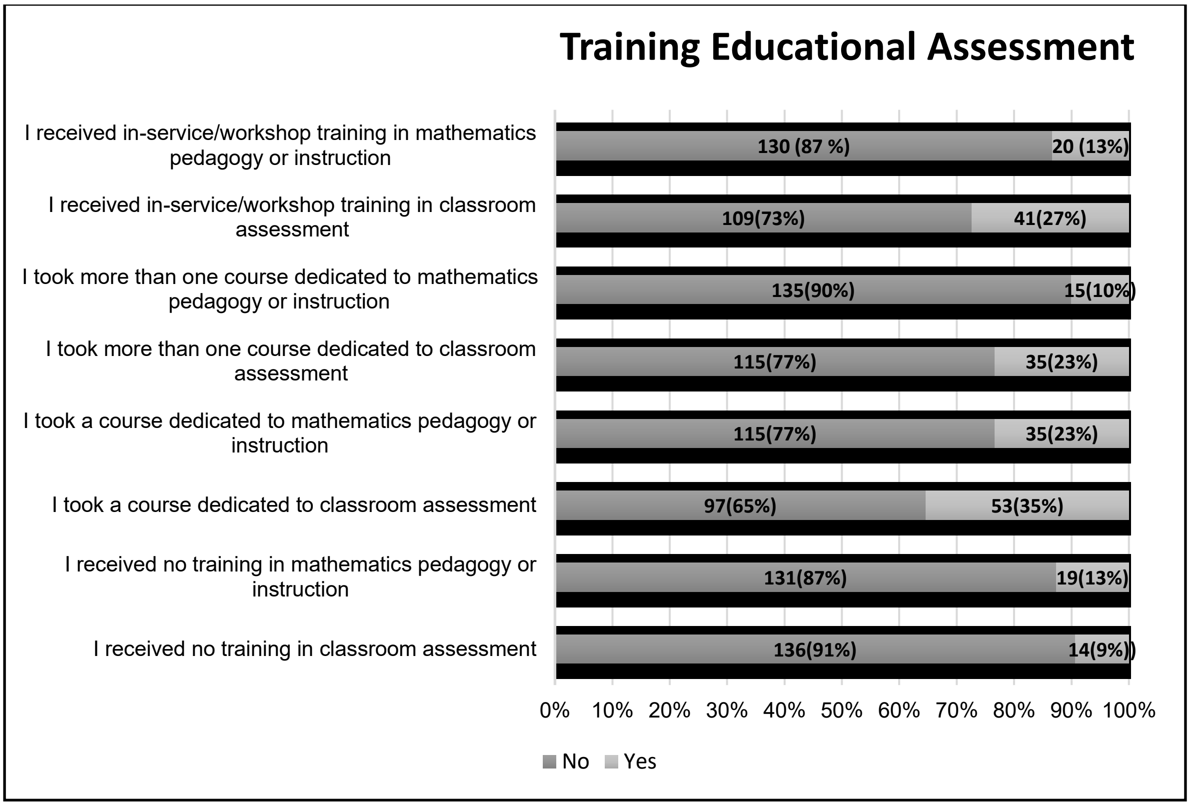


Figure 4.1: Level of the training in Educational assessment as perceived by teachers

As the data reveals, the rate of disagreement of the teachers' perception of in-service training on mathematics pedagogy or instruction and classroom assessment were 130 (86.7%) and 109 (72.7%), respectively. This minimum in-service training through workshops implies that teachers may likely have insufficient knowledge of the transformational classroom instruction embedded in formative assessment practice. The crosstab analysis for qualification and in-service training did not reveal significant results, $\chi^2(4, n=150) = 1.324, p < .857$. Chi-square results to associate qualification and classroom assessment was also insignificant $\chi^2(7, n=150) = 9.623, p < .211$. More teachers (*n*= 115, 76.7%) reported having neither taken more than one course dedicated to mathematics pedagogy or instruction nor a course dedicated to classroom assessment. Crosstab analysis for teaching experience and teachers' training on mathematics assessment in more than one course revealed significant findings χ^2 *(7, n =150) =15.66, p< .028, phi=.028*. Thus, the more experienced teachers, the less likely they were to have taken more than one classroom assessment course.

The findings of this study on demographics (qualification and experience) are inconsistent with the prior reports on teacher qualification (e.g., obtaining a graduate degree) which was found to be related to a higher level of assessment literacy (Hoover, 2009; King, 2010), and schooling experience, which was found to impact teachers' assessment decisions (Campbell & Evans, 2000). However, the current findings do support the basic teacher demographics results for Brown's (2008) study, which found that training experience in assessment did not affect participants.

4.4 Instruments Development

Most studies in education are of the nature of surveys and such studies require the development and application of measurement instruments (Nenty, 2013). Likewise, the current study developed a questionnaire and observation instruments because they can enable the researchers to collect data from many participants over a wide geographical area.

This section provides a detailed outline of the teacher questionnaire and classroom observation instruments used in an attempt to answer the proposed research question of the study in Phase I.

4.4.1 Teacher questionnaire

A review of the related literature was imperative for this study to understand the extent to which teachers engage in true FA practice and teaching of HOTS as well as alter their instructions as per the pupils' needs. This study explored the five FA strategies of classroom practice which included:

(i) learning goals and criteria for success,
(ii) engineering effective classroom discussion (questioning),
(iii) engineering effective classroom discussion (collaboration),

(iv) learning tasks, and

(v) feedback on instructions (Ramsey & Duffy, 2016) and studies on instructional integration for the teaching of HOTS drawn from research on effective FA processes (Heritage, 2010; Kivunja, 2015; WAG, 2010).

For the purpose of this study, a sufficient sample was needed to ascertain the problem context, hence a baseline survey using a questionnaire was appropriate.

A self-administered teacher questionnaire was considered suitable for this study to collect information by asking respondents to complete information on the six variables of FA practices tied to the teaching of the HOTS in mathematics. The decision to use the questionnaire was motivated by the exploratory nature of Phase I of the study. The survey questionnaire was largely planned to gather information about classroom assessment contexts for developing HOTS, explore the five FA domains namely; learning intentions and criteria for success, engineering effective classroom discussion (questioning), engineering effective classroom discussion (collaboration), learning tasks, and feedback on instructions (Heritage, 2010; Kivunja, 2015; WAG, 2010). The sixth variable in the questionnaire inquired about how teachers integrate FA domains and HOTS in the teaching of mathematics.

(i) Development of the draft teacher questionnaire

Based on the literature reviews (Chapter 2), a draft of the closed-ended questionnaire was developed containing statements on FA strategies and HOTS. The research sub-questions (What are the current practices in FA and teaching of HOTS in Botswana primary schools?) was broken down into six domains and aligned with AFL components of the study's conceptual framework (Kivunja, 2015), forming the basis of the questionnaire. Questionnaire statements were adapted from Ramsey and Duffy (2016) as outlined in *Appendix A*. The MSDF's (2016) overall purpose of the study was to help the teachers with the foundation better to understand FA and support them in implementing FA in their classrooms on a routine basis. The data collection instrument for MSDF was designed by Dr Carla C. Johnson, Associate Dean for Research, Engagement and Global Partnerships in the College of Education at Purdue University (MSDF, 2016). The reliability of the instrument was not reported; however, the current study established the internal consistency through Cronbach alpha before analysis of the data.

The MSDF questionnaire was designed for public schools for the USA's large urban areas whose student populations are racially, ethnically and socio-economically diverse. All three districts agreed to participate in the study to learn about formative assessment use and to improve their support to classrooms and schools (MSDF, 2016). The current study also applied a similar context for FA practices since this study intended to use FA intervention to improve teaching and learning of HOTS in mathematics.

However, due to the different geographical settings for the USA and Botswana, some of the items in the instrument were modified to suit the needs of the current study, for instance, the inclusion of items for HOTS. The draft questionnaire was then subjected to the expert judgment of the research supervisor at the University of Pretoria and one lecturer in Research and Evaluation at the University of Botswana. The lecturer from the University of Botswana felt that the statements in the questionnaire were sound and evident that each statement matched its proposed use.

The research supervisor advised considering placing the options next to every statement since participants would not necessarily look at the top of the answer sheet to see what the other options are. She advised that all statements should be clearly labelled with the response options, and that would permit the researcher to be sure that the participants had interpreted the scale correctly. Also, the distinction between "always", "often" and "sometimes" as initially scaled were difficult to measure. To this end, the research supervisor advised making the measure indicators options more tangible for teachers, for instance, that sometimes often be "equal to 2-3 times per week, sometimes equal to weekly and, always equal to every day of the week". Consequently, the draft of the teacher questionnaire was revised accordingly.

(ii) Piloting of the draft and development of the final teacher questionnaire

The revised version of the teacher questionnaire was piloted using two primary school teachers for Standard 4 in one primary school which was not sampled in the current study. The specific objectives of the pilot exercise included assessing the clarity of the statements per category, the unambiguity of the instructions and ensuring that the questions were interpreted in the same way by different participants. The pilot study revealed that there was no evidence that the statements were vague as the two teachers managed to respond to all items in the questionnaire. Hence, the questionnaire was treated as a final teacher questionnaire. The limitation of the piloting was that only two teachers participated in a qualitative analysis of the questionnaire. Statistical evidence was not collected at this point.

The final version of the teacher questionnaire constructs was operationalised using Likert-type scales. A Likert-type scale is a common approach employed to measure a wide variety of latent constructs as it is reliable and easy to use (Churchill & Iacobucci, 2010). According to Adelson and McCoach (2010), scales with a midpoint have greater reliability than even-numbered scales. However, the current study used even-numbered scales (6-point and 4-point Likert-type scale). Ideally, a questionnaire scale is constructed based on the ordinal measurement level. So, placing a neutral/undecided response, say for instance, in a 5-point Likert scale and awarding 3-points for neutral, greater than those who decided, is an inaccurate measurement (Nenty, 2013). This study, therefore, favours an even-point scale, with no room to allow a participant to sit on the fence by selecting a neutral scale option. Despite problematic issues of assuming the middle of the scale to reflect normality (Adelson & McCoach, 2010), this even-numbered scale was followed to find an improved and better way to understand a given social reality or phenomenon, in this case, teachers' understanding toward FA and teaching of HOTS in mathematics.

The questionnaire had six sections. Section A dealt with the demographic and teaching experience of the participants; Section B-E captured the teacher use of FA and behaviours which were characterised as *Not All (0), Once a Week (1), 2-3 Times per Week (2), 4 Times a Week (3) and Always (4)*. The last section (Section F) captured teacher's behaviour on the use of FA in teaching HOTS in mathematics using a 4-point Likert scale: *Strongly Disagree (1), Disagree (2), Agree (3) and Strongly*

Agree (4). The estimated time to complete the teacher questionnaire was 20-25 minutes (see *Appendix A*).

Table 4.3 shows the teacher-related variables that were employed to collect data using the questionnaire and classroom observation. These included factors such as teachers' demographic and teaching experiences, FA strategies (use of learning goal and success criteria, effective discussion through questioning and collaboration, the use of learning tasks in the classroom, effective feedback and lastly the integration of FA in the teaching of HOTS). As can be seen in Table 4.3, the content of the questionnaire and observation tool were derived from the Kivunja (2015) components to maintain the linking of research questions, conceptual framework and data collected.

Table 4.3: Variables for Formative Assessment Practices in the Classroom to enhance HOTS as taken through a Teacher Questionnaire, and Observation

Kivunja Components	**Variable**	**Purpose of the Question**	**Source of information**	**Type of Variable**	**Number Of response categories**
	Demographics and Teaching experience	Provides information on teachers' sex, educational background, teaching experience, teaching level, describe their training in assessing pupils and pupils class achievement.	Teacher Questionnaire	Categorical Ordinal	Between 1 and 5 categories

Kivunja Components	**Variable**	**Purpose of the Question**	**Source of information**	**Type of Variable**	**Number Of response categories**
Learning goals and criteria for success	PART B: Question 1 (a-n) Learning and criteria success	Provides information on the frequency using FA practice when begins with clear learning goals, criteria, and alignment to past and future so that pupils understand and what they should already know, what they are about to learn and how that learning ties together	Teacher Questionnaire Classroom Observation	Categorical Ordinal Categorical Ordinal	Between 1 and 5 categories Between 1 and 3
	PART F: Question 6 (a-o) FA and teaching of HOTS	Provides information on the frequency teachers integrate FA and teaching of HOTS promote the pupils' quality thinking and learning. Pupils test scores for pre-achievement provides evidence of their level of proficiency in mathematical HOTS	Teacher Questionnaire Classroom Observation Pupils' Pre-Achievement	Categorical Ordinal Categorical Ordinal Categorical Ordinal	Between 1 and 5 categories Between 1 and 3 categories Between 1 and 5 categories
Engineering Effective Classroom discussion (Questioning)	PART B: Question 2 (a-g) Questioning	Provides information on the frequency of teacher effectively questioning and use of wait time as well as assess pupils' progress towards instructional goals	Teacher Questionnaire Classroom Observation	Categorical Ordinal	Between 1 and 5 categories

Kivunja Components	Variable	Purpose of the Question	Source of information	Type of Variable	Number Of response categories
Engineering Effective Classroom discussion (Collaboration)	PART C: Question 3 (a-j) Collaboration	Provides information on the frequency of teacher effectively use collaborative techniques to help pupils learning and pupils' ownership of their thinking and work	Teacher Questionnaire Classroom Observation	Categorical Ordinal	Between 1 and 5 categories
Learning tasks	PART D: Question 4 (a-f) Learning tasks in the classroom	Provides information on how a teacher connect to learning goals, clarify learning tasks and use evidence to adjust their instruction as needed both in the moment and in planning for future lessons	Teacher Questionnaire Classroom Observation	Categorical Ordinal	Between 1 and 5 categories
Gain & provide Feedback	PART E: Question 5 (a-i) Feedback	Provides information on teacher assessing pupils' progress during a lesson and design ways to offer individualised feedback	Teacher Questionnaire Classroom Observation	Categorical Ordinal	Between 1 and 5 categories

4.4.2 Classroom observation Instrument

A rising body of literature in educational research studies advise that teaching is the most important school-based determinant of pupils' learning (Hanushek & Rivkin, 2012; Snilstveit et al., 2016). However, the practical identification of effective teaching is not easy and rarely done in practice. More often, when education systems attempt to capture teaching practices, most tools used in low- and middle-income countries fall short on several accounts, as they: (i) measure either the quantity or quality of teaching practices; (ii) do not explicitly focus on teachers' efforts to develop students' socio-emotional skills; (iii) use tools designed for other contexts, which may include irrelevant items or fail to include important ones; and (iv) use tools that are neither evidence-based nor meet essential reliability criteria (Molina et al., 2018, p. 3).

In an attempt to attribute a solution to classroom teaching, Molina et al. (2018) suggested that a classroom observational tool is imperative in response to concerns measuring quality teaching practices and foster the measurement of these teaching practices. Classroom observation is therefore considered as a primarily quantitative technique in which the observer is explicitly counting the frequency and/or intensity of specific behaviours and actions of a particular scene (Guest, Namey & Mitchell, 2013).

The current study too had employed direct classroom observation during the baseline data collection in Phase I as well as in Phase II of this study in mathematics lessons. This classroom observation was considered expedient to the current study as it helped in identifying the bottlenecks in FA delivery, monitor the effectiveness of the practice and focusing efforts to improve teacher practice of HOTS in mathematics. According to Guest et al., (2013, p. 79) "observable behaviour is typically associated with research objectives that require some sort of ordinal data or purely factual description: how often, how many, how intensely, who was there, and the like".

This study adapted a classroom observation rubric from Kanjee (2017), which was then customised to befit classroom observation for FA practice (see *Appendix B*). In classroom observation, the researcher was a non-participant and ensured that his presence in the classroom did not interfere with the teaching and learning process so that data could be collected naturally from the classroom by not coming to the class late, not interacting with pupils, not moving out of the class while lessons were still on and maintaining a normal facial expression (Kanjee & Meshack, 2014, Muucherji & Albon, 2010).

The adapted observational tool was termed Teacher Assessment Practice Observation Schedule (TAPOS). The TAPOS was organised into four features, of which every level attempted to measure the components of FA. Thus:

Beginning of Lesson

(i) The teacher provides a short introduction
(ii) The teacher provides an FA strategy which involves,

- sharing the learning objective and write it on the chalkboard
- determining success criteria for learning objective and write it on the chalkboard,
- choose a mathematical presentation or procedure that is per the FA

During the Lesson,

(i) The teacher provides a task that is focused on enhancing pupils' participation, teacher questioning strategies and interaction which involve problem-solving as well as those questions calling for pupils' HOTS.
(ii) The teacher assesses the pupils' work during the lesson as individuals, pairs and five groups before providing immediate instructional feedback.
(iii) The teacher uses oral feedback, pupils' peer assessment and self-assessment.

At the end of the lesson,

(i) The teacher completes the lesson and determines whether the learning objective has been completed, and check whether assessment criteria have been met.

As part of the classroom observation,

(i) The researcher also reviews the pupils' class exercises with a particular focus on the nature of the task to enhance HOTS, the frequency of such exercises, and the type of feedback that was offered.
(ii) After FA training, the teacher journal reflection was reviewed to determine the implementation of FA.

The behaviours in the TAPOS were characterised either as "Yes" and "No", or "Seen" and "Not Seen" as well as "Often", "Sometimes" and "Not Seen" based on the evidence collected in the lesson observation. These behaviour scores were then translated into a 2-point element scale and a 3-point element scale that quantifies FA practices and teaching of HOTS in mathematics. The series of lesson observations were captured at least twice per teacher in a 30-minute lesson.

4.4.3 Quality criteria for teacher questionnaire using Rasch analysis

The data for the teacher questionnaire and pupils' assessments (discussed in Phase II) were analysed to assess the internal reliability and validity by applying the Rasch statistics framework derived from the Winsteps programme, Windows 7, 64-bit Windows. The Rasch model was developed by George Rasch (1961), a Danish mathematician, to provide interval measures and monitor the adherence of scales to scientific measurement principles (Bond & Fox, 2007; Rasch, 1979; Zhang, Shen & Cannady, 2010). According to Zhang et al. (2010), the Rasch model has sufficient observable statistics for the model parameters and a relatively small sample requirement for parameter estimation hence its use in this study. According to Linacre (2005), the Rasch model is practically the same as the 1PL, except for a few conceptual differences. Rasch analysis is ideal for determining the extent to which items belongs to a single dimension and where items sit within that dimension (Bond & Fox, 2007; Deneen, Brown, Bond & Shroff, 2013).

In summary, the Rasch Measurement Theory (RMT) is a method of analysing response data, in which both the questions on the questionnaire/test and the person taking the test are incorporated into a predictive mathematical model (Karlin & Karlin, 2018). This Rasch process involves converting ordinal data of correct and incorrect responses into interval data for both items and persons. Items are given a difficulty measure, which is a number representing the difficulty of a question, while persons are given a person ability measure, which is a number representing the ability of people in the construct that is being measured (Bond & Fox, 2015; Combrinck 2018).

According to Sadiq, Tirmizi, and Jamil (2015), RMT weighs items based on how many people answered the questions correctly and simultaneously produces difficulty measures for items and persons' ability measures for people. Another important feature of the RMT is the Wright Map, which makes it easier for researchers to improve their instruments. This is accomplished by putting the difficulty level of the items and the ability level of the persons on a shared scale so that the items and persons can easily be compared. Finally, the Rasch measurement model can identify exactly how much multidimensionality is present in an instrument (questionnaire or test), and it is up to the instrument-maker to decide if this amount of multidimensionality is tolerable (Baghaei & Amrahi, 2011; Bond & Fox, 2015; Combrinck 2018; Linacre, 2005; Runnels, 2012;).

The Rasch model is appropriate for many forms of tests and questionnaires which involve dichotomous items marked as right (1) or wrong (0) and polytomous when a single type item with multiple responses is used (Andrich, 1978; Combrinck, 2020; He & Wheadon, 2013; Panayides, Robinson, & Tymms, 2010). According to Bode (2001) if the marks allocated to items vary, then the Partial Credit Model, developed by Masters (1982) is appropriate. For this study, the pupils' tests (discussed in Phase II) and teacher survey instruments involved items that measured more than one score; hence the Partial Credit Model for analysis was found relevant.

Specifically, the Rasch model is a paradigm shift from deterministic models to probabilistic models (Panayides et al., 2010); "ones in which the possible behaviour of a pupil is described by means of a probability that he solves the tasks" (Rasch, 1960, p 11). Rasch's partial-credit model in the Winsteps programme also employs the same property (Linacre, 2016). The logit is defined as *person probability of getting the item correct in a category/person ability* which is equal to *person ability minus Item difficulty* at the intersection of equally probable highest and lowest categories minus *F*, the calibration measurements of the relative category (Bond & Fox, 2015), which is mathematically expressed as

$$Log_e\,(P_{nij}/P_{ni\,(j-1)}) = P_n - D_{ig} - F_{gj}.$$

According to Panayides et al. (2010), the Rasch model is based on three key assumptions: unidimensionality, local independence, and invariance. These three assumptions yielded several criticisms about their appropriateness in using the Rasch measurement approach (Dickson & Kohler, 1996; Goldstein, 1979). However, Andrich (2004) argues that the Rasch Model is fundamentally different from the data-model relationship. All the criticisms of the Rasch utility were drawn from measurement experts in the particular UK at the time only to become orthodox later (Panayides et al., 2010). The major Rasch model assumptions are outlined and discussed below.

The first criticism is the unidimensionality assumption, that the measure should describe a unidimensional construct, that is, a single latent trait and objectively interpretable (Hambleton, 1993; Panayides et al., 2010, Rasch, 1960;). In other words, a test or questionnaire score is validly interpretable only to the extent that it is the variance of the construct under measurement that underlies the variability of all its items. In Rasch analysis, the principal components analysis of the standardised residuals is always more effective at both construct measures and identifying dimensionality than direct factor analysis of the original response-level data (Linacre, 1998; Panayides et al. 2010;). The Rasch measurement model is seen as the “simple and elegant application of IRT” (Gregory, 2007, p. 110) and assumes that all items measure one common trait (unidimensionality) and equally discriminate but differ in difficulty levels (Gregory, 2007; Zhang et al., 2010). Bond and Fox (2007) highlighted unidimensionality, equal item discrimination, and low inclination to guessing as fundamental to the Rasch measurement requirements.

The second assumption is of local independence. Two measures are said to be independent if their scores do not correlate (Linacre, 1996). Two item scores could correlate, but they should correlate locally. Their correlation should be only due to what they share with all the items in the test or instrument and what the test was designed to measure (Andrich, 2004). In the Rasch model, this local independence assumption is attainable through the principal component analysis of the standardised residuals (Linacre, 1998). After the contribution of the measures to the data has been removed, what is left is random, normally distributed noise. If the variance of what they share is removed or held constant, they should not correlate (Inaerm, 2006; Linacre, 1998; Panayides et al., 2010).

The third criterion is the property of invariance. Item invariance stipulates that item (or person) parameters should be independent of the sample (or items) used. According to Panayides et al. (2010), this basic principle of order (or invariance) is an assumption of the Rasch model and the fundamental requirement of measurement. According to Bond and Fox (2007), the Rasch measurement can be used to evaluate possible differential item functioning (DIF), based on the responses of the different groups to specific items and groups of items. Thus, the item invariance criterion can be evaluated using DIF to determine whether item bias is present.

The final criterion is item fit, in other words, whether individual items in a scale fit the Rasch model. There has been a continuing debate around the issue of which is the most appropriate fit statistic to use, what range of fit statistics to be employed when evaluating fit, and how to fit statistics should be interpreted (Karabatsos 2000; Linacre, 2015; Smith, Schumacker & Bush; 1998). However, instead of simply describing the data, the Rasch approach provides the opportunity to understand data by exposure of anomalies, which is the prime function of measurement (Andrich, 2004). Panayides et al. (2010) further explain that the Rasch approach can be used to formalise conditions of invariances, leading to measurement properties. Any data which deviates from the Rasch model also deviates from the requirements of the measurement (Andrich, 2004, Linacre, 1996). When the data does not fit the Rasch model, this is interpreted as an indication that the questionnaire or test does not have the right psychometric properties and hence needs to be revised and improved.

The assumptions mentioned above were applied to the current study to evaluate the quality of the instruments employed. It should be noted that an instrument evaluation involves the use of both quantitative and qualitative analysis in any IRT model of choice. For instance, qualitative analysis for the design and refinement of instruments is recommended and could be crucial for understanding the statistics; for example, misfitting items may have phrasing problems that can be identified through discussion with participants and experts (Combrinck, 2020; Deneen, et al.,2013; Recabarren, Mallinckrodt, & Miles, 2016). For this study discussed both the quantitative and qualitative aspect of evaluating the instrument.

All misfitting items in the baseline questionnaire were removed and not used in the subsequent analysis since they were considered unproductive measurement based on their initial psychometric properties. However, for the pupils' items, misfitting items during the pre-testing were revised and used in post-testing. Table 4.4 shows the summary of Rasch statistics adopted from Combrinck's (2020) table and used to evaluate the teacher survey and pupils' assessment instruments.

Table 4.4: Rasch statistics and interpretation as used in the study

Concept	Interpretation
Item and person data fit model (fit statistics)	Infit Mean Square (MNSQ): 0.5 to 1.5 ideal range, < 0.5 useful but duplicative, 1.5 – 2.0 useful but noisy, >2.0 unexpected responses could be overpowering. Z standardised (ZSTD): Ranges from -2 to 2. Values outside the range considered suspicious. Consider MNSQ first; look only at outlying ZSTD if MNSQ indicates the problem. It is recommended that the INFIT, MNSQ and Outfit Mean Square statistics for all persons and items fall within the accepted range of 0.60 to 1.40 logits and that the acceptable values for infit and outfit values -2 to +2 (p <.05).
Rasch Principal Components (Rasch-PCA) of residuals (Unidimensionality)	The three criteria used to interpret which suggest unidimensionality; (1) At least 50% of the total variance should be explained by the first latent variable/dimension; (2) The first contrast should not have an eigenvalue > 2.0 because an eigenvalue of 2.0 represents the smallest number of items that could represent a second dimension; (3) The ratio of the per cent of raw variance explained by the measures (persons and items) to the per cent of total variance explained in the first contrast should exceed three. (4) If they present a separate construct, analyse those items separately.
Targeting (person/item map)	Numerical values on the extreme left-hand side of the map, which range from -5 to 5, are expressed as log odd unit interval or logit, which is the Rasch Scale's natural unit. Look for an even spread of items and persons which align. Is the mean of the items far from the mean of the person? Also, look at outliners on the map. Well-chosen target tests have item difficulty close to personability.
Local dependence of Items	Pearson correlation coefficient (r); Items should be independent of one another; one item should not lead to an answer of another. Highly correlated items may violate the assumption of local independence. Investigate items where r > .70

Concept	Interpretation
Category and threshold functioning	Graphs: Category probability curves–Categories should be ordered and increase monotonically. A rule of thumb is at least ten observations per category or at least 10% of responses per category. Categories which no one or very few people chose/correctly answered are problematic and could distort data. Disordered thresholds should be investigated and are only a threat when narrow intervals are undesirable.

Adopted from Combrinck (2020, p.13).

Upon successful data cleaning (as described in Section 4.5.2), the data was imported into Winsteps from Excel. The teacher survey baseline data was subjected to quality assurance criteria measures that involved some steps against the Rasch model's assumptions. This study validated the FA (composite and five-scale), and HOTS (composite and three-subscale) scales using the Rasch measurement perspective (Combrinck, 2020; Rasch, 1960). The following considerations apply: (i) response category should balance on the scale, (ii) the items should be unidimensional, (iii) the items should fit the model, as well as vary from easier to harder in their difficulty, the items should be of equal discrimination and items revisions and rejection should be conducted as those items fit the model, (iv) the item and person reliability should be within the desired threshold. The following are the steps involved in the psychometric properties of the teacher formative assessment practice scales:

(i) Step 1: Response category functioning

Rasch analysis was the technique used to analyse the quality characteristics of the teacher survey responses. In the current study, response categories were firstly determined to check if the rating scale was being used in an intended way, by Linacre’s (2002) proposed guidelines. For this study, 50% of the total items did not reach 10% of the observation per category as suggested by the guidelines (Linacre, 2002). As outlined in Appendix Ci and iv, the infit and outfit mean square statistics of some response categories were more than 2.0. Wright Maps with Thurstonian and Item and Multiple Item Characteristics Curves (ICCs) analysis were also further employed for both FA and HOTS items to understand how the raw data was balanced on the scale (See Appendix C ii & iii and Appendix D ii & iii). To get more stable item difficulty estimates, the researcher, therefore, collapsed the categories to a dichotomous and reduced three-polytomous format respectively, as shown in Table 4.5, Appendix C iv and D iv, respectively.

Table 4.5: Categories collapsed to the reduced numbers

Variable	Initial Code	Recoded scale	Description
FA	Not All = 0 Once a Week = 1 2-3 Time per Week = 2 4 Times a Week = 3	The scale values of 0-3 were all recoded to 1	Not Every Day
	Always = 4	The scale value 4, was recoded to 2	Every Day of the Week
HOTS	Strongly Disagree = 1 Disagree = 2	The scale values of 1-2 were all recoded to 1	Disagree
	Agree= 3	The scale value 3, was recoded to 2	Agree
	Strongly Agree = 4	The scale value 4, was recoded to 3	Strongly

As can be seen in Table 4.5, the FA variable scale was collapsed in the dichotomous format as “Not Every Day” (1), and “Every Day of the Week” (2), and these two response categories imitated the Rasch framework (see Appendix C iv).

Similarly, the section (Section F) of the questionnaire, which captured the teachers' behaviour on the use of FA in teaching HOTS in mathematics on a 4-point Likert scale (Strongly Disagree, through Strongly Agree) was also revised. The indiscernible response categories were identified for “Strongly Disagree” (1) and “Disagree” (2). These two categories were also collapsed to form one response category and named “Disagree” (1), while “Agree” (3) and “Strongly Agree” (4) were recoded to 2 – “Agree” and 3 – “Strongly Agree” respectively (see Table 4.5 and Appendix D iv).

Additionally, the scales for FA and HOTS were considered differently, and this distinction was justified through dimensionality analysis of the instrument, which seemed to be multidimensional when analysed on a single scale. To this end, the instrument was split into two subscales, so that FA items and HOTS were handled separately.

(ii) Step 2: Investigated Model Fit of the instrument

This study used the Winsteps 3.75 programme, which provides in-fit and outfits fit mean squares statistics (MNSQ) and standardized fit statistics (ZSDT) (Linacre, 2016). The programme reveals where items and persons fit and where there is a misfit. All statistics are reported in terms of log-odds units and have a range of -5.00 to +5.00 with a mean set at 0.00 and a standard deviation of 1.00 (Bond & Fox, 2015; Boone et al., 2014; Boone & Noltemeyer, 2017; Linacre, 2016). According to Sharif, Hanapi, Nashir, Kob and Abdullah (2019), if the infit or outfit value more than 1.40 logit, then it indicates a potentially confusing/problematic item. If the MNSQ value is less than 0.60 logit, it shows it was too easily anticipated by the respondents (Linacre, 2007).

For this study, the FA (composite scale and five subscales) and HOTS (composite scale and three subscales) were explored for model fit using outfits fit mean squares statistics (MNSQ) and standardized fit statistics (ZSDT) as shown in Table 4.6 and Table 4.7 respectively. Mean square (MNSQ) values outside the range of 0.6 to 1.40 and ZSTD scores outside the +2 to −2 range were identified as potential misfitting items. Table 4.6 shows the FA composite scale, in which eight items were found to exhibit RMM misfit, of which their Outfit MNSQ scores were above 1.40 and Outfit ZSTD 2.7 (> 2.0) as well as Infit MNSQ above 1.18 and infit ZSTD.1.57.

Specifically, *Appendix E* shows item statistics in measure order with the most difficult to endorse items at the top along with the degree item fit to the single-parameter Rasch Model (Deneen, et al., 2013). Thus six items for learning goal and success criteria (that is *LGSC_a; I provide my pupils with learning objectives, LGSC d; I connect each lesson to future learning that will take place, LGSC_j; I discuss with pupils what are they should know by the end of the lesson, LGSC_k; The learning goal(s) for the lesson is connected to country academic standards, LGSC_h; I present the learning goal(s) for the lesson to pupils verbally and LGSC_i; I present the learning goal(s) for the lesson to pupils in writing -e.g., on the board*) were revealed consistently high misfit to the underlying single dimension associated with the remaining items for formative assessment. The correlations for these six LGSC were very weaker and ranged from .28 - .36 than all items fitting the Rasch model.

The six rejected items for LGSC were inspected to identify any cohesion traits, which may have attributed to their different dimension; thus *LGSC_d* and *LGSC_k* was found relatively difficulty to endorse items and such items were related to the extent to which teachers were connecting each lesson either for future learning or reference to the country's academic standards. While the other four rejected formative assessment items (LGSC_a LGSC_h, LGSC_i and LGSC_J) seemed to be the easiest to endorse items and these items were related to the teachers' ability to sharing of learning objectives (See *Appendix E)*.

One item for engineering effective classroom discussion through questioning (*EEQ_f; I use exit ticket to assess pupils in discussion*) was also revealed misfit and rejected from the main single dimension and its correlation was also weak (*r = .32)*. The misfit here is associated highly with difficulty index and thus it was the most difficult item to endorse (that is *b-value*= *2.50*). This might be preliminary evidence that teachers were unfamiliar with some questioning techniques within the formative assessment framework. Lastly, one item for gain and provide feedback (*GPF_b; I review some pupils working during the lesson*) also had an outfit value and was rejected; misfit for the item was likely due to a problem with the phrasing (See *Appendix E)*.

The Rasch analysis results in dropping items or at least significantly modifying items if such items had shown a misfit of the model (Deneen, et al., 2013). For this reason, the current study had dropped all eight items which were identified as a threat to the next level of analysis since they were considered potentially problematic and noisy (Boone et al., 2014). Henceforth, all misfitting items were removed from further analysis and suggested for refinement in the future study.

Table 4.6: Questionnaire constructs on a FA composite and five subscales (5-FA Subscales) Psychometric Evidence in RMM

Parameter	**RMM requirements**	**Formative Assessment (48 items)**	**Formative Assessment (40 items)**	**LGSC- FA Subscale (10 items)**	**EEQ - FA Subscale (6 items)**	**EEC – FA Subscale (10 items)**	**LT- FA Subscale (6 items)**	**GPF – FA Subscale (8 items)**
Model Fit: Summary of Items								
Item mean (SD) logits	0.00	0.00(0.20)	0.00(0.20)	0.00(0.22)	0.00(0.23)	0.00(0.22)	0.00(0.23)	0.00(0.22)
Item Reliability	> .8	.95	.95	.94	.95	.96	.95	.94
Item Separation Index	> 3.0	4.48	4.49	3.88	4.32	4.71	4.32	4.12
Item Model infit MNSQ Range Extremes	0.60 – 1.40	.81 -1.30	-.96 – 1.17	.85 – 1.13	.93 – 1.18	.87 – 1.06	0.93 - 1.18	.88 – 1.05
Item Model Infit ZSTD Range Extremes	-2.0 – 2.0	-2.11 – 3.35	.21 – 1.97	-1.72– .82	-.75–1.57	-1.51 - 0.44	-0.75 – 1.57	-1.28 – 0.51
Item Model Outfit MNSQ Range Extremes	.60 – 1.40	.68 – 1.84	.80 – 1.44	.84 – 1.37	.91 -1.40	.81 – 1.27	.91 – 1.40	.91 – 1.21
Misfitting Item		LGSC_a, d, k, h, i, j; GPF_b EEQ_f	0	0	0	0	0	0
Dimensionality								
Variance accounted for the by 1st Factor	>50 %	31.2%	33.3%	30.8%	31.9%	34.1%	31.9%	30.7%
PCA (Eigenvalue for the 1st)	≤ 2.0	3.86	3.12	1.91	1.53	1.93	1.53	1.71
Local Independence		.60 -.73	.50 - .70	.31 - .37	.36 - .23	.16 - .35	.24 - .36	.19 -.36

Person: Measurement of quality	**RMM requirements**	**Formative Assessment (48 items)**	**Formative Assessment (40 items)**	**LGSC- FA Subscale (10 items)**	**EEQ - FA Subscale (6 items)**	**EEC – FA Subscale (10 items)**	**LT- FA Subscale (6 items)**	**GPF – FA Subscale (8 items)**
Person Mean (SD) Logits		0.28(0.37)	0.21(0.49)	-0.05(0.82)	0.02(1.00)	0.14(0.80)	0.02(1.00)	0.00(0.87)
Person Reliability	>.80 for individual measurement	.91	.90	.50	.13	.56	.13	.40
Person Separation Index	< 2.0	3.16	2.97	1.00	.23	1.12	0.39	.82
Standard Error Measurement	SEM as low as possible	2.86	2.97	1.22	0.99	1.26	0.99	1.16
Person Raw Score reliability	>0.8 for individual measurement	.94	.94	.79	.71	.81	0.71	.76
Global Statistics		χ^2(df=7145) = 7143.34, p= .5033	χ^2 (df=5522) = 5511.47, p= .5374	χ^2(df=1146) = 1142.72, p = .5217	χ^2(df=699) = 696.7479, p = .5169	χ^2 (df=1243) 1236.43, p= .5472	χ^2(df=703) = 696.75, p= .5594	χ^2(df=1004) = 1007.55, p = .4625

Table 4.7: Questionnaire constructs on HOTS composite and three subscales (3-HOTS Subscales) Psychometric Evidence in RMM

Parameter	RMM requirements	HOTS Total (15 items)	HOTS_TrPractice (7 items)	HOTS_CSupport (2 items)	HOTS_EXSpuport (6 items)
Model Fit: Summary of Items					
Item mean (SD) logits	0.00	0.00(0.17)	0.00(0.18)	0.00(0.29)	0.00(0.18)
Item Reliability	> .8	.98	.95	.97	.94
Item Separation Index	> 3.0	6.97	4.30	5.66	4.02
Item Model infit MNSQ Range Extremes	.60 – 1.40	.77-1.21	.80 – 1.12	.98 – 1.01	.80 – 1.22
Item Model Infit ZSTD Range Extremes	-2.0 – 2.0	-2.23 – 1.72	-1.84 – 1.07	-.03 – .10	-1.83 – 1.78
Item Model Outfit MNSQ Range Extremes	.60 – 1.40	.72 – 1.30	.76 – 1.00	.97 – 1.00	.79 – 1.10
Misfitting Item	0	0	0	0	0
Dimensionality					
Variance accounted for the by 1st Factor	>50 %	52.7%	48.5%	62.8%	47.5%
PCA (Eigenvalue for the 1st)	≤ 2.0	2.82	1.84	.0002	1.71
Local Independence	Below <.70	.65 - .74	.56 - .61	.35 -1.00	.32 - .56
Person: Measurement of quality	**RMM requirements**	**HOTS Total (15 items)**	**HOTS_TrPractice (7 items)**	**HOTS_CSupport (2 items)**	**HOTS_EXSpuport (6 items)**
Person Mean (SD) Logits		-0.02(0.53)	-0.03(0.80)	-0.07(2.25)	0.10(0.86)
Person Reliability	>.80 for individual measurement	.87	.75	.00	.77
Person Separation Index	< 2.0	2.60	1.77	.00	1.56

Standard Error Measurement	SEM as low as possible	1.99	1.35	.63	1.25
Person Raw Score reliability	> .8 for individual measurement	.90	.84	.68	.81
Global Statistics		χ^2 (df=2985) = 2978.6935, p= .5291	χ^2(df=*1320*) = *1312.333, p* = *.5543*	χ^2 (df=171) = 167.2981, p = .5645	χ^2(df=1111) 1107.4816, p= .5242

The remaining items for the FA (40 items) and five FA subscales met the RMM fit requirements with Infit MNSQ scores ranging from 0.85 to 1.18, Infit ZSTD scores ranging from −1.72 to 1.57, Outfit MNSQ scores ranging from 0.84 to 1.40 (see Table 4.6). For HOTS composite and three subscale items had fitted the model (See Table 4.7). The investigation on the global statistics reviewed that the data on both composite and subscales for FA and HOTS have met the condition (See Table 4.6 and Table 4.7).

(iii) Step 3: Establish the unidimensionality and Local Independence

As for the internal validity of the instrument, the Rasch modelling term, *unidimensional* means that all of the non-random variance found in the data can be accounted for by a single dimension of difficulty and ability (Bond & Fox, 2015). As in the current study, the questionnaire's unidimensionality was checked through Principal Component Analysis (Rasch-PCA) of residuals in Winsteps software (Linacre, 2016).

The FA composite scale explained 31.20% of the variance and the HOTS composite subscale explained 52% of the variance, with eigenvalues of 3.86 and 2.82 respectively (See Table 4.6 and 4.7). These eigenvalues for the first contrast suggest greater than (2.0) as the expected threshold (Bond X Fox, 2015), hence the subscale for each domain was further explored. For the five FA sub-scales, the percentage of variance accounted for by the first factor ranged from 30.7% to 34.1% with eigenvalues ranging from 1.53 to 1.93 (see Table 4.5). The overall dimensionality of the FA composite scale (40 items) was considered multidimensional since its eigenvalue for the first contrast of 2.82 was more than the recommended threshold of less than two (Bond & Fox, 2015). As for the HOTS composite scale, the percentage of variance accounted for by the first factor ranged from 48.5 % to 62.8% with eigenvalues ranging from 0.0002 to 1.84 (See Table 4.6), hence overall combined, it was also considered unidimensional.

The measure of local independence for items through correlations of the FA composite scale ranged between 0.16 to .37, and according to Winsteps criteria, these are well below .70 (Linacre, 2016), (See Table 4.5). There is evidence of local independence of items, and the invariance of the instrument is not threatened. Similarly, the HOTS composite subscale ranged between -1.00 to 0.61 and met the local independence of items (See Table 4.6).

(iv) Step 4: Investigated internal reliability of the instrument with Rasch Models

Upon completion of the model fit, a reliability analysis was done via the Rasch technique. A reliability index within Rasch the approach, which is commonly known as person separation and reliability, were investigated. Out of 48 items for formative strategies eight were removed due to internal inconsistency of those items, while 15 items for the integration of FA and HOTS were not affected since all the items fit the model.

As shown in Table 4.6, the item reliability values are .95 for the FA composite scale and .98 for the HOTS composite scale, which indicates high levels of consistency (Bond & Fox, 2007). For the FA composite scale, the item separation index was 4.49, and the item mean was 0.0 logits (SD = 0.45) (see Table 4.6). For the FA composite scale, the person separation index was 3.6, and the person mean was .28 (SD = 0.37) (see Table 4.6). The HOTS composite scale revealed that the item separation index was 6.97, and the item mean was 0.0 logits (SD = 0.15) (see Table 4.7). The person separation index was 2.60, and the person mean was -0.02 (SD = 0.53) for the HOTS composite scale.

For the five FA subscales, the item separation indices ranged from 3.88 to 4.71, item reliability ranged from .94 to .96 and the item mean ranged from -.05 (SD = 0.82) to 0.00 logits (SD = 1.0) (see Table 4.6). For the three HOTS subscales, the item separation indices ranged from 4.02 to 5.66, item reliability ranged from 0.94 to 0.97 and the item mean ranged from 0.00 (SD = 0.17) to 0.0 logits (SD = 0.29) (see Table 4.7). As can be seen from Table 4.6 and 4.7, the person separation for both scales (composite and subscales) suggest that the sample was differentiated enough to distinguish between higher and low ability levels.

The Pearson Raw score reliability (deemed equivalent to Cronbach's Alpha coefficient) for the FA composite (.94) and HOTS composite (.94) were reported. The classroom discussion (collaboration) subscale had strong reliability coefficients at .844. The remaining five FA subscales were also within an acceptable range of reliability indicators namely: learning tasks alpha .71 engineering effective classroom discussion (questioning) alpha. 71, learning goal and success criteria alpha .79 and gaining feedback provision alpha .76 (See Table 4.6). As for the HOTS, person raw score reliability for the composite scale was .90 and the three subscales ranged from .68 to .84 (See Table 4.7). The reliability analysis scores for the survey items are good and acceptable, as judged by Bond and Fox (2007) that a reliability value that exceeds .70 is acceptable.

(v) Step 5: Summary of the Psychometrics properties of the survey instrument

The 48 FA and 15 HOTS items were answered by 150 teachers. The RMM fit, dimensionality, as well as person and item reliability, were examined. The instrument evaluation involved an FA composite scale, five FA subscales, a HOTS composite and three HOTS subscales. Eight FA items did not meet the RMM fit requirements, while the remaining FA composite scale (40 items) and five subscale items met the requirements. The model fit test for the HOTS composite and three subscales met the RRM Infit and Outfit MNSQ requirements.

The dimensionality of the FA composite scale and its five subscales were determined via PCA of the residuals within Winsteps. For the FA composite, the first factor's percentage variance was 33.3%, with an eigenvalue of 3.12 (after removing eight misfitting items). The eigenvalue was still more than the acceptable value (<2.0); hence the FA items were not unidimensional and were treated as individual constructs. Thus, five FA subscales were considered multidimensional. The HOTS composite scale for the first factor variance accounted for 52%, with an eigenvalue of 2.82. The recommended values for the percentage of variance accounted for by the first factor is > 50%. Even though the eigenvalue is slightly greater than 2, the percentage variance accounted for indicates strong dimensionality (52.7% variance accounted for by the first factor) for the HOTS composite scale to be considered unidimensional. Moreover, for three HOTS subscales, the curriculum support subscale had the highest percentage of variance accounted for by the first factor at 62.8% with an eigenvalue of 0.002. While HOTS items for the teacher practice scale explained 48.5% variance with an eigenvalue of 1.84, HOTS for the external support subscale accounted for 47.5% variance with an eigenvalue of 1.71. In summary, the HOTS composite and three subscales have met RRM requirements for unidimensionality. Since FA subscales and the HOTS scale exhibited their own separate constructs for dimensionality, the questionnaire was considered multidimensional.

Person reliability results for all the subscales for FA and HOTS were less than optimal. The low person reliability scores can be partially explained by the number of items per FA and HOTS scale (a minimum of two to a maximum of 10 items) (Marsh et al., 2010). In addition, Linacre (2015) explains that Rasch usually underestimates reliability, while Cronbach alpha overestimates it, as seen from the person raw scores (both scales) highly computed by Winsteps. Item reliability was found to be higher above the optimal in both composites and subscales (FA and HOTS items) which indicates that item ordering is very reliable.

4.4.4 Quality criteria for classroom observation instrument

The current study adapted the observation instrument from Kanjee (2017) as described in Section 4.4.2. This observational instrument was developed by experts in the educational assessment field. These are researchers with extensive experience in implementing formative assessment internationally, numeracy specialists and academics based in South Africa and Tanzania who are familiar with local conditions sponsored by the University of Oxford. The tools were used in their research work over three years, beginning in April 2016 (McGrane et al., 2018). This study posed a similar context to employ the instrument, hence the instrument was considered valid. The observation tool was also aligned with the survey questionnaire variables in systemically capturing the FA strategies in the classroom.

4.5 Procedure

The data collection procedure and processing are part of the research methods grand plan for carrying out research exercise (Nenty, 2013). In this survey phase, the researcher, therefore, made it clear and an elaborate description is provided in the next section on the data procedures and preparation involved to ensure a high rate of return and allowed the systemic entrance of data respectively;

4.5.1 Data collection procedure

The researcher delivered the questionnaires directly to schools and collected those questionnaires upon completion in three sub-regions which were within a radius of 50 kilometres from the regional headquarters in Kanye village. For the other two sub-regions, postal mails were used to distribute the questionnaire since those regions were over 100 kilometres from the regional headquarters and the schools are sparsely distributed over the geographical area. This practice was consistent with the ideas of Nenty (2013) that in a wider geographical area, a questionnaire could be sent anywhere by mail, hence every respondent could be reached with relatively little cost.

Before the postal mails were sent to the schools, the researcher telephoned the school heads for all the sampled schools in that two sub-regions respectively. The purpose of the phone calls was to explain the protocol of the study which included the purpose of study, a random selection of schools, targeted Standard 4 teachers, and gate-keeper letters from the Southern regional director's office and research ethics from the Faculty of Education at the University of Pretoria. The content of the mail also included an empty official stamped envelope meant to facilitate the return of the completed questionnaire to the researcher. Follow-ups were made over the phone, to remind the schools to complete and send back the questionnaires.

As for the observation, the researcher arranged appointments with nine purposively random selected schools in which the surveys were done. During the observation visit, the researcher was non-participatory, as explained in the section for the classroom observation tool (Section 4.4.2). The selected teachers for observation were observed at least one per teacher during the baseline survey. All the observed teachers were audiotaped for review and easy reference at a later stage of data analysis.

4.5.2 Data management

The researcher, under the guidance of the research supervisor, established the codebook for the data (questionnaire and observation instruments) and then the researcher entered all the questionnaire data directly into Statistical Package for Social Science (SPSS) version 26 using the correct codes. Accuracy during data entries was met through pre-analysis of 10% of captured data to check for consistency against paper tests (randomly selected), and the error rate below 1%, was deemed to be acceptable (Combrinck, 2018). Cleaning of the data was done by the researcher and assisted by one of the supervisors. Thereafter the questionnaire data would be interpreted based on Rasch logits and all the pre-analysis was done as described in Section (i) and (ii) below.

(i) The Rasch logits scale

The Rasch model converts the item and person estimates to the natural logarithm, a conversion that results in an interval scale (Wright & Stone, 1999). The logits scale ranges from negative infinity to positive infinity. However, most logits scale fall between -5 to +5 on the scale. Rasch logits scales are set to a mean of 0 and a standard deviation of 1. The logits scale may be difficult for people to interpret; however, the current study had found reporting-based on logits/theta as very appropriate to enhance the accuracy in interpretation of the findings (Bond & Fox, 2007; Boone, 2016; Linacre, 2005). This study used two different scales for FA and HOTS items respectively. Thus, the teacher questionnaire used a 2-type rating scale consisting of 40 FA items and a 3-type scale for 15 HOTS items after collapsing some response categories. The use of Rasch logits had accorded an opportunity for comparison on a common scale (Boone, 2016; Linacre, 2005).

(ii) Check for Normality

Having assessed the internal reliability and validity by applying the Rasch statistics framework derived from the Winsteps program described in the previous Section 4.4.3, (Quality Criteria for Teacher Questionnaire), the next step was checking for normality. A test of normality was computed to establish whether the distribution of teachers' scores from the scale deviates from a comparable normal distribution. To this end inputting data and provisional analysis of the variables can be investigated using visual aids (probability plots; normal Q-Q plots) and analytical methods (Kolmogorov-Smirnov/Shapiro-Wilk test) to determine whether the data showed a normal distribution (Field, 2013). A normality check allows the researcher to choose the statistical type to use, either a parametric or non-parametric approach.

The current study explored a normality check through the Kolmogorov-Smirnov test and the Shapiro-Wilk test (K-S) in all six the variables using SPSS. As can be seen from Table 4.8, the K-S test for the FA Strategies and HOTS scores, learning goal and success criteria (LGSC), D (149), = 0.105, *p* =0.000, engineering effective classroom (collaboration, EEC) D (149),= 0.108, *p* = 0.000, engineering effective classroom (Questioning, EEQ), D(149),= 0.132, *p* = 0.001, learning task (LT) D(149),= 0.132, *p* < 0.001, gain and provide feedback (GPF) D(149),= 0.107, *p* = 0.000, and FA and HOTS D(149),= 0.56, *p* < 0.000 were all significantly non-normal. Since all the tests were significant (*p* < .05) and the results illustrate that the distribution for teachers' responses were significantly different from a normal distribution (Field, 2013), the data violated the parametric assumptions.

Table 4.8: Analysis for normality testing of the formative assessment variables as used

Tests of Normality (N=149)						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Engineering effective classroom (collaboration)	.108	149	.000	.955	150	.000
Engineering effective classroom (Questioning)	.132	149	.000	.935	150	.000
Learning goal and success criteria	.105	149	.000	.954	150	.000
Gain and provide feedback	.107	149	.000	.950	150	.000
Learning task	.132	149	.000	.935	150	.000
Higher Order Thinking Skills Tasks	.056	149	.000	.985	150	.107

a. Lilliefors Significance Correction

4.6 Analysis

4.6.1 Descriptive statistics

The Rasch Statistics framework (Logits) derived from the Winsteps program provided a forum for descriptive statistical analysis. Specifically, for this study, the Wright Map and item statistics (measure order) were used to demonstrate the distribution of the teachers' ability and item difficulty on the same logit scale. Lysaght, O'Leary and Ludlow (2017) indicated that the persons' ability is listed on the left side of the map while the item difficulty is on the right side of the map. The higher logits represent the person with a higher ability (left side) and more difficult items (right side) and vice versa.

This study explores the respondents' endorsement towards FA practices to understand their classroom practice better. In this case, a Wright map allows researchers to evaluate how well the items define a variable. The Wright Map also enables researchers to compare the predicted order of the item difficulty (Boone, 2016). Therefore, the current study finds it imperative to explore the Wright map in conjunction with theta (logits) analysis mechanism to determine the teachers' level of FA practices to enhance HOTS in their classroom (Boone, 2016).

4.6.2 Inferential statistics

Given the sample size's nature in Phase I of this study, the data was not normally distributed as attested by normality check in the preliminary analysis. Hence, non-parametric statistics were considered appropriate for the analysis of the data. The data violated the parametric tests' assumptions, among others, the level of measurement (interval or ratio), normal distribution and homogeneity of variance across the groups (Field, 2013).

The normality test's significance permitted Friedman testing to determine whether the teachers' responses on FA strategies and integration of HOTS in teaching mathematics varied by comparing the logits on the common scale. This non-parametric statistical analysis was consistent with the explanation by Field (2013), who advanced Friedman's test as a useful test of the difference between conditions when there are more than two conditions, the same entities have provided scores in all the conditions, and the scores are not normally distributed. Logit or theta for participating teachers' measures from the six FA strategies on the common Rasch scale was subjected to the significance testing. According to Field (2013), Friedman used the test statistics (X^2), degrees of freedom and its significance (0.05 set for this study).

In the case of statistical significance for the Friedman test, a post hoc test is done to examine where the difference occurs, separated by a Wilcoxon signed-rank test. The different combinations of related groups were run through the test. Following the Friedman test, a series of Wilcoxon tests which were extracted from the z scores were used to compare all of the groups (from the six FA strategies on the scale) to determine the magnitude of group differences and to determine the effect size (Field, 2013).

The Wilcoxon signed-rank test was found appropriate for this study as it was used in situations in which there are two sets of scores to compare, and such scores are matched with individuals just once (Field, 2018; Huck, 2014). Additionally, on occasion, like in the case of the current study, a parametric test cannot be used because the values are ordinal or the t-test assumptions are untenable. This study catered for these situations.

Next, the effect size was reviewed as a standardised measure of the size of the effect observed in the study, which can be compared to other studies. According to Field (2018), the effect size is useful since it provides an objective measure of the importance of an effect. A large effect size is a measure of how significant a difference is, while a small effect size suggests less important differences. Cohen (1992) made some broad statements about what constitutes a large or small effect size, as shown in Table 4.9 (below).

Table 4.9: Guideline to interpret effect size as used in this study

Effect size	Magnitude	Interpretation
***r* = 0.10**	Small effect	The effect explains 1 % of the total variance
***r* = 0.30**	Medium effect	The effect accounts for 9 % of the total variance
***r* = 0.50**	Large effect	The effect accounts for 25 % of the total variance

Source: Field (2005)

For this study's effect size, the researcher used these guidelines to assess the effect, and it was calculated based on Field's (2013) suggestion to use SPSS as it converts the test statistics into a *z*-score. The equation to convert a z-score into the effect size estimate, *r*, is as follows (from Rosenthal, 1991, p.19 as cited in Field, 2013):

$r = Z/\sqrt{N}$,

in which Z is the z-score that SPSS produces and *N* is the size of the study (i.e. the number of total observations) on which *z* is based. Since the current study also involved an intervention, it would therefore also apply. The r values were converted to *d* based on the formulae below, and the results were reported based on *d* to fit the notion proposed by Hattie (2013). Thus Hattie's (2013) interpretation of effect size, any intervention higher than the average effect (*d* = 0.40) may be worth implementing as it is beyond normal growth and teaching effects.

Conversions between *d* and *OR* or *r* is done through these formulae.

- $d = \frac{2*r}{\sqrt{1-r^2}}$
- $r = \frac{d}{\sqrt{d^2+4}}$
- $d = \frac{\log(OR)\times\sqrt{3}}{\pi}$
- $log(OR) = d * \frac{\pi}{\sqrt{(3)}}$

4.6.3 Classroom observation

Descriptive statistics were obtained for each of the three elements of the lesson observation that were analysed as per individuals in a given school. The analysis produced the frequency and content analysis of teachers' overall practice to identify who had exhibited the best FA practices. The classroom observation data was computed using Microsoft Excel.

4.7 Findings

This chapter aims to present the findings stemming from data collected in phase 1 through the teacher questionnaire and classroom observation.

4.7.1 Participating teachers in Phase I of the study

This study explored the teachers' use of the FA practice when teaching HOTS in primary school mathematics. Pupils' mathematics underachievement seems to be a challenge across the country, and the pattern is cascading from lower to upper primary schools. Therefore, the current study employed a sample of primary school teachers teaching Standard 4 pupils (lower primary school). As explained in Chapter 3, 150 participants responded to the questionnaire, which included 107 (70.1%) respondents who had attained a diploma, followed by 31 (20.7%) participants with a degree qualification (See Section 4.4).

Nine teachers were randomly selected for pre-classroom observation during Phase I in the Kanye sub-region. All nine participating teachers were females, and their teaching qualifications included Primary School Teaching Certificates (2), Diplomas in Primary Education (4), and a Bachelor of Education (1). All the participating teachers did not specialise in mathematics. Instead, they were specialists in languages, general subjects, and there was a one-degree holder with a physical education specialisation. The participating teachers were very experienced in the teaching profession with their experience ranging from 12 to 27 years. Despite such teaching experience, the participating teachers only taught Standard 4 for a few years, ranging from only two to five years. The class size was generally considered large by international standards (UNESCO, 2010) and the teacher-pupil ratio ranged from 27 to 39 pupils per teacher.

4.7.2 Participating teachers' endorsement of formative assessment and HOTS practices

The teachers' FA practice and HOTS integration were reported using logits to indicate their endorsement of classroom practice strategies when teaching mathematics. Table 4.10 shows the mean, standard deviation, valid N and measure (Rasch logits) on the interval scale created for each of the constructs.

Table 4.10: Person measure statistics for formative assessment practices analysis

	N	Min	Max	Mean	Std. Dev	Measure
Engineering Classroom Collaboration (ECC)	150	1.00	2.00	1.48	.30	.30
Learning Tasks (LT)	150	1.00	2.00	1.54	.31	.29
Engineering Classroom Questioning (EEQ)	150	1.00	2.00	1.62	.28	.29
Gaining & Providing Feedback (GPF)	150	1.00	2.00	1.49	.33	.10
Learning Goal & Success Criteria (LGSC)	150	1.00	2.00	1.54	.28	.09
Higher Order Thinking Skills Tasks (HOTS)	150	1.00	3.00	1.98	.43	.02

Based on descriptive statistics summarised in Table 4.10 above, participating teachers endorsed their classroom practice levels, indicating the easiest statement to agree with the set of statements was HOTS (.02 logit), although this probably did not reflect the reality in the classroom, followed by the use of learning goal and success criteria (LGSC). The areas related to the engineering of effective classroom discussion (questioning) and learning tasks were all rated relatively lower. The engineering of effective classroom discussion through collaboration (.30 logit) was the most difficult FA domain among the participating teachers. Thus, the self-reported results reflect the extent to which teachers had employed the strategies in their classroom. The following section further explores teachers' FA and HOTS practices based on a Wright Map and items logits.

4.7.3 Participating teachers' endorsement of formative assessment and HOTS practices in Wright Rasch Map

For this phase of the study, a variable map was used to demonstrate the distribution of the teachers' agreement and item endorsement on the same logit scale. The variable map for formative assessment (Figure 4.2) and integration of FA and HOTS (Figure 4.3) are discussed separately in this section to make a more detailed analysis through a visual summary of the teachers' agreement and item endorsement as they arise in the survey data (n=150). The level of teacher agreement is listed on the left side of the map, while the item endorsement is on the map's right side. The more difficult items' endorsement (or difficult as agreed with teachers) extent toward the top of the item map, and the least difficult items endorsement (easy as agreed with teachers) are at the bottom. "M" marks the person and item mean logit, "S" is one standard deviation away from the mean logit, and "T" is two standard deviations away from the mean logit (Boone et al., 2014).

Figure 4.2 illustrates the spread of FA strategies across a scale, as one starting at the bottom of the map in which items were easily endorsed by many teachers and indicating the likelihood to be employed routinely in the mathematics classes. The person measure had a mean of 0.21 logit. This positive value implies that the 40 items for FA were not quite challenging for most participating teachers, since the mean value for items, m= 0.00 logits, is just slightly below the person mean logit. Thus, the person measure mean logit suggests teachers' assertion of practising FA strategies in their classroom during teaching and learning.

```
INPUT: 150 PERSON  40 ITEM REPORTED: 150 PERSON  40 ITEM  78 CATS WINSTEPS
--------------------------------------------------------------------------------------------------
MEASURE     PERSON - MAP - ITEM
              <more>|<rare>
   5             ##  +
                     |
                     |
                     |
                 .#  |
   4                 +
                     |
                     |
                  #  |
                     |
   3                T+
                  .  |
                     |
                  #  |  LT4_PupilsUnderstandGoals (LT-d)
                     |
                 .#  |
                     |
   2             .#  +T
               .###  |
                  # S|  GP5_SelfAssessment (GPF-e)
                     |  EC1_SmallGroupWork (EEC-a)
                 ##  |  LG7 Learning goal (LGSC-g)
                .##  |  GP8_PeerAssessFutureInform (GPF-h)
               .###  |  EQ7_AdaptFutureLesssons (EEQ-g)
                        GP7_SelfAssessFutureInform (GPF -g)
   1            ###  +S EC3_PupilsGuideOwnLearning (EEC -c)
                        LG14_PupilsSetGoals (LGSC-n)
                .##  |  LG12_ReferenceGoals (LGSC-l)
                ###  |  GP4_InternalizeFeedback (GPF-d)
                        LG5_CoherentSeq (LGSC-e)
                ###  |  EC5_FacilitateClassDiscuss (EEC-e)
                 .#  |  EC9_MultipleViewPoints (EEC-i)
                        LG15_MultipleOptions (LGSC-o)
                 .# M|  EC2_PartnerWork (EEC-b)
                        EC8_RightAnswerExpected (EEC-h)
                        LG16_PupilsDemonstrateGoals (LGSC-p)
                        LT3_HalfLearnersClearGoals (LT-c)
   0            .##  +M LG6_ConnectionsNew (LGSC-f)
                        LT5_ResponsesAdaptFeedback (LT-d)
                .##  |  EC4_EnableDiscussion (EEC-d)
                        GP9_FeedbackLoops (GPF-i)
                        GP3_RealTimeFeedback (GPF -c)
                        LG13_ShareCriteria (LGSC-m)
            #######  |  EQ6_ExitTickets (EEQ -f)
                        LG2_MeaninigObjectives (LGSC-b)
               .###  |  EC7_EnableGroupWork (EEC-g)
                        LT6_AnalyseResponses (LT-f)
                 ##  |  EQ4_PaceLessson (EEQ -d)
               .###  |
                  #  |  EQ3_AdjustLessonPupilNeeds (EEQ -c)
                        GP1_ReviewSomeWork (GPF-a)
  -1            ###  +S LT2_TasksProgress (LT-b)
                  .  |  EC10_HighExpectations (EEC-j)
                        EQ5_FollowUpQuestions (EEQ-e)
                ### S|  GP2_ReviewDuringLesson (GPF-b)
                 .#  |  LT1_ActivitiesTiedLearningGoal (LT-a)
                     |
                ###  |
                  #  |  EQ2_AskQuestionsPupils (EEQ-b)
                        LG3_ConnectLessons (LGSC-c)
  -2              .  +T EQ1_AskQuestionsLesson (EEQ-a)
                     |
                     |
                  . T|
                     |
                 .#  |
  -3                 +
                     |
                     |
  -4             ##  +
              <less>|<freq>
```

Figure 4.2: A Wright Map of the 150 respondents who responded to the 40 formative assessment items

By observing the items on the right of the map and noting that the item logits above the mean value (0.00 logit) represent the more difficult items, it can be said that the 40 items within the five FA subscales can be categorised as easy or difficult, based on their location on the map. Each domain is discussed and described separately.

(i) Engineering Effective Classroom Discussions through Collaboration (EEC)

Table 4.11 illustrates the rate at which the teachers were endorsing engineering effective classroom discussions through collaboration (EEC) as the most difficult to practice (.30 logit). Seven out of ten items were difficult for the participants because their location was above the mean. This response is an indication that a majority of the participating teachers had difficulty in engineering effective classroom discussion by collaboration. Specifically, teachers found it challenging to get 2-3 pupils to work in small groups (1.50 logit) and allowing pupils to guide their learning (.89 logit) (Table 4.11). The findings suggest that teachers did not emphasise pupils' collaboration regularly to enhance effective classroom discussion. However, most of the teachers also highly endorsed the three remaining items as easy, in particular item 10 (-1.13 logit), in which they have had high expectations for all pupils to succeed.

(ii) Learning Tasks (LT)

Figure 4.2 reports *learning tasks* as the second most difficult strategy employed by the participating teachers (.29 logit). As can be seen from the Wright Map, participating teachers had difficulty agreeing whether all their pupils would understand the lesson's directions (2.46 logit). Item 3 was also located above the mean value at .20 logit, indicating that most teachers were less able to endorse the item, which sought to know if more than half of their pupils were clear about the task and begin work efficiently. On the easier side of the Wright map, item LT_1 was the easiest. This item dealt with the teachers' ability to give pupils some tasks and activities within the lesson as guided by the learning goal. However, in terms of understanding the lesson's directions, more than half of the pupils were more often not clear about the tasks given and could not work efficiently. The findings on learning task strategy indicate that relatively large numbers of teachers rated this strategy as less used in the classroom despite its practicality in empowering pupils' learning.

Table 4.11: Item Statistics for 150 respondents who responded to the 40 formative assessment items

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM	G
29	176	150	2.46	.26	.96	-.19	.81	-.39	.51	.50	87.3	86.7	LT4_PupilsUnderstandGoals	0
36	191	150	1.60	.22	.97	-.24	.84	-.62	.53	.51	78.9	79.2	GP5_SelfAssessment	0
17	192	149	1.50	.22	1.07	.70	1.09	.47	.48	.51	75.9	78.3	EC1_SmallGroupWork	0
38	198	150	1.28	.21	.94	-.61	.89	-.49	.54	.51	76.8	76.5	GP8_PeerAssessFutureInform	0
16	201	149	1.10	.21	1.17	1.67	1.37	1.80	.43	.52	68.1	75.2	EQ7_AdaptFutureLesssons	0
37	203	150	1.07	.21	.84	-1.74	.73	-1.52	.59	.52	78.9	75.0	GP7_SelfAssessFutureInform	0
7	206	150	.94	.20	1.04	.50	1.12	.72	.49	.52	75.4	74.3	LG14_PupilsSetGoals	0
19	206	149	.89	.20	1.10	1.12	1.08	.49	.48	.52	67.4	74.1	EC3_PupilsGuideOwnLearning	0
5	208	150	.86	.20	1.12	1.34	1.07	.45	.47	.52	68.3	73.9	LG12_ReferenceGoals	0
3	212	150	.70	.20	1.14	1.60	1.32	1.81	.44	.52	69.0	73.3	LG5_CoherentSeq	0
35	215	150	.58	.20	.96	-.45	.87	-.83	.54	.52	73.9	73.0	GP 4_InternalizeFeedback	0
21	216	149	.49	.20	1.14	1.61	1.13	.85	.46	.52	66.7	72.9	EC5_FacilitateClassDiscuss	0
8	222	150	.31	.20	.90	-1.23	.83	-1.10	.56	.52	77.5	72.4	LG15_MultipleOptions	0
24	221	149	.29	.20	.82	-2.28	.75	-1.73	.60	.52	78.0	72.5	EC9_MultipleViewPoints	0
9	223	150	.27	.20	.89	-1.29	.80	-1.34	.57	.52	73.2	72.4	LG16_PupilsDemonstrateGoals	0
28	225	150	.20	.20	1.14	1.61	1.19	1.20	.46	.52	64.8	72.2	LT3_HalfLearnersClearGoals	0
18	224	149	.18	.20	1.04	.55	1.00	.08	.50	.52	69.5	72.3	EC2_PartnerWork	0
23	224	149	.18	.20	.86	-1.77	.77	-1.55	.58	.52	79.4	72.3	EC8_RightAnswerExpected	0
30	227	150	.12	.20	.91	-1.08	.88	-.71	.55	.52	73.9	72.1	LT5_ResponsesAdaptFeedback	0
4	230	150	.01	.20	1.06	.77	1.12	.75	.48	.51	72.5	71.9	LG6_ConnectionsNew	0
6	231	150	-.03	.20	1.10	1.22	1.18	1.11	.46	.51	69.0	71.9	LG13_ShareCriteria	0
34	230	149	-.05	.20	.90	-1.17	.82	-1.11	.56	.52	75.2	72.0	GP3_RealTimeFeedback	0
20	231	148	-.14	.20	1.06	.77	.98	-.04	.49	.51	68.6	71.9	EC4_EnableDiscussion	0
39	234	150	-.15	.20	.88	-1.54	.76	-1.51	.57	.51	74.6	71.9	GP9_FeedbackLoops	0
15	234	149	-.21	.20	.92	-.96	.82	-1.02	.55	.51	75.9	72.1	EQ6_ExitTickets	0
1	236	150	-.22	.20	1.17	1.97	1.49	2.53	.42	.51	69.0	72.1	LG3_MeaninigObjectives	0
22	237	149	-.33	.20	1.15	1.69	1.09	.55	.45	.51	66.0	72.6	EC7_EnableGroupWork	0
31	240	150	-.38	.20	.94	-.67	.93	-.32	.53	.51	73.9	72.8	LT6_AnalyseResponses	0
13	242	149	-.53	.20	.82	-2.16	.79	-1.08	.59	.51	79.4	73.6	EQ4_PaceLessson	0
12	250	149	-.86	.21	.97	-.32	.90	-.35	.52	.50	78.7	75.8	EQ3_AdjustLessonPupilNeeds	0
33	250	149	-.86	.21	1.06	.62	.97	-.06	.48	.50	70.2	75.8	GP1_ReviewSomeWork	0
27	254	150	-.96	.21	.80	-2.15	.65	-1.53	.59	.50	83.1	76.5	LT2_TasksProgress	0
14	254	148	-1.13	.22	.96	-.34	1.04	.22	.50	.49	77.9	77.7	EQ5_FollowUpQuestions	0
25	256	149	-1.13	.22	1.10	.92	1.42	1.46	.43	.50	79.4	77.9	EC10_HighExpectations	0
32	261	150	-1.28	.22	1.14	1.21	1.21	.77	.43	.49	77.5	79.1	GP2_ReviewDuringLesson	0
26	262	150	-1.33	.22	.97	-.22	.80	-.63	.51	.49	79.6	79.5	LT1_ActivitiesTiedLearningGoal	0
11	268	149	-1.75	.24	.97	-.19	.77	-.58	.50	.48	84.4	83.4	EQ2_AskQuestionsPupils	0
2	270	150	-1.76	.24	1.09	.65	1.02	.19	.44	.48	83.1	83.5	LG3_ConnectLessons	0
10	271	149	-1.94	.25	.92	-.51	.79	-.45	.51	.48	86.5	85.2	EQ1_AskQuestionsLesson	0
MEAN	229.0	149.5	.00	.21	1.00	.0	.98	-.1			75.1	75.3		
P.SD	23.1	.6	.98	.02	.11	1.2	.20	1.0			5.7	4.0		

(iii) Engineering Effective Classroom Discussions by Questioning (EEQ)

Figure 4.2 reflects that engineering classroom discussion by questioning is the second less endorsed strategy in the classroom in primary schools (Table 4.11). Only one (Item_7) out of six items was found to be the most difficult endorsed by the participating teachers (1.10 logit). Item_7 dealt with using pupils' responses to help teachers in adapting future instruction. This result indicates a need for assistance to improve their questioning techniques for adapting an instruction that would benefit pupils in the classroom. Apart from Item_7, the remaining items were relatively easy and endorsed by the participating teachers, ranging from -1.94 to -.21 logits (See Figure 4.2). These items were, for instance, Item_1, *I ask questions within the lesson to assess the whole group* (-1.94 logit); and Item_2*, I ask questions within the lesson to assess pupils' progress* (-1.75 logit). In other words, the teachers endorsed easier questioning items highly, which could be due to teachers being familiar with curriculum criteria, their frequent use of these strategies or because social desirability was reflected in responding to the items. It seems that a majority of the teachers had no difficulty in procedural questioning in their class; however, they had difficulty in asking questions to guide their instruction and adjust instruction accordingly.

(iv) Gaining and Providing Feedback (GPF)

From Table 4.11, teachers' endorsement towards gaining and providing feedback was relatively higher (.10 logit) than the mean value. The GPF strategy is used to a limited extent in the classroom, except for reviewing pupils' work during the lesson, in which item 2 at -1.28 logit and item 1 at -.86 logit were endorsed as the easiest aspects of feedback provided by the teachers. Most importantly, teachers who agreed less had endorsed relatively little the extent to which they used GPF self-and peer-assessment in the classroom. Items 5, 8, 7, and 4 were related to the concept of self-and peer-assessment and were the most difficult items (Table 4.11). This result may be evidence that teachers are not acquainted with providing pupils with peer and self-assessment feedback or have only minimal skills to accomplish that. According to Heritage (2010) and Kivunja (2015), teachers are ideally supposed to use pupils more often as instructional resources for each other (peer assessment) and use their own assessment (self-assessment) to formulate effective feedback strategies.

(v) Learning Goals and Success Criteria (LGSC)

Items LGSC14, 12, 5, 15 and 16 were considered challenging items because the participating teachers found it difficult to endorse the extent to which they connect each of their lessons to the previous lesson or learning that has taken place, sharing success criteria that they used to determine their success with pupils daily in the lesson; and properly linking learning goals and success criteria (Figure 4.2). For easy items, the participants easily endorsed item 1 (-1.33 logit) and 2 (-.96 logit), which dealt with provision and discussion of the lesson objectives with their pupils, respectively (Table 4.11). These findings suggest that teachers may not have been familiar with facilitating teaching and learning with the aid of an LGSC strategy, particularly establishing success criteria.

(vi) Participating teachers' integration of HOTS

Figure 4.3 illustrates that teacher's integration of FA and HOTS' mean value was slightly higher (m = 0.02, SD = 1.85), including a higher standard deviation than the items' mean and standard deviation (m= 0.00, SD=1.22), respectively. The higher standard deviation for persons implies variance among the participating teachers across the scale of measurement about the integration of FA in the teaching of HOTS in mathematics. Specifically, the participants endorsed most of the HOTS items highly, which was about what they do in the classroom to enhance the teaching of HOTS in mathematics. Among others (see Figure 4.3 and Table 4.12), the easiest items which were endorsed highly include:

- HOTS Item 13: I provide support for pupils to work independently on mathematics problem-solving tasks (-2.608 logit);
- HOTS Item 14: My approach to instruction provides me with opportunities to encourage pupils to be critical thinkers (-1.62 logit); and
- HOTS Item 12: I use different levels of questioning which help pupils to think and reason (-1.47 logit).

```
INPUT: 150 PERSON  15 ITEM REPORTED: 150 PERSON  15 ITEM  45 CATS WINSTEPS
------------------------------------------------------------------------------
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    4            XX  +
                  X  |
                     |
                XXX T|
                     |
    3            XX  +
                     |
                     |
             XXXXXX  |
                     |T
                  X  |
             XXXXXX  |
    2                +  HOT9_TrainingFA (HOT-i)
             XXXXXX  |  HOT10_InstructMaterialsHOT (HOT-j)
                    S|
                  X  |
              XXXXX  |  HOT8_TrainingHOT (HOT-h)
                     |S
         XXXXXXXXXX  |
    1       XXXXXXX  +  HOT6_FormativeAssessHOT (HOT-f)
                        HOT7_AdminSupportHOT (HOT-g)
                     |
           XXXXXXXX  |
             XXXXXX  |  HOT2_PlanFormativeAssess (HOT-b)
                     |
        XXXXXXXXXXX  |
                     |  HOT15_DiagnoseLearningGaps (HOT-0)
    0      XXXXXXXX M+M
               XXXX  |  HOT4_CurriculumGuidesIndiv (HOT-d)
                     |  HOT3_CurriculumGuidesAssess (HOT-c)
          XXXXXXXXX  |
                     |  HOT11_ProblemSolvingHOT (HOT-k)
                        HOT5_InteractAllStudents (HOT-e)
      XXXXXXXXXXXXX  |  HOT1_FormativeAssessUnderstand (HOT-a)
   -1                +
               XXXX  |
                     |S
              XXXXX  |
                     |  HOT12_VaryLevelsQuestioning (HOT-l)
                    S|  HOT14_EncourageCriticalThinking (HOT-n)
      XXXXXXXXXXXXX  |
   -2                +
           XXXXXXXX  |  HOT13_SupportIndependentWork (HOT-m)
                     |
                 XX  |T
                     |
                     |
                  X  |
   -3                +
                     |
               XXXX  |
                    T|
                     |
                  X  |
                     |
   -4                +
                     |
                 XX  |
                     |
                     |
   -5             X  +
               <less>|<freq>
```

Figure 4.3: A Wright Map of the 150 respondents who responded to the 15 HOTS items.

Table 4.12:: Item Statistics for 150 respondents who responded to the 15 HOTS items

ITEM STATISTICS: MEASURE ORDER

ENTRY	TOTAL	TOTAL		MODEL	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH			
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	ITEM	G
9	213	149	2.00	.17	1.00	.05	1.06	.29	.62	.62	70.5	72.5	HOT9_TrainingFA	0
10	229	150	1.84	.17	1.25	1.99	1.18	.91	.56	.64	64.6	70.3	HOT10_InstructMaterialsHOT	0
8	239	149	1.35	.16	1.05	.49	.97	-.08	.64	.66	64.4	67.1	HOT8_TrainingHOT	0
6	255	150	1.02	.15	1.21	1.72	1.30	1.61	.60	.67	59.9	65.4	HOT6_FormativeAssessHOT	0
7	264	150	.92	.16	.89	-.98	.90	-.68	.70	.66	70.7	66.4	HOT7_AdminSupportHOT	0
2	280	150	.55	.16	1.16	1.41	1.21	1.68	.59	.66	66.7	67.0	HOT2_PlanFormativeAssess	0
15	295	150	.14	.16	.91	-.81	.90	-.89	.70	.66	64.6	65.1	HOT15_DiagnoseLearningGaps	0
4	306	150	-.14	.16	.91	-.88	.88	-1.10	.70	.66	71.4	65.7	HOT4_CurriculumGuidesIndiv	0
3	314	150	-.42	.18	1.08	.74	1.04	.39	.60	.64	72.8	71.2	HOT3_CurriculumGuidesAssess	0
5	325	150	-.63	.16	.98	-.19	.99	-.01	.66	.66	63.3	66.1	HOT5_InteractAllStudents	0
11	323	150	-.69	.18	.88	-1.07	.86	-1.16	.69	.64	74.8	70.9	HOT11_ProblemSolvingHOT	0
1	327	150	-.78	.17	1.10	.94	1.07	.58	.60	.64	66.7	69.7	HOT1_FormativeAssessUnderstand	0
12	352	150	-1.47	.17	.85	-1.44	.83	-1.30	.70	.64	75.5	69.6	HOT12_VaryLevelsQuestioning	0
14	355	150	-1.62	.18	.77	-2.23	.72	-2.24	.73	.63	78.9	70.6	HOT14_EncourageCriticalThinking	0
13	369	150	-2.08	.18	1.00	-.01	.93	-.40	.63	.62	70.7	72.0	HOT13_SupportIndependentWork	0
MEAN	296.4	149.9	.00	.17	1.00	.0	.99	-.2			69.0	68.6		
P.SD	46.5	.3	1.22	.01	.13	1.2	.15	1.1			5.1	2.5		

The lower and negative logits for these items (HOTS 13, 14, 1, 11, 5, 3, and 4, respectively) seem to confirm that teachers believed they were integrating FA and HOTS, especially in terms of providing problem-solving and variation in instructional methods. In addition, the HOTS items related to the curriculum were also found below the mean value (0.00 logit), which was -.14 logit for Item 4 and -.42 logit for item 3 (Table 4.12). These two items were also endorsed highly in terms of their curriculum support for FA and teaching of HOTS related tasks, as well as the support of pupils by planning for individualised instruction for those with learning difficulties.

However, the participating teachers found it difficult to agree with HOTS items that dealt with external support. These external support items were located on the top part of the map, indicating that most participating teachers were unable to endorse them (See Table 4.12 and Figure 4.3). These include:

- My school provides me with adequate training on formative assessment (2.00 logit);

- My school provides instructional material for teaching higher-order thinking skills through the use of tasks in the classroom (1.84 logits); and
- My school provides me with training in the setting items for higher-order thinking skills (1.35 logit).

In general, the findings seem to suggest that the teachers could provide a learning environment that supports pupils learning of HOTS in mathematics. However, external support from their schools was limited in training them on FA, setting HOTS items and support from administrators to incorporate FA into teaching. High self-assessment may not accurately indicate the FA implementation level, as the teachers could just be promoting themselves for social desirability, responding to the items.

4.7.4 Participating teachers' survey results on the Rasch Scale

Friedman testing was used to determine whether the teachers' responses for uses of FA strategies and HOTS integration in teaching mathematics varied by comparing the scores on the common scale. Teachers' logits from the six domains on the common Rasch measurement scale were subjected to significance testing. Table 4.13 indicates that there were statistically significant differences between teachers' FA practice and HOTS across the six strategies on the scale (LGSC, EEC, EEQ, LT, GPF, and HOTS, χ^2 [*n*= 150] = 44.52, p <.001). Inspection of the scores and rank means showed *engineering effective classroom discussion by collaboration* had the highest logit median (*Mdn*= 0.58), followed by HOTS (*Mdn = 0.10*) and lastly, three FA strategies were clustered with the lowest logits (*Mdn =- 0.07*).

Table 4.13: Friedman testing of teachers' logits for formative assessment and HOTS

Order	Scale	N	25^{th}	50^{th}	75^{th}	Mean Rank
1	**Learning goal and success criteria (LGSC)**	150	-1.42	.31	.99	2.98
2	**Gaining and providing feedback (GPF)**	150	-1.26	.07	1.30	3.01
3	**Higher Order Thinking Skills Tasks (HOTS)**	150	-1.19	.10	1.13	3.57
4	**Engineering effective classroom (Questioning, EEQ)**	150	-.79	.07	1.91	3.63
5	**Learning task (LT)**	150	-.79	.07	1.91	3.63
6	**Engineering effective classroom collaboration, (EEC)**	150	-1.00	.58	1.09	4.17

$\chi^2 = 44.52$, df=5, p<.000

Table 4.14 illustrates that four pairs had statistically significant differences in teachers' practices. After establishing statistical significance among the six strategies on the scale, the next step was to follow up with post hoc analysis. In this study, post hoc testing was done using Wilcoxon Signed Rank Tests (Pallant, 2010). The post hoc tests found statistically significant differences among the five FA strategies combined with the HOTS domain (see Table 4.14). The effect size correlation ($r = Z/ \sqrt{N}$) was converted to d, Cohen's criteria, to support the interpretation of effect sizes as consistent with Hattie's application in education studies (Hattie, 2013).

Collaboration was found to be the least used FA strategy among the teachers when compared with other strategies. The results suggest that participating teachers found collaboration difficult to use in engineering classroom discussions (Mdn= .58 logit) more often than learning goal and success criteria (Mdn =.31 logit) with medium effect size *(d* =1.6). The effect size for collaboration and learning goals and success criteria was high, and indicated limited use compared with gaining and providing feedback (*d* = 1.2). A significant difference in the logits was also found when gaining and providing feedback was paired with learning task and questioning, which yielded the same small effect size (*d = 0.4*) (See Table 4.14). The teachers' endorsement of *engineering effective classroom (collaboration)* was found to be significantly less used than the learning goal and success criteria, and gaining and providing feedback. This was followed by the teachers' endorsement of cluster strategies, *learning task and questioning* were more used than the *gaining and providing feedback* strategy. This lower level of agreeability regarding the engineering classroom through collaboration by the teachers may be tied to the teachers' ineffective teaching pedagogies, as observed in Chapter 1.

Otherwise, all other pair differences were not statistically-significant different suggesting that the teachers' level of endorsement for those FA strategies and HOTS was similar. The statistically insignificant difference between FA and HOTS suggest that participating teachers provided limited evidence in this study of how they were interactively using the strategies in the classroom to enhance productive learning.

Table 4.14: Wilcoxon signed-rank of formative assessment strategies and HOTS

Variable		**Z**	**Asymp. Sig. (2- tailed)**	**Effect Size (*r*)**	**Effect Size (*d*)**
A	B	A-B			
LGSC	EEC	7.76	0.000**	0.63	1.6
GPF	EEC	6.45	0.000**	0.53	1.2
GPF	EEQ	2.41	0.016**	0.20	0.4
GPF	LT	2.41	0.016**	0.20	0.4
EEQ	EEC	-1.06	0.288		
LT	EEC	-1.06	0.288		
HOTS	EEC	-0.08	0.935		
LGSC	EEQ	-1.88	0.060		
LT	EEQ	-0.00	1.000		
HOTS	EEQ	-0.00	1.000		
GPF	LGSC	-0.25	0.806		
HOTS	GPF	-0.90	0.366		
LT	LGSC	-1.88	0.060		
HOTS	LGSC	-0.57	0.568		
HOTS	LT	-0.57	1.00		

***** p-value < 0.01*** *Significant at the 1% level.*
****p-value < 0.05*** *Significant at the 5% level.*

4.7.5 Summary of the questionnaire survey results

The teacher questionnaire findings revealed that teachers claim to use FA in the classroom. The findings revealed significantly less frequent use of *classroom collaboration* among teachers when compared with other FA strategies such as the *use of learning goal and success criteria (LGSC),* and *gain and provide feedback.* However, the findings explicitly revealed that LGSC was most frequently used by participating teachers to discuss the lesson objectives with their pupils with limited integration of lesson success criteria to aid teaching and to learn productively. Even with gaining and provision of feedback strategy, the teachers endorsed themselves highly for items related to the review of pupils' exercises, rather than in the provision of support through pupils' peer- and self-assessment which are essential aspects of the FA feedback domain. Teachers rated low on all feedback strategies, particularly for self-and peer assessment, including for collaboration (especially items related to pupils *to work in a small group*), as less frequently used in the classroom. The findings indicate that these strategies were possibly challenging for the participating teachers as far as formative assessment practice was concerned. These findings may imply that the strategies are simply very difficult to utilise in the classroom contexts or that the activities were not good assessments, so they did not do them.

A significantly high statistical endorsement for using the questioning and learning task and questioning compared to gaining and provision of feedback may suggest that teachers were using them less frequently than some other concepts, particularly linking learning tasks and feedback for learning improvement and adapting instructions. Moreover, many of the participating teachers did not effectively ask questions, leading to adjustments of their instructions within the lesson. They also appeared not to have used pupils' responses to questions to help them adapt to future instruction. All these findings mentioned above are possible indicators of ineffective classroom discussion by questioning.

The findings for HOTS integration in teaching mathematics was not significantly rated by the teachers compared to FA practice. However, the Wright Map revealed that teachers had endorsed themselves highly on the HOTS items which dealt with practices such as their ability to provide support for pupils to work independently on mathematics problem-solving tasks, and instructional opportunities to encourage pupils to be critical thinkers and use different levels of questioning which will help pupils to think and reason. On the other hand, they endorsed themselves low on the items which call for the support they get from the school as being limited on providing them with training on FA and setting of HOTS items.

To this end, additional data was needed and obtained by using classroom observations for further exploration of the phenomena as to whether teachers had employed FA and HOTS effectively as they seemed to have suggested by high self-assessment in the questionnaire. Given the fact that the study was positioned for teaching practices in low school performance contexts, the teachers may feel vulnerable and defensive reporting in a self-administered questionnaire, and hence these results may reflect unreliable or unrealistic answers (Van Staden & Zimmerman, 2017). Due to potential social desirability when responding, the current study also considered using the additional data collection instrument in the form of some classroom observations and review of pupils' work during the baseline survey (Dunning, Health & Suls, 2004; Van Staden & Zimmerman, 2017). The observation and checking of pupils' work helped in ascertaining the extent of the actual implementation of FA and teaching of HOTS. A mixed-method of data collection was found appropriate to gain complementary findings on the FA practice in the classroom and aid in developing the intervention.

The following section presents and describes the findings of classroom observation during the baseline survey.

4.7.6 Pre-classroom observation results

The classroom observations augmented questionnaire results and were used to understand the use of FA practices better. To this end, the succeeding sections discussed the findings which were observed by the researcher concerning the Teacher Assessment Practice Observation schedule (TAPOS). The observation data was analysed both quantitatively (frequencies) and qualitatively (notes). It is also important to note that observing only one lesson did not allow the researcher to see a broader spectrum of the behaviours. However, for this study, a single observation was done mainly to augment the quantitative survey findings. For this reason, a descriptive frequency count based on the observation scale was used.

(i) Classroom Context

The Standard 4 teachers were observed to determine the level of formative assessment practice in teaching mathematics at the Standard 4-level in the Botswana classroom context. There were nine teachers, and eight were observed while teaching mathematics before and after the intervention. One of the selected teachers could not be observed during the pre-phase of observation due to illness. The class sizes ranged from 27 to 40 pupils per class. During the observational week, seven of the observed teachers (one observation per teacher) were teaching the same topic (money), and only one teacher was teaching "fractions". The teaching pattern indicates that teachers were implementing a common national syllabus and hence their pace of lesson delivery was relatively the same.

A classroom with teaching charts is considered essential to help pupils visualise some concepts (Kanjee, 2017). In all the observed classroom walls, approximately 20% of the displayed charts and pictures were from mathematics and mainly included multiple timetables, shapes, and formulae. None of the classrooms had a data projector or a smartboard. The set-up of the classrooms, in general, were traditional learning environment-oriented, without any electronic devices for learning. The teachers were dependent primarily on the use of chalkboard and manila papers. Technology use and infusion of technology use for formative assessment, in particular, cannot be associated with the observed schools. The context of the learning environment for primary schools in Botswana is yet to achieve the transformed ways of teaching and learning and adopting new technology like smartboard and data projectors.

The lessons were observed based on the weekly lesson timetable, in which mathematics in schools was scheduled either for nine periods or ten periods of 30 minutes of the academic calendar. Weekly, one hour per day is scheduled for mathematics lessons in the ten periods-based timetable (five hours per week). As for nine periods, the lessons were scheduled for the first four days of the week and implemented for an hour each. On the fifth day, the remaining 30 minutes (4 hours 30 minutes) were used. The researcher was a non-participant observer only in each mathematics lesson for an hour. During the observation section, teachers were seen starting their lessons with some previous knowledge learning-related activities. In general, the pupils were not afraid to interact with the teachers, notwithstanding the researcher's presence in their classroom. The language of instruction was English, and teaching in English was a challenge for most teachers. They would sometimes struggle to explain the work or concepts to the pupils and then switch to their home language (mainly Setswana). This practice may also have disadvantaged some pupils with a non-Setswana speaking the home language.

The success of lessons depends mainly on planning and preparation. According to Moloi and Kanjee (2019), lesson planning is integral in the teaching and learning process since "['the] lesson begins when teacher plans and prepares for a lesson - not in the lesson" (Moloi & Kanjee, 2019, p. 12). The weekly planning and preparation for the observed teachers seemed inconsistent with Moloi and Kanjee's (2019) suggestion. As can be seen from the sample in Figure 4.4, the lesson plan did not reflect how each learning goal (LG) would be achieved and did not supply success criteria (SC).

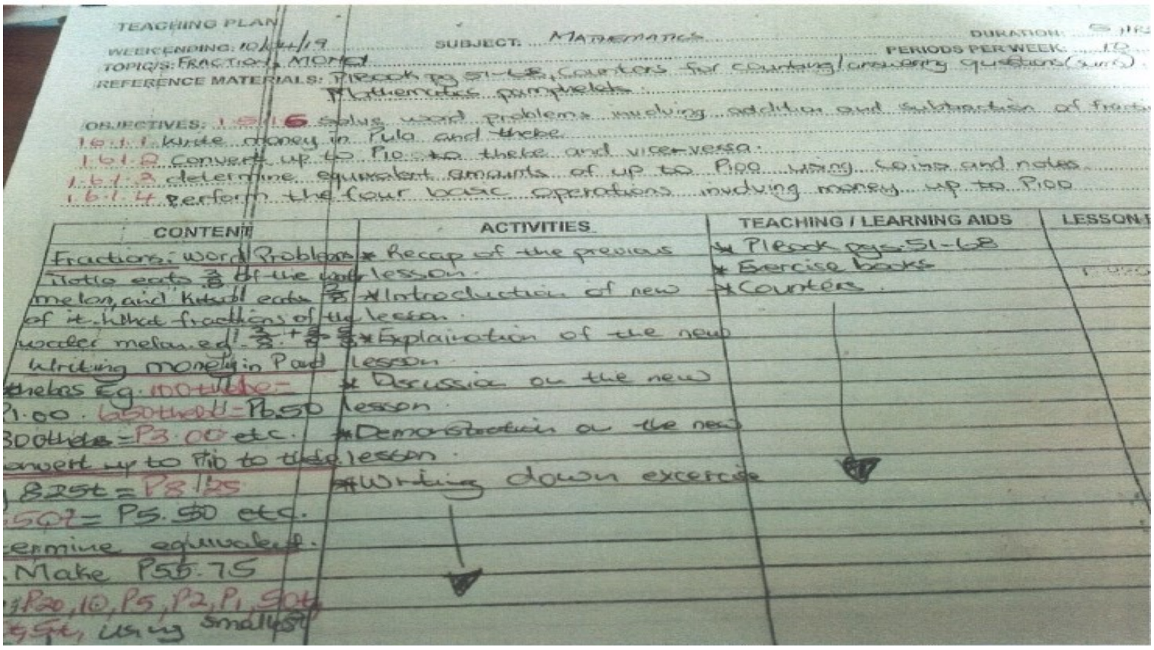


Figure 4.4: Sample of a lesson preparation for the week

The lesson plans did not show how LG and SC would be attained, no key questions were stated, and no FA techniques were planned before lesson presentations of the week. It seems that teachers’ lesson planning and preparations may have been done for purposes other than lesson implementation. It may be that the planning and preparations were conducted for administrative purposes, as was found in South African classrooms (Moloi & Kanjee, 2019).

For the classroom observations, the FA strategies used in this study were incorporated into the lessons' three features, as shown in Table 4.15.

Table 4.15: Teacher assessment practice observation schedule for a lesson

Beginning of the lesson	During the lesson	At the end of the lesson
The teacher was observed introducing learning goal and success criteria**.** **Whether the teacher used words such as:** **"We are learning to" when introducing the lesson objective (LG);** **"What I'm looking for" when introducing the assessment criteria (SC).** **Whether the teacher presented LG and SC;** **Orally,** **Written on the chalkboard,** **Written on the chart,** **Provided in a handout,** **Others.**	**Questioning and interaction**: It is concerned with teachers' way of encouraging participation, the kind of questions the teacher asks and the teacher's technique of inviting answer from pupils. **Learning HOTS tasks** The teacher was observed to see whether: She or he used a HOTS task aligned with the learning goal, used multiple ways of gathering evidence throughout the lesson that were connected to the learning objective, using a diversity of problem-solving tasks, and engaged pupils on previous problem-solving tasks given as homework. **Feedback;** The teacher was observed in respect of giving guidance on pupils' work and responding orally. The teacher was also observed for using peer- and self-assessment during the lesson	**To conclude a lesson**: The researcher observed how the teacher sums up/concludes the lesson. Particular attention was focused on the question; Did the teacher complete the lesson? Have the learning objectives been completed? Have assessment criteria been met?

The TAPOS formed the classroom observation basis and rated the teachers on the extent to which the 28 statements were reflected in their observed teaching and learning practices during mathematics lessons. The rating scale was based on the possible responses, which involved two subset scales. Firstly, the researcher was looking for evidence that it occurred (yes/no), then how frequently (0 to 2). Refer to *Appendix B* to see the observation schedule.

(ii) Learning goal and success criteria at the beginning of the lesson

From the beginning of the lesson, the teachers were observed to rate the extent to which they used the learning goal and success criteria during a single lesson presentation. The findings revealed that teachers tended to present previous learning knowledge and connect that knowledge with the present lesson. It was observed that most teachers did not write the learning goal on the chalkboard but that they would rather write the sub-topics on the board. For example, instead of writing "*we are learning to perform the four basic operations involving money up to P100"*, they would rather say "*we are adding money" or "subtracting money*" depending on the teaching topic at hand. Throughout the observation, the researcher did not see much evidence of sharing the success criteria in any oral presentations, written on the board, chart or provided in a hand-out format (See Table 4.16).

Table 4.16: Relative frequency of observing the use of learning goal and success criteria during pre-observation

FA Strategy	Evidence		
	Pre-classroom observation		
Learning goals[R]	N	Seen (rf)	Not (rf)
The teacher uses words such as **We Are Learning To (WALT)** when introducing the Lesson objective (LO)	8	2(0.250)	5 (0.750)
• Presented orally	8	2(0.250)	5 (0.750)
• Written on the board	8	0 (0.000)	8 (1.000)
• Written on the chart	8	0 (0.000)	8 (1.000)
• Provided a handout	8	0 (0.000)	8 (1.000)
Success Criteria[R]			
The teacher uses words such as **What I'm Looking For (WILF)** when introducing the Assessment Criteria (AC)	8	0 (0.000)	8 (1.000)
• Presented orally	8	0 (0.000)	8 (1.000)
• Written on the board	8	0 (0.000)	8 (1.000)
• Written on the chart	8	0 (0.000)	8 (1.000)
• Provided a handout	8	0 (0.000)	8 (1.000)

Note: R = dichotomious rating; 1= ("Seen"), 0 = ("Not seen"), rf = relative frequency

The following are some examples of developing a practice observed in the LG domain:

Basic practice

The teacher recapped the previous lesson: "I understand that yesterday you did not understand the learning intention, so we are repeating the topic. First, let's list the pula we have" (Participant 1). Teachers' recap is typically a basic way of interacting with pupils in the classroom, the teacher would just pose mainly procedural phrases or statements, and such statements do not necessarily engage pupils in any way.

To developing practice

In this observed instance, the teacher asked, *"What did we talk about yesterday?"* However, the teacher did not expand on such mental maths. They then shifted the lesson to the current topic as an attempt to introduce it. *"Today, we are learning about money". "When adding money, an addition like any other addition we know so far, it's the same. However, with money, we include the sign "t" for thebe or "P" for Pula*[2]*. (Participant 2).* The teachers' recap indicates a developing practice since the teacher begins the lesson, which comprises mainly closed but relevant questions.

Exemplary practice

"Which topic did we do last week?" Whole class "Money". Teacher "Yes money, and this week we are still going to continue learning about money. Today we are still going to learn about money. How do you think money is useful in our life?" (Participant 4). In this lesson, the teacher had incorporated a range of questions and included those which promoted higher-order thinking.

From the examples provided above, evidence was found that one teacher demonstrated an exemplary way of introducing a lesson while the other two teachers are developing a practice that suggests the need for further observation and assistance.

Use of FA during the lesson development

The frequent use of planned questioning and interaction, task-based assessment, and operational feedback is considered an effective way of engineering classroom discussion and enhancing pupils' HOTS and academic performance. During the lesson observations, the evidence of a pupil's self– and peer-assessment was also considered essential to activate pupils as instructional resources for each other and becoming owners of their learning. Classroom observations were made to determine

---

[2] The Botswana Pula is the currency of Botswana. The currency code for Pula is BWP, and the currency symbol is P while Botswana Thebe smallest currency in coin and the currency symbol is T

the application of FA strategies. Therefore, it was used frequently during the lesson implementation. Table 4.17 to Table 4.21 provides an overview of teachers' descriptive statistics regarding their relative use of FA elements during teaching and learning mathematics.

(iii) Engineering effective classroom discussion

Table 4.17 provides evidence that FA was not explicitly observed during lessons attended by the researcher. Table 4.17 shows that half of the teachers observed waited a few seconds before getting a response from pupils during questioning. Besides the use of waiting time, teachers were inconsistent compared with the findings of Wiliam (2011), who classified questioning techniques without eliciting responses from many pupils as undesirable classroom practice. All the participating teachers asked questions, and they allowed pupils to put their hands up (*n*=8). Seven of the eight participating teachers were seen questioning pupils, those who had their hands up, and also asked questions for the “whole” class to respond (six out of eight participating teachers).

Table 4.17: Relative frequency of observing questioning and interaction

Observed behaviour	Not Seen	Sometimes	Often
When the teacher asks questions, pupils put their hands up	0 (0.000)	0 (0.00)	8 (1.000)
The teacher only asks pupils that have their hands up	0 (0.000)	1 (0.125)	7 (0.875)
The teacher involves more than one pupil in questioning	1(0.125)	4 (0.500)	3 (0.375)
The teacher asks questions for the "whole" class to respond	0 (0.000)	2 (0.250)	6 (0.750)
The teacher waits few seconds before response from a pupil	3 (0.375)	4 (0.50)	1 (0.125)
The teacher answers her/his own questions	6 (0.750)	2 (0.250)	0 (0.000)
The teacher uses name/number sticks to select pupils	8 (1.000)	0 (0.000)	0 (0.000)
Pupils use mini-boards during the lesson	8 (1.000)	0 (0.000)	0 (0.000)
Pupils work in groups to guide each other on their learning	3 (0.375)	5 (0.625)	0 (0.000)
Pupils work in groups cooperatively while completing group task	4 (0.500)	3 (0.375)	1 (0.125)
The teacher conveys an attitude of "we all can"	1 (0.125)	7 (0.875)	0 (0.000)
The teacher provides support & encouragement to pupils	1 (0.125)	3 (0.375)	4 (0.500)

Teachers are expected to be systematic when eliciting evidence of learning from all pupils throughout the lesson. Table 4.17 also reveals that during classroom observation, none of the teachers was seen using name/number sticks to select pupils and none of the pupils used mini-boards during the lesson. The findings indicated that the teachers had challenges in engineering effective classroom discussion through pupil interaction. As seen from Table 4.17, a total of five out of eight teachers were observed as sometimes putting their pupils to work in groups to guide each other on their learning. Only half of the teachers were not seen putting their pupils to work cooperatively in groups or while completing a group task, while three out of eight teachers sometimes did it. The results for classroom questioning and interaction suggest that teachers had some limitations in pupils' engagement. Such limited evidence for involving a pupil-centred learning environment may suggest a need for the enhancement of teachers' skills.

(iv) Learning of HOTS task

Table 4.18 reveals how teachers were connecting the learning goals, clarifying learning tasks and using evidence to adjust their instructions as might have been needed in the lesson.

Table 4.18: Relative frequency of observing learning HOTS tasks

Observed behaviour	Not Seen	Sometimes	Often
The teacher uses well-crafted HOTS tasks that are aligned with the learning goal.	7 (0.875)	1 (0.125)	0 (0.000)
All pupils are clear about the HOTS task and can begin work efficiently.	7 (0.875)	1 (0.125)	0 (0.000)
The teacher frequently uses pupils' responses and work to make inferences about progress and adapts instruction accordingly.	4 (0.500)	4 (0.500)	0 (0.000)
The teacher skillfully uses multiple ways of gathering evidence throughout the lesson that are connected to the learning.	4 (0.500)	4 (0.500)	0 (0.000)
The teacher uses multiple approaches to handle problem-solving tasks	8 (1.000)	0 (0.000)	0 (0.000)
The teacher gives homework on problem-solving tasks	8 (1.000)	0 (0.000)	0 (0.000)
The teacher engages pupils on the previous tasks given as a homework	8 (1.000)	0 (0.000)	0 (0.000)

From Table 4.18, the analysis indicates that seven out of eight teachers were not seen using well-phrased HOTS tasks aligned with the learning goal or providing all pupils with clear instruction about the HOTS task at the beginning of the work session during the researchers' classroom observation.

Table 4.18 shows that all the teachers observed had challenges when handling problem-solving tasks, giving homework on problem-solving tasks or engaging pupils on the previous tasks given as homework. The findings seemed to suggest that, in general, the teachers tried to apply a learning task strategy during their lesson by aligning tasks or activities with the learning goals, even though most of the activities were low order thinking oriented.

(v) Effective Feedback

Learning is considered an individual process, so generic feedback to the group does not address the individual pupils' need (Kivunja, 2015). Both oral and written feedback has to be specific to the pupils to move them forward as individuals, and as such, it has to be descriptive, not evaluative (Heritage, 2010). Table 4.19 indicates the teachers assessed pupil progress after giving pupils classwork. They were walking around to check how pupils were doing. A total of six out of eight teachers were seen checking pupils' work, and providing guidance or making comments.

As reflected in Table 4.19, all eight teachers were seen offering oral evaluative feedback, and none of them provided oral descriptive feedback on a specific piece of work during the lesson. These findings suggest that the teachers were familiar with evaluative feedback. Hodgen and Wiliam (2006) discourage teachers from employing evaluative feedback because low achieving pupils in those classes are likely to reduce their effort, interest, and persistence and lose self-esteem to learn. Hence, in general, evaluative feedback is not a supportive learning way of giving feedback to pupils.

Table 4.19: Relative Frequency of observing Oral and Written Feedback

Observed behaviour	No	Yes
After giving pupils classwork, the teacher walks around to check how pupils are doing.	0 (0.000)	8 (1.000)
The teacher provides orally evaluative feedback on a specific piece of work	0 (0.000)	8 (1.000)
When checking pupils' work, the teacher gives guidance or makes comments	2 (0.250)	6 (0.750)
The teacher provides orally descriptive feedback on a specific piece of work	8 (1.000)	0 (0.000)

Additionally, teachers' evaluative feedback was mainly awarding marks and comments that make conscious or unconscious comparisons with others. For instance, in some exercise books which were reviewed, it was revealed that teachers were offering ticks for correct answers and crosses for incorrect answers. Thus, they did not give any guiding comments, for example, "That's a good introduction because you have covered the main step of addition we discussed at the beginning. No which step do you think you missed? Check the previous example."

Similarly, in written feedback, it was observed that teachers were writing evaluative phrases such as "poor" for disapproval, "very good" and "excellent" for approval. One teacher wrote: "pull up your socks" as a comment, shown in Figure 4.5.

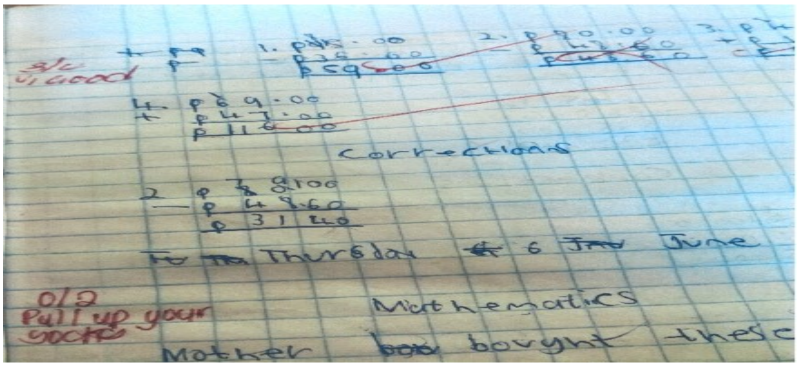


Figure 4.5: A sample of evaluative feedback

The result from the observations and textbook analysis seems to indicate that teachers emphasise evaluative feedback. However, scholars in education discourage judgmental feedback by the teachers based on implicit or explicit norms, affecting how pupils feel about themselves (Hodgen & Wiliam, 2006).

(vi) Peer - and Self-Assessment

Frequent use of peer- and self-assessment is considered to be an effective way for pupils to give feedback to each other and themselves in a purposeful and meaningful way. However, in a traditional classroom, teachers may have too limited FA knowledge to implement peer- and self-assessment daily, or simply very difficult to do in the contexts in which they were observed. Table 4.20 indicates that none of the teachers was observed implementing peer- or self-assessment. In all, a total of 6–8 teachers were not seen using both peer- and self-assessment in the classroom as a way of providing feedback on instruction. This finding corroborates the current study's survey findings, which revealed that surveyed teachers found it difficult to endorse items that dealt with using pupils' peer- and self-assessment daily as evidence for feedback respectively.

Table 4.20: Relative frequency of observing Peer- and Self-Assessment

	Observed behaviour	**No**	**Yes**
Peer Assessment	Pupils are given an opportunity to check their partner's work.	6 (0.750)	2 (0.250)
	The teacher reminds pupils how they should use peer assessment.	8 (1.000)	0 (0.000)
	The teacher visits a few pupils to check how they conduct peer assessment.	8 (1.000)	0 (0.000)
	The teacher gives feedback on how the peer assessments were conducted.	8 (1.000)	0 (0.000)
Self-Assessment	Pupils are given an opportunity to check their work.	8 (1.000)	0 (0.000)
	The teacher reminds pupils how to use self-assessment, e.g., the process and rules are reviewed.	8 (1.000)	0 (0.000)
	A teacher tells pupils to use Success Criteria when checking their work.	8 (1.00)	0 (0.00)
	The teacher visits a few pupils to check how they conduct the self-assessment.	8 (1.00)	0 (0.00)
	The teacher gives feedback on how the self-assessments were conducted.	8 (1.00)	0 (0.00)

(vii) Lesson conclusion by the observed teachers

Table 4.21 indicates that teachers were likely to complete the lesson. However, in half of the observations, the teachers did not check whether the learning objectives had been achieved.

Table 4.21: Relative frequency of how teachers concluded the lesson

Observed behaviour	**No**	**Yes**
Did the teacher complete the lesson?	1(0.125)	7(0.875)
The teacher checks whether the learning objectives have been completed.	4 (0.500)	4 (0.500)
The teacher checks whether the assessment criteria have been met.	8 (1.000)	0 (0.000)

Additionally, none of the observed teachers checked whether the assessment criteria had been met at the end of the lesson. The following examples demonstrate an ineffective practice of ending a lesson as observed by the researcher:

(viii) Incomplete basic practice

The teacher had planned for three learning objectives and she only covered one, while another teacher just switched to the other subject (science) before concluding mathematics:

"Pupils let put aside the mathematics book now, okay. We are moving to another subject, that is science. Okay. We are done with maths". "Sir, [referring to the researcher] we are done with maths". (Participant 1).

"Now that we have finished the group activity, can you collect your textbooks for English? Are we all finished?" Participant 5 replied: "Yes, teacher" [pupils]. "That's good. Let us move on to English" (Participant 5).

From the observations, it seems that teachers have challenges in summarising learning intentions to reinforce learning and check learning gaps before shifting to the other lessons.

4.8 Discussion of the Questionnaire and Classroom Observation Findings for Phase I

This section aims to outline the findings emanating from the data collected through the teacher questionnaire and classroom observation analysis to triangulate results (sub-researcher 1). The discussion provides a synthesis of the teachers' responses and ratings by the researcher to determine any consistency in the findings towards the use of formative assessment in the teaching of HOTS in mathematics. Each domain is discussed separately in this section.

4.8.1 Formative assessment practices

The data collection instrument (both the survey questionnaire and classroom observations) findings showed that teachers found it difficult to explain the connections between new, prior and future learning. Other challenges experienced by teachers included designing coherent sequences of learning that may be preferable to individual lessons. This finding is in agreement with Rahman (2018), who found that teachers mentioned learning objectives before starting a topic. Still, they would not always discuss the topic accordingly; instead, they would present the lesson following their preferences (Rahman, 2018). The teachers and pupils need to be clear on lesson outcomes. The findings of the current study were, therefore, inconsistent with the observation made by Tobey and Goldsmith (2013). They found that sharing Success Criteria (SC) assists pupils to reflect their understanding into the context of the identified goals and monitor their learning progress. The LGSC also allow pupils to receive additional feedback to promote further learning and incorporate the feedback into subsequent work (Heritage, 2010). The Learning Goal (LG) helps the teacher set up some activities to facilitate pupils' understanding of the learning goals through discussion and negotiation (Wiliam, 2011).

The teacher questionnaire findings indicated that teachers claimed to use the learning goals and success criteria in the lesson implementation. Thus, the Wright Map shows that most observed teachers were familiar with connecting a current lesson to the previous lesson. The participating teachers used lesson objectives that could be classified as unproductive. However, during classroom observation, the observed teachers did not share or write the learning goal on the chalkboard, but they did write sub-topics. Secondly, the researcher did not see any evidence of sharing the success criteria as an oral presentation or written on the board/charts. The finding indicates that there were potential challenges in the use of LGSC, and the teachers may not understand the strategy as classified in the current study.

Additionally, the questionnaire results showed that a mean of .09 logit (close to average) of the teachers' endorsement was reported. They simply are centred on the mean of the sample who participated. This result indicates that not all teachers have used LGSC as a strategy in their mathematics lessons, providing particular success criteria (SC). For instance, the researcher did not notice the frequent use of the SC during observations. Nonetheless, teachers and pupils must use SC to strengthen pupils' learning, assess themselves independently and know what is expected. More specifically, Wiliam (2011) justified that SC had to be linked to the learning outcome. It tells pupils what criteria are being assessed to measure whether the learning outcome has been achieved.

As for engineering effective classroom discussion by collaboration, the questionnaire data also indicated that teachers could not guide their pupils on their learning to find solutions to given tasks as a group. The findings of the questionnaire agree with the observation by the researcher. Specifically, teachers identified themselves as challenged by putting 2-3 pupils to work in small groups (1.50 logit) and allowing pupils to guide their learning (.89 logit). Perhaps the participating teachers may have used group work on another day during a classroom discussion in the researcher's absence. The findings for the learning task strategy showed that teachers had some challenges in clarifying learning activities and using evidence of such activities to adjust their instruction as might have been needed during the lesson. Precisely, the observation findings indicated that seven out of eight teachers were not seen using well-phrased HOTS tasks that were aligned with the learning goal or providing all pupils with clear instruction about the HOTS task at the beginning of the work session. This finding is not in line with Kivunja's (2015) and Wiliam and Thompson's (2014) proposals that teachers are required to elicit evidence of learning as the pupils engage with the learning activities. This engagement should include using a variety of strategies to gather evidence on how the pupils are learning and moving towards the defined learning goal.

As revealed by both data collection instruments, the strategy for *engineering effective classroom discussion* through questioning (.29 logit) was more predominately used than *collaboration* (.30 logit). This finding for the questionnaire was inconsistent with the classroom observations. Thus, it was discovered that the observed teachers were asking mainly procedural questions and some questions that assessed certain portions or groups of pupils in classroom discussion. In other words, the teachers' questioning neither targeted nor engaged all pupils. The researcher observed that while teachers asked questions, they allowed pupils to put their hands up, and the teachers mostly asked pupils who had their hands up. However, the teacher asked procedural questions for the "whole" class to respond.

The teachers' classroom discussion implemented were considered ineffective ways of engaging pupils and not in line with this study's theoretical point of departure. The current study's findings do not show evidence of effective questioning techniques as proposed by Wiliam (2011). He recommended increased wait time, no hands up, using name sticks and exit cards. These techniques elicit evidence of learning from pupils as well as assessing pupils' progress towards instructional goals as a whole class. The observed teachers' questioning in the classroom was consistent with what Thompson and Wiliam (2005) noted, namely teachers' questioning skills were too shallow, narrow, or ineffective and focused on grading pupil responses or work only.

Learning tasks are supposed to be the key to elicit learning and ensure alignment with the learning goal. However, the current study's findings for both questionnaire and classroom observations showed a lack of teachers' strength to connect the two concepts. These findings of inadequacy for learning tasks, in turn, may lead to a discrepancy in teaching and learning. This study was unable to corroborate Wylie and Lyon's (2015) findings from a cross-sectional survey that found that most teachers commonly implemented questions or tasks to elicit evidence of learning. These questions or tasks often focused on collecting data from pupils. The teachers' inability to develop learning tasks was likely to weaken pupils’ ability to learn HOTS tasks in the activities explored by the current study.

4.8.2 Integration of formative assessment and HOTS

The use of HOTS with a particular focus on mathematics revealed that it was highly endorsed (the easiest items were rated at -2.08 logit), more than any other FA strategies (the easiest item was rated at -1.94 logit). Most importantly, the teachers have self-reported emphatically about having the ability to provide support for pupils to work independently on mathematics problem-solving tasks, providing instructional opportunities to encourage pupils to be critical thinkers, and using different questioning levels that help pupils think and reason. However, results for classroom observation did not reveal such innovations in those classrooms. Perhaps teachers may have felt vulnerable and defensive reporting in a self-administered questionnaire and hence rated themselves highly, yet in reality, their practices are not reflected accordingly (Dunning et al., 2004; Van Staden & Zimmerman, 2017).

Moreover, the teachers had also reported endorsing some external support factors such as school support less and saw a lack of training and available time as a setback towards the implementation of HOTS. This finding support Masole et al. (2016), who noted that Botswanan schools might not have provided teachers instruction pedagogies or support remedial teaching. So, the current study appeared to confirm that participating teachers may not be fully equipped to enhance pupils’ problem-solving and critical thinking skills to the degree that would assist pupils in learning independently. The findings also support the ideas of Rajendran (2008), who had suggested that HOTS are essential and pertinent to educate the pupils of the 21st century who face complex real-life problems, which often need complex solutions.

Lastly, it is important to note that the survey findings agree with classroom observations, particularly the teachers’ limitation of not having enough time to plan for FA, including higher-order thinking tasks. Even during a classroom observation, the sampled lesson plan mentioned earlier shows the lesson's challenge to deliver HOTS activities. Teachers remain and are expected to be at the centre of planning HOTS activities which involve concept formation, critical thinking, creativity, brainstorming, problem-solving, mental representation, rule use, reasoning, and logical thinking. Likewise, the curriculum must include integrating FA and teaching of HOTS tasks and supporting FA and individualised instruction at a range of grade levels.

4.8.3 Future research

The baseline survey discussion of the findings was underpinned by a mixed-methods research approach but still needs to be interpreted with utmost caution. Firstly, the small cross-sectional survey design only consisted of primary schools' teachers from the Southern Education Region and cannot conclude with confidence about teachers' FA practices and teaching of HOTS in Botswana primary schools. Secondly, despite the observational evidence about teachers' formative assessment practices, which complimented the findings, more observations are needed either by widening the sample across other education regions or including the interview or focus group discussion in the design. Nevertheless, the current findings indicate that the Standard 4 teachers' assessment practice at the regional level could be discussed with some confidence that the sophisticated analysis methods, through Rasch modelling, could permit rating of a limited set of attributes that are representative of the underlying trait and highlight the most problematic items for the researcher to decide whether to either delete, rescore, or reword the collected data.

4.8.4 The practical contribution of Phase I findings to intervention design

Based on the survey data and classroom observation in Phase I, it was concluded that teachers' FA practice was not predicated on the five key strategies. The participating teachers rarely linked the learning goals and success criteria when they were observed. The survey and classroom observations revealed contrasting findings, thus the observed teachers did not *effectively engineer classroom discussions* either through questioning or collaboration. The observed teachers also demonstrated challenges in eliciting evidence of learning through learning tasks that agree with survey findings. The provision of feedback on the instructional activities was self-reported and observed to be a challenge with the primary school teachers. Most importantly, the teachers did not use peer- and self-assessment as central to activating performance improvement during the observations. Due to the teachers' lack of implementing FA practices during classroom observations, it may be necessary to integrate formative assessment and HOTS further in mathematics classrooms.

The results should also be understood within the jurisdiction of Botswana as formative assessment (continuous assessment; CA) is done for the sake of administrative compliance and a record-keeping exercise of individual pupils through quizzes, topic tests, end-of-term tests, and mock examinations. In short, it is regarded as a summative assessment purpose. There is little evidence regarding FA to guide instruction in teaching and learning mathematics, as Black and Wiliam (1998) explained. This assessment practice is closely associated with other African countries like the Tanzanian context to the extent to which CA is used in the education environment, which is supposed to be formative and also contributes to the summative purpose (Kyaruzi, 2017). Such assessment practices can also be transferable in some educational systems that apply for similar assessment programmes, in particular UK colonies or associated territories such as India, Egypt, and Hong Kong (Brown et al., 2019, Kyaruzi, 2017).

Having these findings and jurisdiction assessment practices in mind, in which the Botswana teachers are expected to teach and assess pupils' progress, it's likely to impinge negatively on school learning in general. The context signifies further research that considers a teacher's professional development as far as teaching, learning and assessment are concerned in the 21st century. Such research would possibly support the ETSSP (2015) component of improving teaching and learning at all levels, in particular, the ETSSP emphasises undertaking intensive teacher professional development and developing appropriate assessment patterns to organise school-based assessment and measuring skills and linking classroom assessment with a national assessment in a better way.

Specifically, in Table 4.22, the excerpt from the teacher questionnaire items seemed to justify the impediment of integrating FA and HOTS in teaching mathematics. The results, including the classroom observation, suggest that the teachers had challenges integrating FA and HOTS in the mathematics classrooms and infusing FA and HOTS as consciously self-rated by the participating teachers due to a lack of teacher training on assessment, time for planning and the curriculum limitation to.

Table 4.22: Excerpt from Higher-order Thinking Strategies ranked from most difficult least utilised

Questionnaire Items	**Theta (Logit)**
My school provides me with adequate training on formative assessment (HOTS9)	2.00
My school provides instructional materials for teaching higher-order thinking skills (HOTS10)	1.84
My school provides me with training in setting test items for higher-order thinking (HOTS8)	1.35
I have enough time to gather & act on the FA for higher-order thinking (HOTS6)	1.02
I have admin support to incorporate FA into my teaching of HOTs (HOTS7)	.92
I have enough time to plan FA including higher-order thinking tasks (HOTS2)	.55

In designing the intervention programme, which was named formative assessment for HOTS (FAHOTS), the emphasis was placed on Kivunja's Assessment Feedback Loop (2015), and this was linked to HOTS concepts. These two domains have a common characteristic that suggests that features of FA converge with the features of HOTS with all fitting into Kivunja's Conceptual Framework of the Assessment feedback Loop (2015).

Such an intervention was classified as professional development for in-service training to help teachers implement FA and teach HOTS in mathematics. The workshop-based training was designed and implemented to address the identified context. Thus, after the two days of the training workshop (see Chapter 5, Section 5.3 for a detailed description of the intervention), the teachers were expected to:

- Identify and communicate learning goals and success criteria for a given topic using Bloom's cognitive behaviours (particularly HOTS);
- Recognise the value of aligning assessments with the learning goals;
- Improve pupils' participation and engagement in the classroom to enhance critical thinking;

- Develop higher-order thinking questions following Clarke's (2005) five strategies to improve the use of questioning by changing a recall question into questions that improve pupils' engagement during the lesson and improve the HOTS; and
- Determine ways of providing effective feedback as well as a peer– and self-assessment.

4.9 Formative Assessment for HOTS Intervention

In light of the relatively poor quality of formative assessment for HOTS reported in this chapter, an intervention was designed. In the current study, the researcher designed an interventional programme extracted from Kivunja's Assessment Feedback Loop (2015) and linked it to HOTS; hence it was named formative assessment for HOTS (FAHOTS). This was attempted to answer the SRQ 2; thus, how FA can be supported through intervention to enhance teaching and learning for Standard 4 pupils HOTS in mathematics.

The participating teachers took part in the two-day workshop facilitated by the researcher (see a sample of workshops slides in *Appendix N*). The researcher trained nine teachers about the use of formative assessment to enhance higher-order thinking skills (HOTS) with a particular focus on mathematics.

On Day 1 of the workshop in which teachers were made aware of the FA learning goals (LG) and Success Criteria (SC). Thus, before a teacher could find out what his or her pupils are learning, before giving feedback, before engaging pupils as resources for one another and as owners of their learning*,* both teacher and pupils have to be clear about where they are going (Kanjee, 2017; Kivunja, 2015). Similarly, with the SC, teachers were made cognisant of the fact that SC was a way of linking the learning goal and indicate achievement. Therefore, the SC must be shared with the pupils and must be related directly to the learning outcome and the HOTS task. It has to be explained in a language the pupils can understand. In many instances, the SC is not "owned" by the teacher but can be generated with the pupils. Sharing the SC with pupils allows them to self-edit and self-assess. Thus, take greater ownership of their learning (Kivunja, 2015). The SC help pupils to strengthen their learning, assess themselves independently and know what is expected of them. It takes away the mystery when pupils do not always know what the teacher's criteria for assessing their work are, but by sharing the criteria it becomes transparent (Gardener, 2013; Kivunja, 2015).

The workshop was also used to discuss the value of aligning assessments to the goals with reference to HOTS derived from the Blooms' Taxonomy and learning objectives in the Botswana Standard 4 mathematics curriculum. The discussion on teaching HOTS involved the barriers (for instance lesson planning, the quality of the curriculum materials to teach HOTS tasks, and the time available for planning and reflection) and support (for instance, whether the teachers had all the necessary mathematical classroom resources such as coins, bills, fraction and cubes) as well as the possibility for improvisation and creativity in the classroom.

Day 2 of the workshop focused on the effective questioning that elicits discussion and pupils' higher-order thinking, as well as activating pupils to be instructional resources for each other (peer-assessment) and their own (self-assessment) based on focused, effective feedback strategies (Kanjee, 2017; Wiliam, 2011).

The FAHOTS information was made available to selected teachers who were expected to implement FA circles weekly. To support the teachers during the implementation process, they were urged to embed the model in their daily mathematics curriculum materials. The teacher could use the workshop materials during lesson planning so that their lesson should consider the FAHOTS strategies; (introduce the lesson objective/learning intention, by writing it on the chalkboard and followed by SC which also needs to be written at the corner of the chalkboard. Teacher and pupils; were expected to discuss the LG and SC at the beginning of every lesson and plan sets of questions and HOTS tasks to use in the classroom. These should be aligned to the LG; during lessons, the teacher was expected to apply productive engagement by using the questioning techniques so that all pupils are engaged and participate actively).

Throughout the workshop topics, the researcher integrated training with videos featuring a best-practice example of the FAHOTS, followed by tasks related to those videos. For instance, teachers prepared a few lessons plans according to FAHOTS and received feedback from each other and the researcher. Thereafter, the teachers and researcher further discussed the elementary mathematical procedures and HOTS related tasks that could be best integrated into the instruction of a particular LG and SC. To further support the teachers, they were provided with a detailed handout and PowerPoint slides with examples of lesson plans, videos, Clarke (2005)'s five strategies handout to improve the use of questioning and some practical strategies for effective formative feedback to try out in the classroom. With this training in mind, the teachers were expected to practice FAHOTS in mathematics lessons.

Additionally, the researcher created the WhatsApp social media group to reach teachers for regular supportive text messages on FA and announcement of on-site visit for classroom observation.

To facilitate frequent assessments followed by immediate instructional feedback, the researcher used a FAHOTS intervention which consisted of four daily FAHOTS cycles and a weekly circle as illustrated in Figure 4.6.

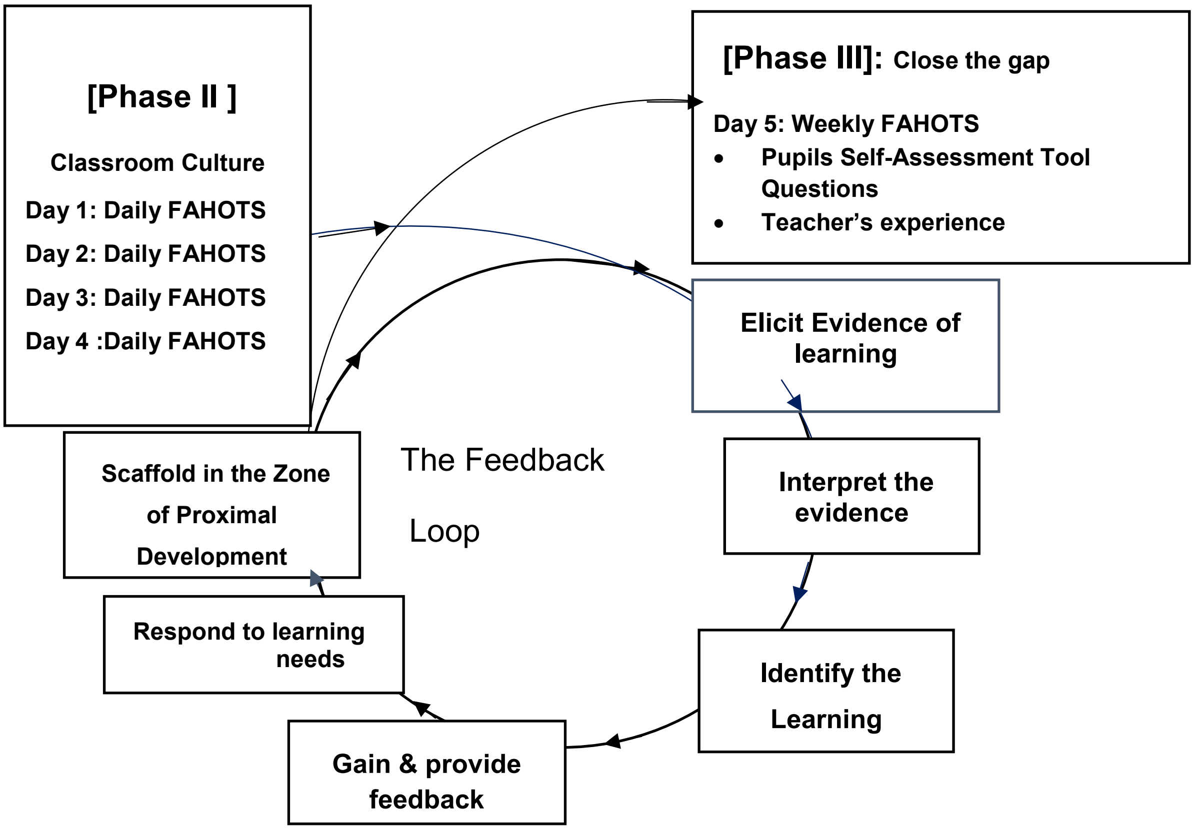


Figure 4.6:The FAHOTS model of four daily FAHOTS cycles and weekly cycle extracted from Kivunja's Feedback Loop

On a five-day timetable (Figure 4.6), the teachers were expected to plan for the daily lesson, which mainly focused on the lesson objective to be taught (LG), SC and assessed at a class level. The teachers had to explain both the LG and SC, at the beginning of the lesson. This was followed by providing a short whole-class discussion instruction that was focused on this LG and used an appropriate basic mathematical procedure or representation and concrete teaching materials.

During, the class discussion teachers should assess the pupils with at least one higher-order thinking question. Hereafter, the teacher assessed pupils' mastery of the LG by giving them either as individuals or as small group of 4-pupils specific tasks related to the LG, such as obtaining change for up to 100 Botswana Pula (BWP) and determining the equivalent amount of money up to BWP 100 using coins and notes.

The teacher observed the pupils while they were working on the given task to permit the teachers to gain more insight into the pupils learning and provide more effective immediate feedback. In this way, effective feedback serves as a good indicator of pupils' progress and help them to identify where learning problems lie and how they can be addressed to improve learning (Kivunja, 2015).

To enhance pupils' classroom participation, the teachers were expected to use Wiliam's (2011) strategies for eliciting evidence of learning and include activities such as questioning, observations of pupil's work, monitoring instructional tasks (for example representations, explanations, performance, problem-solving), mid-lesson checks (for example, thumbs up/down, ABCD cards, whiteboards, traffic lights), exit cards, notes to the teacher and/or quizzes, increased wait time, follow up on answers (correct and incorrect); some closed questions to open questions; asking questions that require pupils to evaluate their work/ or reason or generalise at their current level of mathematical development (Kivunja, 2015; Wiliam 2011). These types of questions can challenge all pupils, but they have a particularly significant role in challenging more mathematically able pupils.

At the end of the week, the teachers assessed the pupils' understanding of the LG using Peer Assessment and Self-Assessment tool Questions to be written in their record book. A pupil's engagement involved setting the criteria for peer assessment activity and clear rules linked to the learning goal. Below were some of the rules which applied pupils for peer assessment to be successful (Davin & Naude, 2017; 111);

- Involve pupils in identifying successful feature for learning (what did your friend do that was really good, e.g. write number symbols legible and neat work)
- Respect the work of others (do not laugh at, tease or embarrass your friend when he or she has made some mistakes)
- When suggesting improvements, think about the learning objective of the activity. (What was your friend supposed to do or achieve with the exercise? What can he or she perhaps try next time that will help him or her to achieve greater success?
- Formulate suggestions positively (do not use negative language such as "poor work". (Davin & Naude, 2017, p. 111).

The example of a peer assessment Rubric depicted in Table 4.23 was adopted and recommended by teachers to suit the need of the assessment activity.

Table 4.23: Example of a peer assessment rubric for mathematics

Criteria for peer assessment **Content Area**: Numbers and Operations **Topic**: Solve word problems involving addition and subtraction		Tick the applicable box
What I like about my friend's work	Neat	[ ]
	Easy to read	[ ]
Next time my friend	Use a sharp pencil	[ ]
	Show the working for the answer	[ ]
	Write the following number symbols correctly	[ ]

Source: (Davin & Naude, 2017).

Figure 4.7 (below) is a self-assessment as an example of problem-solving in a mathematics tool that was adopted for another mathematical content area, e.g. pattern, function and measurement (Davin & Naude, 2017; p. 112).

At the end of the Peer and Self-Assessment tool Questions, teachers were expected to analyse the records of the pupils' work to decide which pupils needed some more instructional feedback and remedial work.

Mathematics self-assessment: Problem Solving Standard 4

Name: ______________________ **Date**: _____________________

New Mathematics words that I have learned this week

____

I used the following mathematical symbols to help me calculate:

____

Next time when I solve the same kind of problem, I will try to do the following:

Use a counting frame____

Draw pictures____

Use a number line____

Break down number in tens and units___________________________________

Figure 4.7: Example of a self-assessment in Mathematics
Adapted from Davin and Naude (2017)

5 CHAPTER FIVE: DESIGN OF MATHS TESTS AND PUPIL MATHS PERFORMANCE BEFORE AND AFTER INTERVENTION

In Phase I, Chapter four, a baseline survey of this study focused on investigating the current practices in FA and teaching of HOTS in Botswana primary schools in an attempt to inform Phase II. Based on the understanding of the FA practice findings, Phase II of the study aimed at developing an intervention that infuses FA strategies to enhance the teaching of HOTS in mathematics.

The outline of this chapter includes Section 5.1 which describes the design used in Phase II, Section 5.2 details the sampling employed. Section 5.3 entails the process involved in pupils' assessment instrument development while Section 5.4 and Section 5.5 cover the procedures and analysis employed in Phase II of the study respectively. The last three sections present the findings (Section 5.6), findings (Section 5.7) and discussion (Section 5.8).

5.1 Research Design

In this study, as mentioned in Chapter 3, Section 3.2.2, the mixed method design was employed as a sequential embedded approach with a single-group design. Specifically, a single-group pre-test, intervention (training) and post-test design was employed as shown in Figure 5.1. This design involved outcome measures that provided a basis for comparing an intervention from time one to the next, that is from a baseline to the post-intervention (Kratochwill & Levin, 2010).

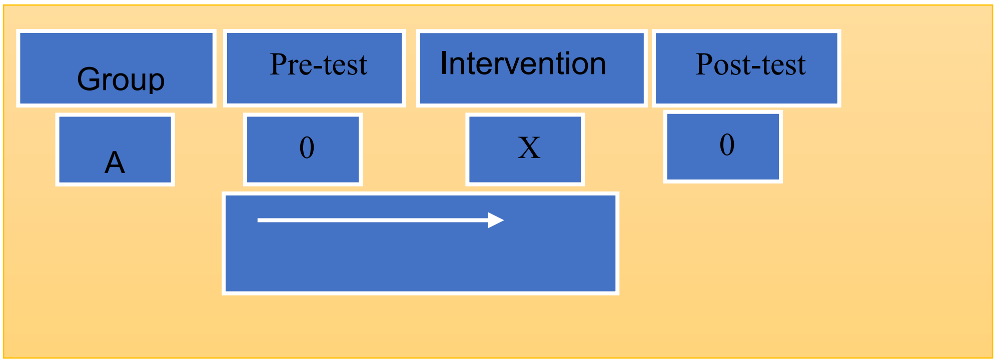


Figure 5.1: Single-group pre-test-post-test design.
Adapted from McMillan and Schumacher (2006)

For Phase II of the study, there was no control group to compare with the experimental group because the nine classes were ethically supposed to receive equivalent instruction. In this case, it implies that the experimental group was not randomly selected, rigorously controlled and extraneous variables were possible; therefore, the quantitative part of this phase of the study was not a true experiment. In other words, convenient samples of participants were used to address the contextual problem identified in Chapter 1 (McMillan & Schumacher, 2006).

Some important considerations that could be threats to this phase of the study's validity as depicted in Figure 5.1. Firstly, the measurement procedure's possible effect on the manipulated variable's construct validity was a very important consideration. Even the historic seminal work by Campbell (1975) affirmed that several other factors apart from the manipulated variable may contribute to the change in scores. Observed changes may be due to reactive effects, not the manipulated variable (Kerlinger & Lee, 2000). However, for this study, an intervention focused on formative assessment for HOTS (FAHOTS) as a manipulated variable and was appropriately defined based on meaningful constructs that represent the focus of the intervention effort. The intervention programme was developed by the researcher (See details in Section 5.3) as informed by Kivunja's assessment feedback loop (Kivunja, 2015) and the baseline survey findings. Many scholars have used and statistically affirmed FA strategies and associated feedback loops in various contexts as providing a means of collecting data that constitutes evidence of pupils learning, helps to identify gaps between what pupils know and what they should know and provides guidance on how to respond to pupils' learning tasks (HOTS) to close the knowledge gap (Black & Wiliam, 2003, 2006; Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006; Kanjee, 2017; Kivunja, 2015; Kyaruzi et al., 2020; Kyaruzi et al., 2020; Ramsey & Duffy, 2016; Weurlander, Soderberg, Scheja, Hult, & Wernerson, 2012; Wiliam, 2013; Wiliam & Thompson, 2014).

Secondly, history and maturation are also possible threats to the single-group pre-test-post-test design's validity when the dependent variable is unstable, particularly because of maturational changes (Cook & Campbell, 1979; Kerlinger & Lee, 2000). Kerlinger and Lee (2000) explain that mental age increases with time, which can easily affect achievement. People can learn in any given interval, and this learning may affect dependent variables. The longer the time interval in the investigation, the greater the possibility of extraneous variables influencing dependent variable measures (Shadish, Cook & Campbell, 2002). However, McMillan and Schumacher (2006) suggest that if the time between the pre-test and post-test are not too long or increasing, it may no longer be a threat. They suggest that if the time between the pre-test and post-test is relatively short (at least two or six weeks), then maturation is not a threat. Thus, to reduce the chance of maturation, the researcher minimised the intervention time between the pre- and post-test, which was eight weeks, and implemented all the tests between 20 May and 20 September 2019. In general, Botswana has evidence of pupils' low mathematics achievement in both national examinations and international large-scale assessment (BEC, 2017; Masole et al., 2016); hence history and maturation may not be a major threat in the current study. Though to see if teacher professional development help pupils would need much longer time.

There are other possible threats apart from those mentioned above, which may affect the intervention's implementation integrity (or "treatment fidelity" (Hagermoser, Sannetti & Kratochwill, 2005) as typically scheduled to determine whether the actual intervention was implemented as intended throughout the experiment. However, it is important to note that Phase II of the study, is embedded in the mixed-method approach in a complementary manner to compensate for its weakness (Gray, 2014). For this reason, a single-case intervention approach in the current study was assumed to have minimised some threats to the validity of inferences concerning the researcher's interpretation of the effects observed and analysed. For instance, some insights into what teachers in the schools were doing during the mathematics lessons, the baseline survey and pupils pre- and post-test was helpful to determine the level of FA practices and pupils' achievement respectively during Phase I of the study. After that, the FAHOTS condition was made available to the intervention schools. Some schools were also uncertain about taking part in a study and engaging in observations, testing, and interviews as they would get disrupted in their daily delivery. However, the researcher guaranteed teachers of a minimum disruption during the project period and provided low impact intervention to be able to observe the teachers in class. This intervention embedded design discussed in this chapter, guided by the following sub-research questions (SRQ);

- SRQ 3. What are the current pre-intervention levels of mathematics achievement in HOTS items for Standard 4 pupils?
- SRQ 4. To what extent does the FA strategies intervention enhance Standard 4 pupils' HOTS academic achievement in mathematics when comparing the pre- and post-intervention?

The following section first explains how the participants were selected, and specify their characteristics. Next, it describes the outline of the condition of the intervention. Finally, it presents the instruments and data analysis methods that were used.

5.2 Sampling and Participants

This section discusses the two sampling units for Phase II.

5.2.1 First-stage sampling unit

At this stage of sampling, the unit was focused on getting pupils to be exposed to the pre-achievement assessment and for the researcher to observe the teacher's administration of FA as part of the Phase II exploration of this study. Here, in the already sampled schools, the researcher purposive randomly sampled nines schools, in which the pupils' schools were selected. Each sampled school has a list of eligible classrooms for the Standard 4 pupils which were picked randomly using an online Number Generation. Thus, one class was selected in each school.

5.2.2 Second-stage sampling unit

The second stage sampling unit was focused on the pupils within the classrooms. This study addressed all the pupils who have enrolled in Standard 4 levels. Once a classroom is sampled, the whole intake of such a Standard 4 class and pupils participated in the study that is usually referred to as a cluster. According to Nenty (2013), clusters are representatives of the population under study. For example, classrooms are sections of the population, which has a close semblance in characteristics to the entire population of such clusters, in this case, the classes of Standard 4 pupils.

The sampled teachers who have already answered the survey questionnaire that was administered as part of Phase I were followed through their selected classes of pupils in respective schools. A Standard 4 class was selected purposively random, and the teacher of such a class was also eligible to participate in the study. The class size ranged between 17 to 38 pupils per classroom of the Standard 4 levels depending on the size of the school. Pupils' responses were gathered from each script and processed through excel. Data from pupils who did not participate in both pre- and post-test administration were removed from the analysis. Complete data from 272 pupils were obtained. Matched data totalled for each school is outlined in Table 5.1. The matched data sets were used to determine each pupil's individual ability, as well as the difficulty of the items. The average ability level for each participating school was also calculated.

Table 5.1: A sample of Standard 4 pupils involved in the study

School	Pre- and post-Test		
	Male	**Female**	**Total**
303	13	16	29
305	9	16	25
306	14	18	32
307	16	7	23
308	20	18	38
309	12	12	24
314	17	14	31
319	19	14	33
320	20	17	37
Subtotal (n)	140	132	272

The design of the study was a single-group pre-test-post-test. It assumed some level of similarities which included, all schools being from one sub-region, the age equivalent of cohort pupils for Standard 4 level. The schools used a common national Standard 4 mathematics syllabus and sat for the common end of term achievement tests, and all the teachers at least had a diploma qualification in education. The aforementioned features possibly guarantee the participating schools to be similar and to some extent, enhanced the internal validity of the study.

Notwithstanding the sampling process in Phase II, it is also interesting to note that despite the purposive random assignment of the classes to participate in the intervention, some schools' management influenced the selection in the current study. During the school debriefing with the individual school management, they had preferred and insisted that those classes in which pupils were weak mathematically should be given a priority. They have considered the intervention as opportunities for teacher training and professional development that justified the involvement implied by this request. However, the ministry didn't isolate schools and classrooms from any regular contact.

The other factor that influenced class selection for the intervention was the teachers' level of seniority. All classes for the head of departments did not part-take in the intervention, that is only senior teachers and teachers participated in the study. In such situations, the assignment of classes to the intervention was generally based on the weaker classes, especially where the school had more than one cohort class and with seniority of the teachers as well as those teachers who needed to be assisted in teaching mathematics.

Figure 5.2 illustrates the diagrammatic layout of Phase II. As can be seen in Figure 5.2, reflects the stages which were employed. Thus, Figure 5.2 constituted selected schools participating in the study. Next, Figure 5.2 shows the nine teachers selected who were trained and then implemented the FA in their mathematics lessons, as well as the participation of their pupils in the pre-achievement test. The last stage involved pupils' post-achievement test and semi-interview of the seven teachers, and only two teachers missed the interview during the survey due to illness (See the detailed presentation in Chapter 6, Phase III).

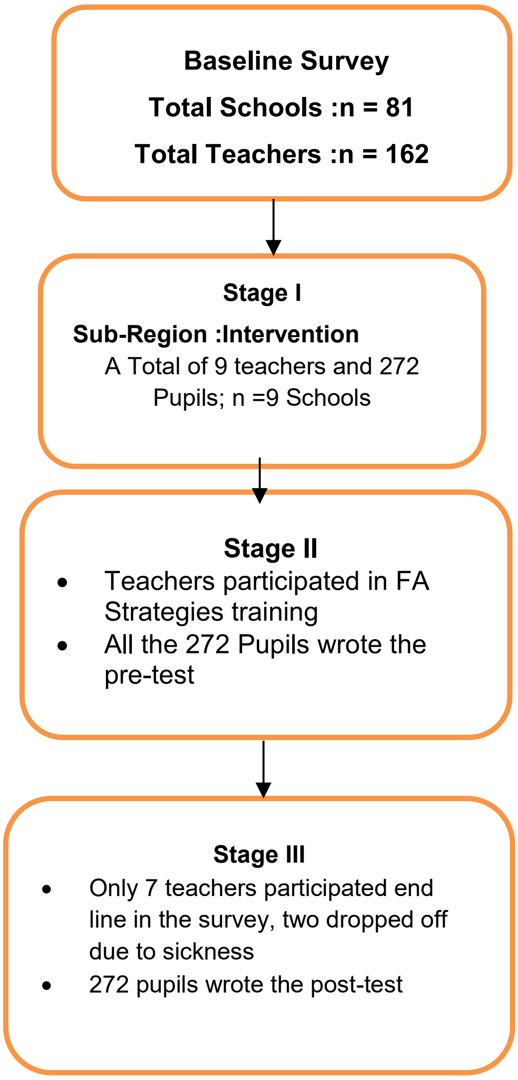


Figure 5.2: Participants through single group pre-test-post-test design of the study

5.3 Test Instruments' Development

This section discusses test instruments which were used in the study. The discussion involves the role of the pre-test and post-test, development and the quality criteria of the instruments.

5.3.1 The role of pre- and post-assessment

The pre- and post-assessment for Phase II of the study was used as assessment tools to compare the gain scores of the pre- and post-achievements of pupils in mathematics. The gain scores were expected to reveal the knowledge acquired by each school as well as the extent to which the FA intervention enhanced pupils' achievement from pre- to post-testing. For this study, the pre/post-HOTS items were adapted from the Botswana Examination Council (BEC) tests as they were already available and administered nationally instead of developing the researcher's new tests. While taking the pre-test at the beginning of the intervention, pupils were not expected to know the answers to all of the questions, however, they were expected to utilise previous knowledge to predict rational answers. The pre-test was also considered as a baseline study to ascertain pupils' weakness and strength in HOTS items before the intervention. According to Guskey (2018), pre-test or pre-assessment in research studies is to establish a baseline of performance from which pupils' growth or learning can be gauged.

When taking the alternative equivalent test called a post-test at the end of the interventional period, pupils should be expected to answer more questions correctly based on an increase in knowledge and understanding to claim possible effects of the intervention. The intended purpose of the post-test was to be administered to the same pupils after the intervention. The pupils wrote the end-line test to determine the influence of the teachers' FA instruction on pupils' outcomes in HOTS items by comparing it with the written answers and results of the baseline test. For this reason, the post-test had a similar version and an equal number of items with the pre-test provided. Ten (10) common items from the pre-test were retained in the post-test to ensure the similarity of the two tests. This common-item approach linked the two tests and accorded a comparison of the performance of pupils on different tests represented on a common scale.

5.3.2 Pre- and post-assessment Instruments

It is always the purpose of an assessment exercise that determines the subject matter contents to be considered, the specific learning objectives to be measured, as well as nature and difficulty level of the assessment items. Generally, the two purposes of testing are to measure maximum performance and to measure typical behaviour (Moyo & Nenty, 2017). The practical framework within which cognitive educational objectives could be organised and measured are provided by Bloom's taxonomy (Bloom, Engelback, Trust, Hill & Krathwohl, 1956 as cited in Nenty & Umoinyang, 2004). Bloom's taxonomy involves those objectives that have to do with the development of intellectual or cognitive abilities and skills and categorises behaviour into six hierarchical levels from the simple to the most complex.

For this study, the researcher adopted the HOTS items from BEC, which were developed based on the original Bloom taxonomy's Levels and measured the curriculum societal needs for Botswana mathematically. The test items were collected from BEC in 2015, 2016, 2017 and 2018, for the Standard 4 Attainment tests. The items were constructed to collect data for BEC standardised examinations which are administered to all candidates across the country.

At BEC, the Directorate of Product Development and Standards (PD&S) is mandated to direct the development and implementation of operational policies, strategies, processes, and procedures for the development of examinations, assessments, and operational documents. Also, the PD&S ensured that appropriate quality and standards were maintained for school and any other examinations and assessments. Based on the Botswana Standard 4 mathematics curriculum outcomes, PD&S designs and develops assessment and examination programmes to determine whether the intended outcomes are achieved as well as to provide evidence on areas needing improvement based on analysis of the results. The examination data assists in determining pupils' levels of achievement. It is also used to determine whether the system is achieving its educational goals and can also be used to benefit curriculum development processes important for continuous improvement.

Test development for BEC involves recruitment of the specialists who are required to have experience and expertise in the subjects for primary schools teaching, as well as knowledge of setting assessments for classroom use and systemic testing. Those specialists then design test drafts that are aligned to the learning outcomes of the curriculum. After the development of the test drafts, the items are subjected to trial testing. This testing is an exercise in which procedures and instruments are pretested on a proportion of the target population. In a typical trial test, instruments are administered to a small group of individuals that has similar characteristics to the target population. Through trial testing, it is therefore ensured that appropriate questions are asked, a suitable data collection is used, and the right data is collected.

According to BEC, the objectives of trial testing are:

- Identify and get rid of mistakes that may compromise the quality of live instruments.
- Make corrections and necessary adjustments and revisions to the procedures.
- Determine how long it may take to collect information using the instrument in question.
- Establish how easy or difficult it may be to answer questions or supply the required information.
- Improve the usefulness, reliability, and validity of data collected for performance measurement.

BEC is responsible for setting, administering, and marking the examinations at the end of both primary and junior secondary schooling. Therefore, BEC ensures that setting, administration, and marking are done following the set standards for adequate validity and reliability. Since the scoring or marking of Standard 4 attainment tests is done at the school level, BEC prepares the detailed scoring guides. The scoring guides entail every item in the examination and explanations for the allocations of marks for each constructed-response item. Similarly, this study had applied the same scoring guides, and hence such scores would be obtained from BEC items which were assumed to be valid, reliable, and technically sound and addressed the competencies that had been defined for the Botswana curriculum regarding Standard 4 HOTS in mathematics.

(i) Description of the HOTS items instrument used

Items in the BEC examinations entail the content for Standard 4 of the syllabi including items for HOTS. For this reason, the researcher analysed all the specific learning objectives from the syllabi, according to Bloom’s Taxonomy and reported only those objectives which measure HOTS. Table 5.2 reflects the specific learning objectives for Botswana Standard 4 mathematics syllabi which are expected to measure HOTS.

Table 5.2: Botswana Standard 4 Mathematics specific objectives HOTS levels

Topics and Sub-topics	Specific Learning Objectives for HOTS	Bloom's Taxonomy
Module 1: Numbers & Operations	Pupils should be able to	
1.1 Sorting and classification	***No HOTS specific learning objective was identified***	
1.2 Number	1.2.1.2 use place value of up to Thousands	Application
	1.2.1.6 generate patterns involving even, odd number and multiples of 2, 3, 4, 10 up to 100	Synthesis
1.3 Addition and Subtraction	1.3.1.3 solve word problems involving addition and subtraction	Application
1.4 Multiplication and Division	1.4.1.7 carry out multiplication and division in box notation	Application
	1.4.1.8 solve word problems involving multiplication and division	Application
1.5 Fractions	1.5.1.3 order set of fractions with the same denominators in terms of value.	Analysis
	1.5.1.6 solve word problems involving addition and subtraction of fractions with the same denominators.	Application
1.6 Money	1.6.1.3 determine the equivalent of up to P100 using coins and notes	Application
	1.6.1.4 perform the four basic operations involving money up to P100	Synthesis
	1.6.1.5 obtain change from up to P100	Analysis
	1.6.1.6 make simple bills sum up P100	Synthesis
	1.6.1.7 solve word problems involving money up to P100	Application
Module 2: Geometry		
2.1 Angles	2.1.1.1 draw two straight lines that meet/cross each other to show angles	Analysis
Module 3: Measures		

3.1 Length	3.1.1.5 discover the formula for finding the perimeter of a square and rectangle	Application
3.2 Area	3.2.1.1 discover the formula for finding areas of square and rectangle	Application
	3.2.1.2 find areas of squares and rectangles in cm^2 using formula	Application
3.3 Capacity	3.3.1.1 use standard instruments and units to measure capacity	Application
	3.3.1.2 measure, compare and record capacity of the given container in l and ml	Application
3.4 Mass	3.4.1.1 use standard instruments to measure the mass of objects in g and kg	Application
	3.4.1.3 use the four basic operations in solving mass problems.	Application
3.5 Time	3.5. 1.4 read and use simple time -tables involving same day and single journey	Application
Module 4: Problem Solving		
4.1 Games and Puzzles	4.1.1.1 play games	Application
	4.1.1.2 solve puzzle	Application
4.2 Investigations	4.2.1.1 carry out simple investigations involving numbers, geometry, and measures	Application
	4.2.1.2 conduct a simple research project	Application
Module 5: Statistics		
5.1 Graphs	5.1.1.1 Conduct a simple survey on emerging issues and the environment	Application
	5.1.1.2 Represent data in a tabular form	Comprehension
	5.1.1.3 draw pictographs and bar graphs using information from the survey	Analysis
	5.1.1.4 Interpret information given in the graph	Comprehension
	5.1.1.5 find mode from a given data	Comprehension

The Botswana Standard 4 curriculum objectives contain thirty objectives for measuring HOTS (BEC, 2002). Specifically, the curriculum revealed twenty objectives for application, four for analysis, three for synthesis and three tied to knowledge objectives (BEC, 2002). This result revealed that the evaluation domain is not represented in the curriculum, and this proportion of the domains may likely reflect in item development at the classroom level. Following the analysis of the objectives based on Bloom's Taxonomy, the researcher reviewed items from the Botswana Standard 4 national past examination papers and adapted all the items which were measuring HOTS. These items were used to assemble the pre-test.

A complete draft test was assembled and comprised of HOTS items that required the awarding of partial credit and gathering of non-dichotomous data as used by BEC. The item analysis matrix of the draft test was done to ascertain the extent to which items covered all five domain contents: number and operations, geometry, measures, problem-solving, and statistics, as reflected in the Botswana Standard 4 syllabi. The process of item validation involved the consultation of the two primary school teachers in Botswana and the examiner at BEC. The Angoff method (1971) was used to determine the level of agreement among the reviewers. Thus, an estimation of how well pupils would perform per item was developed and rated through a probability borderline scale ranging from 0 to 1.0 for each item by three judges. If a judgement for an item differed by 20% or more, those judges who provided the high and low scores might lead a discussion of their rating (Yudkowsky, Park & Downing, 2020). For this study, the Angoff rate and passing score[3] were 0.59 (59.0%), and judgement did not differ by 20% across the reviewers (*Appendix F*). This validation of items and input from research supervisors was imperative for bettering the final draft of the test.

(ii) The pilot of the draft test

[3] The case passing score (percent) is the average of passing scores for all items. In this study, the Angoff ratings for a 13-item for HOTS rated by three Angoff judges that is (7.70/13) *100%=59.00%.

Given that the composition of the draft test was HOTS oriented items from BEC past examination, which already have gone through the process of test trials and hence pilot in this study was done mainly for instructional purposes. The pilot study was intended to assess whether the draft of assembled test items was clear, well arranged and time allocated to the draft text as well as to ascertain the suitability for the current study. Creswell (2012) advises that participants whose characteristics are similar to the sample will ensure successful pilot results. As for this study, the pilot involved two pupils who were currently doing Standard 4 and from different education region beyond the researcher's current investigation setting. Pupils were writing the test under the timed condition and invigilated by their parents who happened to be teachers in primary schools. The results of the pilot items revealed that some items were halfway solved and some not attempted at all by the pupils within 60 minutes of the allocated time for the whole test draft.

Since the items were adapted from the national examination, they were treated as raw items, hence statistical analysis was not involved during the pilot phase. The test pilot results focused on scrutinising the level of pupils' attempts for each item, revised and replaced items, particularly those the pupils did not attempt, and the time allocated was revised from 60 to 75 minutes. In general, the whole paper's font sizes were increased so that the test looked as user-friendly as possible to enable pupils to answer what was asked. The test was also aligned to the pupils' reading level of problem-solving-related items since the pupils graduated from Standard 3 and advanced to Standard 4 levels (in their second term of years). Finally, most of the items were selected from the curriculum, which had already been covered in the first term of the calendar and those items related to contents covered in Standard 3.

(iii) Description of the final version of the pre-assessment

The final version of the pre-test mathematics achievement was assembled. It consisted of 13 items which were organised into the five domains as prescribed in the Botswana Standard 4 syllabus and measured pupils' HOTS. Each of the 13 items consisted of open-ended questions or problems that merited partial credit when marking following the BEC's scoring guidelines. The total number of items per question ranged from one to three, making a total of 21 pre-test items (see *Appendix Gi*).

Since the items were adapted from the BEC examination, so to say, the marks assigned for each item was also not altered from the original format. Thus, the BEC marks allocation was deemed to be consistent towards the rigour of the work involved to attain the correct answer and complexity of the concept assessed. The total composite marks for the whole paper was 40. Table 5.3 shows the marks across the test items aligned according to sub-topic, specific learning objectives, and cognitive behaviours.

Table 5.3: Higher-order thinking skills from Botswana Examination Council previous years papers

Question	Marks Allocation	Sub-topic	Specific Objective for HOTS	Level of Bloom Taxonomy
Q1	Q1a =2 Q1b =2	1.2 Number	1.2.1.6	Synthesis
Q2	Q2 =4	1.2 Number	1.2.1.6	Synthesis
Q3	2	3.1 Length	3.1.1.5	Application
Q4	2	3.4 Mass	3.4.1.3	Application
Q5	2	1.3 Addition & Subtraction	1.3.1.3	Application
Q6	2	1.4 Multiplication & Division	1.4.1.7	Application
Q7	2	1.5 Fraction	1.5.1.6	Application
Q8	2	1.6 Money	1.6.1.4	Synthesis
Q9	Q9a = 2 Q9b = 2	1.6 Money	1.6.1.6	Synthesis
Q10	Q10a = 1	5.1Graphs	5.1.1.3	

Question	Marks Allocation	Sub-topic	Specific Objective for HOTS	Level of Bloom Taxonomy
	Q10b = 2			Analysis
	Q10c = 2			
Q11	1	4.1 Games & Puzzles	4.1.1.2	Application
Q12	Q12a = 1	3.5 Time	3.5.1.4	Application
	Q12b = 2			
	Q12c = 2			
Q13	Q13a = 1	1.6 Money	1.6.1.7	
	Q13b = 2			Application
	Q13c = 2			
Total	= 40			

All the test items were aligned to the component of the study's conceptual framework, which focused on teaching HOTS in mathematics. Table 5.4 shows the blueprint of the pre-test items tied to the content domains and Bloom's taxonomy as a way to validate the measures of the HOTS learning outcomes per domain assessed and prescribed in the Botswana Standard 4 syllabi. Table 5.4, shows that operation and numbers items contributed 62.5 % by score weights, followed by 22.5% for measures and 12.5% for statistics. Only 2.5 % was for problem-solving score weight, while there were none for geometry. These score weights distribution for the Standard 4 examinations do not fairly cover all domains in the curriculum. In other words, BEC items did not fairly balance the setting of mathematics items for the national attainment examination at the Standard 4 level.

Moreover, as observed in the table of item specification, the items were comprised of 12 items (54.5 %) for application, three items (13.3%) for analysis and seven items (31.8%) for synthesis. The proportion of items for HOTS seems to be dominated more by the application level of cognitive behaviour than other levels (analysis, evaluation and synthesis). This notion is consistent with Nenty and Umoinyang (2004) who posited that ranking and weighting items should be based on the level of cognitive behaviour whose development is deemed a priority given the class or age-level of pupils for which the test is intended. As for the targeted participants in this study, these are pupils in their last year of lower primary school, and it is a foundation level; hence Nenty and Umoinyang suggested that the test at this level should place emphasis on the development of comprehension/understanding and application/problem-solving. The blueprint in Table 5.4 is therefore appropriate for Standard 4 pupils as it is subjugated to the application of HOTS in Mathematics.

Table 5.4: Specification for the development of a 22-item on Higher Thinking Skill for Standard 4 Mathematics from Botswana Examination Council previous years papers

Subject Matter Content	**Level of Higher Order Thinking Cognitive Behaviour**					
	Application	Analysis	Synthesis	Evaluation	Total	Score Weights
Number & Operation	Q5, Q6, Q7, Q13a, Q13b, Q13c		Q1a, Q1b, Q2, Q8, Q9a, Q9b		12	25 (65.5%)
Geometry						0 (0.0%)
Measures	Q3, Q4, Q12a, Q12b Q12c				5	9 (22.5%)
Problem Solving	Q11				1	1 (2.5%)
Statistics		Q10a, Q10b, Q10c			3	5 (12.5%)
Total	12	3	7		21	40 (100%)

(iv) Description of the final version of the post-assessment

The post-test was also assembled on a similar version following the pre-test procedure (see *Appendix Gii*). Thus, all the test development procedures also included the extraction item from BEC past examinations, items aligned to HOTS of the Blooms' taxonomy, experts' consultation and align the items to the objectives of Botswana Standard 4 syllabi were met during the drafting of the post-test.
For this study, it should be noted that the test instrument was not designed with anchor items, instead, the researcher assembled items from available instruments (standard 4 previous examination papers for mathematics), and selected the best functioning items as anchors during pre-test results analysis. Thus, a total of ten items were retained from the pre-test as anchor items. According to Sinharay, Haberman, Holland and Lewis (2006), anchor items should focus on (a) content representation, (b) have the same difficulty as the total tests, and (c) the spread of item difficulties should be less than that of the total tests. In addition, the range of item difficulties for anchors is expected to be -2.52 to +3.85 (Sinharay et al., 2006). After psychometric analysis of the pre-test items, the anchor items were selected (see Section 5.4.3, Quality assurance of pre- and post-assessment instrument). The assessment framework is provided in *Appendix I*.

Table 5.5 summarises anchor items which were selected based on the analysis of the model fit (Outfit MNSQ -0.81–0.95) and level of difficulty (ranged from -1.69 to 3.15) in the Winsteps framework during pupils' pre-test result. Seven out of ten anchor items were from numbers and operations, while three items came from statistics content. Thus geometrics, problem-solving and measures' contents were not represented in the anchoring process. However, the HOTS items were covered relatively well for all the cognitive domains, that is five items for synthesis, three items for analysis, and three application items. These anchor items functioned relatively well for this study, given the proportion of items represented in pretesting. It is interesting to note that no one of the pupils in the test got a score above 0.00 but 5 items are much harder than that as anchor items. Also, weakest students core -3.00 but no items below -2 as anchors. These items have no threat to validity since the items were within the curriculum context and during post-test, the content would have been covered fairly.

Table 5.5: Test item for selected as anchor item from pretesting results

ENTRY NUMBER	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
Q13a	3.15	.54	.68	-.64	-.61	1.11
Q13b	2.5	.41	.97	.04	.44	-.90
Q10c	2.2	.37	.93	-.13	.62	-.53
Q10b	1.13	.25	1.01	.09	.72	-.68
Q9b	1.07	.25	.79	-1.2	-.64	1.98
Q8	.16	.19	.97	-0.2	.84	-.68
Q6	-.57	.16	.98	-.19	.93	-.42
Q9a	-.57	.16	.90	-1.31	-.81	1.31
Q10a	-1.07	.15	.95	-.77	.95	-.41
Q1b	-1.69	.09	.93	-.84	-.81	1.05

N= 272

5.3.3 Quality criteria for pupils' pre- and post-assessment instrument

Before delving any further with analysis, it is in the interest of the best practice to determine the quality criteria of the data collection instrument. In this study, some quality criteria were used to assess the pre-test and post-test items. Those measures were cumulative category probabilities, model fit, reliability indices, unidimensionality, and Different Item Functioning (DIF).

(i) Cumulative category probabilities

The first step involved was the diagnosis of the response category function. For this diagnosis, all items were evaluated. Those that did not satisfy the requirement of INFIT and OUTFIT MNSQ, greater than 1.3 and unweighted percentages less than 10%, were identified and collapsed (see *Appendix Hi & iv*). Wright Maps with Thurstonian, item and multiple item characteristics curves (ICCs) analysis were also further employed to understand how the raw data was balanced on the scale (See *Appendix Hii & iii)*, hence some items categories which did not function well were collapsed.

(ii) Model fit analysis

The MNSQ infit and outfit values within the Rasch framework are also used to check the item and person fit of the model. Bond and Fox (2012) assert that for the polytomous scale, like for this current study, the acceptable MNSQ outfit value has to be in the range of 0.6–1.40. These figures suggest that any item identified below and above the range has to be removed from the analysis as it is considered a misfit. In this study, three items (Q11, 12c, and 12b) from the pre-test were found out of range, while the remaining items were productive measurements (See Table 5.6).

Table 5.6: Item model fit for pre-test results for pupils' achievement in HOTS

		MODEL	INFIT		OUTFIT	
ITEM	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD
Q12c	4.01	.60	1.13	.40	3.34	3.00
Q12b	.18	.14	1.28	3.10	1.99	4.40
Q11	.71	.16	1.07	.70	1.55	2.10
Q4	-2.71	.10	1.11	2.40	1.38	4.00
Q7	-2.64	.10	1.19	4.10	1.32	3.50
Q2	-.03	.13	1.00	.00	1.24	1.50
Q5	-1.8	.10	1.05	1.20	1.11	1.50
Q13c	2.70	.33	1.10	.40	.70	-.50
Q1a	-1.55	.07	1.07	1.30	1.08	1.00
Q3	-.13	.13	.98	-.30	1.04	.30
Q13a	3.15	.54	.68	-.64	-.61	1.11
Q13b	2.50	.41	.97	.04	.64	-.90
Q10a	-1.07	.15	.95	-.77	.95	-.41
Q1b	-1.69	.09	.93	-.84	-.81	1.05
Q8	.16	.19	.97	-.20	.84	-.68
Q10b	1.13	.25	1.01	.09	.72	-.68
Q10c	2.20	.37	.93	-.13	.62	-.53
Q6	-.57	.16	.98	-.19	.93	-.42
Q12a	3.30	.43	.90	-.10	.79	-.20
Q9a	-.57	.16	.90	-1.31	-.81	1.31
Q9b	1.07	.25	.79	-1.20	-.61	1.98
MEAN	*0.00*	*.17*	*1.00*	*.10*	*1.09*	*.20*
SD	*1.79*	*.13*	*.11*	*1.60*	*.61*	*2.20*

As for the post-test, only two test items (Q3 and Q7) were misfit (See Table 5.7). These misfitting items were dropped from the next analysis. Even though some items were identified as a misfit, in general, the global fit statistics were not statistically significant, and this finding suggests in general that the data fitted the Rasch model.

Table 5.7: Item model fit for post-test results for pupils' achievement in HOTS

INPUT: 272 PERSON 27 ITEM REPORTED: 272 PERSON 27 ITEM 57 CATS WINSTEPS 4.7.1.0

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM	G
19	63	272	.33	.16	1.17	1.95	1.74	3.37	A .26	.40	76.1	79.3	Q3_POST	0
22	176	272	-1.96	.14	1.35	4.91	1.43	3.18	B .21	.46	59.2	73.4	Q7_POST	0
23	225	272	-3.17	.18	1.15	1.41	1.27	1.22	C .28	.40	83.1	84.6	Q11A_POS	0
25	72	272	.11	.15	1.10	1.29	1.24	1.42	D .32	.41	75.7	77.1	Q12A_POS	0
18	213	272	-.53	.11	1.15	1.93	1.23	2.49	E .46	.55	62.5	60.7	Q2_POST	0
24	216	272	-2.91	.17	1.04	.44	1.14	.75	F .38	.42	82.0	82.1	Q11B_POS	0
12	23	270	1.72	.23	1.09	.57	.96	.00	G .27	.30	90.4	92.0	Q2_PRE	0
11	177	272	-.37	.10	1.08	1.03	1.06	.47	H .54	.57	62.5	57.8	Q1A_PRE	0
16	153	272	-1.50	.14	1.04	.73	1.07	.69	I .43	.46	68.8	70.7	Q7_PRE	0
13	31	263	1.31	.21	1.04	.34	.75	-.78	J .33	.33	88.6	88.9	Q3_PRE	0
26	82	272	-.12	.15	1.03	.47	1.00	.07	K .40	.42	71.7	74.7	Q12B_POS	0
14	154	271	-1.53	.14	.98	-.28	1.02	.25	L .47	.46	70.5	70.8	Q4_PRE	0
1	223	272	-.81	.09	1.00	.03	.99	.00	M .61	.61	55.9	56.0	Q1B_ANCH	0
7	21	271	1.84	.24	.99	.01	.77	-.56	N .31	.29	92.6	92.7	Q10B_ANC	0
17	106	272	-.61	.14	.98	-.27	.91	-.83	m .47	.45	71.0	70.9	Q1A_POST	0
3	42	272	.93	.18	.93	-.62	.82	-.66	l .41	.36	86.4	85.5	Q8_ANCHO	0
2	66	272	.25	.16	.92	-1.05	.84	-.91	k .47	.40	79.4	78.5	Q6_ANCHO	0
10	7	269	3.12	.40	.92	-.13	.68	-1.20	j .28	.20	97.8	97.4	Q13B_ANC	0
4	66	272	.25	.16	.91	-1.09	.79	-1.26	i .47	.40	81.6	78.5	Q9A_ANCH	0
8	9	271	2.84	.36	.91	-.21	.68	-.75	h .27	.22	97.0	96.7	Q10C_ANC	0
6	86	271	-.20	.15	.89	-1.70	.80	-1.51	g .51	.43	76.0	73.8	Q10A_ANC	0
27	26	272	1.57	.22	.88	-.71	.53	-1.62	f .41	.31	91.5	91.0	Q13C_POS	0
21	184	272	-2.13	.15	.87	-2.04	.83	-1.34	e .54	.45	79.8	74.7	Q5_POST	0
20	229	272	-3.31	.18	.86	-1.23	.68	-1.49	d .50	.39	87.5	85.8	Q4_POST	0
5	22	271	1.78	.24	.83	-.98	.66	-1.80	c .43	.29	93.0	92.3	Q9B_ANCH	0
15	107	271	-.63	.14	.83	-3.22	.73	-2.63	b .58	.45	75.6	70.8	Q5_PRE	0
9	4	270	3.74	.52	.74	-.46	.67	-1.35	a .31	.16	98.5	98.5	Q13A_ANC	0
MEAN	103.1	271.2	.00	.19	.99	.0	.90	-.2			79.8	79.8		
P.SD	77.2	1.8	1.84	.09	.13	1.5	.32	1.5			11.8	11.5		

(v) Reliability analysis

Table 5.8 also shows high item reliability for both instruments, namely pre-test .98 and post-test .99, respectively. These findings for internal consistency of the items making up the instrument was high and that items were consistent with each other. According to Bond and Fox (2012), the accepted item reliability cut-off is at 0.7. The persons and item separation of both pre-test and post-test data are summarised in Table 5.8. They show that the item separation for the pre-test and post-test data was greater than 2 for the purpose of this analysis (Bond & Fox, 2012). However, the person separation for the two instruments were less than 2 for both analyses.

Table 5.8: Item and person reliability for pre- and post-test analysis

Time	Variable	Reliability index	Separation	OUTFIT MNSQ	OUTFIT ZSTD
Pre-Test	Item	.98	8.05	1.09	.20
	Person	.71	1.57	.98	.70
	Person raw Score	.79			
Pro-test	Item	.99	8.41	.90	-.18
	Person	.79	1.95	.90	.05
	Person raw Score	.89			

These results mean that the assessment instruments had excellent item reliability. The items (for pre-test and post-test) could be separated into nearly eight groups according to pupils' responses. The scale of measure is probably low because the people are very much grouped between -3.00 and 0.00 logits, whereas the items spread from -1.6 to 2.6. On the other hand, the person reliability was just marginally acceptable, and the pupils could be separated into almost two groups by the items in both assessment instruments (See Table 5.8). Perhaps the intervention increased the spread of the scores. According to https://www.winsteps.com/winman/reliability.htm, person separation is used to classify people. Low person separation (< 2, person reliability < 0.8) with a relevant person sample implies that the instrument may not be sensitive enough to distinguish between high and low performers. For this study, both the tests could discriminate the sample into at least 2-3 enough levels. However, more items may be needed.

Moreover, the Winsteps "person reliability" is equivalent to the traditional "test" reliability. Low values indicate a narrow range of person measures or a small number of items. To increase person reliability, testing persons with more extreme abilities (high and low), and lengthening the test. Improving the test targeting may help slightly. In this study, the person raw scores reliability for both pre- and post-test were well above the acceptable threshold (Bond & Fox, 2007).

(vi) Unidimensionality and local independence

The test of unidimensionality of the two assessments was conducted through a Principal Components Analysis (PCA) of Rasch residuals in the Winsteps software (Linacre, 2016). According to Boone, Staver and Yale (2014), the Rasch-PCA residuals are used to detect more than one factor that can explain the response structure (i.e., unidimensionality) by comparing the differences between the observed and the expected responses. Linacre (2016) recommends researchers should use the eigenvalue of the first contrast of less than 2.0, and the amount of the variance explained by measures should be greater than 20% as an acceptable criterion for establishing unidimensionality. In this study, the raw variance was explained by the Rasch measures for both assessments which were 44.90% and 39.10%, respectively (Table 5.9). The results indicated that the mathematics assessment instrument was likely to be underpinned by a single dimension, consistent with the assessment design intent. Additionally, the local independence of the test items was also independent of one another; no item for anyone answer led to an answer on another. Thus, all items correlation pairs were less than 0.7, (Pre-test; -.14 - .25 and post-test -.14 - .43).

Table 5.9: Dimensionality analysis of pupils' pre- and post-test results

	Pre-test results			**Post-test results**		
Variance	**Eigenvalue**	**Observed**	**Expected**	**Eigenvalue**	**Observed**	**Expected**
Raw variance explained by measures	17.0801	44.90%	44.90%	19.5957	42.10%	43.10%
Raw variance explained by persons	3.9749	10.40%	10.50%	5.97890	12.80%	13.10%

	Pre-test results			Post-test results		
Variance	**Eigenvalue**	**Observed**	**Expected**	**Eigenvalue**	**Observed**	**Expected**
Raw Variance explained by items	13.1052	34.40%	34.50%	13.6168	29.20%	29.90%
Raw unexplained variance	21.000	55.10%	55.10%	27.00	57.90%	57.90%
Total raw variance in observations	38.0801	100.00%	100.00%	46.5957	100.00%	100.00%
Factors	**Eigenvalue**	**Observed**	**Expected**	**Eigenvalue**	**Observed**	**Expected**
Unexplained variance in 1st contrast	1.7758	4.70%	8.50%	1.9024	4.3%	7.0%
Unexplained variance in 2nd contrast	1.5148	4.00%	7.20%	1.6441	3.7%	6.1%
Unexplained variance in 3rd contrast	1.4659	3.80%	7.00%	1.4282	3.2%	5.3%
Unexplained variance in 4th contrast	1.2827	3.40%	6.10%	1.4212	3.2%	5.3%
Unexplained variance in 5th contrast	1.2689	3.30%	6.00%	1.3003	2.9	4.8
Raw unexplained variance (total)	21.000	100.00%	100.00%	27.000	100.00%	100.00%

(vii) Differential Item Functioning

Assessing Differential Item Functioning (DIF) is one of the vital indicators of quality consideration for instrument fairness. Thus, DIF tends to determine test-takers' ability in some measured latent trait that has different probabilities of achieving a correct response to an item (Moyo & Nenty, 2017; Wang, 2008). The DIF magnitude translates the extent to which the item parameter differs across the groups under comparisons such as gender, location, item content, and socioeconomic status (Moyo & Nenty, 2017). According to Boone et al. (2014), the Rasch model has a modern and more effective approach for DIF analysis using the Mantel-Haenszel probability. All the items from the two instruments (p-values for Mantel-Haenszel test) had no significant difference for DIF for gender (*Appendix Ji & Jii*). On this basis, it is concluded that items from the pre- and post-assessment instruments indicated no significant difference in DIF by gender. This finding implies that the pre-test and post-test assessments are fair to both boys and girls who participated in the achievement tests.

5.4 Procedure

Data processing in Phase II involved data collection procedure, scoring and analysis as described in the subsequent section.

5.4.1 Data Collection Procedure

At the beginning of school term 2, in May 2019, the pupils in all selected classes took a paper-and-pencil mathematics pre-test on the selected HOTS items adapted from Standard 4 attainment tests covered in their curriculum. The test administrator from each of the sampled school administered the instruments in their schools. Conversely, to enhance the credibility and security of the data, the teachers from an upper primary in the same schools were the test administrators. Pupils were instructed to attempt all the items to the best of their ability and were assured that the feedback of the test results to be provided. The pupils were also told that the test results feedback was to inform their teachers' classroom practice.

Those test administrators collected the instruments for each day from the school head's office and opened them in the presence of the pupils. The test administrators supervised the actual process of administration and finally sealed off the work of the pupils before taking it back to the school head. The scripts were collected from each school by the researcher for scoring.

The professional development workshop on the 16–18 May 2019, which was attended by nine teachers from the intervention, is described in the section on the FAHOTS condition. Firstly, the teachers were given two weeks to plan and trial the FAHOTS in the classroom. During the first and second weeks of June 2019, the researcher visited those teachers to coach on the job, once during the daily mathematics lesson. After these two weeks, the teachers were expected to implement the FAHOTS by themselves. However, the schools also started to write the end of term tests and eventually closed for vacation. The researcher maintained continuous communication with the teachers through a WhatsApp group by reminding them of the FAHOTS condition.

In the second week of the third term of the school year, the researcher started observing the teachers, twice during daily mathematics lessons and once during a weekly Self-Assessment tool Questions. After eight weeks of intervention, pupils took a paper- and pencil mathematics post-test covering HOTS items from the learning goal taught during the year. Once again, the test was administered according to the standardised testing procedures.

5.4.2 Scoring procedure of the tests

As explained in the test development sections (Section 5.3), the composition of tests were HOTS questions oriented, adapted from BEC and consisted of open-ended questions that merited partial credit when scoring. Due to this nature of the tests, the researcher made use of the BEC scoring guidelines to enhance consistency when evaluating pupils' responses. Yearly, each examination developed by BEC has detailed scoring guides with rubrics and explanations for the allocations of marks for each constructed-response item. Since the researcher was the sole examiner of all papers for two tests, the scoring guides from BEC was frequently used as a reference document during marking. Additionally, to BEC scoring guidelines, the following validity indicators of what was being measured were considered:

(i) The researcher scored every pupil's response to one item at a time. That is score or grade item after item and not pupil after pupil. This scoring procedure was done to avoid the undue positive or negative influence which the quality of response to one item might have on the evaluation of the next item for the same pupil. Scoring all pupils' responses to one item only at a time will also lead to a thorough and uninterrupted familiarity with the standard and scoring criteria used for that item. This process tends to make scoring easier and to encourage the award of scores that are in direct relationship to the quality of responses.

(ii) As much as possible, each item for all pupils was scored uninterrupted at a sitting. The researcher scheduled his time to enable him to score one item in all the scripts without interruption. When the researcher took a healthy break for scoring, on resumption, he re-familiarises himself with the scoring standard and criteria by re-reading a random sample of some of the scripts already scored in the light of the scores awarded, before he went on to score the remaining scripts.

(iii) While scoring responses to a given item, the researcher strived to avoid the formulation of a temporary mindset which may cause him to score either too leniently or too harshly. Hence, efforts were made to maintain a high level of consistency and uniformity in scoring and avoid the tendency to be influenced by the quality of the response to an item on a previous script while scoring subsequent ones.

(iv) Scored each script anonymously without knowledge of the pupils whose script were being scored. This scoring procedure aims at preventing the natural conscious or unconscious tendency to allow the opinion which the scorer has about the testee, or the relationship between them, from influencing the scoring procedure. Thus, biasing the grade or score awarded to the response (Nenty, 2010).

5.4.3 Data management procedure

Through the assistance of the designed assessment framework, all the items from pre- and post-test were entered directly into Statistical Package for Social Science (SPSS) using the correct codes. While data for post-observation was entered into excel. Next, data was cleaned by running some basic descriptive statistics to established the neatness of the data, and any error identified was rectified accordingly.

The two similar mathematics tests were subjected to the Rasch Partial-Credit Model as a way to conduct psychometric analysis of the data set using the Rasch Software program Winsteps (Linacre, 2016). The other procedures involved include linking the two tests and check for normality respectively.

(i) Linking of the pre-test and post-test

Linking the measurement scale through common items makes it possible to express test-takers' performance on the same measurement scale (Boone, 2016). The intended purpose of the current study was to measure growth in pupils' knowledge over time. By using Rasch techniques, it enabled the two tests to be expressed on the same scale. Thus, both tests had ten-item anchors that were presented across forms and served as reference points, such that pupils' performance was expressed on a single scale. Table 5.10 shows that half of the anchor items followed the expected pattern in which Standard 4 pupils performed better in the post-test than in the pre-test assessment, and this finding implies that anchor items data fitted the expectation of the measurement model showing progression. The other half of the anchor items were showing a reversal pattern, that is, pupils performed better in pre-test than in post-test. The descriptive statistics identify fives items as being problematic. These items include item 6, 9a, 9b, 10b, and Q13b (showed reversal of expected pattern) identified as potentially not adequately assessing change, and consequently require rephrasing

Table 5.10: Item means based on item difficulty, logit unit and mean the difference

Items	Pre-test logit (May 2019)	Post-test logit (September 2019)	Mean Difference
Q1b	-1.69	-1.49	.20
Q6	-.57	-.87	-.30
Q8	.16	.91	.75
Q9a	-.57	-.77	-.2
Q9b	1.07	.71	-.36
Q10a	-1.07	-.77	.30
Q10b	1.13	0.83	-.30
Q10c	2.2	2.23	.03
Q13a	3.15	1.88	-1.27
Q13b	2.50	3.77	1.27

Further analysis was conducted to determine the extent to which anchor items were correlated. Thus, the anchor items' difficulties with their independent Rasch analysis measures on the scale of -5 to +5 using Winsteps were exported for both the pretest and post. The independent Rasch analysis of the anchor items was significantly and highly positively correlated with the item difficulties measured from pretest to post-test, *r=.913, p<.001*(See Figure 5.3). As can be seen from Figure 5.3, the anchor-item measures from pretesting and post-testing were plotted within 95% confidence interval lines (except item 13, which is lying slightly outside the 95% confidence interval) in Winsteps using Pearson's correlation coefficient. This Rasch analysis reveals that the anchor items were very stable and a productive measurement in monitoring pupils' achievement.

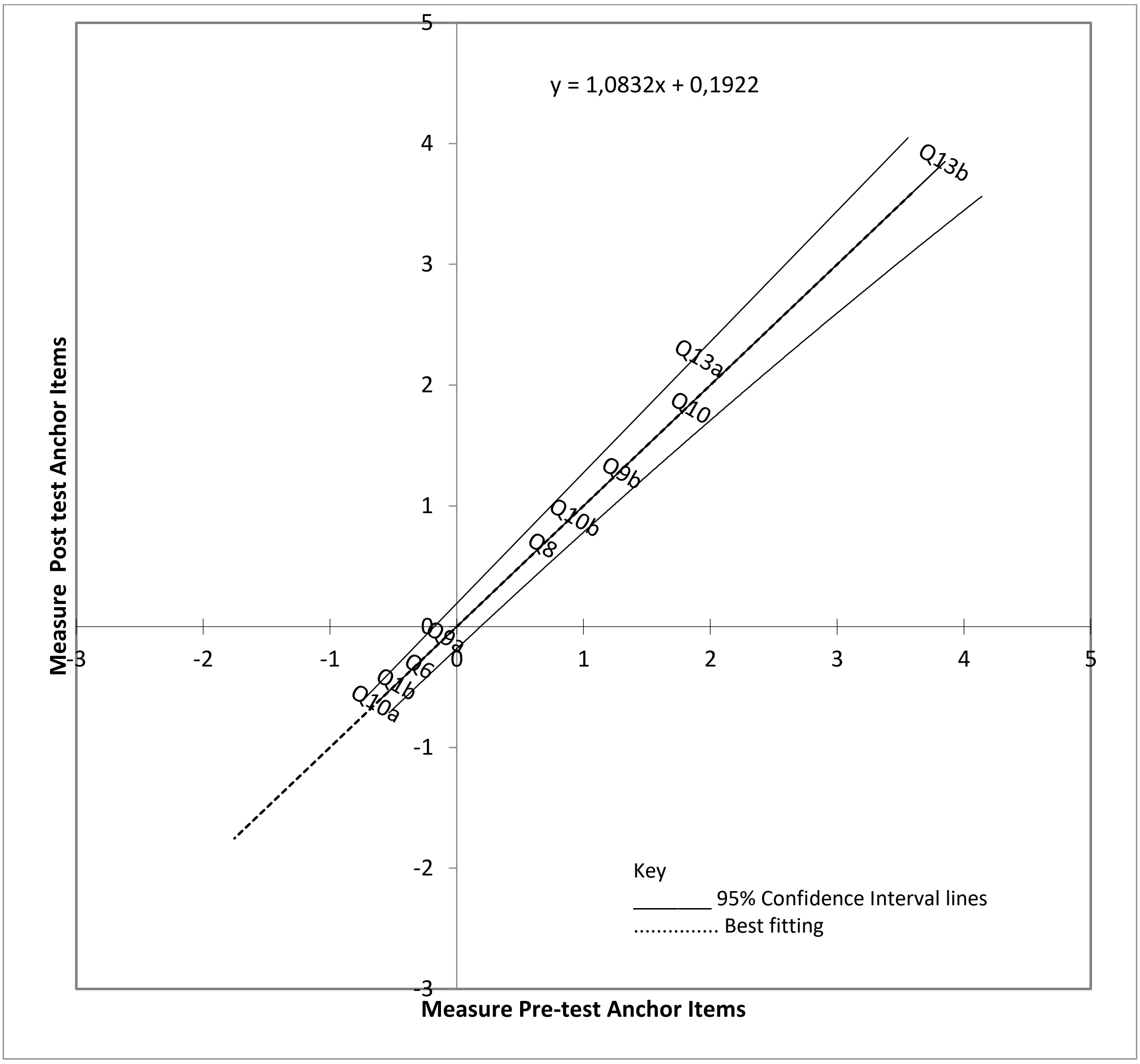


Figure 5.3: Correlation of anchor item-measures based on the Rasch analysis

(ii) Check for normality

Kolmogorov-Smirnov statistics was used to assess the normality of the data. Table 5.11 indicated that both the measures, Pre-test and Post-test Rasch scores were statistically significant (both the Kolmogorov-Smirnov Shapiro-Wilk, sig. less than 0.05) hence the normality was not assumed. Next, the data was subjected to non-parametric statistics for inferential analysis as informed by the pre-condition for assessing the normality.

Table 5.11: Normality analysis for pre-test and post-test Rasch logit

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Pre-test Rasch logits	.082	271	.000	.974	271	.000
Post-test Rasch logits	.058	271	.000	.982	271	.000
a. Lilliefors Significance Correction						

5.5 Analysis

For any measurement of change to have an unambiguous numerical representation and a substantive meaning, the pupil's achievement measures are estimated and compared within a common frame of reference encompassing both times points (Anselmi, Vidotto, Bettinardi, & Bertolotti, 2015; Combrinck, 2018). This process is called stacking, which involved measuring persons in a more similar frame of reference, thus stacking persons from Time 1 and Time 2 in one data set. In such a frame of reference, test instrument changes were controlled by having some anchor items and threshold measures equal in the two-time points (Anselmi et al., 2015; Combrinck, 2018). In this regard, the pupil pre-test data from a time point was analysed to obtain the pupils' performance measures for that time point. Then, the data from the other time point is analysed by anchoring the item and threshold measures to the values estimated in the previous analysis. This procedure would provide a set of pupil performance measures for the new time point (Time 2), comparable with the previous ones (Time 2).

Item and person logits were employed descriptively using the Rasch scaled data. One sample non-parametric statistic was also executed to determine the level of pupils' mathematics performance for HOTS items during pre-test assessment (Field, 2013). Specifically, the Wright Map analysis was further used to explore pupils' ability and item difficulty on the same logit scale. Wright Map analysis was employed since it can describe the person' ability as listed on the left side of the Map while the item difficulty is on the right side of the map (Lysaght, O'Leary & Ludlow, 2017). As for the post-observation data, mainly relative frequency was employed as a complementary finding for this Phase.

5.6 Findings

This section discusses the findings for Phase II of the study. During this phase of the study, the mathematics achievement assessment was administered to a sample consisting of 272 Standard 4 pupils from nine schools in the Kanye education sub-region at the beginning of Phase II. The items were adapted from the Botswana Examination Council (BEC) tests, which established the reliability and validity of the questions. And these indices were confirmed as discussed in Section 5.4.3 (Quality criteria for pupils' pre- and post-assessment instruments). A pre-test was used even though some aspects of the curriculum had not been taught, however, pupils should be able to utilise previous knowledge from previous standards to predict answers. The pre-test was also considered a baseline to ascertain pupils' strengths and weaknesses in HOTS items before the intervention.

The following section describes the Standard 4 pupils' mathematics achievement in the pre- and post-test results obtained during the implementation of the intervention.

Section 5.6.1, which deals with pupils' mathematics achievement by interpreting the pretesting findings while Section 5.7.1 looks at the pre- and post-test comparison of the findings. Section 5.8 concludes the chapter phase by looking into the discussion of the findings.

5.6.1 Pupils' Mathematics Achievement in Pre-test findings

The test scores for pupils in all nine schools were interpreted based on Rasch logits for accurate analysis of the data. The item difficulty is on a logit scale which ranges from -5 to +5 using Winsteps. Table 5.12 indicates participants' counts and THE average logits of pupils' mathematics HOTS achievement during the pre-test.

Table 5.12: Person measures for the HOTS achievement scores for pre-test

No	School ID	N	Logits	Std.Dev
1	303	29	-.617	1.59
2	208	38	-.908	1.37
3	319	33	-1.188	1.29
4	314	31	-1.363	1.49
5	309	24	-1.536	1.8
6	307	23	-2.151	1.49
7	306	32	-2.182	1.55
8	305	25	-2.264	1.31
9	320	37	-2.435	1.48
	All Pupils	272	-1.605	1.59

Table 5.12 reveals that all the nine schools had negative mean logits ranging from -2. 435 to -0.617. As can be seen in Figure 5.4, the Wright Map for the pre-intervention results for pupils in all nine schools (n= 272) achieved a very low mean, a standard deviation higher (*M*= -1.605, *SD*= 1.59) than the items (*M*= 0.00, SD=0.27). One sample, in the non-parametric test, was also found to be significantly below the mean value, t (271) = 89.000, p = .000. All the schools are well below the mean item difficulty. It suggests the test was not well-calibrated to pupil actual ability. However, school 1 and 2 were better than schools 6 to 9 inclusive. The nine schools were negatively and significantly spread below the mean value, indicating that most pupils performed poorly in the pre-test HOTS mathematics (See Figure 5.4).

Figure 5.4 also presents the result of the items for the pre-test graphically. Thus, item 12 (time) and 13 (money) are the top items, appearing above the measured value of the scale, indicating that the participating pupils had a challenge in dealing with those items about time and money concepts. Item 1 was also located above the mean value, and it dealt with concepts of *numbers and operations*, the participating pupils also had some difficulties in handling it. On the other hand, Q4, Q7, Q5, Q1b, Q10a and Q1a were easier items for the participating pupils because they were located below the Wright Map's right side. Questions 4 and 7 were the easiest items, and this implies that pupils had no difficulty dealing with concepts measuring through mass and application of fractions (see Figure 5.4).

In general, the findings reflect that almost all participating pupils were located on the bottom left of the Wright Map, indicating that the pupils performed poorly on the pre-test (See Figure 5.4). These findings for both persons and items had enabled the researcher to determine pupils' knowledge, skills, or dispositions before the intervention. The pre-test findings suggest that before the intervention, the pupils' mathematics achievement was very low and hence provided the researchers with baseline data from which pupils' learning progress could be established.

```
INPUT: 272 PERSON  21 ITEM REPORTED: 272 PERSON  21 ITEM  44 CATS WINSTEPS 4.7.1.0
-----------------------------------------------------------------------------------
MEASURE     PERSON - MAP - ITEM
                <more>|<rare>
    3                 +   Q12c (Measures)
                      |
                      |   12a (Measures)
                      |
                      |   Q13b (Number & Operation)
                      |
    2                 +
                      |
                      |
                      |S  Q13c (Number & Operation)
                      |
                   .  |   Q11    Q10c (Problem solving & Statistics respectively)
                      |
    1                 +
                   .  |
                      |   Q13a (Number & Operation)
                      |
                   .  |   Q10b (statistics)
                      |   Q9b (Number & Operation)
                      |   Q2
    0              .  +
                     T|   Q3 (Measures)
                 .##  |M  Q12b (Measures)
                      |
                      |   Q8 (Number & operation)
                 .##  |
                      |
   -1           ####  +
                      |
                ####  |   Q6    Q9a
                   . S|
                 ###  |
                   .  |   Q10a
   -2      .########  +   Q1a
                   .  |S
              .#####  |
                      |   Q5
         ###########  |   Q1b
                     M|
          ##########  |
   -3                 +
                      |
           .########  |   Q7
                      |   Q4
                      |
         ###########  |
                     S|
   -4                 +T
                      |
                      |
                      |
       .############  |
                      |
                      |
   -5                T+
                      |
   -5       .#########  +
                <less>|<freq>
```

Figure 5.4: Wright Map for pupils' mathematics achievement during the pre-test

5.7 Post-intervention findings

An attempt made to compare the pupils' pre- and post- assessment in Phase III of this study to compare the gains attained in mathematics achievement after the intervention took place. The pre/post-HOTS items were adapted from the Botswana Examination Council (BEC) tests, as already explained. Having determined the pre-test results as a baseline, a post-test was also done at the end of the interventional period (eight weeks), and pupils were expected to answer more questions correctly based on an increase in knowledge and understanding. Ten well-functioning common items from the pre-test were retained in the post-test to ensure the similarity of the two tests, the common items linked the two tests and provided a comparison of the performance on the common scale.

The nine teachers who took part in the two-day workshop facilitated by the researcher were expected to focus and implement their classroom lessons instruction within FAHOTS when teaching mathematics were also post-observed. The researcher further made some visits to observe teachers' implementing of FAHOTS and provided some guidance. For this reason, a comparison of both findings for pre- and post-classroom observation and pupils' HOTS achievement at different times was imperative to determine the level of gain and corroboration among the different findings. This approach of comparing baseline and post data is appropriate when measuring the same person to establish an indication of growth availed (Combrinck, 2018; Smith & Stone, 2009).

Having established the common scale through Rasch using Winsteps software, the logits (theta) mean comparison was used to evaluate the difference from Time 1 to Time 2 in pupils' mathematics achievements in HOTS. The Rasch logits, including aids anchor items (equating items), made it possible to compare gain in pupils' achievement. The equating through the anchor items process was conducted to establish comparable scores, with equivalent meaning, on different versions of test forms of the same test; it allows the scores to be used interchangeably (Sinharay et al., 2012).

That is stacking was done to measure a person measures in a similar point of reference as described in Section 5.5. Thus, stacking of Time 1 and Time 2 in one data had accorded a Rasch analysis to run on the complete stacked data, with the items and threshold measures anchored at the values that were estimated on the data (Anselmi, Vidotto, Bettinardi, & Bertolotti, 2015; Combrinck, 2018). From here a summary of the indices and Wright Map for the staked analysis were provided and finally, a Wilcoxon Signed Rank Test was then computed to evaluate differences in pupils' ability between pre- and post-intervention (stacked data) and the results were interpreted based on Rasch logits, including the effect size (Pallant, 2010).

As for classroom observation, the researcher used only descriptive data on the basis of frequency for content analysis to compare the participating teachers' practice before and after the FAHOTS intervention.

5.7.1 Pupils' mathematics achievement of stacked data analysis (pre- and post-test findings)

(i) Description of Stacking data for Rasch analysis

As noted in Section 5.3.3, thus Figure 5.5 depicts Rasch analysis of the anchor items was highly positively correlated with the item difficulties measured from pretest to post-test, $r=.913$, $p<.001$. Thus, greater measures of the anchor items indicate the stable and productive measurement in monitoring pupils' achievement. A summary of the indices for the stacked analysis from Winsteps was also presented in *Appendix L*. Thus person separation index was 2.02 (slightly above the cut-off point of 2) while reliability for the persons was exactly at the cut-off point of acceptable .80 (Boone et al., 2014). The item separation indices were also above satisfactory levels (Linacre 2016).

(ii) Description of the spread of item difficulty for mathematics in stacked data results

The Wright item maps are a visual summary of the relationship between item difficulties and learner proficiencies as they emerge in the test. The items are ranked from the easiest (at the bottom) to the most difficult (at the top). The item map in Figure

5.5 illustrates the spread of item difficulty across the cognitive domain of mathematics constructs for the post-assessment test including the anchor items.

In Figure 5.5 the item threshold mean value is set at zero. The estimated pupils' proficiency values in post-assessment have a negative lower mean than the item mean, which suggests that the test is targeted higher than their proficiency. Specifically, Figure 5.5 reveals that Q2 in pre-test means it was higher than the post Q2. Six items had higher scores in the pre-test, and 11 items had higher scores in the post-test. All the anchor items are in the same location as when selected much earlier. Many of the pupils are positioned in the bottom half of the map, indicating that their proficiency is low in comparison to the items and many of the pupils in post-mathematics assessment would still have experienced the test as a difficult one. For instance, most of the anchor items were still positioned on the top of the map, indicating overall that items are not well-targeted to the proficiency of the pupils and it was predictable from the pre-test.

The pupils' proficiency in all nine schools were captured using a person measure from Winsteps software analysis. The use of the Rasch Measurement Model had enabled an opportunity to apply the items and threshold calibrations from the independent analysis of the baseline that is pre-test of pupils' mathematics achievement to the pre-test and post-test results.

INPUT: 272 PERSON 27 ITEM REPORTED: 272 PERSON 27 ITEM 57 CATS WINSTEPS

```
--------------------------------------------------------------------------------------------
MEASURE     PERSON - MAP - ITEM
               <more>|<rare> |Pre           |Post          |Anchor
    4                +T                                     Q13A_ANC
                     |
                     |
                  .  |                                      Q13B_ANC
                     |
                     |                                      Q10C_ANC
    3             .  +
                     |
                  .  |
                     |S                                     Q10B_ANCHOR Q9B_ANCHOR
    2                +        Q2_PRE
                     |                       Q13C_POST
                     |
                    T|        Q3_PRE
                 ##  |
                .##  |                                      Q8_ANCHOR
                     |
    1            .#  +
              .####  |
                     |                       Q3_POST
               ####  |                                      Q6_ANCHOR  Q9A_ANCHOR
               .### S|                       Q12A_POST
              #####  |M                      Q12B_POST
                     |                                      Q10A_ANCHOR
    0   .##########  +        Q1A_PRE
        .##########  |                       Q2_POST
         .#########  |        Q5_PRE         Q1A_POST
                     |                                      Q1B_ANCHOR
           .#######  |
          .########  |
   -1               M+
          #########  |
        ###########  |        Q4_PRE;Q7_PRE
                     |
         .#########  |S
                     |                       Q7_POST
          #########  |                       Q5_POST
   -2             .  +
         . ######### S|
                     |
                     |
            #######  |
                     |                       Q11B_POST
                  .  |                       Q11A_POST
   -3        .###   +                        Q4_POST
                     |
                     |
                .## T|T
                     |
   -4                +
                     |
                ###  |
                     |
   -5                +
               <less>|<freq>
```

Figure 5.5: Person-Item measure for post-test Mathematics

(iii) Comparison of pre- and post-testing of pupils' achievement findings in mathematics

Table 5.13 indicates participants' counts and average logits mathematics achievement at different times. As shown in Table 5.13, overall the pupils in the pre-intervention test (n= 272) obtained a lower mean value (*-1.61 logit*) than when participated in the post-intervention (*n* = 272) test *(-.61 logit*). A closer look at Table 5.13 reveals that all schools had slightly performed better during post-test than in pre-test.

Table 5.13: Pupils' mathematics achievement scores analysis for the pre and post-test

School ID	N	Pretest Logits (SD)	Post Logits (SD)	Rasch Learning Gain (Θ_{post}) – (Θ_{pre})	Z Wilcoxon Signed-Rank	Sig. (2-tailed)	Effect size (r)	Effect size (d)
303	29	-.62 (1.59)	-.24 (0.95)	.38	1.44	.150		
308	38	-.91 (1.37)	-.24 (1.13)	.67	3.67	.000	0.60	1.5
319	33	-.19 (1.29)	-.46 (0.91)	.73	3.21	.001	0.56	1.6
314	31	-.36 (1.49)	-.52 (1.38)	.84	3.74	.000	0.67	1.8
309	24	-1.54 (1.80)	-.88 (1.34)	.66	2.83	.005	0.56	1.6
307	23	-2.15 (1.49)	-.89 (1.17)	.26	4.03	.000	0.84	3.1
306	32	-2.18 (1.55)	-.89 (1.04)	.29	4.64	.000	0.82	5.0
305	25	-2.26(1.31)	-.94 (1.50)	1.32	4.10	.000	0.82	5.0
320	37	-2.45 (1.48)	-.64 (1.29)	1.80	5.26	.000	0.86	3.4
All Pupils	272	-1.61 (1.59)	-.61 (1.21)	1.00	11.40	.000	0.70	3.0

The Wilcoxon Signed-Rank test was utilised to compare the pre-test and post-test results of pupils' mathematics achievement (Table 5.13). The analysis shows that pupils' mathematics achievement findings recorded a significantly higher gain of average achievement values from the pre-test (*1.61 logit, SD=1.59*) to post-test (*-.61 logit, SD= 1.21*), *z = 11.40, p = .000*, with a large effect size (*r* = 0.68). All the schools significantly gained from pre-test to post-test results, except for school (303), which was not significantly gained (p =.150). The effect sizes for all statistically significant gain in pupils' achievements were so large including for the overall gain. According to Hattie (2013), "any intervention higher than the average effect (*d* = 0.40) is worth implementing". For this reason, the analysis of the results for the current study shows that the effect size is large and beyond the expected average growth. This result indicates a significant large improvement in pupil's performance in the post-intervention, and one can associate the improvement with the attribute of using the FA strategies. Since teachers who participated in the intervention condition could have assimilated some strategies well which enhanced their teaching and learning of mathematics HOTS concepts (as discussed in the next pre- and post-observation comparison section).

5.8 Discussion

5.8.1 Pre-Intervention mathematics achievement

The pre-intervention test results investigated the pupils' levels of mathematics achievement in HOTS items for Standard 4 pupils before their teachers implemented the intervention (sub-research question-3). The research question sought to determine the baseline of the pupils' achievement before their teachers were exposed to the intervention. The pupils' pre-test achievement in mathematics was negatively and significantly low when compared to the Rasch mean value, thus indicating low performance. This finding generally supports previous studies and national results patterns on Botswana pupils' mathematics low achievements (BEC, 2008, 2017; MOESD, 2016; Tabulawa, 2014). More specifically the finding is in line with the TIMSS 2015 study that observed that the Botswana pupils who took part in the assessment were unable to apply HOTS, such as critical thinking and problem-solving in mathematics (Masole et al., 2016). These results also support debates about the inadequacy of assessments and examinations in some African countries about the limited measure of HOTS (Allen et al., 2016; Chapman &Schnyder, 2000; Mainali, 2012; Mitana et al., 2018). In particular, a study by Mitana et al. (2018) found that Ugandan's assessment seemed to be merely measuring superficial learning that would hardly enable the country's achievement of its National Vision or the Sustainable Development Goals (SDGs).

More explicitly, this finding of a low mathematics achievement by pupils in Standard 4 is a replication of a thirteen-year-old BEC (2007) study. That study assessed the level of pupils' performance and determined that over 50% of the pupils did not reach the low-performance level of the benchmark performance levels developed by the BEC and teachers. According to the criteria they developed, only 22.5% of the pupils qualified for progression in mathematics (2007). Most pupils could not score 50% of the available marks.

5.8.2 Pre and Post-Intervention comparison of mathematics achievement

Sub-research question 4 sought to determine the extent to which the intervention has enhanced Standard 4 pupils' HOTS academic achievement in mathematics when comparing the pre- and post-intervention. The mathematics achievement findings revealed significantly different scores, which indicated how much pupils gained as far as learning was concerned. Note that the test in its entirety required use of three different HOTS skills, so the gain is in the field of HOTS. The significant gain in pupils learning can be associated with the intervention. The magnitude of the effect size was beyond the expected growth, an effect size greater than Hattie's (2013) hinge-point in the case of the intervention. This evidence corroborates the observational findings in Phase II of that study that teachers indeed did something different, and it was a sign of change in the classroom teaching and learning of mathematics.

It is important to note with the results for this study that a statistically significant gain was observed for pupils' achievement in mathematics in the tests which were constituted of HOTS items. Once again, it should be noted that the finding of this study was in line with Kiplagat’s (2016) study, which was concerned with rethinking primary school mathematics teaching and concluded that FA classroom strategy improved pupils’ achievement. More explicitly, the current results corroborate the most recent study by Chemeli (2019b), who found a positive impact of the five key FA strategies on pupils’ achievement in mathematics instruction.

The current study's findings were consistent with Ballan (2012), who employed a mixed-method intervention study on assessment for learning (formative assessment) in mathematics education in which the intervention was found positively significant. More specifically, the study’s findings aligned well with Balan and Metcalfe (2012) who found positive results by employing formative assessment intervention in mathematics education to problem-solving for students. In addition, this study findings agree with (2015), who found a strong significant difference in mathematics scores for exposed pupils in formative assessment.

Chemeli (2019b) and WAG (2010) noted that the integration of FA could promote pupils' quality of learning HOTS. The current study results suggest that the classroom integration of FA and HOTS can encourage pupils to learn by promoting, among others, active engagement, critical thinking, active listening, asking questions, summaries and explaining their understanding, as confirmed by the interview findings. Thus, FA's learning environment was aided by a classroom that also embraced sensitivity, constructive engagement, and reflectivity so that pupils felt safe to make mistakes (Heritage, 2010).

5.8.3 Methodological limitations

The results of this phase of the study are interpreted in the light of some limitation. Firstly, the test instrument was not designed with anchor items; instead, the researcher assembled items from available instruments (Standard 4 previous examination papers for mathematics) and selected the best functioning items as anchors during pre-test results analysis. Secondly, the test reported the sub-scales, but the test was not factor structure analysed or tested for sub-scale reliability with classical or Rasch modelling. The researcher assumed the items in each topic area were sufficient since items were adapted from previous BEC examinations.

6 CHAPTER SIX: IMPACT OF THE INTERVENTION ON TEACHER PRACTICES, ATTITUDES AS DEMONSTRATED IN MULTIPLE STUDIES (INTERVIEWS, OBSERVATIONS, AND PUPIL WORKBOOK)

In Chapter five, an interventional phase formed part of Phase II of this study aimed at supporting FA to enhance teaching and learning for Standard 4 pupils HOTS in Mathematics. Based on the implementation of the FAHOTS in attempts to enhance the teaching of HOTS in mathematics, it was also equally imperative to evaluate and reflect on the implementation of the intervention as well as its possible impact (Phase II). Phase III mainly looks into the findings based on the experiences of teachers participating in the intervention.

This chapter covers a brief description of the research design used in this phase (Section 6.1), and the participants involved (Section 6.2). A procedure and analysis employed cover Section 6.3 and Section 6.4, while the chapter is concluded with Section 6.5 to present the findings. Section 6.6 deals with a discussion and conclusion based on the findings.

6.1 Research Design

The implementation of the FAHOTS as an effort to enhance the teaching of HOTS in mathematics was evaluated. The intention of this phase is not to describe teachers' FAHOTS practices, but rather to provide an overview of teachers' experiences in using it as a tool for evaluating the collected data through pre- and post-classroom observations and lastly analysis of data collected through interviews. This evaluation approach was aligned to the conceptual framework that provided this study with some lenses to obtain a holistic view on how teachers' assessment finds their practice situated within Kivunja's (2015) power of Assessment Feedback Loop (AFL) model as a guidance and departure point. Specifically, as shown in Figure 6.1, this study's intervention was customised to a weekly frame cycle-like model to illustrate that FAHOTS is a continuous process, integrated into daily classroom instruction to close the gap. The study attempted to determine the extent to which the knowledge gap could be closed using both pupils' assessment data, observation data and teachers' experiences upon completion of the workshop and implementation of the FAHOTS intervention in their classroom.

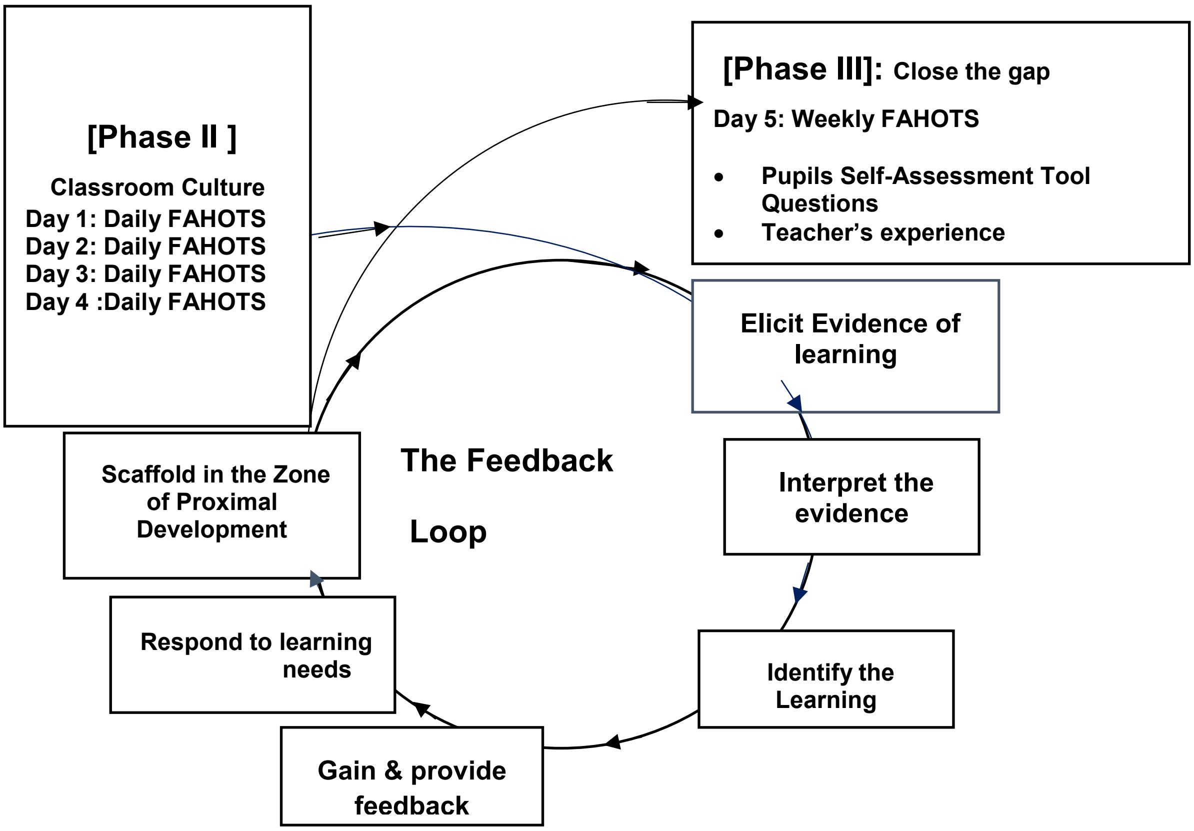


Figure 6.1: The FAHOTS model of four daily FAHOTS cycles and weekly cycle extracted from Kivunja's Feedback Loop.

Precisely the pupil's assessments and participating teachers' experiences/reflections were addressed and were guided by the following research sub-questions (SRQ):

- SRQ 5. To what extent does the FA strategies intervention enhance Standard 4 teachers' teaching of HOTS in mathematics when comparing the pre- and post-observation?
- SRQ 6. What are teachers' experiences following the FA strategies intervention and mathematics teaching on the pupils learning outcomes?
- SRQ 7. What are teachers' reflections following the FA strategies intervention and mathematics teaching of pupils' HOTS?

6.2 Participants in Phase III of Study

Phase III of the study was a further exploration following the intervention that comprised the pupils' achievement gain in mathematical HOTS items, classroom observations and teachers' experience and reflection of the FA practice through the interviews. In Chapter 5, pupils' mathematics achievement assessment from a sample of Standard 4 pupils, consisting of 272 pupils, was compared from pre-test to post-test in nine schools for Kanye education sub-region. This chapter compares teacher observation and interview data pre- and post-intervention. This data provided complementary insights into implementing the FAHOTS intervention to enhance HOTS teaching in mathematics in eight weeks.

The participants for this phase of the study remained the same teachers involved in the intervention. It has to be kept in mind that these teachers were from one region of the research site and were selected purposively randomly during classroom observations in Phase I of the study. That is, participating teachers were involved in the completion of the teacher questionnaire before being observed. They were recruited through their schools' management to participate in workshop training. They implemented the FAHOTS intervention and were observed post-intervention. For this reason, these teachers were followed purposefully to get their viewpoints concerning the intervention. Only seven teachers were interviewed, while the other two were on sick leave during the scheduled interview week.

6.3 Procedure

The following sections discuss the procedures involved in Phase III of this study;

6.3.1 Teacher interview procedure

An interview was adopted in this phase of the study. The interview is useful in trying to gain insight into some elements of human experience beyond the basic facts of who, what, when and where. Instead, it is always in part or whole, about how and why, helping the researcher to understand their interviewees' views of the process (Guest et al., 2013). Generally, interviews are distinguished based on how structured or unstructured they are (Gay, 2011). This study employed one-to-one open-ended questions and semi-structured interviews. According to Guest et al., (2013), the flexibility of the semi-structured interview allows the interviewer to pursue a series of less structured questioning and also permits the exploration of spontaneous issues raised by the interviewee to be explored. This exploratory nature of the semi-structured interview motivated the researcher to apply it in Phase III of the study.

Phase III of the study used the interview to make a follow-up, probing and prompting questions based on the principle of allowing the interviewee control over the interview process (Ryan, Coughlan, & Cronin, 2009). The focus is on permitting the interviewee to tell his/her own story rather than answer a series of structured questions. In this regard, Standard 4 mathematics teachers were followed up, following the post-classroom observation and intervention to explain their own experiences of teaching under the embedment of FA strategies in mathematics lessons. The interview schedule determines the structure of the interview, the nature of the research and the aims and questions of the study (Ryan et al., 2009). In relation to this study, the interview was intended to address the research sub-question 6, and 7 which focused on teachers' experience and reflection of FA strategies intervention and mathematics teaching of pupils' HOTS.

The interview questions were aligned with Kivunja's Model, and an assessment feedback loop formed the basis of the semi-structured interview questions. The draft schedule was subjected to the expert judgment of the research supervisor and a PhD candidate student in assessment and quality assurance at the University of Pretoria. The supervisor advised that the items must also be aligned to the research design, conceptual framework and elicit the explanation validly measure the research sub-question under-investigation. The PhD candidate student advised that interview questions were too wordy and were therefore revised accordingly. The input of the expert judgments was used to revise the interview guide questions (See *Appendix K*).

Basically, the interviews were done to determine whether their experiences and reflections had some impact on the pupils learning of mathematics as a result of the FA integration in teaching of HOTS in their classrooms. Halfway through the third term of the school year (September 2019), the researcher interviewed the teachers from the intervention condition during a scheduled time after teaching hours. Those interviews were conducted during teachers' free teaching periods. The audio-recorded interviews were taken from each teacher lasting about half an hour each. Probing and prompting questions were facilitated to acquire insight into the teacher's implementation of FA strategies and teaching of HOTS concepts.

6.4 Analysis

6.4.1 Interview analysis

The interview data was from open-ended items and was analysed qualitatively, using thick descriptions to capture respondents' views and organising them into themes. In this case, questions were focused on formative assessment practice, strategies and their impacts on pupils' learning and understanding of HOTS mathematics, techniques for pupils' engagement in the classroom, strategies used to improve questioning, pupils' engagement and higher-order thinking skills.

For this study, all interviews were audiotaped and transcription was done as a pre-requisite to thematic data analysis by a researcher (Creswell & Clark, 2008). Thematic analysis is the process of identifying patterns or themes within qualitative data. According to Maguire and Delahunt (2017), thematic analysis is not tied to a particular epistemological or theoretical perspective. This perspective makes it a very flexible method, a considerable advantage given the diversity of work in educational research, hence it was found appropriate for the current phase of the study (Maguire & Delahunt, 2017).

In an attempt to analyse qualitative data, Braun & Clarke (2006) provide a six-phase guide as a useful framework for conducting this phase of the study's qualitative data analysis.

Step 1: Become familiar with the data

The first step in any qualitative analysis is reading, and re-reading the transcripts. In this case, reading was done to get very familiar with the entire body of data before the researcher went any further. At this stage, it was useful to make notes and jot down early impressions.

Step 2: Generate initial codes

In step two, an inductive approach was implemented by initiating data code or categorised for analysis, without fitting it to a pre-determined coding frame in the Microsoft Excel worksheet.

Step 3: Search for themes

Following the migrated data to a Microsoft Excel worksheet, the researcher then generated a single column consisting of all comments. The data was initially reviewed to remove any duplicate entries arising from recording the same points during the interview. The data was then analysed with a view to assigning thematic areas. Each cell was reviewed and assigned to a thematic area, to which a cell colour code was applied. The codes were examined and some of them fitted together into a theme. At the end of this step, the codes were organised into broader themes that seemed to say something specific about this research question and themes that were predominately descriptive (Braun & Clarke, 2006).

Step 4: Review themes

This phase involved reviewing, modifying and developing the preliminary themes that were identified in Step 3. The attempt at this point was to gather all the data that is relevant to each theme. Reading through the data associated with each theme it was possible to consider whether the data did support the themes as far as the context of the entire data set was concerned (Maguire & Delahunt, 2017).

Step 5: Define themes

The current study made a final refinement of the themes and the aim to be consistent with Braun and Clarke (2006), such that the ‘essence’ of what each theme is about.

Step 6: Write-up

The current study was reported based on the subsequent themes that emerged from teachers’ responses to implement formative practice and reflections of the practical steps for enhancing pupils HOTS using FA strategies as revealed in the analysis.

Figure 6.2 shows the exploratory approach of the qualitative data collection and analysis employed during this phase of the study (see *Appendix M*).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

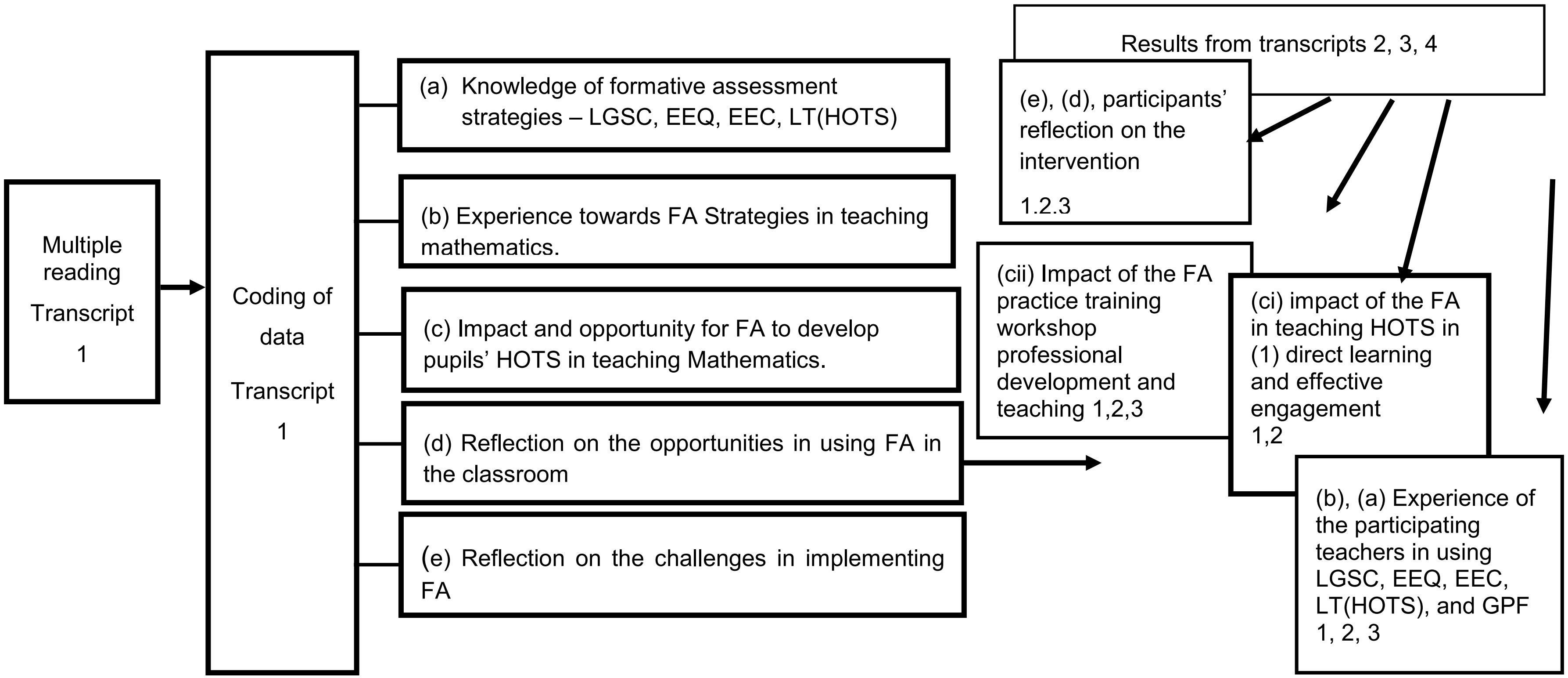


Figure 6.2:Process of the thematic analysis of data using the FAHOTS.

Based on Kivunja's Assessment Feedback Loop (Kivunja, 2015c)

6.4.2 Post Intervention Observations

After six weeks into intervention practice, the same nine teachers who had participated in professional development, coached and implemented were post observed twice in their classroom while they were teaching mathematics. The researcher used the same instrument for pre-observation, as explained in Chapter 4. That is, the participating teachers were also involved in pre-observation. This arrangement allowed an opportunity to compare the teachers' formative assessment practices pre- and post-observation, respectively.

6.5 Findings

6.5.1 Classroom observation in pre- and post-findings

(i) Teachers' use of learning goals, Success Criteria, Questioning and Learning Tasks

Frequency use of learning goal (LG) and Success Criteria (SC), questioning and interaction, learning (HOTS) tasks and feedback is considered to be effective in enhancing pupils' performance. It is for the interest of this study that after the professional development, (FAHOTS), the teachers were expected to apply these elements more frequently during their lessons than before the intervention. For Table 6.1, the proportion of using learning goals and success criteria has to some extent changed when comparing the pre-observation to post-observation. Neither during the pre-observation nor post-observation were the observed teachers seen using learning goals and success criteria on a chart or handout.

Table 6.1: Rating frequency of the use of learning goal and success criteria within the FAHOTS as observed by the researcher pre- and post-observations

FA Strategy	Evidence					
	Pre-classroom observation			Post-classroom observation		
Learning goals[R]	N	Seen (rf)	Not (rf)	N	Seen (rf)	Not (rf)
The teacher uses words such as "We Are Learning To" (WALT) when introducing the Lesson objective (LO)	8	2 (0.250)	6 (0.750)	8	5 (0.625)	3 (0.375)
• Presented orally	8	2 (0.250)	6 (0.750)	8	5 (0.625)	3 (0.375)
• Written on the board	8	0 (0.000)	8 (1.000)	8	4 (0.500)	4 (0.500)
• Written on the chart	8	0 (0.000)	8 (1.000)	8	0 (0.000)	8 (1.000)
• Provided a handout	8	0 (0.000)	8 (1.000)	8	0 (0.000)	8 (1.000)
Success Criteria[R]						
The teacher uses words such as "What I'm Looking For**" (WILF)** when introducing the Assessment Criteria (AC)	8	0 (0.000)	8 (1.000)	8	5 (0.625)	3 (0.375)
• Presented orally	8	0 (0.000)	8 (1.000)	8	5 (0.625)	3 (0.375)
• Written on the board	8	0 (0.000)	8 (1.000)	8	4 (0.500)	4 (0.500)
• Written on the chart	8	0 (0.000)	8 (1.000)	8	0 (0.000)	8 (1.000)
• Provided a handout	8	0 (0.000)	8 (1.000)	8	0 (0.000)	8 (1.000)

Note: R = dichotomous rating; 1= ("Seen"), 0 = ("Not seen"), rf = relative frequency

Table 6.2 provides an overview of the first and third quartiles, and the median scores of the observed teachers in both conditions (pre- and post-observation), regarding their use of questioning, was classified as ineffective and effective. The results of the pre-observation indicate that the teachers were inclined towards ineffective questioning (Mdn = 2.00) more than effective questioning techniques (Mdn = 1.00). This result is perhaps indicating a traditional classroom discussion in which teachers focus much of their attention on the pupils who raise their hands when wishing to attempt a question before them. Wilcoxon signed-rank tests show that there was no statistically significance for ineffective ($z = 1.823$, $p = 068$) and effective questioning ($z = 1.414$, $p=.157$) from pre-observation to post-observation respectively (See Table 6.2). These results indicate that the participating teachers did not change their way of questioning, even after the FAHOTS intervention; thus, most of the teachers remained ineffective as far as questioning skills are concerned. Evidently, the descriptive value for effective questioning remained unchanged (Mdn=1) from pre- to post-observation (See Table 6.2).

Table 6.2: Rating results for using questioning within FAHOTS as observed by the researcher during pre- and post-observations

Questioning and interaction	Pre- observation				Post-observation					
	N	Q1	Mdn	Q3	N	Q1	Mdn	Q3	Z	Asymp. Sig. (two-tailed
Ineffective Questioning and interaction[R]	8	2.000	2.000	2.000	8	.000	1.00	2.000	1.823	.068
• When the teacher asks questions, pupils put their hands up										
• The teacher only asks pupils that have their hands up										
• The teacher asks questions for the "whole" class to respond										
• The teacher answers her/his questions										
Effective Questioning interaction[R]	8	.000	1.00	1.000	8	.250	1.00	2.000	1.414	.157
• The teacher involves more than one pupil in questioning										
• The teacher uses name/number sticks to select pupils										
• Pupils use mini-boards during the lesson										
• Pupils work in groups to guide each other on their learning										
• Pupils work in groups cooperatively while completing the group task										

Note: R = Rating format: min = 0 ("not seen"), max = 2 ("often"); Q1= Quartile 1, Mdn (Median)= Quartile 2, and Q3 = Quartile 3

As for Learning tasks, Table 6.3 shows median pre- and post-observation scores for the observed teachers. A closer look at the results reveals that none of the teachers applied the strategy in pre-observation while only one teacher applied it during post-observation (Mdn=1). However, participating teachers did not differ significantly in their use of learning tasks (HOTS), $Z = 1.890$, $p = .059$, indicating insufficient evidence for its utility in the classroom.

(ii) Peer and self-assessment pre- and post-observation

Table 6.4 shows (next page) the observations recorded on feedback practice provision in the observed teachers' classrooms during pre- and post-observation. With oral feedback, an improvement was seen in post-observation. After giving pupils classwork, a 0.75 proportion of teachers walked around to check how pupils were doing compared with the pre-observation (0.00). A ratio of 0.75 of the participating teachers was also seen checking pupils' responses, providing some gathered feedback or some comments when compared with the pre-observation. As for peer assessment, the participating teachers still had some challenges using the strategy (See Table 6.4). Even for self-assessment, Table 6.4 depicts that the proportion using the strategy ranged from 0.00 to 0.5 for pre- and post-observations, respectively.

Table 6.3: Rating results for using learning tasks within FAHOTS as observed by the researcher during pre- and post-observations

Learning tasks (HOTS)	Pre- observation				Post-observation					
	N	Q1	Mdn	Q3	N	Q1	Mdn	Q3	Z	Asymp. Sig. two-tailed
• The teacher uses well-crafted HOTS tasks that are aligned with the learning goal • All pupils are clear about the HOTS, they can begin work efficiently. • The teacher frequently uses pupils' responses and work to make inferences about progress and adapts instruction accordingly. • The teacher skillfully uses multiple ways of gathering evidence throughout the lesson that are connected to the learning. • The teacher uses multiple approaches to handle problem-solving tasks • The teacher gives homework on problem-solving tasks • The teacher engages pupils on the previous tasks given as a homework	8	.000	.000	.000	8	25	1.000	1.000	1.890	.059

Note: R = Rating format: min = 0 ("not seen"), max = 2 ("often"); Q1= Quartile 1, Mdn (Median)= Quartile 2, and Q3 = Quartile 3

Table 6.4: The frequency and use of feedback within the FAHOTS as observed by the researcher during pre- and post-observations

	Pre-classroom observation			Post-classroom observation		
	N	Yes (rf)	No (rf)	N	Yes (rf)	No (rf)
Oral feedback[R]						
After giving pupils classwork, the teacher walks around to check how pupils are doing.	8	0 (0.000)	8 (1.000)	8	6 (0.750)	2 (0.250)
The teacher provides oral evaluative feedback on a specific piece of work	8	0 (0.000)	8 (1.000)	8	5 (0.625)	3 (0.375)
When checking pupils' work, the teacher gives guidance or makes comments	8	2 (0.250)	6 (0.750)	8	7 (0.875)	1 (0.125)
The teacher provides oral descriptive feedback on a specific piece of work	8	8 (1.000)	0 (0.000)	8	5 (0.625)	3 (0.375)
Peer Assessment[R]						
Pupils are given an opportunity to check their partner's work.	8	6 (0.750)	2 (0.250)	8	4(0.500)	4 (0.500)
The teacher reminds pupils how they should use peer assessment.	8	8 (1.000)	0 (0.000)	8	3 (0.375)	5 (0.625)
The teacher visits a few pupils to check how they conduct peer assessment.	8	8 (1.000)	0 (0.000)	8	3 (0.375)	5 (0.625)
The teacher gives feedback on how the peer assessments were conducted.	8	8 (1.000)	0 (0.000)	8	3 (0.375)	5 (0.625)
Self-Assessment [R]						
The teacher reminds pupils how to use self-assessment, e.g. processes and rules are reviewed.	8	8 (1.000)	0 (0.000)	8	4 (0.500)	4 (0.500)
A teacher tells pupils to use Success Criteria when checking their work.	8	8 (1.000)	0 (0.000)	8	3 (0.375)	5 (0.625)
The teacher visits a few pupils to check how they conduct the self-assessment.	8	8 (1.000)	0 (0.000)	8	4(0.500)	4 (0.500)
The teacher gives feedback on how the self-assessments were conducted.	8	8 (1.000)	0 (0.000)	8	4(0.500)	4 (0.500)

Note: R = dichotomic rating; 1= ("Yes"), 0 = ("No"), rf = relative frequency *(iii) Mathematics Exercises and Type of Feedback for pre- and post-observation*

Table 6.5 shows sampled workbooks of the lower performing and higher performing pupils after the lesson observations. The attempt to collect additional data from sampling workbooks was meant to explore the level of learning tasks that involved problem-solving (HOTS) and lower-order thinking skills (LOTS) and the types of feedback given. Table 6.5 below illustrates data from post-observation for a small proportion of exercises (min = 10, max= 25), for LOTS (*M* = 18.63; *SD* = 5.53) to HOTS (*M* = 5.75; SD = 1.58). Even in pre-observation, it shows that for every 12 exercises which constitute the LOTS items (*M* = 31.38; SD = 8.72), only one item was for HOTS (*M* = 2.62; SD = 0.518). A Wilcoxon Signed Rank Test from pre-observation *(M* = 31.38; SD = 8.72) to post-observations (*M* = 18.63; *SD* = 5.53) for LOTS exercises was found significantly reduced (*Z= 2.313, p=.021)*, with a large effect size (r= *.818*). Regarding the rate of giving HOTS, exercises had increased significantly from pre-observation to post-observation (*Z= 2.388, p=.017)*, with a large effect size (r= *.844*).

Pre and Post feedback in all the reviewed exercises significantly changed for Lower and Higher-order oriented tasks from pre to post-intervention. For example, before the intervention comments included: “Pull up your socks”, “This is poor work”, “See me”. After the intervention, feedback comments included: “Excellent work”, “Very good attempt”.

Table 6.5: Rating results of pupils' learning exercises during pre- and post-observation as observed by the researcher

	Observation 1 January-May 2019					Observation 2 June –October 2019					Difference		
Standard 4 workbook exercises	N	Min	Max	Mean	SD	N	Min	Max	Mean	SD	Z	Asymp sig. (2 two-tailed) *p*	*R*
Low Order Oriented exercises	8	22	48	31.38	8.717	8	10	25	18.63	5.528	2.313	.021	.818
Higher-Order Oriented exercises	8	2	3	2.63	.518	8	3	8	5.75	1.581	2.388	.017	.844
Low and High Order Oriented combined	8	25	51	34.25	8.779	8	16	32	24.38	5.878	1.827	.068	

In general, a shift of fewer LOTS, and more HOTS was observed. The acquisition of these skills will result in explanations, decisions and products that are valid within the context of available knowledge and experience, and promote continued growth in higher-order thinking, as well as other intellectual skills. In the same vein studies that have been published are much more focused on generating aspects of FA in enhancing pupils' achievement and learning improvement as well as indicating the effect of the intervention on generating HOTS. Hence, it is difficult for the researcher to find studies that specifically focus on the percentage of exercises needed for HOTS to be effective. However, the large effect size of HOTS exercises from pre-observation to post-observation reflects desirable effort in the classroom practice to enhance learning HOTS among pupils. In addition, based on the foregoing analysis (See Table 6.5), types of class exercises were common to all pupils in this way and did not show any evidence of pupils' remedial and enrichment activities respectively unless separate workbooks were used. It also appears that the majority of the teachers' feedback is still inclined to be based on evaluative feedback.

6.5.2 Thematic Analysis of the Interview Data

The data was interpreted and analysed following the thematic analysis processes as described in Chapter 3 of this study. Thus, thematic data analysis was performed by the researcher who read the data several times, coded each response, revised the codes and formed themes. For this phase of the study, thematic analysis was considered a flexible method for qualitative data analysis which emphasises the development of the codes as an attempt to generate themes on the questions answered by the participants' viewpoints.

Following the data coding process, six themes emerged from the raw data which included participants' experience with the intervention, the impact of the strategies and the workshop, pupils' reaction towards learning, barriers and support and a general reflection on the intervention. These themes were discussed and linked (where possible) to the constructs of Kivunja's assessment feedback loop conceptual framework to identify possible growth areas following the implementation of the intervention programme. Sub-themes emerged from some main themes. Under the theme impact of the strategies, it became clear that some participants who employed FA strategies did so to direct learning and effective engagement, management of the class size and workloads. The teachers perceived the workshops as professional development and linked them to the training workshop for FA in the classroom. The majority of the participating teachers reported that the implementation of FA helped to improve pupils' motivation in learning while the other teachers considered FA as strategies that provide an opportunity for learning improvement. These comments of participants were listed under the theme of *pupils' reaction towards learning*. Some participants experienced some challenges during the implementation of the FA while other participating teachers had support from school management.

(i) Participating in teachers' experience with the intervention

The teachers were clear and consistent about what constituted their FA and how they used those strategies in the classroom. The participants seemed to suggest that they have integrated at least three of the five formative strategies well into their mathematics teaching. Those strategies were *learning goals and success criteria*

(LGSC), *engineering effective classroom discussion through questioning* (EEQ), *engineering effective classroom discussion through collaboration* (EEC*), learning task* (*LT) and feedback through based instruction as well as pupils' peer- and self-assessment (GPF).*

Only one participant (participant 1) reported applying all five of the strategies in the class, and she seemed to understand the relationship tied to the strategies. In describing which strategies, she employed, Participant 1 indicated that "*I have used all of them because there is no way how one can separate them in a lesson. Because they complement each other during teaching and learning*".

As for the use of *learning goal and success criteria,* it was reported to be a common strategy among five of the participants. The teachers attempted to write the learning goal and success criteria on the chalkboard or manila folder and discussed with the pupils that which were useful in terms of getting a clear direction for learning, clear instructions, what was expected from pupils, procedures to solve a given problem, and to make learning easier for most of the pupils. As one of the participants revealed:

> *The LG, I would write it on the chart and then displayed it on the chalkboard, and discuss it with pupils, it makes the pupils know what is going to be achieved during the lesson, what they are going to do in the lesson, so the SC I would also display them there on the chalkboard* (Participant 5).

To this effect, the teachers had considered sharing the *learning goal and success criteria* as the most effective way of starting the lesson implementation.

Engineering effective classroom discussion through questioning was reportedly used by five of the teachers as another common strategy. As one of them (Participant 2) said, "*during questioning, I always discourage my [pupils] to raise their hands, so I use pupils' name card to pick the [pupil] randomly".* Thus, the participants identified questioning techniques that enabled them to engage all the pupils, among them were hands down, thinking time, pair and share, and exit card or whiteboard as opposed to the traditional methods of engaging pupils.

One of the participants further said to have exit cards during a group activity as a way of *engineering effective classroom discussion through collaboration.* Thus, she issued some problem-solving tasks during a mathematics lesson and asked pupils to handle the activities in small groups using their exit cards. *Gaining and providing feedback* as a strategy was mainly employed by two participants and there was evidence from some participating teachers that they provided feedback by commenting on pupils' responses as either evaluative or descriptive oriented feedback.

(ii) Impact of formative assessment strategies on classroom practice

(a) Direct learning and effective engagement

Some participants noted that FA strategies helped indirect learning and effective engagement. The teachers said they attempted to write the learning goal and success criteria on the chalkboard or manila folder and discussed with the pupils. The teachers said that the strategies were useful in terms of getting clear directions for learning, clear instructions, what is expected from pupils, procedures to solve a given problem, and making learning easier for most of the pupils.

According to participant 5, sharing LGSC was the most effective way of starting the lesson implementation, since both the teachers and pupils have to be clear about the goals of learning to build confidence among the pupils. In her own words, she said "*I explain each of the SC so that they may know how they are going to achieve the LG, how they are going to achieve the learning goal, they become confident*" (Participant 5). Participant 5 also attested that "*explaining the SC together with the LG, it directs both the teacher and [pupils]. You have to know what is expected from them and or what they should do*"

The teachers said they were creating effective classroom discussions through questioning, activities, and tasks that offered the right type of evidence on how pupils are progressing to the espoused learning goals. The teachers explained the need to plan for higher-order questions as a way of stimulating focused learning and animated discussion. Five of the participants identified questioning techniques as enabling them to engage all the pupils, among them were hands down, thinking time, pair and share, and exit card or whiteboard.

During questioning, one participant discouraged her pupils from raising their hands, and use the pupils' name card to pick the pupil randomly. According to the second participant, this "*technique made my pupils very alert, prepared and showed a high level of concentration during our class discussion, maybe would be his or her turn to answer the question*". Participant 3 also supported no hand up strategy because it "*helped me to engage all pupils. I find it also important to plan for higher-order questions in class*" This mentioned technique of questioning inculcates the sense of pupils' high concentration, alertness, preparedness, and improved thinking capacity during the classroom discussion.

Five of the participating teachers said they used whiteboards or exit cards and found it to be an effective way to address pupils' concerns either individually or with the whole class. Unlike the traditional methods of engaging pupils, exit cards give ongoing feedback and help the teacher make decisions about instructional techniques, pacing, or classroom management. Teachers said that the use of exit cards provided an equitable way for all pupils to share their thinking, and they can open up opportunities to elicit further learning dialogue with pupils who may hesitate to share concerns face-to-face. With the exit cards:

> *"It helps [pupil] to work independently when given a task and likely to improve their thinking capacity" (Participant1; Participant2).*
>
> *"It does help the teacher to identify where the problem is" (Participant 4; Participant 7).*
>
> *"It is easy to spot a learning opportunity" (Participant 3).*

It emerged that teachers used exit cards during a group activity in which pupils were given the learning tasks during a mathematics lesson.

> *The activity on the exit card/whiteboard as a group, as I used it, I observed that pupils became active and at the end of the lesson wanted to have a share of the activity or to be involved in that particular lesson. Besides, I used it so that they come together and so*

> *that when given individual work, at least they have grasped something from other learners. Like I said earlier, I was also giving pupils some group tasks which involved HOTS. This was used to help them think deeply before answering the question and practice thinking independently as problem solvers* (Participant 4).

Teacher 4 said she found that effective learning tasks motivate pupils to engage with the material at their own pace and level of difficulty, and provide information about pupils' understanding and skill attained.

With this current study, two teachers said they used feedback which entailed commenting on pupils' responses as either evaluative or descriptively oriented. Participant 4 identified peer assessment as a vital aspect of giving feedback during classroom discussion.

> *As for peer [assessment], by giving the pupil activity, and I then asked to mark for each other's work, and thereafter we go through the exercise at the board. This kind of activity, assisted my pupils to get feedback from their peers as well as from the teacher if all the pupils did not get it correct* (Participant 4).

Feedback is essential for pupils to understand how well they are learning and to identify the next steps that move in their learning forward (Kivunja, 2015).

(b) Management of the large class size and workloads

More often than not, teachers in developing countries mention the large class size and heavy workloads as impeding them from doing effective teaching in their classroom. However, in the current study, participants said that the implementation of FA in the classroom addresses problems associated with large class size or heavy workloads. In this finding, three out of seven participants indicated that they viewed FA as a pupil-centred approach as opposed to teacher-centred.

Since the exposure to the FA training workshop, the two teachers said they felt that their teaching approach was transformed towards a more pupil-centred style. One of the participants said, "*I find it to be pupils' approach; hence pupils spend time in engaging activities"*. The participants reported better lesson planning, and hence the pupils spend time engaging in activities, helping one another and improving the management of large class sizes. Also, six out of seven of the participants reported that effective implementation of FA was viewed as a reduction in their teaching workloads as quoted below;

> *Our classes, in general, are very big, but formative assessment, did reduce the workloads, because most of the time, I was using pupils centred activities particularly with the exit card, for pupils to show that they understood and made progress to the next activity or objectives as well as using SC to arrive at the solution. So, they work on the exit card and show me their answers. I just give them instructions, then use the card to work either as an individual or group and provide the answer. The formative assessment did assist me in managing the large class* (Participant 7*).*
>
> *"I don't even feel a workload" (Participant 1).*
>
> *"Even though the class size is big, the use of formative assessment made it much easy to handle the workload" (Participant 5).*
>
> *"And the workload is not bad, particularly with the implementation of formative assessment" (Participant 2).*
>
> *"With formative assessment, it helps me to manage the large class size, unlike when I do not use it. Nowadays I give pupils' clear instructions and expectation, then they work. Unlike me talking the whole day" (Participant 3).*
>
> *"Before I was workshopped, the class load was too much; however, after I have been introduced to FA, my workload has been reduced as most tasks are done by the pupils and they understand quickly" (Participant 4).*

Despite the positive affirmation of the FA to eliminate large class size and workloads, one of the participating teachers revealed a challenge related to the preparation of the learning materials which seemed to be a problem due to strenuous lesson preparation in the unusual formats. However, the teacher said she found ways to handle the situation, and ultimately the situation improved.

(iii) Impact of the formative assessment practice training workshop

(a) Professional Development

The teacher's instructional development focuses on the teachers' experience with the FA with which they asked for their opinions as to whether it instilled professional development. The findings showed that teachers supported FA practice in the classroom, and they perceived it as professional development. They perceived that FA offered much help to both teacher and the pupils in the teaching and learning environment. As for teachers, they perceived FA as a tool to manage time, a signal to direct teaching and learning of mathematics, adding value to their style of teaching, and broaden the minds of teachers.

> *It broadens our minds in such a way that before you give the [pupils] a topic you have to go through the topic, you have to search about the topic before handy, to check the materials to be used, the things to be used and what to bring in the classroom* (Participant 6).

The teachers' experience with FA also seems to suggest that they were exposed to systematic planning when teaching mathematics. Participant 7 said that:

> *It is a professional development, okay, the plan is well done and it is completely different, it is not like the way we used to do it. Thus, we used to keep everything in our mind as teachers, hoping that you will share it with pupils and then failing to share the whole concepts, but nowadays, I plan and write it for [pupils] to read it together with them [during the lesson implementation]* (Participant 7).

Participants said that FA is a helpful teaching pedagogy that encourages the pupils to develop in-depth thinking. For instance, one of the participants highlighted that "*it is a* *professional development because it helped me so much since we started using the FA, there is much development in the pupils, and they can even answer structured questions and think reasonably deep*" (*Participant* 5).

(b) Training workshop enhanced formative assessment practice among teachers

In general, the teachers' experience with a training workshop was perceived as useful and had helped them in the implementation of formative assessment. Participant 1 regarded the workshop as professional in-service training. She further explained that such a workshop should be taken seriously, since most of the teachers, especially in primary school, lacked specialised training in teaching mathematics. Participant 1 said that most primary school teachers seemed not to give pupils engaging activities and time to practice learning mathematics, especially HOTS. She said that the workshop was a good empowerment strategy and assisted her a lot because she would spend time planning HOTS learning activities for pupils and assisted pupils based on the formative assessment strategies and techniques.

Teacher 2 has considered the two-day workshop as professional development and to help to guide teachers on teaching pedagogies with formative assessment and said that: "*it added value to my teaching style*". The workshop could enable self-assessment regarding daily teaching and empowering teachers with some formative assessment strategies for teaching and learning HOTS in mathematics.

(iv) Pupils' reaction and improvement in learning

Participants highlighted the impact on pupils as positive. All seven teachers said that the pupils' reactions ranged from; excitement, building confidence and increased attention. To reaffirm the aforesaid reactions, one of the participating teachers revealed that:

> *I think my [pupils] were so interested and as you know young people are so interested in new things. However, interestingly, my [pupils] are now interested in doing things on their own, interested in hands-on activities, whereby everyone is given time to do a particular task as I said with exit cards and that every [pupils] is so engaged in the*

> *classroom and then show their solution to the teacher, it makes the [pupils] so interested* (Participant 7).

Most of the participants said that they saw improvement in mathematics performance, including for HOTS items. The observed improvement was identified across all the pupils of different abilities, including the very low achievers. For instance, according to one of the participants:

> *Even those pupils who got "D", when I paged their scripts, I observed that they were just lazy to add up things, but in general you can see that they knew the method for solving it, though the answer was wrong. That the pupils knew the necessary steps, which to me was an improvement in terms of understanding the question* (Participant 5).

The participants collectively said they observed a substantial improvement in their pupils' mathematics achievement even though some pupils were not improving. Five of the participants outlined the level of improvement for their pupils in mathematics on average as they observed pupils' marks from the first term through to the third term of the academic year calendar:

> *"as I speak if you take my result analysis, my pupils have made an average of 90% in Maths" (Participant 1).*

> *"They are doing well, you remember last time we sent the pupils results to you. They got 90%". (Participant 3).*

> They have improved, even though two or three still lag, however, in general, the whole class has improved in maths and even in other subjects, at the end of the first term "*their average in maths was 29% and then end of the second term they moved to 95.1*%"(Participant *4).*

> *"A lot of improvement in maths was 68% in the first term and at the end of the second term, it was 78 %" (Participant 5).*

> *"My pupils managed to score 97% this term in Maths and that's a great improvement*" (Participant 6).

(v) Participants' support and challenges faced with formative Assessment Implementation

The participants said they felt supported by their school management. For instance, the schools' management allowed them to participate in all the phases of the study, some visited their classroom to observe the learning process, and in general, the management was supportive.

The main problem advanced by the participating teachers during the implementation phase of FA among others include:

- Lack of technology such as computers and pupils' tablets had hampered teacher's innovation to explore some engaging activities;
- Lack of learning materials, the teachers had improvised by using the manila folders instead of the whiteboards and bought the whiteboard markers for the pupils; and
- In some classes, many pupils were low achievers, and some teachers had challenges to implement some of the FA particularly for those pupils who were unable to read hence peer-and self-assessment was difficult to affect.

(vi) Participants' reflection on the intervention

The intervention was introduced to the participants to help bridge the gaps in FA practice. All of the participants said that they found the FAHOTS intervention to be very relevant to their daily classroom assessment, and those strategies would be useful across standard levels in primary schools. However, the researcher's multiple roles as a principal investigator, interventional developer, workshop facilitator and evaluator during the study may anyhow have an impact on the findings of the study, since participants may give overly positive feedback. Like most of the participating teachers, Participant 1 said:

> *I think it should be done or used in all schools and starting at the foundation level or phase and going up the level so that pupils would acquire the key skills for HOTS or innovative skills at an early age and*

become independent thinkers as they grow up. And with formative assessment is very easy (Participant 1).

I recommend that it should start from the lower class and goes to the upper classes. If we start from the lower primary school class, those pupils gain and be familiar with the strategies to the high level or upper primary schools. And again, pupils will get used to being actively engaged across all levels (Participant 4).

I recommend it to other teachers even in my school, they are saying I should put up a workshop for them so that they may use it even in other subjects because it showed that the strategies worked very much. Because, I also taught the other class of Standard 4, what I was doing with this class, I also taught the same thing and so it shows that it worked. I, therefore, recommend that formative assessment should be rolled to the whole standards (Participant 5).

6.6 Discussion, conclusion and recommendation for practice

The study's present phase III aimed to determine the possible effect of a FAHOTS intervention used by Botswana Standard 4 mathematics teachers' (BS4MT) to enhance Standard 4 pupils' HOTS. The phase consisted of three questions of which the results are discussed separately in the next sections.

6.6.1 Classroom observations

Sub-research question-5 investigated if Standard 4 teachers' exposure to the FAHOTS intervention has enhanced their teaching of HOTS in mathematics when comparing the pre- and post-observation. The classroom observation results indicated that teachers employed FAHOTS to some extent. The most prominently used strategies were using learning goals, engineering effective classroom discussions on questioning, and learning tasks to elicit evidence of learning HOTS concepts. This positive development in teaching and learning is a desirable experience for the participating teachers, suggesting abandoning teaching to the test, as observed in

Chapter 1 of this study. The workbooks also revealed significant HOTS tasks compared from pre-observations to post-observations, as reported in chapter 5. The findings indicate that participating teachers may have assigned some HOTS learning task as often as taught during the workshops.

A change in pupils' exercises is a good development; however, the extent to which pupils were assisted with feedback to improve learning is important. The feedback findings from pupils' workbook do not seem to help the pupils to learn. More specifically, comments like "pull up your socks", and "very poor" were still noted against the incorrect solution during post-observations. Moreover, evidence of peer- and self-assessment was limited during post-observation, just like in pre-observation. These results do not support previous studies on teachers' monitoring, scaffolding and feedback delivery practices (Lizzio & Wilson, 2008; Kyrauzi, 2017; Van de Pol et al., 2014). Specifically, the results do not replicate the findings by Van de Pol et al., (2017), who found that exposing teachers to training on how to support pupils learning improved their diagnostic strategies and the quality of their support to pupils.

Additionally, the current results were inconsistent with Chemeli (2019a), who found inadequate support concerning training and resources in implementing FA strategies in mathematics instruction. Despite the current study's findings that demonstrated some limited evidence that participating teachers had on FA, the observed significant change in pupils' workbook indicates an impact on the transformation of teaching and learning of mathematics, particularly the HOTS.

6.6.2 Teachers' experiences and reflections

Sub-research question 6 investigated the participating teachers' experiences following the FAHOTS intervention and mathematics teaching on the pupils learning outcomes. The participating teachers revealed some positive experiences with the FAHOTS intervention during interview findings. The participants affirmed that they had implemented at least three to five of the formative assessment strategies. Such interview findings corroborate the classroom observations and found some evidence that teachers have employed FAHOTS, particularly using learning objectives, engineering effective classroom discussions on questioning, and learning tasks to elicit evidence of learning HOTS concepts, respectively.

This positive development in teaching and learning is a desirable experience for the participating teachers to suggest abandoning the notion of teaching to the test, as observed in Chapter 1 of this study. Moreover, the workbooks revealed that HOTS tasks were assigned as often as taught during the workshops, which indicates possible intervention impact, though it may require long-term follow-up and support for teachers. These findings are consistent with Accado (2017) who identified teacher training in professional development learning communities, in which the schools and the government supported and implemented the formative assessment as a way to improve pupils' performance. Moreover, the study's findings are also in line with Ballan (2012), who employed a mixed-method intervention study on assessment for learning (formative assessment) in mathematics education, in which the finding was positively significant.

On the other hand, a most recent study by Chemeli (2019a) investigated the Kenyan teachers' support for effective implementation of the FA strategies in mathematics instruction, and the findings revealed inadequate support concerning training and resources. However, the current study's finding demonstrated some limited evidence that teacher training in FA impacted the transformation of teaching and learning of mathematics, particularly the HOTS. The participating teachers said they were in favour of the FAHOTS intervention, as revealed in the interview findings. This response may have been a socially desirable response during the interviews, but teachers did not exhibit some of these strategies during the classroom observations. However, the teacher support for in-service training of the FAHOTS fits well with Botswana ETSSP's (BOT, 2015) proposed teacher training development in the school-based assessment.

The last sub-research question (SRQ 7.) investigated teachers' reflections following the FAHOTS intervention and mathematics teaching of pupils' HOTS. The results from interviewed teachers were positive towards using the FAHOTS in their classrooms. The participating teachers suggested that such professional development should be cascaded to all subjects and across different primary school levels. This result indicates that FAHOTS intervention could be used as a strategy to enhance teaching and learning of HOTS, but more evidence is needed for the long-term implementation (Black & Wiliam, 2003; WAG, 2010).

The results are in line with Melani's (2017) findings where an in-depth case revealed that, when provided with specific information about FA through staff development, teachers became more positive towards such assessment, and their implementation skills were greatly improved. When teachers are trained, they can support the FA as a method for monitoring pupils. The findings support Accado (2017), who found that mathematics teachers who received professional development (PD) training in FA provided by a university faculty with expertise in mathematics education used instructional methods to improve pupils' learning. The findings in the current study were also similar to the positive factors noted by Chemeli (2019b). Chemeli found that formative assessment strategies eased the teacher's workload, raised pupils' attitudes and interest, improved pupils' critical thinking, and teachers and pupils enjoyed using formative assessment strategies.

6.6.3 Theoretical contribution

(i) Formative Assessment

It should be noted that participating teachers and pupils, in general, seemed weak in teaching pedagogies and were underachievers, respectively. Therefore, positive evidence may also be associated with this sequential embedded approach with a single-case study but should be discussed cautiously. In particular, the use of learning goals and success criteria, questioning, change in classroom exercises (increase on HOTS tasks in learning mathematics) and general positive reactions of pupils evidenced during post-observation can provide support for the improvement in classroom instruction and pupils' learning experiences. Standard 4 teachers perceived formative assessment practice as a strategy to enhance the teaching of HOTS in mathematics and to improve pupils learning outcomes, which is in line with the theory of formative assessment as proposed by Black and Wiliam (2003).

(ii) Teacher professional development

The findings show that the participating teachers implemented some elements of FAHOTS to a certain extent during mathematics lessons. Classroom observations allowed the researcher to assess the Kivunja's Assessment Feedback Loop (Kivunja, 2015) being used. Some teachers were seen introducing lesson objectives or learning intentions and success criteria, including writing them either on the chalkboard or manila paper and discussing them with pupils after the intervention. Some teachers also planned at least one objective oriented question before the lesson and implemented effective questioning techniques such as name sticks, waiting time, no hands up, and mini-boards during the lesson engagement for mathematics. Evidence of feedback was also observed, though it was very limited and not as effective as desired. Observing the classroom changes may support the training workshop, but it should be noted that it is unknown whether the changes will last long-term. However, it seems hard to determine every strategy's effect and how it influenced mathematics achievement. The classroom observation results replicate the previous studies focused on teachers' professional preferences for FA practices (Kyaruzi, 2017; Milani 2017). Hence the Kivunja framework is helpful for teacher professional development.

(iii) Rasch modelling.

Rasch analysis is ideal for determining the extent to which items belong to a single dimension and where items sit within that dimension. Advancing in sound measurement and sophisticated psychometric methods such as Rasch Modelling provides an enhanced understanding of research results and positive effect in the pupil and teacher evaluation instrument refinement in education and other disciplines. In the current study, the reliability and validity inferences were drawn and fair measurement comparisons and identification of areas for enhancement were attained through Rasch Modelling.

6.6.4 Limitation and implication

Even though the study systematically developed the intervention, the results should be interpreted, bearing in mind that it was only a sequential embedded approach with a single-case study, which may raise some concerns about the validity.

Firstly, the lack of a control group has limited the study's causal findings. The current study's findings may mean it was by chance that the researcher worked with really enthusiastic, smart, and motivated participants. Additionally, the tests were still too hard for the participating pupils, so the researcher cannot confidently tell if the results would be better with better-calibrated tests. Having mentioned these factors, it is also imperative to note that the teachers' reflection on the experience had augmented the intervention's impact on pupils' learning of HOTS in mathematics. Secondly, as pointed out previously, a lack of control on participants' allocation, random assignment and the limited number of sampled schools and teachers mean that this study's results cannot be generalised to other settings in Botswana and beyond.

Moreover, the eight teachers involved were mainly from the junior position of teaching rank (only two senior teachers), experienced some challenges in teaching mathematics, and most of their pupils were underachievers since their school management specifically selected them to be assisted through the intervention. In addition, the test instruments used were not specifically designed with anchor items, hence to some extent limited the test linking approach.

6.6.5 Recommendation practice

Based on this study's findings on the possible effect of a FAHOTS intervention, the Botswana Standard 4 mathematics teachers (BS4MT) were encouraged to use the intervention circle in every mathematics lesson to enhance Standard 4 pupils' HOTS. As indicated in the findings, teachers and pupils have to be clear regarding the learning goals. This clarity will expedite feedback and engage pupils as resources for one another. A change in pupils' exercise findings means a teacher should be encouraged in aligning assessments to the goals with HOTS, derived from Bloom's Taxonomy and the learning objectives in the Botswana Standard 4 mathematics curriculum. The findings show that the FAHOTS intervention improved teachers' teaching of HOTS, but it is unclear which strategies enhanced the results. Future studies should investigate the effectiveness of the FAHOTS intervention in other domains.

Specifically, it should investigate issues such as (i) questioning that elicits discussion and stimulates pupils' higher-order thinking, and (ii) activating pupils to be instructional resources for each other (peer-assessment) and themselves (self-assessment). This should be based on focused, effective feedback strategies aligned to pupils' higher-order thinking in mathematics concepts. Lastly, it is paramount to investigate pupils' experiences about their teachers' implementation of FAHOTS intervention to determine whether their teachers' assessment practice had changed for the betterment of pupils' learning. Based on the pupils' findings, their performance was very poor and confirmed the school management selection criteria imposed on the researcher to assist them with an intervention to improve their achievement. From post-test classroom observations, the findings show some indications that participating teachers had implemented some of the interventional strategies while teaching mathematics. Thus, participating teachers were observed to use some types of formative strategies even though the implementation of the practice was uneven. Some changes of the teachers from their traditional teaching practice of issuing exercises by including more HOTS task was a positive development in teaching and learning of mathematics. Perhaps, a stepping stone series of introductions is imperative in future, so to stage a series of workshops, each introducing one strategy and not introducing the next until evidence of practice proved that they were doing that thing and it influenced pupil's achievement positively.

It should be noted that even after workshop training, most of the participating teachers still preferred traditional methods of lesson planning, and only a few made some attempts to plan within the formative assessment format. However, though the researcher had encouraged the participating teachers to apply the formative plan during preparation, he did not impose the new lesson plan format without the Ministry of Basic education's input and approval. The effort made by the researcher was only limited to training, encouraging teachers to integrate formative assessment as much as possible in teaching mathematics, sharing all the training material, visiting their classrooms for observations and motivational support. The researcher also opened the WhatsApp group for communication, support materials and general discussion on the intervention.

A follow-up study was needed if such intervention could be paired to determine any gain or improvement following the FAHOTS intervention. Thus, it was recommended that the findings from teachers' pre-observation be compared with post-observation with an active control group. Similarly, the findings for pupils' pre-assessment should be compared with the post-assessment with the control group. Lastly, these phase II results also indicated that the eight participating teachers should be interviewed to determine their experiences and reflections towards FAHOTS intervention.

7 CHAPTER SEVEN: GENERAL DISCUSSION AND CONCLUSIONS

The rationale for this thesis was to investigate twofold in three empirical studies: (1) to explore FA practices in the Botswana context (Phase I reported in Chapter 4), and (2) to propose a possible effect of an intervention for BS4MT's use of the FA strategies in the classroom to enhance pupils' HOTS (Phase II reported in Chapter 5 and Phase III reported in Chapter 6).

The following sections provide a summary of the three phases with their findings, followed by a merged (integrated) discussion and conclusions for each phase of the thesis. The chapter concludes with a reflection on the conceptual framework, methodological limitations, conclusions and recommendations, new knowledge, innovations, insights and further research. The foundation of the design for this PhD thesis took the sequential mixed-method data collection from each phase to complement findings to the research questions. Data was collected from the participants as discussed in Chapter 4 and 5. A single-group pre-test, FAHOTS intervention (training teachers) and post-test matched measures design was employed.

7.1 Summary of the studies

7.1.1 Chapter 4: Baseline & Intervention

A cross-sectional survey explored the current practice in FA and teaching of HOTS in Botswana primary schools. A formative assessment-oriented questionnaire was submitted to a sample of 150 Standard 4 teachers from a sample of 81 schools in the five Southern Education Sub-Regions using stratified random sampling. Thereafter, the researcher followed up with an observational exploration inquiry of teachers' practice in nine schools' classrooms. The teacher survey was measuring five FA strategies for classroom practice which included: learning goals and criteria for success, engineering effective classroom discussion (questioning), engineering effective classroom discussion (collaboration), learning tasks, and feedback on instructions (Ramsey & Duffy, 2016), and explored studies on instructional integration for the teaching of HOTS drawn from research on effective FA processes (Heritage, 2010; Kivunja, 2015; WAG, 2010). All the subscales were adapted to suit the Botswana context. The data was analysed using Rasch modelling (questionnaire) and descriptive relative frequency counts (classroom observation). The Rasch Measurement Model fit, dimensionality, as well as person and item reliability, were examined. One hundred and fifty (150) teachers answered the 48 FA and 15 HOTS items. Eight FA items did not meet the RMM fit requirements, while the remaining FA composite scale (40 items) and five subscale items met the requirements. Since the FA and HOTS scales exhibited different levels of dimensionality, the questionnaire was considered multidimensional. Item reliability was higher above the optimal in both composites and subscales (FA and HOTS items), indicating that item ordering is very reliable.

The Rasch logits results revealed participating teachers' integration of HOTS as the most used strategy followed by the use of learning goal and success criteria (LGSC). Rasch logit analysis further indicates that engineering effective classroom discussion through collaboration was the least used strategy by the teachers compared to others. No significant variation was found between the HOTS and FA strategies. More specifically, the Wright map's findings had spelt out that the teachers endorsed highly on HOTS items which assessed their practice. In contrast, external support and curriculum-oriented support were endorsed low, indicating less support given to them. The survey and observation data were strongly similar. What the teachers said they found hardest was indeed seen least in the classrooms. However, the easiest things in the survey were still not very strongly present. So, there might be some problems as revealed in the baseline findings, which include:

(i) **The FA questionnaire had limited discrimination**. Most teachers reporting that they use all of the FA strategies. This response pattern may have indicated social desirability responding.

(ii) **Higher Order Thinking Skills Strategies (HOTS) was highly endorsed for teachers' practice but less endorsed for external support such as training on FA.** This response also indicates the social desirability of responding to the questionnaire. The pre-workbook analysis and pre-classroom observations findings confirmed that teachers did not use HOTS as claimed.

The two data set were inconsistent on the findings of HOTS practice; thus, the teachers have self-assessed highly on providing support for pupils to work independently on mathematics problem-solving tasks, instructions that provide opportunities to encourage pupils to be critical thinkers and use different levels of questioning which help pupils to think and reason. On the other hand, the observation findings showed none of the teachers providing instructions that enhance critical thinking. However, they could have used those instructions in the absence of the researcher. Likewise, the sampled pupils' workbooks revealed very limited mathematics exercise-related problem-solving. These findings support the social responding notion that may have influenced the participants or teachers may have felt vulnerable and defensive reporting in a self-administered questionnaire and may have rated themselves highly regarding the practice yet, in reality, these practices are not reflected accordingly (Dunning, Health & Suls, 2004; Van Staden & Zimmerman, 2017). Based on these ineffective and inconsistent findings associated with FA and teaching HOTS, a conclusion was reached to effect an intervention accordingly as an attempt to empower teachers to adjust their assessment practice.

Phase I of this thesis explored teachers' assessment practice regarding the teaching of HOTS in mathematics. Previous scholars also support FA as a systematic process to continuously gather evidence about learning for pupils and teachers (Black & Wiliam, 1998). The findings from the two data sets (questionnaire and classroom observation) resembled the situation in which participants' FA practices were investigated before the intervention. The findings are consistent with the evidence of FA practices identified with the teachers that were ineffective in terms of the use of learning goal and success criteria, classroom questioning and collaboration, and use of learning tasks in general. The findings also agreed on the feedback being the least used strategy, in particular regarding self-and peer assessment. These results are inconsistent with the literature which supports FA as to how a teacher identify a pupil's current level of learning and how to adapt lessons to help the pupil reach the desired learning goal (Black & Wiliam, 1998, Hattie & Timperley, 2007; Heritage, 2010; Stiggins, 2009).

The researcher designed an intervention extracted from Kivunja's Assessment Feedback Loop (2015) and linked it to HOTS; hence it was named formative assessment for HOTS (FAHOTS). The participating teachers took part in the two-day workshop facilitated by the researcher. The nine participating teachers were exposed to two-day workshop training and then observed teaching mathematics (pre-observation, post-observation). The training constructs included sharing of lesson objective and success criteria, effective classroom questioning and collaboration, and aligning learning of tasks to include HOTS related tasks, which were considered the most important strategies in Kivunja's (2015) model with which to elicit evidence of learning. The use of a wide range of assessment tasks ensured that every pupil had an opportunity to demonstrate what she/he had learnt.

7.1.2 Chapter 5: Pupil Achievement

The sampled teachers who had already answered the survey questionnaire administered as part of Phase I were followed through their selected classes and pupils in the respective schools. A Standard 4 class was selected based on the observed teacher, and such a class was also eligible to participate in the study. Most importantly, the pupils' acquisition of HOTS in mathematics was measured based on learning outcomes prescribed in the Standard 4 national curriculum. The existing items for HOTS from BEC's previous examinations were used to measure pupils' achievement for the first time (pre-test) and the second time (post-test). The pupils' test was developed by adopting the BEC mathematics items from the Standard 4 curriculum, which were supposed to measure HOTS constructs across Bloom's taxonomy's higher cognitive domains.

Hence, the test instruments that constituted HOTS items were also well-intentioned to be explored in association with Kivunja's (2015) model. The Rasch modelled analysis of pupils' matched pre-test and post-test measures revealed significant results indicating a gain in performance. After pupils' pre-test administration, the teachers were encouraged and supported to apply the FAHOTS in teaching mathematics. The

272 pupils' pre-test results indicated very low performance, which meant their cohort's performance was similar to national norms in mathematics. When compared, the observed pupils' workbook data showed a significant difference between pre-post measures in pupils exercises for HOTS tasks. Descriptive feedback on the observation data revealed developing evidence of teachers using the strategies. Specifically, the teachers were seen using learning goal and success criteria, some wrote them on the chalkboard and briefly discussed them with their pupils. Additionally, some observed teachers applied various effective questioning techniques such as exit cards, no hands-up and follow-up on answers. It was concluded that the intervention highly improved the Standard 4 teachers in teaching mathematics. Based on this finding, a recommendation to investigate the pupils' pre-test and post-test measures in mathematics, pre- and post-observation, and teachers' experiences and reflections on FAHOTS intervention was made.

The findings on the pupils' achievement showed that their mathematics achievement in post-intervention has significantly improved on their mathematics achievement in pre-intervention (effect size, r = .70). The effect size's magnitude was beyond the expected growth, an effect size greater than Hattie's (2013) hinge-point in an intervention. The pupils' assessment findings corroborate the observational findings for pre- and post-observation of the reviewed pupils' workbooks which showed a significant, positive change in mathematical HOTS learning tasks. The results support previous studies which show that teachers' formative assessment practice can be improved after exposure to an intervention (Accado, 2017; Box, 2015; Chemeli, 2019; McManus, 2008; Melani, 2017).

7.1.3 Chapter 6 Impact of FAHOTS on Teachers and Teaching

Next, the participating teachers were interviewed to investigate their experiences and reflections following the FAHOTS intervention.

The data sources (i.e. the observed teaching, the student workbooks, the interviews), showed a change in the teachers after the FAHOTS intervention. The triangulated data showed a beneficial and significant impact in this study. Thus, the observed records in pupils' workbooks were significantly different from pre-observation to post-observation. The descriptive findings on the observed teachers showed some elements of change in their FA practice, for instance, the sharing of learning goals and success criteria, and questioning techniques. This complemented the test score findings of change and gave the triangulation outcome on the causal impact of the intervention. Moreover, the interview results showed that teachers' experience and reflection were positive towards the intervention and change in some classroom assessment practice. Hence all those changes support a claim that this study intervention worked.

In conclusion, the exploratory case studies showed the possibility of enhancing teaching, learning and improving pupils' mathematics achievement in HOTS and, more importantly, validated evidence that the intervention worked. Teach teachers, and they will learn, as observed:

(i) **Developing classroom use of FA strategies was observed after the intervention–** specifically in relation to the frequent use of planned questioning and interaction, task-based assessments and effective feedback as a way of determining the effectiveness of engineering classroom discussion and enhancing pupils' HOTS. The pupils' achievement gain can be associated with the intervention, however, with some cautions.
(ii) **Teachers were likely to assign HOTS tasks to pupils in the classroom or for homework despite the intervention.** Workbook analysis revealed that teachers did assign some HOTS type tasks after the intervention and supported positive affirmations during the interviews.

The pupils' significant quantitative achievement between pre- and post-intervention was confirmed by the qualitative observation findings that the participating teachers were certainly employing some of the strategies but did not effectively support feedback on instruction. The pre- and post-observation findings on feedback were ineffective and weak. These results do not support previous studies on exposing teachers to training for monitoring, scaffolding and feedback delivery practices to support pupils learning improved by their diagnostic strategies and the quality of their support to pupils (Kyaruzi, 2017; Lizzio & Wilson, 2008; Van de Pol et al., 2014).

However, during the interview findings, findings emerged collectively from the participating teachers in which they confirmed the benefits of FA concerning pupils' effective participation in classroom engagement, energising pupils' interest in learning and improved achievement by pupils. In support, all the participating teachers highlighted the positive impact on pupils' reactions: excitement, building confidence, ever alertness, enjoyment and feeling free. They became open to the teacher, to one another, and they assisted each other. Similarly, the participating teachers outlined the level of improvement for their pupils in mathematics on average after the intervention phase.

Most importantly, some participants have already shared materials with other teachers. During the interview, the participants were very particular that the training workshop should be extended to other teachers and other subjects across the standards. One of the participants (*Participant 1*) asserted that:

> *I think formative assessment should be done or used in all schools and starting at the lower primary school and going up the standards so that pupils would acquire the key skills for HOTS or innovative skills at an early age and become independent thinkers as they grow up* (Participant 1).

This reflection from a teacher strongly supports that FA practice can impact transforming classroom teaching and learning to be pupil-centred. Engaging pupils encourage the room for in-depth thinking that pupils should also be exposed and not just to remembering factual knowledge. They should also use such knowledge to solve problems, analyse, create and evaluate. The support for FA was viewed as professional development. The participating teachers jointly acknowledged the training workshop as having added value to their lesson planning and directional teaching practice. They claimed not to have not possessed such knowledge before the training workshop. The teachers' reflection at this point seemed to indicate that in-service training in formative assessment is obliged to address the challenge for teaching HOTS.

7.1.4 The power of Context

The education system has failed year in and year out to equip a significant proportion of Botswana pupils with a basic achievement level, as repeatedly shown in both national and international standardised examinations. The Revised National Policy in Education RNPE (1994) is yet to be fully implemented twenty-six years after its inception. It has long recommended the introduction of continuous assessment (CA) as a component deemed necessary for formative assessment in schools. However, teachers in different schools have differently revealed some challenges and limited knowledge of formative assessment and teaching of HOTS in mathematics. Obviously, many pupils would have limited learning opportunities to raise their achievement if the schools make little effort to provide pupils with a conducive learning environment to explore problem-solving concepts in mathematics.

In Phase II (reported in Chapter 5), the pupils' pre-achievement findings were measured negatively in respect of logit (theta), below a mean value. This finding concerning pupils is consistent with the TIMSS 2015 study that had identified the Botswana pupils as unable to apply HOTS such as critical thinking and problem-solving in mathematics (as discussed in Chapter 1). Nationally, mathematics is also among the low performing subjects by pupils in primary schools through to secondary schools in Botswana on standardised assessments implemented by the Botswana government (BEC, 2017). The international and national patterns of underachievement highlight the need for the country's education system to implement focused interventions. Against this background, Botswana teachers' effectiveness and the availability of resources are among the pedagogical issues that emerged in the current study are aligned with other previous studies (Fetogang, 2015; Masole et al., 2016).

(i) **Teachers perceived challenges to implementing FA strategies** to the extent that during the interviews, the teachers said that they lacked time, training and support to use FA strategies effectively in the classroom.

The low pupils' achievement in HOTS may also have emanated from the curriculum. Thus, the items adapted from the BEC were application-oriented and mainly numbers and operations. There was little to no items identified for synthesis, analysis and evaluation, thus indicating Botswana's mathematics curriculum for Standard 4 was to some extent limiting the teaching of HOTS. Certainly, if the curriculum does not cover much domain HOTS, perhaps teachers not teaching HOTS is rational. The curriculum shortcoming may be one reason why Botswana’s pupils have lower achievement than international cohorts. For instance, the education systems in some Asian countries such as Hong Kong, Malaysia and Japan had long ago revised their curriculum in a deliberate, concerted effort to integrate HOTS as the main component for pupils’ learning. This integration was done to develop schooling for children who are critical and creative in thinking and are on par with global needs (Balakrishnan et al., 2016).

Schools’ management who espoused and selected the weaker classes of pupils to be given a chance to participate in the intervention contributed to the low performance in pupils’ pre-intervention findings. However, this selection of pupils might still have contributed to the validity of the intervention if any improvement by pupils was noted. The results for the comparison of the pre- and post-observation descriptively revealed positive developing evidence of FA practices among teachers. These results were consistent with the FA strategies implementation assumption that teachers are required to attempt to address all five broad formative assessment strategies in their classroom. Still, the specific techniques used within each strategy are up to the individual teacher (Education Endowment Foundation, 2018). It can be concluded that participating teachers for Standard 4 may have improved their assessment practice to be formative to enhance learning.

7.2 Reflection on the conceptual framework

The current study used a framework that was enveloped in the post-positivism paradigm. Pupils’ assessment zone of proximal development uses scaffolds and emphasises knowledge and understanding of pupils’ context in mathematics instruction (Heritage 2010; Kivunja 2015). As indicated in Chapter 2, Kivunja’s Assessment Feedback loop (AFL) model underpinned the conceptualisation and design of this study. It most closely supported preliminary ideas and appeared highly relevant to already existing classroom assessment mathematics achievement literature. Kivunja (2015) confirmed that the AFL model is an integral pedagogy of teaching, learning and assessment, which constitutes step by step consideration of nine essential processes. FA was infused to provide information on learning progress to pupils and the teacher, and to improve teaching and learning. Similarly, in this study, the AFL approach was used and regrouped into three phrases to address all the study's research questions as shown in Figure 7.1.

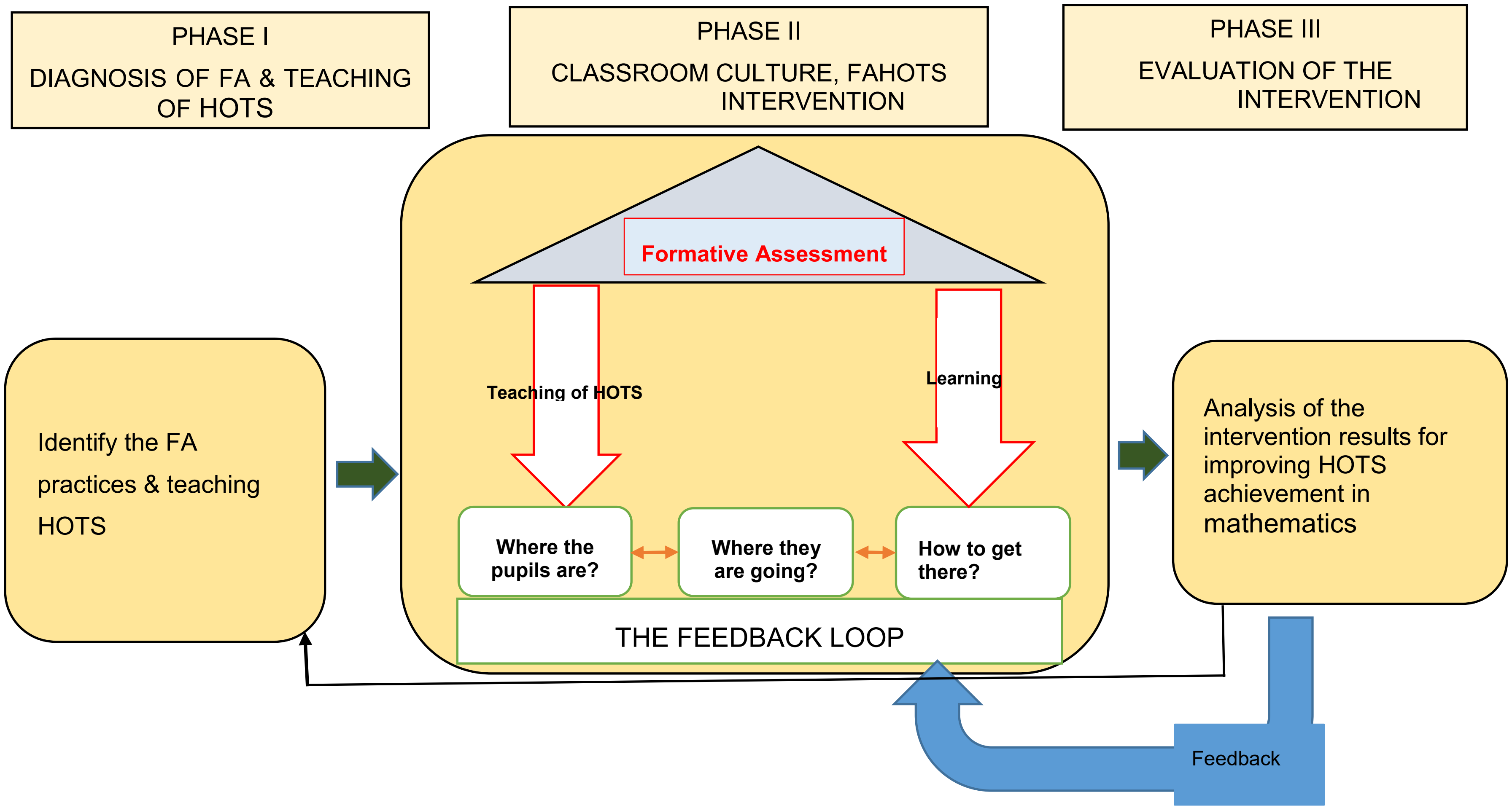


Figure 7.1: FA Feedback Loop to Enhance Pupils' Higher-Order Thinking Skills Model

Figure 7.1 provides the current study with an insight into the problem and an understanding of the impact of teachers' incorporation of FA in teaching HOTS among the Standard 4 pupils in learning mathematics.

The modification of the framework was done to simplify the explanation of the assessment processes in simple diagrammatical languages, as described in the following section.

7.2.1 Phase I: Teachers’ formative assessment and teaching of HOTS

Figure 10.1 of this Phase I of the study looked into the teaching and learning problem context, which seems to be orientated to teaching to the test or public examinations (Fetogang, 2016; 2017). It was set out as a theoretical lens to explore the Botswanan primary schools’ practice to determine the learning goals, learning progression and criteria for success (Process 1). This phase of the study customised a process as used by Kivunja to determine the teaching in the context of FA and HOTS (Research-sub question 1).

It has to be kept in mind that Phase I of the study was motivated by the purposeful diagnosis of classroom culture to ascertain the links between instruction and how the pupils are engaged concerning the instruction and meeting the pupils' learning outcomes. With reference to the results, as discussed in Phase I, the classroom teachers’ FA practice was not consistent with Kivunja’s Assessment Feedback loop, but rather indicated the fragmentation of classroom assessment practice as a point of departure. It was found that participants varied significantly among themselves in certain aspects of the framework for the formative practice in mathematics. Such a teacher assessment practice may have contributed to a lower pre-test achievement, but it cannot be confirmed with confidence as there was no control group. The baseline findings suggest that classroom pedagogies were not completely explained by the Kivunja model, instead their teaching approach was just didactically oriented. Therefore, this ineffectiveness of using FA allowed an opportunity for professional development training for the teacher.

7.2.2 Phase II: Classroom culture with formative assessment intervention

The outcomes of the survey were used to inform the teacher empowerment FAHOTS intervention to improve classroom assessment practices and teaching of HOTS. The classroom culture (teacher assessment, pupils and peer assessment) for teaching and learning was designed and embedded within the AFL as a theoretical foundation to guide teacher training development. The BS4MT were then empowered with professional training using FAHOTS as an intervention, emphasising the Feedback Loop outlined in Figure 7.2.

The exposure of BS4MT to the feedback loop (Process 2 Process 7, as detailed in Chapter 2) provided a means of gathering data that constituted evidence of learning. This evidence helped to identify gaps between what the pupils know, what they should know, and how to respond to pupils' learning of the need of HOTS to "close the gap" (Kanjee, 2017; Kivunja, 2015). It was found that after training, teachers were enabled to design learning activities within the FAHOTS intervention which activated effective classroom engagement and enhanced deeper learning activities for mathematics. In this way, the Kivunja model served as a good indicator of pupils' progress and helped teachers to identify where learning problems lay, and how they could be addressed to improve learning (Kivunja, 2015).

Figure 7.2 below shows a weekly FAHOTS feedback loop which participating teachers employed as assessment practice through a variety of learning activities, including clarifying and sharing of the learning goal and success criteria, the use of effective questioning, providing opportunities for self-and peer-assessment, and identifying the next steps in learning during daily lesson implementation for mathematics.

Model results (as discussed in Chapter 5) seemed to agree with the expected function of the conceptual framework used. The evidence suggests that FAHOTS impacted both teachers and pupils in terms of effective teaching and pupils' acquisition of HOTS in mathematics, respectively. Kivunja's (2015) model can be used to propel achievement in mathematics from the classroom observation data, and a significant gain of pupils' achievement was an indicator that FAHOTS might contribute to the achievement. However, longitudinal studies utilising control groups are needed to confirm the findings.

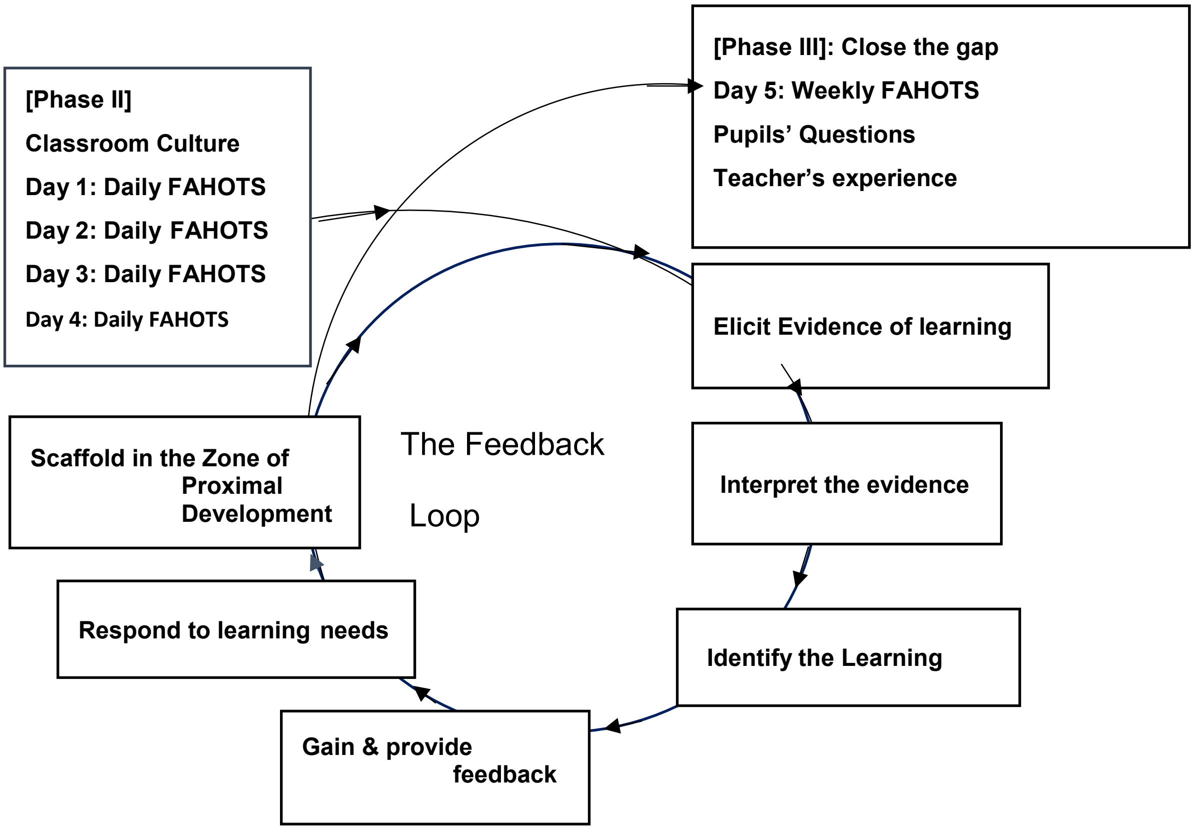


Figure 7.2: FA Feedback Loop to enhance pupils' Higher-Order -Thinking Skills model

Moreover, based on what the researcher saw, perhaps the teachers did not implement all of the aspects of FA that was expected from them upon the 2-days PD. Additionally, it is also worth noting that the pupils still did not cope with the hard items. This result suggests that the solution requires more than two days of PD and eight weeks of intervention. There is a need for a staged series of workshops and sequenced PD arising from this. A series of workshops which should only introduce one strategy at a time until some evidence in practice revealed its success, and then move to the next strategy, as outlined below;

(i) Get everyone to work on Learning Goals and Success Criteria

(ii) Once that is embedded, work on the use of effective questioning, collaborative learning inquiry and learning task (HOTS) integration

(iii) Last work on PASA

7.2.3 Phase III: Evaluating classroom culture with formative assessment intervention

As used by Kivunja's model, closing the gap (Process 8 as discussed in Chapter 2) is the final element in the FA process. It forms the beginning of the next set, in which this study attempted to evaluate the experience of the participating teachers as an attempt to complete the circle and start a new FA cycle, as shown in Figure 7.2.

In this phase of the study, evaluation of the FAHOTS intervention constituted teachers' experience in applying mathematical HOTS related concepts, which were acquired through independent classroom tasks, direct and indirect instructions, interactive instructions and experiential learning. These activities made up the remainder of the study. The evaluation ended with teachers' reflections following the FA strategies intervention and teaching of HOTS concepts in mathematics. Based on the findings, the Kivunja model fitted the concerns and needs of teachers to improve achievement. As discussed in Chapter 2, through Kivunja's Model, teachers' reflections seemed to confirm that formative assessment implementation can inform their classroom pedagogy for teaching mathematical HOTS concepts.

In conclusion, Kivunja's Assessment Feedback loop was used to explain the phenomenon as it supported preliminary ideas about formative assessment and teaching of HOTS in mathematics to raise achievement. All three proposed phases of the study fitted the conceptual framework, and overall, the model provided sufficient support for the implementation of the intervention. The adjustment of the assessment processes into three phases was a suggested arrangement that made it easier to explain and interpret the flow of events to assist teachers in understanding and implementing the formative assessment in the Botswanan classroom environment. It has to be kept in the mind the FAHOTS is a pedagogical model, not inherently connected to teaching mathematics. However, perhaps at the Standard 4 level of primary schooling, it may provide sufficient classroom pedagogy. Similarly, this intervention is also relevant when dealing more deeply with mathematical thinking at higher school levels since FA assists the mathematics teachers when dealing with errors, responding to pupils appropriately, and managing the (lack of) connection of classroom work to mathematics among pupils (students). So, recommending that teachers should be encouraged to use errors in mathematics learning tasks (problem-solving tasks) or assignments to monitor, scaffold, (i.e., improve learning, improve the quality of feedback practices (i.e., “feedback delivery” and “promoting feedback seeking”) provided insights into mathematics teachers’ FA practice.

7.3 Methodological Limitations

The current study chose and employed a sequential embedded mixed method design as a methodological strategy and different data collection methods to answer the posed questions. In this case, the study entrenched and employed a survey questionnaire, classroom observations, pupils’ mathematics assessments, post-observations and interviews to explore the formative assessment to enhance teaching HOTS in mathematics. Such a broader approach of investigations on the phenomenon of interest are not without limitations that may impinge and restrict some conclusions.

7.3.1 Challenges related to design

Even though the cross-sectional survey entrenched some observational data, still the findings cannot be generalised beyond this sample, as the first limitation of the study. Secondly, a single-group design was used, without a control group, therefore the conclusions of the study are limited. The conditions were not randomly selected nor rigorously controlled. Extraneous variables may account for some of the changes observed. Additionally, some school management wanted the teachers with the weakest classes in mathematics achievement to participate in the study. Their preference to include the lowest-achieving classes in the intervention may be one reason the pre-test achievement was so low. However, the study's internal validity was enhanced to some extent during the assignment of the participants, and through triangulation on the causal impact of the intervention that was claimed. Additionally, Rasch modelling of the quantitative analysis (in Phase 1, teachers' survey and pupils' assessment in Phase II) enhanced the accuracy of evaluating the instruments' functioning. Additionally, the audiotaped interview (Phase III) also enhanced the authenticity of the findings. However, the results should be interpreted with some caution because the researcher's multiple roles as a principal investigator, interventional developer, facilitator and evaluator during the study might have impacted the findings of the study. In future, it would be imperative to include a research design that randomly assigned the experimental and control group as well as involving some trained research assistant in some stages.

7.3.2 Sample size and data analysis

Firstly, the low response rate of postal questionnaires during Phase I reduced the sample size and introduced a response bias in the study findings. Despite all mailed questionnaires being accompanied by a covering letter and including a stamped, addressed envelope, the response rate was minimal. Only 15% of participants from the two participating sub-regions returned the questionnaires. The attempt by the researcher to randomly distribute more questionnaires directly to be collected at later dates as a strategy increased the effective response rate to 95%. The study findings had overrepresented some teachers in three sub-regions while under-sampling the participants in the other two sub-regions; hence the results could not be compared across the five education sub-regions.

7.3.1 Psychometric properties of measurements

Statistical analysis procedures require enough representation of responses per category to balance the participants' responses for the items, particularly the even scales. For this thesis, none of the scales used balance well, affecting almost 50% of the total items in each scale due to the small samples. A decision was made to collapse some categories to enhance the balancing of responses on scales. With Rasch modelling data, measures are built upon the best judgments of spread items (evenness steps), the reduced error of measurement (precision), probability and improbability (fit) of item and person values to that expected from the model and overall reliability (noise) to accord an accurate interpretation. Against this backdrop, caution should be employed when discussing the comparison and integration of findings. Nevertheless, Rasch modelling some of the teacher variables (in particular of the quantitative survey) and all the pupils' assessments provide some confidence in evaluating the instruments' functioning and enhanced accurate interpretation of the findings. In future, for a sample of fewer than 300 participants, an odd Likert scale would be employed while a large sample would use an even Likert-scale instrument.

7.4 Conclusions and Recommendations

During the past decade, debates in Botswana has centred on pupils' low achievement in basic numeracy, literacy, and life skills. In response to the situation, the government has invested considerable funding into education expenditures. Notably, since 2014, the Ministry of Basic Education has been leading with more than 30% of the government's total national budget, but still, pupils' achievement remains low. In return, the different stakeholders still call on the government to address the education system's inadequacy. Anecdotal issues in the public domain include class size, crowded summative assessments, the collapse of in-services structures and function, reduced contact time due to unresolved work hours, automatic pupil progression and administrative crisis in schools.

7.4.1 Main conclusions

The studies presented and discussed in this thesis corroborated the status quo in Botswana. The baseline for teachers' practices and pupils' results found them aligned with previous research about poor teaching pedagogy in the classroom and performance respectively. Formative assessment is an active and intentional learning process that partners the teacher and the pupils to continuously and systematically gather evidence of learning with the express goal of improving pupil's learning outcomes. However, baseline research evidence in this thesis showed pupils did not benefit from the teachers' pedagogy practices in the classroom.

The preceding description of formative assessment reflects the objective of guiding instruction and assisting pupils in learning HOTS; therefore, an intervention FAHOTS was created. In addition, a curriculum aligned pre- and post-test was created aligned to the curriculum and equated to each other. Case studies in this thesis showed that interventions could improve teaching and learning; as the teachers' teaching and attitudes improved. There was evidence of patterns of change as the teachers started to use some new strategies in teaching mathematics. Moreover, the teachers made some efforts to plan objective questions for HOTS through improved questioning techniques and better engagement of all pupils. Similar to pupil's achievement, this thesis showed that the pupils' pre- and post-achievement gains in mathematics reflected the quality of instructional strategies that teachers employ in the classroom. Most importantly, it has to be considered that mathematics emphasises pupils' ability to develop and apply mathematical thinking to solve a range of problems in everyday situations, which has improved after the intervention.

As revealed in this thesis, contextual factors that constrain improvement include resources and support to continue using the FAHOTS intervention. More specifically, it is important to note that the Standard 4 mathematics curriculum (as a reference resource for the teachers) has some significant limitations in addressing all cognitive development levels based on Bloom's taxonomy. The curriculum mainly focuses on applying basic mathematics, and little or nothing is covered beyond synthesis and evaluation. Therefore, if the curriculum at the teachers' disposal is incapable of infusing the other higher level of skills, it is not surprising that they cannot teach such levels. Then Botswanan pupils cannot fairly compete in large-scale international assessments.

It should be noted further that the evidence in this thesis did not determine if there is a need for a sequence of future professional development (PD) to change teacher practice gradually. This thesis also failed to establish if the intervention works with higher ability pupils or if the intervention works with higher standard (grade) levels. Additionally, this thesis did not determine if changes in policy or curriculum would improve teaching and outcomes more than PD.

However, the participating teachers' reflections supported this thesis's findings that rolling out FA is a desirable endeavour. The participating teachers suggest that formative assessment is needed in mathematics for Standard 4, but it should also be rolled out to all subjects in different standards and across primary schools in

Botswana. The study's findings came when the education system of Botswana is struggling to raise the pupils' achievement. The findings came from reflections of the participating teachers that mainly focused on formative assessment enhancement to learning mathematical HOTS concepts. It became apparent that the majority of the pupils seemed to have demonstrated a change in their motivation, in response to the questions and attempted to reason rationally and critically, though the pupils' achievement remained low.

The attempt of the thesis, in general, has been typically consistent with the approach for the AFLA project, which focuses on the use of assessment for improving learning outcomes in the core curricular area of numeracy. The findings of this thesis apply in particular in challenging educational settings described as schools with harsh realities, with large classes and few resources (Aga Khan University, 2017; Kanjee, 2017, Kyaruzi, 2017). The current study was also kept in mind with such observed pupils' deficiency in numeracy, especially in the HOTS domain and it was postulated that FA could be a possible solution as confirmed by the pockets of findings in this study.

7.4.2 Recommendations

This thesis provided some insight into Botswana teacher' experience of formative assessment in Standard 4 mathematics with a focus on HOTS. The research has some implications for theory, practices, policy and future research as identified in the next section.

(i) Theoretical implications

This thesis aimed at investigating twofold issues in three empirical phases: (1) to explore FA practices in the Botswana context, and (2) to propose a possible effect of an intervention for BS4MT's use of the FA strategies in the classroom to enhance pupils' HOTS. It is extensively reported that if FA is well implemented and well perceived by pupils, it can improve pupils' learning (Black & Wiliam, 1998, 2009; James & Pedder, 2006; Kyaruzi, 2017; Wiliam, 2011; Wiliam et al., 2004). FA strategies help to identify gaps between what pupils know, and they should know and determine how to respond to pupils' learning tasks (HOTS) to close the gap (Black &

Wiliam, 2003, 2006; Hattie & Timperley, 2007; Wiliam, 2013; Wiliam & Thompson, 2014).

The results in Chapter 6 showed some indications that participating teachers had implemented some of the interventional strategies while teaching mathematics. The teachers were still developing and occasionally used some strategies such as learning goal and success criteria in their lesson introduction, then questioned and interacted with their pupils through eliciting learning tasks and feedback, which is consistent with the theory proposed by Kivunja (2015) and Heritage (2010). These findings support previous work on primary teachers' FA-focus and assessing the learning of pupils in mathematics (McGrane et al., 2018; Kanjee, 2017; Kivunja, 2015; Kyaruzi et al., 2020; Aga Khan University, 2017). More specifically, the findings support Wylie and Lyon (2015), who found that teachers frequently implemented only one FA practice associated with the learning goal. The teachers scored significantly lower on the quality of implementation scale for this FA strategy than the other FA strategy practices implemented after two-year of professional development.

The Chapter 5 results showed a statistically positive impact in the pupils' workbooks, indicating that the number of HOTS tasks given to pupils' post-intervention increased. Such findings imply that teachers changed their traditional practice of issuing exercises and support. Nenty and Odili's (201 2) theoretical submission that teachers who possess detailed conceptual knowledge and apply the levels and sub-levels of Bloom's taxonomy of human cognitive behaviour can subsequently influence their teaching and learning outcome demands. Thus, the findings presented in Chapter 3 indicated that pupils' mathematical achievement has improved. These results have been associated with the teachers' implementation of HOTS by planning and engaging pupils in teaching and learning processes of mathematics, which are essential to change the stigma of the difficulty of mathematics (Abdullah et al., 2017; Serin, 2013; Brown 2002b; Brown & Hattie, 2004).

The results of Chapter 4 and 6 showed the participating teachers had limitations in using peer- and self-assessment (PASA) during classroom activities. These findings confirm the conclusions from previous studies based on various themes (such as improvement, accountability, social interaction, and accuracy) which found that teachers were challenged to plan or structure the activities (Wylie and Lyon, 2015). These studies suggested that teachers must be provided with concrete instruction for PASA and carefully manage interpersonal issues for successful implementation (Harris & Brown 2013; Noonan & Duncan, 2005; Suurtamm, 2004).

Several studies have investigated teacher-focused FA practices, including video applications (Kyaruzi, 2017) and everyday classroom observations (Accado, 2017; Akom, 2010; Chemeli, 2019b; Kanjee, 2017). However, only a few studies have been done in the African context, and Botswana has had minimal exposure to such interventional inquiries. From Chapter 6's results, participating teachers supported the professional development and recommended implementing the FA strategies to improve pupils' performance. These results, concerning support for FA through staff development, are consistent with studies that showed that the teachers became more favourably disposed towards such assessment and their implementation skills improved greatly (Accado, 2017; Kyaruzi, 2017; Chemeli, 2019b).

(ii) Practical implications

This thesis has demonstrated the effect of an intervention concerning teachers' use of classroom FA strategies to enhance pupils' HOTS. The practical implications for this thesis are based on the major findings across the case studies. In Chapter 4, it was noted the teaching of higher-order thinking skills (HOTS) in mathematics needs to be integrated as early as possible into the education system. In this way, the researcher suggests Standard 4 teachers build pupils' numeracy reasoning right from the early years of schooling and facilitate learning through effective questioning and using mathematics learning tasks from their local surroundings, enhancing HOTS as per Bloom's Taxonomy guidance. Secondly, the Standard 4 teachers are urged to plan for every lesson with some planned key questions that direct every learning goal and success criteria.

(iii) Policy implications

The findings in this thesis involve policy, professional development and teacher education, as noted in Chapter 4. Thus, professional development is recommended in future to improve teachers' integration of FA and HOTS linked as FAHOTS. The results in Chapter 5 and 6 indicate that teachers were likely to assign HOTS tasks to pupils, and demonstrated growing classroom use of FA strategies were observed after the intervention. Deep knowledge and skills are required of non-subject specific pedagogical knowledge and subject-specific knowledge to explain different representations of a subject in a mathematical idea for teachers to implement FA strategies effectively. It is recommended that teachers need specialisation in education, specialised teaching, subject-oriented support, and networking clusters across primary school levels to implement such formative assessment effectively.

Customised professional training on the essentials of questioning based on Bloom's Taxonomy and teachers' guides and practices on asking HOTS questions based on Bloom's Taxonomy is recommended for both in-service and pre-service teachers. Specifically, staged and sequenced PD training should include:

a) identifying and communicating learning goals and success criteria for a given topic using Bloom's Taxonomy (particularly HOTS in analysis, synthesis, and evaluation), including the value of aligning assessments with the learning goals;
b) developing higher-order thinking questions following Clarke's (2005) five strategies to improve the use of questioning by changing recall questions into questions that improve pupils' engagement during the lesson and improve the HOTS and; and
c) exposure to self and peer-assessment (PASA).

The use of FA practices should be extended beyond daily traditional classroom activities and work towards improving pupils HOTS engagement in the classroom or for homework in information technology-based developments of interactive teaching and learning materials. Thus, it is recommended that teachers be taught the use of formative assessment and feedback practices that can advance pupils' activities to go hand in hand with the endeavour to capture, visualise and provide feedback on complex cognitive, emotional and behavioural patterns.

Training is recommended for heads of schools and the pedagogical teaching staff in each school to consider formative assessment as a strategy for teaching and learning, not just an assessment. Thus, the Ministry of Basic Education for Botswana is encouraged to equip teachers with professional development on the effective assessment and practices from the Continuous Assessment (CA) programme. The teacher training institutions for both primary and junior secondary school teachers should equip pre-service mathematics teachers with formative assessment to foster an in-depth understanding of concepts to improve mathematics performance at both levels. The Teacher Training and Development Department should provide in-service mathematics teachers, Heads of Departments, School Heads, Principal Education Officers, curriculum planners and government officials with formative assessment strategies and processes in different settings. Such settings include workshops, conferences, presentations and publications for government and other key stakeholders in education to enhance the marketing of the practice that can help pupils perform well in both Standard 4 attainment examinations, PSLE and JCE levels and large-scale international assessments.

This preceding recommendation is consistent with the view of the Ministry of Basic Education, Botswana, a currently proposed reform of the primary school assessment over the next three years in line with the provisions of the recently developed draft General Education Curriculum and Assessment Framework (GECAF). According to Basic Education Minister, Mr Fidelis Molao, one of the critical decisions yet to be made in considering the draft GECAF related to the discontinuation of the Primary School Leaving Examinations (PSLE), which was in line with international trends, is to replace it with a robust school-based assessment system (BOPA, 2020). "This will entail giving more time to learning than teaching with an emphasis on continuous standardised school-based assessment for progressive learning," (BOPA, 2020).

The thesis has also provided strong evidence in support of and added value to the Botswana ETSSP (BOT, 2015) plan as discussed in Chapter 1, which proposed in-service training of teachers in school-based assessment in relation to the development, implementation, assessment management, and monitoring tools as well as procedures for assessment in the classroom as an attempt to raise pupils' achievement in schools and therefore assist in the situation.

(iv) Future research

The study provided evidence on the possible effect of formative assessment to enhance HOTS mathematics teaching using teachers and pupils in the intervention. However, the study employed research that used a limited sample; hence the results could not be generalised to all the schools but provided a fundamental basis upon which further studies could be built. It is suggested that further research should be conducted in the following areas:

a) Including a larger sample size, which will enable the results to be generalised to all schools and different contexts with confidence;
b) Additionally, a similar study with an active control group should be done to understand classroom practices accordingly. The pupils should also be involved in surveys for both pre- and post-interventions to determine a change in some factors such as motivation/attitude, pupils' engagement in HOTS and instructional feedback on mathematics that can be examined further to determine the effect of the intervention.
c) A follow-up study can be done comparing the intervention and control pupils' level of achievement in the post-achievement test. This follow-up could provide evidence of the intervention's predictive validity and the attainment of standards set for performance in mathematics.

Additionally, the design of a new questionnaire or improving the existing one can be enhanced by applying Rasch models at each instrument design stage to facilitate refinement, use and accurate reporting. Thus, the principles of measurement within these approaches provide research findings accurately considering reliability and validity, since self-report is deemed to be susceptible to social desirability because of "the tendency on the part of individuals to present themselves in a favourable light" (Van Staden & Zimmerman, 2017). Additionally, the socially desirable responses may be reduced by ensuring the confidentiality and anonymity of teacher responses, gathering data longitudinally rather than just at one point in time, and gathering data from more than one source.

7.4.3 New knowledge, innovations and insights

This study provided the Botswana Teachers' experience towards FA practice and the integration of HOTS in teaching Standard 4 mathematics. The mixed-methods exploration was done to get rich data for Botswana Teachers' FA practice and teaching of HOTS data as used by teachers to corroborate the use of teaching strategies and compare the data of pupils' achievement in mathematics.

The mixed approach inquiry provided an opportunity to gauge the extent of different FA practices. Thus, during Phase I of studying the teachers, the researcher came to realise that sharing of learning objectives as one of the FA strategies was socially more desirable to all participating teachers than any other. The participating teachers could not interpret and implement FA strategies as an interrelated classroom practice to yield meaningful pupil-centred learning approaches. Instead, teachers were placing a high value on the procedural type of questions, which in most cases, did not measure the learning goal and success criteria and did not even engage all pupils in the classroom. In addition, poor pupil achievement is a measurable indicator that may point to a poor-quality education system and facilitate the widening gap towards an unattainable knowledge-based economy.

Having been exposed to the intervention, the participating teachers had different perceptions towards FA and the integration of HOTS. Despite that, the teachers were still developing the context of FA; however, they seemed to promote the quality of FA and HOTS in relation to:

- pupils' actively engaging in lessons from the very start to the end;

- encouraging teachers' and pupils' exploration of practical learning opportunities and experiences from improvised materials;
- pupils' encouragement to think, question and talk;
- reduction of teachers' workload due to activities that are pupil-centred;
- teachers and pupils actively listening, asking questions, summarising and explaining their understanding;
- encouraging group talk and collaboration through articulation–using appropriate vocabulary, pupils clarify their learning;
- teachers and pupils playing a key role in mediating learning experiences through active listening, asking appropriate questions, summarising, explaining and understanding; and
- a constructive and reflective learning environment so that pupils feel safe to make mistakes.

The link between HOTS and FA suggests that a teacher who properly implements the FA strategies can also effectively influence pupils' acquisition of HOTS thinking, learning, and mathematics achievement. Equally, the participating teachers had some challenges with the new approach. They asserted that implementation of FA required instructional resources like pupils' whiteboards, exit cards and ICT related teaching aids (for instance, projectors, computers and internet connectivity) to benefit fully from the strategies. In all cases observed, the classrooms were traditionally oriented to chalk and blackboard. For all the participating teachers, lesson planning differentiated between their traditional common format and new approaches within FA, which also remained a challenge, and was often beyond their control. So, for example, teachers could not fully adopt the new intervention into their daily classroom practice since they were torn between what their employer expected of them and the alignment of the new FA practice intervention. Additionally, a lack of specialisation in teaching at the lower primary school was found to be a contributing factor since all the observed teachers were not specialised in mathematics.

Moreover, specifically, the Botswana Standard 4 mathematics curriculum does not seem to account for its second aim which intends to develop in all pupils the "*inquiry skills, creativity, critical thinking and problem-solving ability"* (BEC, 2002). From the review analysis of the curriculum, it was identified that the objectives were inclined to measuring knowledge and application cognitive domains for basic mathematics. Certainly, the application of sequential series of numbers and operations in problem-solving tasks throughout the curriculum does not seem like higher- order-cognition in the absence of analysis, synthesis and evaluation in varied content and context.

Understanding these challenging trends towards implementing FA in primary schools can help national and local policymakers create the structure and support that can lead to the development and implementation of school-based management of FA in the classroom across school systems. The current study aids the possibility of revising the mathematics curriculum and training on quality assessment in African education systems to enhance teaching and learning standards for numeracy. Thus, there is a need for multiple stakeholders to work together to create conducive environments for FA strategies to be utilised as well as alignment with policy and curriculum needed.

The current study remains a case providing evidence to improve mathematics achievement in primary education in Botswana. As already stated in Chapter 2, Botswana ETSSP’s (BOT, 2015), poor teaching and curriculum delivery, and the decline of educational quality in core subjects such as mathematics and science were identified as critical challenges in the Botswana primary schools’ context. Therefore, it is hoped that the implementation of the recommendations indicated in this study will assist in addressing the challenge that the findings of this study have highlighted.

7.5 Closing Thoughts

The pupils' low achievement in numeracy has been a topical debate by the Botswana education stakeholders, at least for a decade now. Interestingly, it is still the same education system entrusted with the power to assist the pupils in attaining basic numeracy. Despite Botswana being an upper-middle-income country, the country's education and the economy are seen as not closely interrelated in any way. Ideally, economic development provides the resources to sustain educational growth and creates opportunities for profitable work. In return, education supplies the skills and the specialised manpower that the economy requires and develops abilities useful to the self-employed in their occupation (BOT, 1977).

The foregoing anecdotal evidence has a bearing on the education processes of any country. For instance, the current study has attempted to demonstrate an intervention's possible effect to enhance HOTS in teaching mathematics. The ultimate goal for teaching pupils the HOTS is to equip them with 21st-century skills which are centred on innovative thinking and problem-solving skills to embrace the 4 Cs - “super skills” (Creativity, Communication, Critical Thinking, and Collaboration). However, the current national curriculum for a lower primary school in Botswana has not been transformed, as it only outlines the aims and specific learning outcomes without provision for varied and appropriate learning opportunities. The omission of the learning opportunities, and more specifically, the teaching of HOTS, is enough evidence to predict that pupils would be unable to handle HOTS related tasks.

Moreover, the deterioration of the pupils' mathematics achievement demands empirical attempts at determining possible factors that might contribute to this problem in education. For instance, the implication of the mathematics gap achievement may depend upon the teacher's abilities to interpret and teach the curriculum explicitly to the required level. In the Botswana context, however, there is no specialised teaching at primary schools. Most teachers are expected to teach all the subjects, which compromises the quality of teaching and learning of mathematics.

In addition, the Botswana Government, through ETSSP reform, recognises the curriculum framework that takes full account of global trends in education where learning is more focused on learning outcomes and developing skills. Thus:

> *There are initiatives underway to introduce different approaches, along the lines of an outcomes-based approach to learning. Plans are underway to achieve a better balance between subjects, time, content, skills, national and school-based assessment through a curriculum review and the establishment of a national curriculum framework.* (BOT, 2015; p.30)

The current study can impact the curriculum, teaching, learning, and HOTS assessment in mathematics for primary schools. Further research is needed to continue promoting the strategies and solutions identified for the future teaching of mathematics in Botswana aligned to an outcome-based approach to learning.

8 REFERENCES

Abdullah, A. H., Mokhtar, M., Halim, N. D. A., Ali, D. F., Tahir, L. M., & Ilapsed, U. H. A. (2017). Mathematics Teachers' Level of Knowledge and Practice on the Implementation of Higher-Order Thinking Skills (HOTS) *EURASIA Journal of Mathematics Science and Technology Education*, *13* (1), 3-17.

Accardo, A. (2017). Monitoring Student Learning in Algebra. *Mathematics Teaching in the Middle School, (22)*6, 352-359.

Abd-El-Fattah, S. M. (2015). Rasch Rating Scale Analysis of the Arabic Version of the Physical Activity Self –Efficacy Scale for Adolescents: A Social Cognitive Perspective. *Psychology, 6*, 2161-2180. http://dx.doi.org/10.4236/psych.2015.6 16213. [2018, 28 September].

Adams, C. (2011). *Teaching Problem Solving Skills in Math by Engaging Higher Order Thinking.* Masters Research Project, St. Mary's College of Maryland.

Adedoyin, O. O., & Chilisa, B. (2004). Assessing the role of teachers in the educational access of students to mathematics at the junior secondary school level in Botswana. In *22nd AEAA Conference: The role assessment in the implementation of national education policy*, 13 - 17 September 2004. Gaborone International Convention Centre.

Adelson, J. L., & McCoach, D. B. (2010). Measuring the mathematical attitudes of elementary students: The effects of a 4-point or 5-point Likert-type scale. *Educational and Psychological Measurement, 70*(5), 796–807. https://doi.org/10.1177/001316441036669

Aga Khan University. (2017). Assessment for Learning in Africa Project Launch and Public Seminar: Improving Pedagogy and Assessment for Numeracy in Foundation Years. *Institute for Educational Development Aga Khan University for Education, Salama House,* January 24-27, 2017, Dar es Salaam.

Aggarwal, R., Nagpurnanand, P., & Manju, P. (2002). Institutional Allocation in Initial Public Offerings: Empirical Evidence. *Journal of Finance, 57*, 1421-1442, Available at SSRN: https://ssrn.com/abstract=292106. [2018, 22 April].

Ajeigbe, Taiwo & Afolabi, Eyitayo. (2014). Assessing Unidimensionality and Differential Item Functioning in Qualifying Examination for Senior Secondary School Students, Osun State, Nigeria. *World Journal of Education. 4*(10), 30-.54.

Akir, O., Eng, T, H., & Malie, S. (2012). Teaching and learning enhancement through outcome-based education structure and technology e-learning support. *Procedia Social and Behavioural Sciences*, *62*, 87–92.

Akom, G. V. (2010). *Using formative assessment despite the constraints of high stakes testing and limited resources: A case study of chemistry teachers in Anglophone Cameroon* (Doctoral dissertation, Western Michigan University). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3470394).

Allen, R., Elks, P., Outthred, R., and Varly, P. (2016). Uganda's Assessment System: A Roadmap for Enhancing Assessment in Education: An Assessment Report.

Al-Wassia, R., Hamed, O., Al-Wassia, H., Alafari, R., & Jamjoom, R. (2015). Cultural challenges to implementation of formative assessment in Saudi Arabia: An exploratory study. *Medical Teacher*, *37*(Sub1), S9–S19. https://doi.org/10.3109/0142159X.2015.1006601.

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). A *taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Boston, MA: Allyn & Bacon.

Anderson, C., & Palm, T. (2017). The impact of formative assessment on student achievement: a study of the effects of changes to classroom practice after a comprehensive professional development programme. *Learning and Instruction, 49*, 96 –102.

Andrich, D. (1978). Relationships Between the Thurstone and Rasch Approaches to Item Scaling. *Applied Psychological Measurement*. 2(3):451-462. doi:10.1177/014662167800200319

Andrich D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms? *Medical care*, *42*(1 Suppl), I7–I16. https://doi.org/10.1097/01.mlr.0000103528.48582.7c

Andrich, D. (2016). Georg Rasch and Benjamin Wright's Struggle with the Unidimensional Polytomous Model with Sufficient Statistics. *Educational and Psychological Measurement, 76(5),* 713–723.

Anselmi, P., Vidotto, G., Bettinardi, O., & Bertolotti, G. (2015). Measurement of change in health status with Rasch models. Health and quality of life outcomes, 13, 16. https://doi.org/10.1186/s12955-014-0197-x.

Arase, A., Kamarudin, N., & Hassan, A. (2016). The Development of Students' Capabilities in Higher Order Thinking Skill (HOTS) Through Science Education *Jurnal Pemikir Pendidikan, 2016*, (7),1–18.

Armitage, A. (2007). *Mutual research designs: redefining mixed methods research design.* Retrieved on July 20 2020 fromhttp://www.leeds.ac.uk/bei/Education-line/browse/all_items/167799.html.

Arnold, J. C., Boone, W. J., Kremer, K. & Mayer, J. (2018). Assessment of competencies in scientific inquiry through the application of Rasch Measurement Techniques, *Education Sciences*, *8*(184), 1-20.

Avedia, A. (2014). *Survey Design*. Boston, MA: Harvard law school.

Babbie, E. R. (2009). *The Practice of Social Research* (12th Edition). Belmont, CA: Wadsworth Publishing. ISBN-13: 9780495598411.

Bachor. D. G., & Anderson, J. O. (1994). Perspectives on Assessment Practices in the Elementary Classroom in British Columbia, Canada. *Assessment in Education: Principles, Policy & Practice*, *1*(1), 63-94.

Baghaei, P., & Amrahi, N. (2011). The effects of the number of options on the psychometric characteristics of multiple-choice items. *Psychological Test and Assessment Modeling*. 53. 192-211.

Balakrishnan, M., Nadarajah, M. G., Vellasamy, S., & George, E. G. W. (2016). Enhancement of Higher Order Thinking Skills among Teacher Trainers by Fun Game Learning Approach. *World Academy of Science, Engineering and Technology International Journal of Educational and Pedagogical Sciences. 10*(12), 2954 – 3958.

Balan, P., & Metcalfe, M. (2012). Identifying teaching methods that engage entrepreneurship students. *Education and Training*, *54*(5), 368-384.

Baird J., Andrich, D., Hopfenbeck, T.N. & Stobart, G. (2017) Assessment and learning: fields apart? *Assessment in Education: Principles, Policy & Practice, 24*(3), 317-350. https: doi.org/10.1080/0969594X.2017.1319337.

Baird, J. (2010). Beliefs and practices in teacher assessment. Editorial. *Assessment in Education: Principles, Policy & Practice, 17*(1), 1–5.

Bambale, A. J. (2014). Research Methodological Techniques as a Model for Quantitative studies in Social Sciences. *Journal of Economics, Management and Trade*, 862-879.

Biemer, P. P. L. (2003). *Introduction to survey quality*. Hoboken, NJ: John Wiley

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25. https://doi.org/doi:10.1080/0969594x.2010.513678

Bennett, R., & Gitomer, D. (2009). Transforming K–12 Assessment: Integrating Accountability Testing, Formative Assessment and Professional Support. 10.1007/978-1-4020-9964-9_3.

Berry, R. (2011). Assessment trends in Hong Kong: seeking to establish formative assessment in an examination culture. *Assess. Educ. Policy Principle. Practice.* 18, 199–211. https://doi.org/10.1080/0969594X.2010.527701g.

Berry, R., & Adamson, B. (Eds.). (2011). *Assessment reform in education: policy and practice* (Vol. 14). Berlin: Springer Science & Business Media.

Biesta, G., & Burbules, N. C. (2003). *Pragmatism and educational research*. Lanham, Md: Rowman & Littlefield Publishers.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). Assessment for Learning: Putting It into Practice Black, P., & Wiliam, D. (1998) 'Assessment and Classroom Learning, Assessment in Education: Principles, Policy & Practice, 5:1, 7-74.

Black, P., & Wiliam, D. (2006) Developing a theory of formative assessment. In J. Gardner (Ed.), Assessment and learning (pp. 81-100). London: Sage. https://doi.org/10.1080/0141192032000133721

Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 81-100). London, UK: Sage

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31. https://doi.org/10.1007/s11092-008-9068-5.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York, NY: David McKay.

Bode, R. (2001). Partial credit model and pivot anchoring. *Journal of applied measurement, 2*, 78-95.

Bond, T. G., & Fox, C. M. (2007). Apply the Rasch Model: Fundamental Measurement in the Human Sciences (3rd Ed). Mahwah, NJ: L. Erlbaum

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

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (Ed. 3). New York, NY: Routledge.

Bonner, S. M. (2016). Teachers' perceptions about assessment: Competing narratives. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 21-39). New York: Routledge.

Boone, W., J. (2016). *Rasch Analysis for Instrument Development: Why, When, and How?* Bethesda, MD: CBE- Life Sciences Education, 15:rm4, 1-7.

Boone, W. J., Staver, J. R. & Yale, M. S. (2014). *Rasch analysis in the human sciences.* London: Springer.

Boone, W., J & Noltemeyer, A. (2017). Rasch Analysis: A Primer for School Psychology Researchers and Practitioners. *Cogent Education. 4.* https://doi.org/10.1080/2331186X.2017.1416898.

Bopa (2020). Ministry reforms primary school curriculum, Botswana daily news- July 28, 2020. http://www.dailynews.gov.bw/news-details.php?nid=57442

Botswana Examination Council. (BEC). (2008). Standard 4 Assessment Report 2007. Gaborone.

Botswana Qualifications Authority No 24. (BQA). (2013*). Botswana Qualifications Authority Act (Act No. 24 of 2013).* Gaborone. Government Printers.

Box, C. (2015). A Case Study of Teacher Personal Practice Assessment Theories and Complexities of Implementing Formative Assessment. *American Education Research Journal, (52)*5, 234-267.

Braun, H., & Kanjee, A. (2006). Using assessment to improve education in developing nations. In H. Braun, A. Kanjee, E. Bettinger, & M. Kremer (Eds.), *improving education through assessment, innovation, and evaluation. Cambridge,* MA: American Academy of Arts and Sciences.

Brown, G.T.L. (2002a). Student beliefs about learning: New Zealand students in Year 11. Academic Exchange Quarterly. 6. 110-114.

Brown, G. T. L. (2002b) Teachers' conceptions of assessment. Unpublished doctoral dissertation, University of Auckland.

Brown, G. T. L. (2008). Assessment literacy training and teachers' conceptions of assessment. In C. Rubie-Davies & C. Rawlinson (Eds.), Challenging Thinking about Teaching and Learning (pp. 285-302). New York: Nova Science.

Brown, G. T. L., & Abdulnabi, Hasan. (2017). Evaluating the Quality of Higher Education Instructor-Constructed Multiple-choice Tests: Impact on Student Grades. Frontiers in Education. 2. 10.3389/feduc.2017.00024. model. International Journal of Quantitative Research in Education, 1(3), 297-315.

Brown, G. T. L., Gebril, A., & Michaelides, M. P. (2019) Teachers' Conceptions of Assessment: A Global Phenomenon or a Global Localism. *Front. Educ*. 4:16. doi: 10.3389/feduc.2019.00016.

Brown, G. T. L., & Hattie, J. A. (2012). The benefits of regular standardized assessment in childhood education: Guiding improved instruction and learning. In S. Suggate & E. Reese (Eds.), Contemporary debates in childhood education and development (pp. 287-292). London: Routledge.

Brown, G. T. L., & Remesal, A. (2017). Teachers' conceptions of assessment: Comparing two inventories with Ecuadorian teachers. *Studies in Educational Evaluation*. 55. 68-74. 10.1016/j.stueduc.2017.07.003.

Buchholtz, N. F., Krosanke, N., Ryan B. Orschulik, A. B., & Vorholter. K. (2018). Combining and integrating formative and summative assessment in mathematics teacher education. *ZDM, 50*, 715 – 728.

Campbell, D. T. (1975). "Degrees of Freedom" and the Case Study. *Comparative Political Studies*, *8*(2),178-193. https://doi.org/10.1177/001041407500800204

Campbell, C. (2013). Research on teacher competency in classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 71–84). Thousand Oaks, CA: SAGE Publications, Inc.

Campbell, C., & Evans, J. A. (2000). Investigation of preservice teachers' classroom assessment practices during student teaching. *The Journal of Educational Research, 93*(6), 350-355

Carless, D. (2011). *From Testing to Productive Student Learning: Implementing Formative Assessment in Confucian-Heritage Settings*. London: Routledge.

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*, 395-407. https://doi.org/10.1080/03075071003642449

Carnoy, M., Chisholm, I., & Chilisa, B. (2012). *The Low Achievement Trap: Comparing Schooling in Botswana and South Africa. Southern and East Africa*: Pretoria: HSRC Press. Retrieved from www.hsrcpress.ac.za.

Chapman, D., & Snyder, C. (2000). Can high stakes examinations improve instruction? Re-examining Conventional Wisdom. *International Journal of Educational Development, 20*(6), 457-474.

Chemeli, J. (2019a). Impact of the five key formative assessment strategies on learner's achievement in mathematics instruction in secondary schools: A case of Nandi County, Kenya. *International Academic Journal of Social Sciences and Education, 2*(1), 212-229.

Chemeli, J. (2019b). Teachers support on the effective implementation of the five key formative assessment strategies in mathematics instruction in secondary schools: A case of Nandi County, Kenya. Elixir Soc. Sci. 128 (2019) 52954-52959.

Chee-Cheong, C. (1999). Public examinations in Hong Kong. *Assessment in Education: Principles, Policy & Practice*, *6*(3), 405-417.

Chen, M. H. (2016). Theoretical Framework for Integrating Higher-Order Thinking into L2 Speaking. *Theory and Practice in Language Studies*, *6*(2), 217-226.

Chidozie, C. C., Yusri, K., Muhammad, S., Wilfredo, L.H., (2014). Implementing higher-order thinking skills in teaching and learning of design and technology education. *International Seminar Proceedings on Technical and Vocational*.

Churchill, G. A., & Iacobucci, D. (2006). *Marketing research: methodological foundations*. New York: Dryden Press.

Clarke, S. (2005). *Formative assessment in action: Weaving the elements together*. London, Great Britain: Hodder and Stoughton.

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

Cohen, J. (1992). Statistical Power Analysis. *Journal Indexing and Metrix, 1*(3), 98-101. https://doi.org/10.1111/1467-8721.ep10768783.

Coleman, J. (1966*). Equality of educational opportunity.* Washington, DC: National Center for Educational Statistics.

Combrinck, C. (2018). *The use of Rasch Measurement Theory to address measurement and analysis challenges in social science research*. Unpublished Doctoral thesis. Pretoria: University of Pretoria.

Combrinck, C. (2020). Is this a useful instrument? An introduction to Rasch measurement models. In S. Kramer, S. Laher, A. Fynn, & H. H. Janse van Vuuren (Eds.), *Online Readings in Research Methods*. Psychological Society of South Africa: Johannesburg. https://doi.org/10.17605/OSF.IO/ BNPFS.

Cook, A. (2016). Exploring learners' management of career-related transitions through self-construction (PhD proposal). In K. Maree (Ed.), *first steps in research (2nd Ed*). (pp. 338–377). Pretoria, South Africa: Van Schaik.

Cook, M. (2016). *Personnel Selection: Adding Value Through People - A Changing Picture.* John Wiley & Sons.

Cook, T. D., & Campbell, D. T. (1979). Causal inference and the language of experimentation. In T. D. Cook & D. T. Campbell (Eds.), *Quasi-experimentation: Design & analysis issues for field settings (*pp. 1- 36). Boston, MA: Houghton Mifflin.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*(1), 7. https://doi.org/10.7275/jyj1-4868.

Creswell, J. W. (2003). *Research designs: Qualitative, quantitative, and mixed methods approaches* (2nd Ed). Thousand Oaks, CA: Sage.

Creswell, J. W. (2012). Qualitative inquiry & research design: Choosing among five approaches (4th Ed). Thousand Oaks, CA: Sage.

Creswell, J.W. (2014). *Research design.* (4th Ed). Los Angeles: SAGE Publications Inc.

Creswell, J.W. & Plano Clark, V.L. (2017). *Designing and conducting mixed methods research*, 3rd ed. Thousand Oaks, CA: Sage.

Darling-Hammond, L. (2004). Standards, Accountability, and School Reform. Teachers College Record. *Teachers College Record*, *106*(6), 1047-1085. https://doi.org/10.1111/j.1467-9620.2004.00372.x.

Davin, R., & Naude, M. (2017). *Assessment in the Foundation Phase*. Pretoria, Van Schaik publisher.

De Kleijn, R. A. M., Mainhard, M. T., Meijer, P. C., Brekelmans, M., & Pilot, A. (2013). Master's thesis projects: Student perceptions of *supervisor feedback, Assessment & Evaluation in Higher Education, 38*(8), 1012-1026. https://doi.org/10.1080/02602938.2013.777690.

Deneen, C. C., Brown, G. T. L., Bond, T. G., & Shroff, R. H. (2013). Telling the difference: A first evaluation of an outcome-based learning innovation in teacher education. Asia-Pacific Journal of Teacher Education, 41(4), 441-456. doi:10.1080/1359866X.2013.787392.

Denzin, N. (2012). Triangulation 2.0. *Journal* of mixed methods research, 6(2), 80-88. https://doi.org/10.1177/1558689812437186.

Dibu-Ojerinde, O. O. (2005). Formative assessment for learning: A survey of teachers' practice in selected private secondary schools in Osun State, Nigeria. *International Journal for Learning, 12*(8), 355–360.

Dunn, K., & Mulvenon, S. (2009). A Critical Review of Research on Formative Assessment: The Limited Scientific Evidence of the Impact of Formative Assessment in Education. *Practical Assessment, Research, and Evaluation*, *14*(1), 7. https://doi.org/10.4324/9780203462041_chapter_1.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological science in the public interest*, *5*(3), 69-106.

Earl, L. M., & Katz, S. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for, as and of learning*. Retrieved from: http://www.edu.gov.mb.ca/k12/assess/wncp/full_doc.pdf.

Edmonds, R.R. (1979). Effective Schools for the Urban Poor. *Educational leadership*, *37*(1), 15-24

Education Endowment Foundation (2018) *Embedding Formative Assessment* Retrieved from https://educationendowmentfoundation.org.uk/public/files/EFA_evaluation_report.pdf. [2018, 3 May].

Entwistle, N. J. (2000). Promoting Deep Learning through Teaching and Assessment: Conceptual Frameworks and Educational Contexts. In *Proceedings of the Teaching and Learning Research Program (TLRP) Conference, Leicester, 9-10 November 2000*. http://www.tlrp.org/acadpub/Entwistle2000.pdf. [2018, 6 March]

Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. Review of Educational Research March 2013, Vol. 83, No. 1, pp. 70–120 https://doi.org/10.3102/0034654312474350. http://rer.aera.net.

Fahim, M., & Masouleh, N. S. (2012). Critical thinking in higher education: A pedagogical look. *Theory and Practice in Language Studies*, 2 (7), 1370-1375. Doi: 10.4304/tpls.2.7.1370-1375.

FAST & SCASS. (2008, October). Attributes of effective formative assessment. *Paper prepared for the Formative Assessment for Teachers and Students State Collaborative on Assessment and Student Standards of the Council of Chief State School Officers*. Washington, DC: Council of Chief State School Officers.

Feilzer, M. Y. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*, *4*(1), 6-16.

Fetogang, E. B. (2015). Assessment Syllabus and the Validity of Agricultural Education. *Research Journal of Educational Sciences, 3(*5), 1-9.

Fetogang, E. B. (2016). Teaching to the test: assessment syllabus and validity of English education in Botswana secondary schools. *European Journal of English Language Teaching, 1*(20), 1-16.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics and Sex and Drugs and Rock "N" Roll*, 4th Edition, Sage, Los Angeles, London, New Delhi.

Field, A. (2005) Discovering statistics using IBM SPSS Statistics: (2nd Ed). London, Sage.

Firn, G. (2016). *From Formative Assessment to Informative Assessing in the Math Classroom: preventing the failure to learn*. Retrieved from https://fs24.formsite.com/edweek/images/WP-DreamBox--From_Formative_Assessment.pdf. [2018, 8 May].

Firestone, M. (2013). *Constructivism in psychology: Definition, theories & approaches*. Retrieved from http://study.com/academy/lesson/constructivism-in-psychology-definition-theories-approaches.html. [2018, 28 June].

Flewelling, G., & Higginson, W. (2002). *A Handbook on Rich Learning Tasks: Realising a Vision of Tomorrow's Classroom.* Ontario: Centre for Mathematics, Science and Technology Education, Queen's University.

Foong, P. Y. (2000). Open-ended Problems for Higher-Order Thinking in Mathematics. *Teaching and Learning*, *20*(2), 49-57

Forehand, M. (2010a). Bloom's taxonomy. *Emerging perspectives on learning, teaching, and technology*, *41*(4), 47-56.

Forehand, M. (2010b), Emerging Perspectives on Learning, Teaching, and Technology, Global Text, Michael Orey. (Chapter 3). Retrieved from https://textbookequity.org/Textbooks/Orey_Emergin_Perspectives_Learning.pdf

Foster, G. (2019). Changing New Zealand Education to be 21st Century Ready • Does the Tomorrow's Schools Review give the right direction? *Consultancy Paper 7 January 25th, 2019, Education Futures 21C.*

Foy, P. & Joncas, M. (2003). *Progress in International Reading Literacy Study 2006: School Sampling Manual*. IEA: Statistics Canada.

Fulmer, G., Lee, I. H., & Tan, K. (2015). Multi-level model of contextual factors and teachers' assessment practices: an integrative review of research. *Assessment in Education: Principles, Policy & Practice, 22*, 475 - 494.

Gall M. D. (1984). Synthesis of research on teachers' questioning. Educational *Leadership, 42* (3), 40-47.

Gattullo F. (2000). Formative assessment in ELT primary (elementary) classrooms: an Italian *case study. Language Testing, 17(2), 278-288. https*://doi.org/10.1177/026553220001700210

Gerryts, E. W. (2018). *Enhancing the Employability of Young Adults from Socio-Economically Challenged Contexts*. Unpublished PhD thesis. University of Pretoria.

Ghasempour, Z., Kashefi, H., Bakar, M. N., & Miri, S. A. (2012). Higher Order Thinking via Mathematical Problem Posing Tasks among Engineering Students. *ASEAN Journal of Engineering Education, 1*(1), 41-47.

Gibbs, G., & Simpson, C. (2003). Measuring the response of students to assessment: The assessment experience questionnaire. *Paper presented at the 11th Improving Student Learning Symposium, Hinckley, UK, 2003.*

Gichuhi, C. (2014). *Teachers 'competence in Tests Construction within Blooms Taxonomy for Effective Learning Assessment: A Case Study of Kikuyu District, Kiambu County.* Unpublished Thesis, University Of Nairobi.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction. 20*, 304-315. https://doi.org/10.1016/j.learninstruc.2009.08.007.

Glazer, N. (2014). Formative Plus Summative Assessment in Large Undergraduate Courses: Why Both? *International Journal of Teaching and Learning in Higher Education,* 26(2), 276-286.

Goethals, P. L. (2013). The Pursuit of Higher-Order Thinking in the Mathematics Classroom: A Review, *Center for Faculty Excellence, United States Military Academy, West Point, the* Government of Botswana. (BOT). (1977). *Education for Kagisano: Report of the National Commission on Education*. Gaborone: Government Printer.

Government of Botswana. (BOT). (1994). *The Revised National Policy on Education. Government Policy Paper.* Government Paper No 2 of 1994. Gaborone: Government Printer.

Government of Botswana (BOT). (2010). *Revised National Population Policy.* Gaborone: Government Printer.

Government of Botswana (BOT). (2013). Botswana Qualifications Authority Act. Gaborone: Government Printer.

Government of Botswana (BOT). (2015). *Education & Training Sector Strategic Plan*. (ETSSP 2015-2020). Gaborone, Botswana. Government Printer. Retrieved from https://www.gov.bw/sites/default/files/2020-03/ETSSP%20Final%20Document_3.pdf. [2018, 23 March].

Government of Botswana (BOT) (2016a). *National Credit Qualification Framework.* Gaborone Printer.

Government of Botswana (BOT). (2016b). *Vision 2036 – Achieving Prosperity for All.* Government Printer, Gaborone: Botswana.

Gray, D. E. (2014). *Doing Research in the Real World*. London, SAGE.

Gregory, R. J. (2007). BB\

(5th ed.). Boston, MA: Pearson International.

Guest, G., Namey, E. E., & Mitchell, M. L. (2013). *Collecting Qualitative Data: A Field Manual for Applied Research.* Los Angeles, Sage.

Guskey, T. R. (2015). Does Pre-Assessment Work? *Educational Leadership, 75*(5), 52-57.

Hambleton, R. K. (1993). Translating achievement tests for use in cross-national studies. *European Journal of Psychological Assessment, 9*(1), 57–68.

Hanushek, E. A., & Rivkin, S. G. (2012). The Distribution of Teacher Quality and Implications for Policy. *Annual Review of Economic, 4*(13), 1-57.

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: the role of feedback's perceived usefulness. *Educational Psychology, 34*(3), 269-290, https://doi.org/: 10.1080/01443410.2013.785384.

Harris, L., & Brown, G. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, *36,* https://doi.org/101-111. 10.1016/j.tate.2013.07.008.

Harrison, C. (2005). Teachers developing assessment for learning: Mapping teacher change. *Teacher Development, 9*(2), 255–263.

Hattie, J. (2009). Visible learning: A synthesis of meta-analyses in education. London: Routledge.

Hattie, J. (2013). *Visible Learning for Teachers: Maximizing Impact on Learning*. Milton Park, Abingdon, UK: Routledge.

Hattie, J. A., & Brown, G. T. L. (2004). Cognitive processes in asTTle: The SOLO taxonomy (asTTle Tech. Rep. #43). Retrieved from Wellington, N.Z. :https://easttle.tki.org.nz/content/download/1499/6030/version/1/file/43.+The +SOLO+taxonomy+2004.pdf.

Hattie, J. A., & Purdie, N. (1998). The SOLO model: Addressing fundamental measurement issues. In B. Dart & G. Boulton-Lewis (Eds.), Teaching and learning in higher education (pp. 145-176). Melbourne, Aus.: ACER.

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Herman, J., Osmundson, E., Dai, Y., Ringstaff, C., & Timms, M (2015). Investigating the dynamics of formative assessment: relationships between teacher knowledge, assessment practice and learning, *Assessment in Education: Principles, Policy & Practice*. https://doi.org/10.1080/0969594x.2015.1006521.

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 89*(2), 140–146. Retrieved December 2, 2020, from: http://www.pdkintl .org/kappan/k_v89/k0710her.htm.

Heritage, M. (2010), *Formative Assessment: Making It Happen in the Classroom.* Corwin Press, Thousand Oaks, CA.

Heritage, M., Kim, J., Vendlinski, T. Herman, J. (fall, 2009). From evidence to action: a seamless process in formative assessment? *Education Measurement: Issues and Practice*, *28*(3), 24-31.

Herman, J., Osmundson, E., Dai, Y., Ringstaff, C., & Timms, M. (2015). Investigating the dynamics of formative assessment: Relationships between teacher knowledge, assessment practice and learning. *Assessment in Education: Principles, Policy & Practice, 22*(3), 344– 367. https://doi.org/10.1080/0969594x.2015.1006521.

He, Q., & Wheadon, C. (2013). The effect of sample size on item parameter estimation for the partial credit model. *International Journal of Quantitative Research in Education*, *1*(3), 297-315.

Higgins, S. E. (2007). Learning skills and the development of learning capabilities. Policy and Practice Information and Coordinating Centre. Durham University. Retrieved from https://www.academia.edu/4276605/Learning_skills_and_the_development_of_learning_capabilities. [2019, 20 February].

Hodgen, J. & Marshall, B. (2005). Assessment for learning in English and mathematics: A comparison. *The Curriculum Journal*, *16*(2), 153-176.

Hogden, J., & Wiliam, D. (2006). *Mathematics Inside the Black Box: Assessment for Learning in the Mathematics Classroom.* London, UK: Granada Learning.

Hoover, N. R. (2009). *A descriptive study of teachers' instructional use of student assessment data. Unpublished doctoral dissertation*. Richmond, VA: Virginia Commonwealth University.

Houts, C. R., Edwards, M. C., Wirth, R. J., & Deal, L. S. (2016). A review of empirical research related to the use of small quantitative samples in clinical outcome scale development. *Quality of Life Research, 25*(11), 2685-2691. https://doi.org/doi:10.1007/s11136-016-1364-9.

Huck, S. W. (2014). *Reading statistics and research*. London: UK, Pearson Education Limited.

Hunter, A., & Brewer, J. (2015a). Designing multimethod research. In S. Hesse-Biber & R. B. Johnson (Eds.), *The Oxford Handbook of multimethod and mixed methods research inquiry* (pp. 616–623). Oxford, England: Oxford University Press

Hunter, A., & Brewer, J. (2015b). Conundrums of multimethod research. In S. Hesse-Biber & R. B. Johnson (Eds.), *The Oxford Handbook of multimethod and mixed methods research inquiry* (pp. 185–205). Oxford, England: Oxford University Press.

James, M., & Pedder, D. (2006). Beyond method: assessment and learning practices and values. *The Curriculum Journal, 17*(2), 109-138. 109-138. https://doi.org/10.1080/09585170600792712.

Jones, M., Alexander, J., & Estell, D. (2010). Homophily among peer groups members' perceived self-regulated leaning. *The Journal of Experimental Education, 78*, 378-394. https://doi.org/10.1080/00220970903548020

Jonsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*, *14*(1), 63-76.

Kanjee, A. (2009). Enhancing teacher assessment practices in South African schools: Evaluation of the assessment resource banks. *Education as Change, 13*(1), 73–89.

Kanjee, A. (2017). Assessment for learning in Africa. *Assessment for learning strategies workshop Handouts and school-based project at Ikeleng Primary School, Winterveld from 18th February 2017 to 29th July 2017.*

Kanjee, A., & Meshack, M. (2014). South African teachers' use of national assessment data. South African Journal of Childhood Education. 4. 90-113. 10.4102/sajce. v4i2.206.

Kapambwe, W. M. (2010). The implementation of school based continuous assessment (CA) in Zambia. *Educational Research and Reviews, 5*(3), 99–107.

Karabatsos G (2000) A critique of Rasch residual fit statistics. *Journal of Applied Measurement, 1*(16),152-176.

Kennedy, D. (2007). *Writing and using learning outcomes: a practical guide,* Cork, University College Cork.

Kennedy, K. J., Chan, J. K. S., Fok, P. K., & Yu, W. M. (2008). Forms of assessment and their potential for enhancing learning: Conceptual and cultural issues. *Educational Research for Policy and Practice, 7*(3), 197-207.

Kerlinger, F. N., & Lee, H. B. (Eds.). (2000). *Foundations of Behavioral Research*. Orlando, FL: Harcourt College Publishers.

King, J. D. (2010). *Criterion-referenced assessment literacy of educators. Unpublished doctoral dissertation.* University of Southern Mississippi, Hattiesburg, MS.

King, P. E., Schrodt, P., & Weisel, J. J. (2009). The Instructional Feedback Orientation Scale: Conceptualizing and validating a new measure for assessing perceptions of instructional feedback. *Communication Education, 58*(2), 235-261. https://doi.org/10.1080/03634520802515705

Kiplagat, P. (2016). Rethinking primary school mathematics teaching: A formative Assessment Approach. *Baraton Interdisciplinary Research Journal*, *6,* 32-38.

Kivunja, C. (2015a). Why Students don't Like Assessment and How to Change Their Perceptions in 21st century pedagogies. *Creative Education, 6,* 2117- 2126.

Kivunja, C. (2015b). Exploring the Pedagogical Meaning and Implications of the 4Cs "Super Skills" for the 21st Century through Bruner's 5E Lenses of Knowledge Construction to Improve Pedagogies of the New Learning Paradigm. *Creative Assessment in Education, 6,* 224-239. http://dx.doi.org/10.4236/ce.2015.62021.

Kivunja, C. (2015c). *Teaching, learning and Assessment*: Steps towards Creative Practice. Melbourne: Oxford University Press.

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*(4), 344–348. https://doi.org/10.1016/j.learninstruc.2009.08.005

Koloi-Keaikitse, S. (2012). *Classroom assessment practices: a survey of Botswana primary and secondary school teachers.* (Unpublished Doctoral Dissertation, Ball State University, Muncie, Indiana).

Kostons, D., Tamara, G., & Fred, P. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. Learning and Instruction. *Learning Instructions, 22*. https://doi.org/10.1016/j.learninstruc.2011.08.004.

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods 15*(2):124-44. DOI: 10.1037.

Kuze, M. W., & Shumba, A. (2011). An investigation into formative assessment practices of teachers in selected schools in Fort Beaufort in South Africa. *Journal of Social Sciences, 29(*2), 159–170.

Kyaruzi, F. (2017). Formative assessment practices in mathematics education among secondary schools in Tanzania. (Ph.D. Unpublished thesis), Ludwig Maximilians Universitat, Munich, DE). Retrieved from https://edoc.ub.uni-muenchen.de/25438/7/Kyaruzi_Florence.pdf.

Kyaruzi, F., Strijbos, J. W., Ufer, S., & Brown, G. T. L. (2018). Teacher AFL perceptions and feedback practices in mathematics education among secondary schools in Tanzania. *Studies in Educational Evaluation, 59*, 1–9. https://doi.org/10.1016/j.stueduc.2018.01.004

Kyaruzi, F., Strijbos, J.-W., Ufer, S., & Brown, G. T. L. (2019). Students' formative assessment perceptions, feedback use and mathematics performance in secondary schools in Tanzania. *Assessment in Education-Principles Policy & Practice, 26*(3), 278-302. https://doi.org/10.1080/0969594x.2019.1593103

Kyaruzi, F., Strijbos, J.-W., & Ufer, S. (2020). Impact of a short-term professional development teacher training on students' perceptions and use of errors in mathematics learning. In *Frontiers in Education* (Vol. 5, p. 169). Frontiers https://doi.org/10.3389/feduc.2020.559122.

Leahy. S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership*, *63*(3), 18 -26.

Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & Quantity*, *43*(2), 265-275.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125*(2), 255-275. https://doi.org/10.1037/0033-2909.125.2.255;

Linacre, J. M. (1996) The Rasch Model cannot be "Disproved". *Rasch Measurement Transactions, 10*(3) 512-514.

Linacre, J. M. (1998). Detecting multidimensionality: Which residual data type works best? *Journal of Outcome Measurement*, 2, 266–283.

Linacre, J.. (2002). A User's Guide to Winsteps: Rasch-Model Computer Program. Available at https://www.winsteps.com/winman/copyright.htm

Linacre, J. M. (2002). What do Infit and Outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16,* 878.

Linacre, J. M. (2005). Rasch dichotomous model vs. One-parameter Logistic Model. *Rasch Measurement Transactions, 19,* 1032.

Linacre, J. M. (2006). Data variance explained by measures. *Rasch Measurement Transactions, 20.*

Linacre, J. M. (2007). *A user's guide to WINDTEPS Rasch-model computer programs*. Chicago, Illinois: MESA Press.

Linacre, J. M. (2016). Winsteps® Rasch measurement computer program User's Guide. Beaverton, Oregon: Winsteps.com.

Lizzio, A., & Wilson, K. (2008). Feedback on assessment: students' perceptions of quality and effectiveness. *Assessment and Evaluation in Higher Education, 33*(3), 263-275. https://doi:org/10.1080/02602930701292548.

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in educational research: From theory to practice* (Vol. 28). Hoboken, NJ: John Wiley & Sons.

Longo, L. (2013). *Teaching philosophy: an inquiry into truth, relativism and practice by giuseppe longo*. Unpublished Master's thesis. University of Toronto.

Lonka, K. (2018)*. Phenomenal Learning from Finland.* Edita, Finland: Edita Publishing. International.

Lubisi, R. C., & Murphy, R. J. L. (2002). Assessment in South African schools. *Assessment in Education: Principles, Policy & Practice*, *9*(2), 255-268.

Lysaght, Z., O'Leary, M., & Ludlow, L. (2017). Measuring Teachers' Assessment for Learning (AfL) Classroom Practices in Elementary Schools. *International Journal of Educational Methodology, 3*(2), 103-115. https://doi.org/10.12973/ijem.3.2.103.

MacDonald, C. (2012). Understanding Participatory Action Research: A qualitative Research Methodology Option. *Canadian Journal of Action Research, 13(*2), 34-50.

McManus, S. (2008). Attributes of Effective Formative Assessment-Paper prepared for the Formative Assessment for Teachers and Students (FAST)-State Collaborative on Assessment and Student Standards (SCASS) of the Council of Chief State School Officers (CCSSO). *Washington, DS: The Council of Chief State School Officers (CCSSO)*.

Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholar. *All Ireland Journal of Teaching and Learning in Higher Education (ALSHE-J), 8*(3), 3351-33514.

Majid, U., Ennis J, & Bhola M. (2017). The role of meaning in life in adjustment to a chronic medical condition. *EC Psychology and Psychiatry*, *7*(12), 1023-1030.

Maree, K. (Ed). (2008). *First steps in research*. Pretoria, South Africa: Van Schaik.

Mareka, G. (2015). *Comparative analysis of the predictive validity of Botswana 2012 and 2013 JCE results using 2009 and 2010 PSLE grades in mathematics.* Unpublished Master thesis. University of Botswana.

Masole, T.B., Gabalebatse, M., Guga, T., Pharithi, M., & BEC (2016). *TIMSS 2015 Encyclopaedia Botswana. TIMSS & PIRLS. International study center.* Lynch School of Education, Boston College.

McGrane, J., Hopfenbeck, T. N., Halai, A., Katunzi, N., & Sarungi, V. (2018). Psychometric research in a challenging educational context. *2nd International Conference on Educational Measurement, Evaluation and Assessment [ICEMEA 2018]*. Oxford University Centre for Educational Assessment.

Mchazime, H. (2003). *Integrating primary school curriculum and continuous assessment in Malawi.* Washington, DC: American Institutes for Research.

McMillan, J. H., (2001). Teacher classroom assessment and grading practices decision making. *Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA*, 2001.

McMillan, J. H*., & Schumacher, S. (2006). Research* in education Evidence-based inquiry (6th Ed). Boston, MA Allyn and Bacon.

Melani, B. (2017). Effective Use of Formative Assessment by High School Teachers. *Practical Assessment, Research & Evaluation, (22)*8, 10.

Merriam, S. B. (2009). *Qualitative Research a Guide to Design and Implementation* 2nd Ed, San Francisco, CA 9403-1741: Jossey-Bass.

Mertens, D. (2010). Transformative Mixed Methods Research. *Qualitative Inquiry, 16*. 469-474. https://doi.org/10.1177/1077800410364612.

Ministry of Education. (2002). *Lower primary school syllabus: Standard one to four*. Gaborone: Government Printer.

Ministry of Education and Skills Development. (2005). *Upper primary school syllabus: Standard five to seven.* Gaborone: Government Printer.

Ministry of Education Skills and Development MoESD. (2014): *Study on Declining Results – Investigating Senior Secondary Education.* Gaborone: Government Printer

Ministry of Education and Skills Development. (2016). *Request for quotation for the evaluation of declining results in basic education sector (primary, junior and senior secondary) since 2007 to date- 2015*. Gaborone, Botswana: Government printer.

Ministry of Finance and Development. (MFDP). (2018). *About our Country*. Retrieved from https://www.gov.bw/about-our-country. [2018, 29 August].

Miettinen, R. (2006). Epistemology of transformative materials activity: Dewey's pragmatism and cultural-historical activity theory. *Journal for the Theory of Social Behaviour*, *36* (4), 389 – 408. https://doi.org/10.1111/j.1468-5914.2006.00316.x

Miles, J., Mallinckrodt. B., & Recabarren, D. A. (2016). Reconsidering Focus Groups and Rasch Model Item Response Theory in Instrument Development. *The Counselling Psychologist*, *44*(2), 226-236. https://doi.org/10.1177/0011000016631121.

Miske, S. (2003). *Proud pioneers: Malawian teachers implement continuous assessment in primary school classrooms*. Washington, DC: American Institutes for Research.

Mitana, J. M. V., Muwagga, A. M., and Ssempala, C. (2018). Assessment of higher order thinking skills: A case of Uganda Primary Leaving Examinations. *African Educational Research Journal*, *6*(4), 240-249.

Mohamed, R., & Lebar, O. (2017). Authentic Assessment in Assessing Higher Order Thinking Skills. *International Journal of Academic Research in Business and Social Sciences, 7*(2), *466-476.*

Molina, E., Fatima, S.F., Ho, A., Hurtadod, C. M., Wilichowksi, T., & Pushparatnam, A. (2018) Measuring Teaching Practices at Scale Results from the Development and Validation of the Teach Classroom Observation Tool. *Policy Research Working Paper 8653, World Bank Group.*

Mokotedi, R. T. (2013). Beginning Primary School Teachers' Perspectives on the Role of Subject Specialization in Botswana Colleges of Education: Implications for the Professional Development of those who did not specialise in Languages (English and Setswana). *International Journal of Scientific Research in Education, 6*(1), 88-99.

Monyaku, B., & Mmereki, O. A. (2011). *The SACMEQ III Project in Botswana: A Study of the Conditions of Schooling and Quality of Education*. Retrieved 28 October 2018 from The Southern and Eastern Africa Consortium for Monitoring Educational Quality: http://www.sacmeq.org/sacmeqmembers/botswana/sacmeq-reports.[2018, 17 March].

Motswiri, M. J. (2004) *Supporting chemistry teachers in implementing formative assessment of investigative practical work in Botswana.* Unpublished PhD thesis. University of Twente, Netherlands.

Mosothwane, M. (2014). Mathematics Curriculum Reforms in Botswana: An Historical Perspective International Journal of Innovative *Research & Development, 3*(13), 36-43.

Moyo, S.E., & Nenty, H. J. (2017) Dimensionality analysis of student's performance in 2013 Botswana General Certificate of Education (BGCSE) agriculture examination: Implication for Differential Item Functioning. *International Science Congress*, 5 (1), 0508-2321.

Moyosore, O. A. (2015). The effect of formative assessment on students' achievement in secondary school mathematics. *International Journal of Education and Research, 3*(10), 481-490.

Morgan, D. L. (1996). Focus Groups. **Annual Review of Sociology, 22**, 129-152. https://doi.org/10.1146/annurev.soc.22.1.129.

Morgan, D. L. (2007). Paradigms lost and pragmatism regained methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, *1*(1), 48-76. Morse.

Mullis, I. V. S., Martin, M. O., Foy, P. & Arora, A. (2012). *TIMSS 2011 International results in mathematics. Chestnut Hill, MA: TIMSS & PIRLS*. International Study Center, Lynch School of Education, Boston College. Retrieved 2 January 2019, fromhttp://timssandpirls.bc.edu/timss2011/downloads/T11_IR_Mathematics_FullBook.pdf. [2018, 3 March].

Murphy, P. K. (2003). The philosophy in thee: Tracing philosophical influences in educational psychology. *Educational Psychologist, 38*(3), 137-145. https://doi.org/10.1207/S15326985EP3803_3

Murray, E. C. (2011). *Implementing Higher-Order Thinking In Middle School Mathematics Classrooms*. Doctor of Philosophy, Athens, Georgia.

Mukherji, P., & Albon, D. (2010). *Research Methods in Early Childhood: An Introductory Guide*. London: SAGE Publications Ltd.

Narciss, S. (2008). *Feedback Strategies for Interactive Learning Tasks*. *Learning and Instruction*. Technische Universitaet Dresden, Germany.

Department of Curriculum Management, Malta, Singapore. (June 2014). *Mathematics a revised syllabus for primary schools.* Department of Curriculum Management, Malta, Singapore.

Nenty, H. J. (2008). *Constructing, Administering, Scoring and Interpreting from Educational Instruments*. Educational Foundations, University of Botswana, Gaborone, Botswana.

Nenty, H. J. (2010). Gender-bias and human resources development: Some measurement considerations. *Ilorin Journal of Education, 29*, 13-26.

Nenty, H. J. (2013). *Fundamentals of quantitative research in education*. Educational Foundations, University of Botswana.

Nenty, H. J., Adedoyin, O. O., Odili, J. N. & Major, T. E. (2007). Primary teacher's perceptions of classroom assessment practices as means of providing quality primary/basic education by Botswana and Nigeria. *Educational Research and Review*. *2* (4), 74-81.

Nenty, H. J., & Umoinyang, I. E. (2004). *Principles of test construction*. University of Calabar. Helimo Associates, Calabar.

Nenty, J. N., & Odili, J. N. (2012) Skill development gap and applications of Bloom taxonomy-related assessment techniques among teachers in Delta State of Nigeria. *International Review of Social Sciences and Humanities*. *2*(2), 77- 87.

New Zealand Ministry of Education (1994). *Assessment: Policy to Practice.* Wellington, NZ: Learning Media Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218.

Nieuwenhuis, J. (2016). Introducing qualitative research. In K. Maree (Ed.), *First steps in research* (50–131). Pretoria, South Africa: Van Schaik Publishers.

Nitko, A.J. & Brookhart, S.M. (2007). *Educational Assessment of Students (*5th edition). Upper Saddle River, N.J.: Pearson Merrill Prentice Hall.

Nkate, J. (2008, June 17). Botswana registers poor qualitative achievements in education. *Sunday Standard, June 17, 2008, p. 7*.

Noonan, B., & Duncan, C. R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research & Evaluation*, *10*(17), 1–8.

Nwafor, A. O. (2013). Fraudulent Trading and the Protection of Company Creditors: *The Current Trend in Company Legislation and Judicial Attitude, 42*(4), 297-323.

Organisation for Economic Co-operation and Development (OECD). (2013). PISA 2019 Results. Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm. [2020, 30 May].

Ottevanger, W., Akker, J., & Feiter, L. (2007). *Developing science, mathematics, and ICT education in Sub-Saharan Africa*. Washington DC: The World Bank.

Ornstein, A. C., & Levine, D. U. (2008). *Foundations of Education*, (8th Ed), Boston, MA: Houghton Mifflin.

Pallant, J. (2010). *SPSS Survival Manual. A step by step guide to data analysis using SPSS (4th Ed).* Australia, McGraw-Hill Education.

Panhwar, A. H., Ansari, S., & Shah, A. A. (2017). Post-positivism: An effective paradigm for social and educational research. *International Research Journal of Arts & Humanities (IRJAH)*, *45*(45).

Panizzon, D., & Pegg, J. (2008). Assessment practices: Empowering mathematics and science teachers in rural secondary schools to enhance student learning. *International Journal of Science and Mathematics Education, 6,* 417–436.

Partnership for Assessment of Readiness for College and Careers (PARCC). (2010, June).*The PARCC application for the race to the top comprehensive assessment systems competition.* Retrieved from http://www.fldoe.org/parcc/pdf/apprtcasc.pdf. [2019, 26 January].

Patel, V. L., Yoskowitz, N. A., & Arocha, J.F., (2009). Towards effective evaluation and reform in the medical education: A cognitive and learning sciences perspective. *Advanced Health Science Education, 14*(5): 791 -812.

Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A Framework for Conceptualizing and Evaluating the Validity of Instructionally Relevant Assessments. *Educational Psychologist*, *51*(1), 59-81. https://doi.org/10.1080/00461520.2016.1145550.

Perry, L. (2013). Review of Formative Assessment Use and Training in Africa. International *Journal of School & Educational Psychology, 1*(2), 94-101. Http://www.dx.doi.org/10.1080/21683603.2013.789809. [2019, 20 January].

Pieger, E., Mengelkamp, C., & Bannert, M. (2016). Metacognitive judgments and disfluency - Does disfluency lead to more accurate judgments, better control, and better performance? *Learning and Instruction*. *44,* 31-40. https://doi.org/10.1016/j.learninstruc.2016.01.012.

Pinckok, N., & Bandt, W.C. (2009). *Connecting formative assessment research to practice.* Washington, D.C: Learning point Associates.

Polit, F, D. & Beck T, C. (2018). *Essential of Nursing Research*. (9th Ed). Philadelphia: Lippincott Williams & Wilkins.

Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Popper, K. R. (1966). The open society and its enemies, Volume 2 (Hegel and Marx). New York, NY: Routledge & Kegan Paul.

Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. *assessment and evaluation in higher education, 33*(2), 143-154. https://doi.org/10.1080/02602930601127869.

Preece, P. F. W., & Skinner, N. C. (1999) The gender gap and discriminating power in the National Curriculum Key Stage three science assessments in England and Wales. *International Journal of Science Education, 21*(9), 979-987, https://doi.org/10.1080/095006999290264

Pryor, J., & Akwesi, C. (1998). Assessment in Ghana and England: Putting reform to the test of practice. Compare: *A Journal of Comparative Education, 28(*3), 263–276.

Race, P. (2009). *Designing assessment to improve physical sciences learning*. Hull, UK: Higher Education Academy.

Raboijane, B. M. (2005). *Mathematics Teachers' Understanding of Alternative Assessment as Applied in Junior Secondary Schools in Gaborone (Botswana).* Unpublished thesis. School of Education of the University of the Witwatersrand, Johannesburg, South Africa.

Rahman, Md., M. (2018). Exploring Science Teachers' Perception of Classroom Assessment in Secondary Schools of Bangladesh. *European Journal of Education Studies, 4*, (9), 139-160, 2018. Available at SSRN: https://ssrn.com/abstract=3299956. [2018, 6 June].

Rajendran, N. S. (2008). *Teaching & Acquiring Higher Order Thinking: Theory and Practice*. Tanjong Malim: Universiti Pendidikan Sultan Idris.

Rakoczy, K., Klieme, E., Leiß, D., & Blum, W. (2017). Formative assessment in mathematics instruction: Theoretical considerations and empirical results of the Co2CA project. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.). *Competence assessment in education: Research, models, and instruments* (pp. 447–467). Berlin: Springer.

Rakoczy, K., Harks, B., Klieme, E., Blum, W., & Hochweber, J. (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning and Instruction, 27,* 63-73.

Ramos, J. L. S., Dolipas, B. B., & Villamore, B. B. (2013). Higher Order Thinking Skills and Academic Performance in Physics of College Students. A Regression Analysis. *International Journal of Innovative Interdisciplinary Research, 4*, 48-60.

Ramsey, B., & Duffy, A. (2016). Formative assessment in the classroom: Findings from three districts. *Michael and Susan Dell Foundation and Education*.

Raosoft. (2004). http://www.raosoft.com/samplesize.html.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago, MI, The University of Chicago Press

Rasch, G. (1979). Letter from George Rasch to Ben Wright – personal communication, 19 October. Retrieved from http://www.raschmeasurement

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (Expanded, Ed.). Chicago: University of Chicago Press.

Reckase, M. D. (1979). *Development and application of a multivariate logistic latent trait model.* Syracuse University, Unpublished doctoral dissertation.

Romm, N. R. A. (2020). Justifying Research as Conscious Intervention in Social and Educational Life: Activating Transformative Potential. *Educational Research for Social Change,* *9*(2), 1-15. http://dx.doi.org/10.17159/2221-4070/2020/v9i2a1

Rubie-Davies, C. M. (2006), 'Teacher expectations and student self-perceptions: Exploring relationships. *Psychology in the Schools 43*(5), 537–52.

Rubie-Davies, C. M. (2007), Classroom interactions: Exploring the practices of high and low-expectation teachers. *British Journal of Educational Psychology*,77(2), 289–306.

Rubie-Davies, C. M., Hattie, J., & Hamilton, R. (2006). Expecting the best for students: Teacher expectations and academic outcomes. *The British journal of educational psychology*, *76*, 429-44. *https://doi.org/*10.1348/000709905X53589

Runnels, J. (2012). Using the Rasch model to validate a multiple-choice English achievement test. *International Journal of Language Studies*, RRR*6*, 141-153.

Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, *37*(1), 15–24. https://doi.org/doi:10.1016/j.stueduc.2011.04.003.

Ryan, F., Coughlan, M., & Cronin, P. (2009). Interviewing in qualitative research. *International Journal of Therapy and Rehabilitation. 16*, 309-314. https://doi.org/10.12968/ijtr.2009.16.6.42433.

Rycroft-Smith, L. (2017). *Realising potential in Mathematics for all.* London, The mathematics Association.

Sebatane, E. M., Chabane, C. M., & Lefoka, J. P. (1992). *Teaching and learning strategies in Lesotho: An empirical perspective on primary school classrooms*. Ottawa: International Development Research Centre.

Sadiq, M., Tirmizi., S. H., & Jamil. M. (2015). Using Rasch Model for the Calibration of Test Items in Mathematics, Grade-9. *Journal of Research and Reflections in Education*, 9(2).

Saunders, M., & Thornhill, A. (2012). Research methods for Business students. 6th edition.

SCASS. (2017). *Revising the Definition of Formative Assessment*. Retrieved from https://www.michigan.gov/documents/mde/New_FAST_SCASS_definition_paper_June_27_2017_601106_7.pdf. [2018, 17 March].

Sengul, C., & Ustundag, T. (2010). Physics teachers' levels of critical thinking attitudes and the place of critical thinking in the organised activities. *Hacettepe University Faculty of Education Journal*, *36.*

Serin, O. (2013). The critical thinking skills of teacher candidate. Turkish Republic of Northern Cyprus sampling. *Egitim Arastirmalari-Eurasian Journal of Educational Research, 53*, 231-248.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Shah, M. W. (2017). *Exploring feedback practices of primary teachers on students' homework in a public school in Gilgit Baltistan* (Unpublished master's dissertation). Aga Khan University, Karachi, Pakistan).

Sharif, M., Hanapi, Z., Nashir, M., Kob, C., & Abdullah, A. S. (2019). Construct Validity and Reliability of Manufacturing Industry: A Rasch Measurement Model Approaches for Pilot Study. I*nternational Journal of Academic Research in Business and Social Sciences*, *9*(3), 654-66.

Shavelson, R. J. (2008). Guest editor's introduction [Editorial]. *Applied Measurement in Education, 21*(4), 293–294. https://doi.org/10.1080/08957340802347613

Shepard, L. A. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher, 29*(7), 4-14. https://doi.org/10.3102/0013189X029007004.

Shepard, L. A. (2006). Classroom assessment. In *Educational measurement*, 4th ed., ed. Brennan, Robert L., 623–46. Westport, CT: Praeger.

Shepard, Lorrie A. (2008). Formative assessment: Caveat emptor. In *The future of assessment: Shaping teaching and learning*, ed. Dwyer, Carol Anne, 279–303. New York, NY: Lawrence Earlbaum

Shen, C. (2002) Revisiting the relationship between students' achievement and their self-perceptions: a cross-national analysis based on TIMSS 1999 data, *Assessment in Education, 9*(2), 161–184.

Sinharay, S., Haberman, S., Holland, P., & Lewis, C. (2006). A Note on the choice of an Anchor Test in Equating. *ETS Research Report Series. 2012.* https://doi.org/10.1002/j.2333-8504.2012.tb02296.x.

Smith, R. M., Schumacher, R. E., & Bush, M. J. (1998). Using item mean squares to evaluate fit to the Rasch Model. *Journal of Outcome Measurement, 2*, 66-78.

Smith, C., Stone, R., & Comings, J. (2012, March). *Literacy policy and practice in Ethiopia: Building on the TELL program and EGRA results.* Washington, DC: American Institutes for Research.

Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., and Jobse, H. (2016). The Impact of Education Programmes on Learning and School Participation in Low and Middle-Income Countries: A Systematic Review Summary Report. *Systematic Review Summary 7, International Initiative for Impact Evaluation, London.*

Spiller, D. (2015). *Principles of assessment for learning*. University of Waikato, Teaching Development Unit, Wahanga Whakapakari Ako. Retrieved from www.waikato.ac.nz/tdu/pdf/Booklets/2015/AssessmentPrinciples.pdf. [2019, 28 January].

Statistics Botswana. (2012). *Primary Education Statistics 2012*. Gaborone: Statistics Botswana.

Statistics Botswana. (2015). *Pre and Primary School Education.* Stats Brief 2015: Gaborone Statistics Botswana.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, *83*(10), 758–765.

Stiggins, R. J. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, *87*(4), 324–328.

Stiggins, R. J. (2006). What a Difference a Word Makes: Assessment "for" Learning Rather than Assessment "of" Learning Helps Students Succeed. *Journal of Staff Development*, *27*(1), 10-14.

Strijbos, J.W, Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*(4), 291-303. https://doi.org/10.1016/j.learninstruc.2009.08.008.

Strijbos, J. W., Pat-El, R. J., & Narciss, S. (2010, May). Validation of a (peer) feedback perceptions questionnaire. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell, & T. Ryberg (Eds.), *Proceedings of the 7th International Conference on Networked Learning* (pp. 378-386) [Electronic Proceedings, ISBN 978-1-86220-225-2]. Aalborg, Denmark: Aalborg University.

Supramani, S. (2006). Penyoalan Guru: Pemangkin Pemikiran Aras Tinggi Murid. *Jurnal Pendidikan*. 225-246.

Suurtamm, C. (2004). Developing authentic assessment: Case studies of secondary school mathematics teachers' experiences. *Canadian Journal of Science,* *Mathematics and Technology Education, 4*(4), 497–513. Https://doi.org/10.1080/14926150409556630.

Swaffield. S. (2011). Getting to the heart of authentic Assessment for Learning, Assessment in Education: *Principles, Policy & Practice, 18*,(4), 433-449. https://doi.org/10.1080/0969594X.2011.582838

Tabachnick, B., & Fidell, L. S. (2007). *Using Multivariate Statistics*. (Vol. 5, pp. 481-498). Boston, MA: Pearson.

Tabulawa, R. T. (2014, February 14). Generation fail. *The Botswana Gazette, 14 February 29, 2019, p. 3.*

Tajudin N, M. (2015). Mathematical knowledge and higher order thinking skills for teaching algebraic problem solving. *Proceedings of SOCIOINT15- 2nd International Conference on Education, Social Sciences and Humanities, Istanbul Turkey.*

Taras, M. (2005). Assessment—summative and formative—some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466–478.

Teddlie, C., & Tashakkori, A. (2009). *The foundations of mixed methods research: Integrating quantitative and qualitative techniques in the social and behavioural sciences.* Thousand Oaks, CA: Sage.

Tengku, Z. T. Z., Ramlee, M., & Abdul R., H. (2009). Pengetahuan Pedagogi Isi Kandungan Guru Matematik bagi Tajuk Pecahan. *Kajian Kes di Sekolah Rendah. Pendidikan Malaysia, 34*(1), 131-153.

Thompson, R. A. (2008). *Early attachment and later development: Familiar questions, new answers.* In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (p. 348–365). The Guilford Press.

Thomas, A. & Thorne, G. (2009). *How to Increase Higher Order Thinking, Methane, LA. Center for Development and Learning.* Available from http://www.readingrockets.org/article/34655.

Thompson, T. (2008). Mathematics Teachers' Interpretation of Higher-Order Thinking in Bloom's Taxonomy. *International Electronic Journal of Mathematics Education. 3*(2), 96-109.

Thompson, M., & Wiliam, D. (2008). Tight but loose: a conceptual framework for scaling up school reforms. In E. C. Wylie (Ed.), *Tight but loose: scaling-up teacher professional development in diverse contexts* (RR-08-29, pp. 1-44). Princeton, NJ: Educational Testing Service.

Tobey, C. R. & Goldsmith, L. (2013). *Developing self-regulated learner: The critical role of feedbac*k. ASCD express.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroo*m. Buckingham, UK: Open University Press.

Torrance, H., & Pryor, J. (2001). Developing Formative Assessment in the Classroom: Using action research to explore and modify the theory. *British Educational Research Journal*, *27*(5), 615-631.10.1080/01411920120095780.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2010). *World Data on Education.* Retrieved from http://www.ibe.unesco.org/fileadmin/user_upload/Publications/WDE/2010/pdf -versions/Botswana.pdf?&session-id=9bef60f240054b618f9f4af6e3d00ddf. [2018, 28 October].

UNICEF. (2019). *Public Expenditure Review of Basic Education Sector in Botswana.* Gaborone: UNICEF.

Van den Berg, M., Bosker, R. J., & Suhre, C. J. M. (2018). Testing the effectiveness of classroom formative assessment in Dutch primary mathematics education. *School Effectiveness and School Improvement, 29* (3), 339-361. Https://doi.org/ 10.1080/09243453.2017.1406376.

Van de Pol, J., Volman, M., Oort, F., & Beishuizen, J. (2014). Teacher scaffolding in small group work: An intervention study. Journal of the Learning Sciences, 23(4), 600-650.

Van Staden S., & Zimmerman L. (2017). Evidence from the Progress in International Reading Literacy Study (PIRLS) and How Teachers and Their Practice Can Benefit. In: Scherman V., Bosker R.J., Howie S.J. (Ed) *Monitoring the Quality of Education in Schools.* Sense Publishers, Rotterdam.

Van Staden, S. (2010). *Reading between the Lines: Contributing Factors that Affect Grade 5 Learner Reading Performance.* Unpublished PhD thesis. University of Pretoria.

Venson- Moitoi, P. (2014, May 15). Mathematics, science vital to development. *Botswana Daily News, 15 May 2019. p. 2.*

Venkatesh, V., Brown, S. A., & Sullivan, Y.W. (2016). Guidelines for Conducting Mixed-methods Research: An Extension and Illustration. *Journal of the Association for Information Systems. 17*, 435-495. https://doi.org/10.17705/1jais.00433.

Venkatesh, V., Brown, S, A., & Sullivan, Y. W. (2014). Guidelines for Conducting Mixed-methods Research: An Extension and Illustration. *Journal of the Associations for Information Systems. 7*(17), 436 – 494.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children, 23*(3), 34-41.

Wade, S. E., & Moje, E. B. (2000). The role of text in classroom learning. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 609- 627). Mahwah, NJ: Erlbaum.

Waismann, F. (2011). *Causality and logical positivism. In Humanities, Social Science and Law.* Berlin: Springer eBooks.

Wang, W. C. (2008). Assessment of differential item functioning. *Journal of Applied Measurement, 9(*4), 387-408.

Webb, M., & Jones, J. (2009) Exploring tensions in developing assessment for learning, *Assessment in Education: Principles, Policy & Practice, 16*,2, 165-184. https://doi.org/10.1080/09695940903075925.

Weinstein, R.S. (2002), Overcoming Inequality in Schooling: A Call to Action for Community Psychology. *American Journal of Community Psychology*, *30*(1), 21-42. . https://doi.org/10.1023/A:1014311816571.

Welsh Assembly Government (2010). Howe to develop thinking and assessment for learning in the classroom. Retrieved from https://www.gwegogledd.cymru/wp-content/uploads/2018/04/How_to_develop_thinking_and.pdf. [2018, 13 May].

Weurlander, M., Söderberg, M., Scheja, M., Hult, H., & Wernerson, A. (2012). Exploring formative assessment as a tool for learning: Students' experiences of different methods of formative assessment. *Assessment & Evaluation in Higher Education*, *37*(6), 747-760.

Wiliam, D. (2006). Formative Assessment: Getting the Focus Right'. *Educational Assessment*, *11*, 283-289.

Wiliam, D. (2007). Keeping learning on track: Formative assessment and the regulation of learning. Making Mathematics Vital: Proceedings of the Twentieth Biennial Conference of the Australian of the Mathematics teachers.

Wiliam, D. (2010a). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (eds.), *Handbook of formative assessment* (pp. 18–40). The UK. Abingdon: Routledge.

Wiliam, D. (2010b). The role of formative assessment in effective learning environments. *The nature of learning: Using research to inspire practice*, 135-155. https://doi.org/10.1787/9789264086487-8-en.

Wiliam, D. (2011*). Embedded formative assessment: Practical strategies and tools for K–12 teachers.* Bloomington, IN Solution Tree.

Wiliam, D. (2013). Assessment: The bridge between teaching and learning. *Voices from the Middle*, *21*(2), 15.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11*(1),. 49-65. https://doi.org/10.1080/0969594042000208994.

Wiliam, D., & Thompson, M. (2008). Tight but loose: a conceptual framework for scaling up school reforms. In E. C. Wylie (Ed.), *Tight but loose: scaling up teacher professional development in diverse contexts* (Vol. RR-08-29, pp. 1-44). Princeton, NJ: Educational Testing Service.

Wiliam, D., & Thompson, M. (2014). Integrating assessment with instruction: what will it takes to make it work? In C. A. Dwyer (Ed.). *The future of assessment: shaping teaching and learning* (pp. 53-82). New York, N.Y: Routledge.

Wilkins, J. L. (2008). The relationship among elementary teachers' content knowledge, attitudes, beliefs, and practices. *Journal of Mathematics Teacher Education, 11*(2), 139–164.

Willis, J. (2012). Foundational Issues: Postpositivist and Critical Perspectives. *Foundations of Qualitative Research: Interpretive and Critical Approaches, 67–94.* https://doi.org/10.4135/9781452230108.n3*.*

Wood, D. & Wood, H. (1988). *Questioning Versus Student Initiative, in J.T. Dillon (Ed.) Questioning and Discussion: a multidisciplinary study*. Norwood: Ablex.

Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice,* *22*(1), 140–160. Https://doi.org/10.1080/0969594X.2014.99041

Yang, Y. T. & J. Gamble. (2013). Effective and practical critical thinking-enhanced EFL instruction. *English Language Teaching Journal, 67*(4), 398-412. Doi: 10.1093/elt/cct038.

Yee, M. H., Jailani, M. Y., Widad, O., Razali, H. & Tee, T. K. (2011). The Effectiveness of Higher Order Thinking Skills for Generating Idea among Technical Students. *Recent Advances in Educational Technologies*.

Yee, H., Othman, W., Yunos, J., Tee, T., Hassan, R., & Mohaffyza, M. (2011). The Level of Marzano Higher Order Thinking Skills among Technical Education Students. *International Journal of Social Science and Humanity*, *1*(2), 121. . https://doi.org/10.7763/IJSSH.2011.V1.20.

Young, M, E., & Ryan, A. (2020). Postpositivism in Health Professions Education Scholarship. *Academic Medicine*, *95*(5), 695-699.

Yudkowsky, R., Park, Y. S., & M. Downing. S. M. (2020). *Assessment in Health Professions Education*. 2nd(ed). New York, Routledge.

Zhang, O., Shen, L., CRyandy, M. (2010, April). Polytomous IRT or testlet model: An evaluation of scoring models in small testlet size situations. *Paper presented at the Annual Meeting of the 15th International Objective Measurement Workshop, Boulder, CO*.

9 APPENDICES

9.1 *Appendix A*. Chapter 4 (Phase I) - Teacher questionnaire

Baseline Survey Questionnaire

Technical stakeholders in education have raised concerns about the little attention paid to the use of formative assessment in monitoring teaching and learning at standard four (4) primary school level. The sole purpose of this questionnaire is to examine the extent to which Botswana standard 4 mathematics teachers' implement (use) the formative assessment strategies in the classroom assessment practices to enhance pupils' higher-order thinking skills. The results of the study will advance our understanding of teachers' classroom assessment needs. All responses are anonymous and your participation in this study is completely voluntary, there are no benefits or disadvantages from your participation, and you are free to discontinue at any time. However, your participation is very much appreciated and will assist in the education processes for Botswana.

Informed Consent for Participation

Dear Colleague/ Participant

I consent to my participation in the research being conducted by Sello Editor Moyo, PhD Student from the University of Pretoria, on the title research: *Botswana teachers' Experiences in Formative assessment in Standards 4 Mathematics. T*he researchers have explained the purpose of the study, the procedures that will be followed, and the amount of time it will take. I understand the possible benefit of my participation. I know that I can choose not to participate in this study without incurring a penalty. I am aware that if agree to participate, I am free to withdraw from the study at any time, my identity will not be revealed and there will be no penalty. I am also aware that before I can participate in the study, the researcher must obtain approval from the University of Pretoria Research Ethics.

Date:________________ Signed:_________________________

Section A: Demographic Information and Teaching Experience

Directions: **Please place a tick where applicable.**

1. **Gender**
   Male _______ Female ______

2. **Your educational background**

[ ] Certificate

[ ] Diploma

[ ] Bsc/BA Degree

[ ] Masters' Degree

[ ] Other please specify_________________

3. Your teaching experience _____ in years
4. What standard do you teach currently? _______________
5. Do you teach mathematics as a core subject? Yes [ ] No [ ]
6. Which of the following best describes your training in assessing pupils learning (choose all that apply)?

[ ] I received no training in classroom assessment,

[ ] I received no training in mathematics pedagogy or instruction

[ ] I took a course dedicated to classroom assessment

[ ] I took a course dedicated to mathematics pedagogy or instruction

[ ] I took more than one course dedicated to classroom assessment

[ ] I took more than one course dedicated to mathematics pedagogy or instruction

[ ] I received in-service/workshop training in classroom assessment

[ ] I received in-service/workshop training in mathematics pedagogy or instruction

7. I believe the average achievement of pupils in my Standard 4 mathematics class

Far above average	Above average	Average	Below average	Far below average
[ ]	[ ]	[ ]	[ ]	[ ]

Section B: Professional Practice in Classroom Assessment

Use of Formative Assessment: Part A – Scale 0 (Not at all), 1 (once a week), 2 (2-3 times per week), 3 (4 times a week), 4 (always)

1. In my class . . .						
		Not at all	**Once per week**	**2-3 times per week**	**4 times a week**	**Every day of the week**
a.	I provide my pupils with learning objectives.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
b.	I discuss the meaning of learning objectives with my pupils.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
c.	I connect each lesson to the previous lesson or learning that has taken place	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
d.	I connect each lesson to future learning that will take place	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
e.	I design coherent sequences of learning rather than individual lessons.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
f.	I explain to pupils the connections between new, prior and future learning.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
g.	I have a learning goal(s) for the lesson.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
h.	I present the learning goal(s) for the lesson to pupils verbally.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
i.	I present the learning goal(s) for the lesson to pupils in writing (e.g on the board).	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week

j.	I discuss with pupils what are they should know by the end of the lesson	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
k.	The learning goal(s) for the lesson is connected to country/local academic standards.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
l.	I reference the learning goal(s) multiple times within the lesson.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
m.	I share with pupils the criteria that will be used to determine their success in the lesson	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
n.	I have pupils participate in developing the criteria for success.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week

o.	I provide pupils with multiple options to demonstrate their learning.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
p.	I have pupils demonstrate an understanding of the criteria for success.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week

Use of Formative Assessment: Part B – Scale 0 (Not at all), 1 (once a week), 2 (2-3 times per week), 3 (4 times a week), 4 (always)

2. In my class . . .

		Not at all	**Once per week**	**2-3 times**	**4 times**	**Every day**

				per week	a week	of the week
a.	I ask questions within the lesson to assess the whole group progress.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
b.	I ask questions within the lesson to assess individual pupils' progress.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
c.	I make adjustments to instructions within the lesson based upon pupils' responses.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
d.	I ensure the pace of the lesson provides adequate wait time for pupils to respond to questions.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
e.	I use follow-up questions when engaging pupils in the discussion.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
f.	I use exit tickets to assess pupils learning.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
g.	I use pupils' responses to questions to help me adapt to future instruction.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week

Use of Formative Assessment: Part C – Scale 0 (Not at all), 1 (once a week), 2 (2-3 times per week), 3 (4 times a week), 4 (always)

3. In my class . . .

	Not at all	**Once per week**	**2-3 times per week**	**4 times a week**	**Every day of the week**

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

a.	I have pupils work in a small group with 2-3 other pupils.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
b.	I have pupils to work with a partner.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
c.	I allow pupils to guide their learning.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
d.	I enable pupils to engage in discussion regarding the lesson in small groups.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
e.	I facilitate pupils in discussion regarding the lesson as a whole class.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
f.	I include both individual and group assignments when group work is used.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
g.	I enable pupils to learn from each other when they engage in group work.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
h.	I expect pupils to find the right answer to a teacher provided a problem.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
I.	I encourage pupils to consider multiple viewpoints or approaches in the problem-solving task.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
j.	I have high expectations for all pupils to succeed.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Use of Formative Assessment: Part D, Scale 0 (Not at all), 1 (once a week), 2 (2-3 times per week), 3 (4 times a week), 4 (always)

4. In my class . . .

		Not at all	**Once per week**	**2-3 times per week**	**4 times a week**	**Every day of the week**
a.	The tasks and activities within the lesson are directly tied to the learning goal(s).	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
b.	The tasks and activities within the lesson provide evidence of pupils' progress towards learning goal(s).	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
c.	More than fifty per cent of pupils are clear about the task and begin work efficiently.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
d.	All pupils understand the directions for the lesson.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
e.	Pupils' responses provide me with evidence for adapting instruction within the lesson.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
f.	I analyse pupils' responses and work to identify patterns of understanding/misunderstanding within the lesson.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week

Use of Formative Assessment: Part E– Scale 0 (Not at all), 1 (once a week), 2 (2-3 times per week), 3 (4 times a week), 4 (always)

5. In my class . . .						
		Not at all	**Once per week**	**2-3 times per week**	**4 times a week**	**Every day of the week**
a.	I review all pupils work during the lesson.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
b.	I review some pupils work during the lesson.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
c.	I provide real-time feedback on pupils works to all pupils.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
d.	I provide pupils with opportunities to internalise feedback and apply it in a meaningful way.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
e.	I use the pupils' self-assessment.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
f.	I use the pupils' peer-assessment.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
g.	I use evidence generated through pupils' self-assessment to inform future teaching and learning.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
h.	I use evidence generated through pupils' peer-assessment to inform future teaching and learning.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week
I.	I generate feedback loops during classroom discussion where one question leads to the elaboration and further questioning to build the discussion.	Not at all	Once per week	2-3 times per week	4 times a week	Every day of the week

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Use of Formative Assessment in the teaching of higher-order thinking skills:

Part F – Scale 1(strongly disagree), 2(disagree), 3 (agree), 4 (Strongly agree)

6. In my mathematics class…

		Strongly disagree	disagree	Agree	Strongly Agree
a.	I understand what formative assessment is and how to use it for higher-order thinking tasks.	Strongly disagree	disagree	Agree	Strongly Agree
b.	I have enough time to plan a formative assessment that includes higher-order thinking tasks.	Strongly disagree	disagree	Agree	Strongly Agree
c.	The curriculum I use includes integration formative assessment and teaching of higher-order thinking skills related task.	Strongly disagree	disagree	Agree	Strongly Agree
d.	The curriculum I use support formative assessment and individualized instruction at a range of grade levels.	Strongly disagree	disagree	Agree	Strongly Agree
e.	My approach to instruction provides me with ample opportunities to interact with all of my pupils and act on formative assessment data.	Strongly disagree	disagree	Agree	Strongly Agree
f.	My class periods provide enough time to gather and act on the formative	Strongly disagree	disagree	Agree	Strongly Agree

	assessment for higher-order thinking skills				
g.	I have administrator support to incorporate formative assessment into my teaching of higher-order thinking skills	Strongly disagree	disagree	Agree	Strongly Agree
h.	My school provides me with training in setting test items for higher-order thinking skills	Strongly disagree	disagree	Agree	Strongly Agree
i.	My school provides me with adequate training on formative assessment	Strongly disagree	disagree	Agree	Strongly Agree
j.	My school provide instructional materials for teaching higher-order thinking skills to support remedial teaching.	Strongly disagree	disagree	Agree	Strongly Agree
k.	I assess pupils' problem-solving skills through the use of tasks in the classroom.	Strongly disagree	disagree	Agree	Strongly Agree
l.	I use different levels of questioning which help pupils to think and reason.	Strongly disagree	disagree	Agree	Strongly Agree
m.	I provide support for pupils to work independently on mathematics problem-solving tasks.	Strongly disagree	disagree	Agree	Strongly Agree
n.	My approach to instruction provides me with opportunities to encourage pupils' to be critical thinker.	Strongly disagree	disagree	Agree	Strongly Agree
o.	I know how to use data to diagnose underlying learning gaps and identify	Strongly disagree	disagree	Agree	Strongly Agree

	lessons and instructional strategies appropriate to help pupils catch up.				

[The End]

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

9.2 *Appendix B*. Chapter 4 (Phase I) - Observation Schedule (Pre and post)

FA2019: Teacher assessment practices

Observation Schedule

The purpose of the observation is to record teachers' use of assessment during lessons.

Date:	Sub-region:	District:
Standard:	LoLT:	Subject/Phase:
Lesson Topic:		
Duration of the lesson:	Start time:	End time:
Name of School:		
Name of Teacher:		
Name of Observer:		

i. Ask for a copy of the **lesson plan** that the teacher plans to use.
ii. Ask for **5 exercise books of low-**performing pupils and **5 high-**performing pupils from the 2019 class. **Select** one book for low-performing pupils and one book for a high-performing learner to complete Section 13.

Attach a copy# of the lesson plan / a formative Assessment Preparation Schedule to this document ***(# Take a picture of it; ideally not reflecting names)***

Copy of lesson plan attached#	Yes	No

Description of the classroom context.

Description	Yes, No or N/A & Relevant Comments
iii. Number of pupils	
iv. Classroom walls have relevant (phase, grade and subject appropriate) wall charts, pictures, etc. *(If necessary, estimate any relevant percentages out of the total on display to indicate the extent of availability.)*	
v. The classroom has a data projector	
vi. The classroom has a SmartBoard	
vii. Any other information	

BEGINNING OF LESSON

1. Record how the teacher started the lesson. Observe whether the teacher linked it to previous lessons or the pupil's previous experience. Did the teacher start with mental maths or daily reading activity (*only applicable to Foundation Phase / Standard 3 lessons*)? How did learners respond?			
	Developing	effective	Exemplary

2. Does the teacher introduce the lesson objective	Yes	No	3. Does the teacher use Assessment Criteria in the introduction	Yes	No

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

If YES, **rate it** and please write the Learning objective using the EXACT words of the teacher				If YES, **rate it** and please write the Success Criteria using the EXACT words of the teacher			
	Developing	effective	Exemplary		developing	effective	Exemplary

Please indicate if you found any evidence for the activities listed below, and mark: either *'Seen', or 'Not Seen', or as otherwise relevant.*

4. Formative Assessment					
a) The teacher uses words such as **We Are Learning To (WALT)** when introducing the Lesson objective (LO)				Seen	Not Seen
b) The LO is:	Presented orally	Written on the board	Written on chart	Provided in a hand-out	Other:
c) The teacher uses words such as **What I'm Looking For (WILF)** when introducing the Assessment Criteria (AC)				Seen	Not Seen
d) The AC are:	Presented orally	Written on the board	Written on chart	Provided in a hand-out	Other:

DURING THE LESSON

What materials does the teacher use? How does the teacher encourage participation? What kind of questions did the teacher ask? Did the teacher invite answers from the pupils'? Specific evidence of pupil-pupil interaction in the classroom? Are pupils given a chance for discussions?			
5. (Questioning and interaction)	**Often**	**Some times**	**Not Seen**
a) When the teacher asks questions, pupils put their hands up.	2	1	0
b) The teacher only asks pupils that have their hands up.	2	1	0
c) The teacher involves more than one pupil to answer a single question.	2	1	0
d) The teacher asks questions for the "whole" class to respond.	2	1	0
e) The teacher waits a few seconds before getting a response from a learner.	2	1	0
f) The teacher answers her/his questions.	2	1	0
g) The teacher uses name/number sticks to select pupils.	2	1	0
h) Pupils use mini-boards during the lesson.	2	1	0
i) Pupils work in groups to guide each other on their learning	2	1	0

j) Pupils work in groups cooperatively while completing a group task	2	1	0
k) The teacher conveys an attitude of "we all can"	2	1	0
l) The teacher provides appropriate support and encouragement to pupils	2	1	0

8. Learning HOTS tasks (Implemented)	**Often**	**Some times**	**Not Seen**
a) The teacher uses well-crafted HOTS tasks that are aligned with the learning goal.	2	1	0
b) All pupils are clear about the HOTS task and can begin work efficiently.	2	1	0
c) The teacher frequently uses pupils' responses and work to make inferences about progress and adapts instruction accordingly.	2	1	0
d) The teacher skillfully uses multiple ways of gathering evidence throughout the lesson that are connected to the learning.	2	1	0
e) The teacher uses multiple approaches to handle problem-solving tasks	2	1	0
f) The teacher gives homework on problem-solving tasks	2	1	0
g) The teacher engages pupils on the previous tasks given as a homework	2	1	0

9. Oral Feedback	Yes	No
a) After giving pupils classwork, the teacher walks around to check how pupils are doing.	Seen	Not Seen
b) The teacher provides orally evaluative feedback on a specific piece of work	Seen	Not Seen
c) When checking pupils' work, the teacher gives guidance or makes comments	Seen	Not Seen
d) The teacher provides orally descriptive feedback on a specific piece of work	Seen	Not Seen

10. Does the teacher use **Peer assessment** during the lesson? If yes, complete Question 10 (a-d), else go to Question 11	Yes	No
a) Pupils are allowed to check their partner's work.	Seen	Not Seen
b) The teacher reminds pupils how they should use peer assessment.	Seen	Not Seen
c) The teacher visits a few pupils to check how they conduct peer assessment.	Seen	Not Seen
d) The teacher gives feedback on how the peer assessments were conducted.	Seen	Not Seen

11. Does the teacher use **Self-assessment** during the lesson? If yes, complete Question 11 (a-e)	Yes	No
a) Pupils are allowed to check their work.	Seen	Not Seen
b) The teacher reminds pupils how to use self-assessment, e.g. process and rules are reviewed.	Seen	Not Seen
c) The teacher tells pupils to use Success Criteria when checking their work.	Seen	Not Seen
d) The teacher visits a few pupils to check how they conduct the self-assessment.	Seen	Not Seen
e) The teacher gives feedback on how the self-assessments were conducted.	Seen	Not Seen

END OF LESSON

12. Indicate how the teacher ends her/his lesson. How does the teacher sum up/conclude the lesson? Does the teacher refer to the lesson objectives?

a) Did the teacher complete the lesson?	Yes	No

b) The teacher checks whether the **Learning objectives** have been completed.	Seen	Not Seen
c) The teacher checks whether the **Assessment Criteria** have been met.	Seen	Not Seen

d) Please note any **other interesting observations** you made in this lesson.

13. Written Feedback: Select the 2019 exercise <u>book/s</u> of ONE high- and ONE low-performing pupil. (<u>Note</u>: There could be more than 1 book for a subject.)

After the lesson, review the pupil work STARTING from the first lesson.

Count the number of ticks, crosses and signatures seen and write down, **EXACTLY**, any comments/symbols that the teacher wrote in the pupil's book.

11a. Low-performing pupil

No of ticks______ **No of crosses ______** **No of signatures___** **Number of stamps ______**

(Note what the stamps refer to)

How many exercises were completed in total? ______ (Do not count "Corrections")

Types of Feedback (comment)

a. Descriptive____________________________________

b. Evaluative ____________________________________

With reference to Standard 4 work, indicate how many exercises covered including the HOTS task:

Low order oriented ____________ **Higher Order oriented _______________**

Low order and higher Order combined ___________

Comments/symbols seen

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

11b. High-performing Pupils

No of ticks______ No of crosses ______ No of signatures___ Number of stamps ______

(Note what the stamps refer to)

How many exercises were completed in total? ______ (Do not count "Corrections")

With reference to standard 4 work, indicate how many exercises covered including the HOTS task:

Types of Feedback (comment)
a. Descriptive__________________________________
b. Evaluative __________________________________

With reference to Standard 4 work, indicate how many exercises covered including the HOTS task:

Low order oriented ___________ Higher Order oriented______________

Low order and higher Order combined __________

Comments/symbols seen

"End"

9.3 *Appendix Ci*. Chapter 4 (Phase I)- FA items category functioning before rescoring

1	ITEM	CODE	VALUE	SCORE	UNWTD	UNWTD %	WTD	WTD %	AVGE MEAS	P.SD MEAS	S.E. MEAS	INFIT MNSQ	OUTFIT MNSQ	PTMA	LABEL
42	9	0	0	0	21	14	21	14	0.398	0.443	0.099	1.071	1.910	-0.3311	LGSC_i
43	9	1	1	1	11	7.3333	11	7.3333	0.520	0.409	0.129	1.214	1.112	-0.171	LGSC_i
44	9	2	2	2	9	6	9	6	0.996	0.607	0.214	2.445	4.452	0.0556	LGSC_i
45	9	3	3	3	20	13.3333	20	13.3333	0.722	0.462	0.106	1.645	0.817	-0.1006	LGSC_i
46	9	4	4	4	89	59.3333	89	59.3333	1.044	0.547	0.058	0.890	0.937	0.3674	LGSC_i
47	10	0	0	0	4	2.6667	4	2.6667	0.580	0.214	0.124	2.693	1.978	-0.0835	LGSC_j
48	10	1	1	1	2	1.3333	2	1.3333	1.072	0.167	0.167	6.915	3.942	0.041	LGSC_j
49	10	2	2	2	9	6	9	6	0.427	0.216	0.076	0.971	0.567	-0.1947	LGSC_j
50	10	3	3	3	17	11.3333	17	11.3333	0.746	0.418	0.105	1.262	0.954	-0.0771	LGSC_j
51	10	4	4	4	118	78.6667	118	78.6667	0.928	0.604	0.056	1.246	1.126	0.1939	LGSC_j
52	11	0	0	0	16	10.6667	16	10.6667	0.609	0.564	0.146	1.515	3.584	-0.1567	LGSC_k
53	11	1	1	1	8	5.3333	8	5.3333	0.747	0.498	0.188	1.843	2.577	-0.0506	LGSC_k
54	11	2	2	2	21	14	21	14	0.778	0.558	0.125	1.523	2.183	-0.0645	LGSC_k
55	11	3	3	3	25	16.6667	25	16.6667	0.724	0.403	0.082	1.100	0.620	-0.1135	LGSC_k
56	11	4	4	4	80	53.3333	80	53.3333	1.004	0.597	0.067	1.215	1.292	0.2494	LGSC_k
57	12	0	0	0	14	9.3333	14	9.3333	0.313	0.444	0.123	0.906	1.136	-0.3111	LGSC_l
58	12	1	1	1	14	9.3333	14	9.3333	0.572	0.477	0.132	1.277	1.320	-0.166	LGSC_l
59	12	2	2	2	29	19.3333	29	19.3333	0.647	0.382	0.072	0.805	0.810	-0.1893	LGSC_l
60	12	3	3	3	35	23.3333	35	23.3333	0.856	0.484	0.083	1.042	0.978	-0.0129	LGSC_l
61	12	4	4	4	58	38.6667	58	38.6667	1.195	0.557	0.074	0.869	0.901	0.4497	LGSC_l
62	13	0	0	0	12	8	12	8	0.323	0.333	0.100	0.956	0.938	-0.2805	LGSC_m
63	13	1	1	1	11	7.3333	11	7.3333	0.387	0.363	0.115	0.842	0.732	-0.2361	LGSC_m
64	13	2	2	2	22	14.6667	22	14.6667	0.562	0.415	0.091	0.835	0.875	-0.222	LGSC_m
65	13	3	3	3	24	16	24	16	0.818	0.432	0.090	0.595	0.907	-0.0389	LGSC_m
66	13	4	4	4	81	54	81	54	1.115	0.559	0.063	0.960	0.944	0.4624	LGSC_m
67	14	0	0	0	21	14	21	14	0.499	0.400	0.090	1.213	1.309	-0.2606	LGSC_n
68	14	1	1	1	13	8.6667	13	8.6667	0.469	0.366	0.106	0.820	0.759	-0.2148	LGSC_n
69	14	2	2	2	27	18	27	18	0.628	0.551	0.108	1.252	1.477	-0.197	LGSC_n
70	14	3	3	3	33	22	33	22	0.868	0.349	0.062	0.646	0.513	-0.0018	LGSC_n
71	14	4	4	4	56	37.3333	56	37.3333	1.219	0.579	0.078	0.907	0.915	0.4699	LGSC_n
72	15	0	0	0	3	2	3	2	0.708	0.412	0.292	1.654	1.860	-0.0401	LGSC_o
73	15	1	1	1	10	6.6667	10	6.6667	0.393	0.395	0.132	0.957	0.890	-0.2217	LGSC_o
74	15	2	2	2	31	20.6667	31	20.6667	0.437	0.350	0.064	0.630	0.541	-0.3842	LGSC_o
75	15	3	3	3	34	22.6667	34	22.6667	0.797	0.404	0.070	0.751	0.640	-0.0682	LGSC_o
76	15	4	4	4	72	48	72	48	1.163	0.572	0.068	0.895	0.897	0.4905	LGSC_o
77	16	0	0	0	6	4	6	4	0.496	0.293	0.131	1.319	1.292	-0.1326	LGSC_p

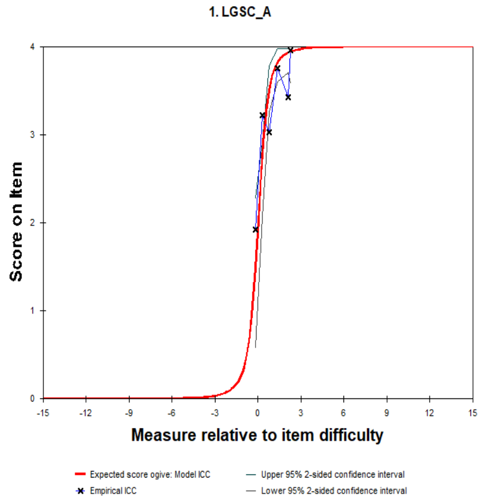
9.4 *Appendix Cii*. Chapter 4 (Phase I)- Rasch-Thurstonian thresholds for FA items before rescoring

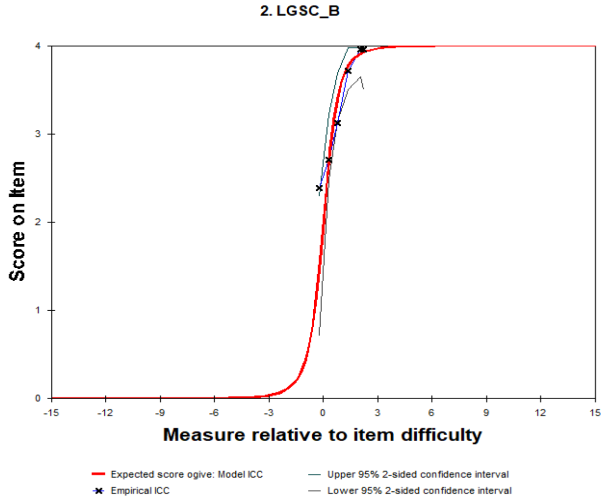
```
INPUT: 150 PERSON  48 ITEM REPORTED: 150 PERSON  48 ITEM  251 CATS WINSTEPS 4.7.1.0
--------------------------------------------------------------------------------

MEASURE    PERSON - MAP - ITEM - 50% Cumulative probabilities (Rasch-Thurstonian thresholds)
          <more>|    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17   18   19   20
    5           +
                |                     EE.5
                |
                |
                |
                |
                |
                |
    4           +
                |
                |
                |
                |
                |
                |
                |
    3           +
                |
                |
                |                     EE.5 EE.6 EE.7 EE.8 EE.9 E.10 E.11 E.12 E.13 E.14 E.15 E.16 E.17 E.18 E.19 E.20
             ## |
             ## |
                |
    2          T+                LT.4
              . |
            .## |                EE.4
             ## |
            .## |           EE.4
             # S|           EE.3 GP.4
          ##### |                GP.4
                                 EE.4
                                 GP.4
         .##### |T               EE.4
                                 LG.4
                                 LG.4
    1    ###### +      EE.2 LT.3 LG.4
                            GP.3 GP.4
                                 EE.4
                                 LG.4
     ######### M|      EE.3 EE.3 LG.4
                            GP.3 EE.4
                                 LG.4
                                 LT.4
                                 EE.4
        .###### | EE.1           EE.4
                                 LT.4
```

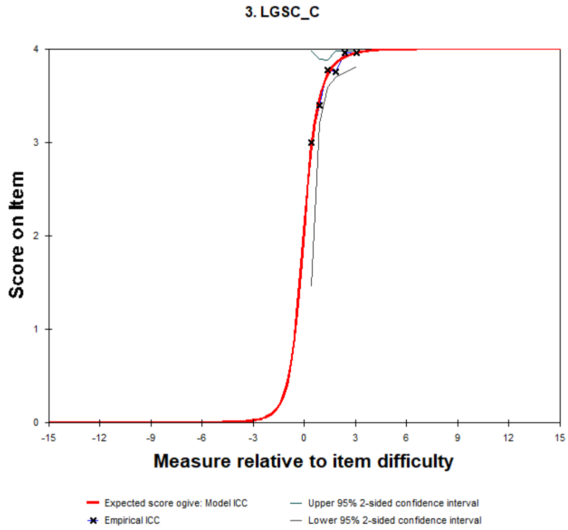
```
                                      LG.4
                                      LG.4
                                      LG.4
                                      GP.4
        .######  |S             GP.3  GP.4
                                GP.3  EE.4
                                LG.3  LG.4
                                LG.3  GP.4
                                      EE.4
                                      LG.4
                                      EE.4
      #########  |              LG.3  LG.4
                                LG.3  LT.4
                                EE.3  LG.4
                                EE.3  EE.4
         ###### S|        GP.2  EE.3  GP.4
                          GP.2  LG.3  EE.4
                                LG.3
                                LG.3
                                LG.3
                                LG.3
            .##  |        LG.2  LG.3  LG.4
                          LG.2  GP.3
                                LG.3
                                EE.3
                                GP.3
                                LG.3
                                LG.3
                                EE.3
                                GP.3
                                LT.3
            ###  |        LG.2  EE.4  EE.4
                          GP.2  LT.4  LT.4
                          LG.2  GP.4
                          GP.2  EE.3
                          LG.2  LT.3
                          LG.2
                          LG.2
                          LG.2
                          LG.2
0            ##  +M LG.1  LG.2  EE.3  LG.4
                    LG.1  GP.2  EE.3
                          LG.2  GP.3
                                LT.3
                 |  LG.1  LT.2  LG.3  EE.4
                    LG.1        EE.3  LG.4
                                EE.3
             .# T|  LG.1  LG.2  LG.3
                          EE.2  EE.3
                          EE.3  LT.3
                          EE.4
                          LG.2
                          EE.2
                 |  LG.1  GP.3
                    LG.1  EE.2
                    GP.1  GP.2
```

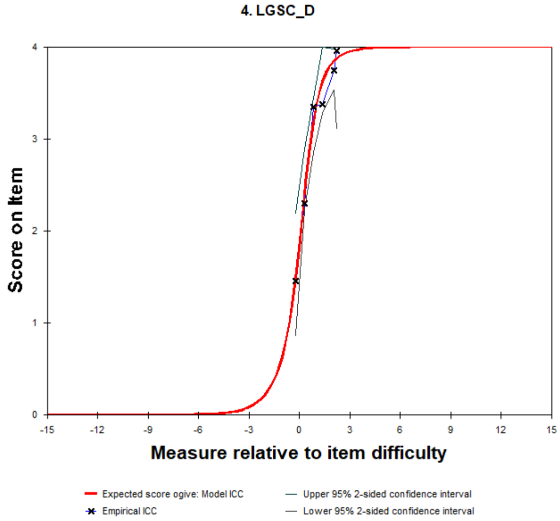
```
                       LG.1  EE.2
                       GP.1  EE.2
                       GP.1  EE.2
                       LG.1
                    |S EE.2  GP.2  EE.3
                       GP.1  LG.2
                       LT.1  LG.2
                       EE.1  EE.2
                             GP.2
                             EE.2
                             LT.2
                    |  LG.1  EE.2  LG.3
                       GP.1  LT.2
                       LG.1  EE.2
                       LG.1  LT.3
                             EE.2
                    |  LG.1  LT.2
                       GP.1  EE.2
                       EE.1
                       EE.1
                       LG.1
                    |  EE.1  LG.2
                       EE.1  LT.2
                       EE.1
                       EE.1
                       EE.3
  -1                +T GP.1
                    |  LG.1  EE.2
                       LT.1
                    |  EE.1
                       LT.1
                       LG.1
                       LT.1
                    |  GP.2
                       EE.1
                       EE.2
                    |  EE.1
                       EE.1
                       GP.1
                       EE.1
                    |
                    |
  -2                +  EE.1
                       LT.1
                    |
                    |  LT.2
                    |
                    |
                    |
                    |
                    |
  -3                +
               <less>|   0     1     2     3     4     5     6     7     8     9    10    11    12    13    14    15    16    17    18    19
EACH "#" IS 2: EACH "." IS 1
```

9.5 *Appendix Ciii*. Chapter 4 (Phase I)-FA- Multiple ICCS be before Rescoring

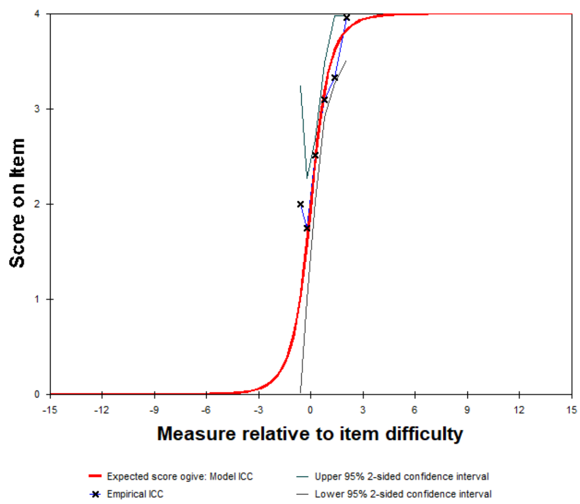




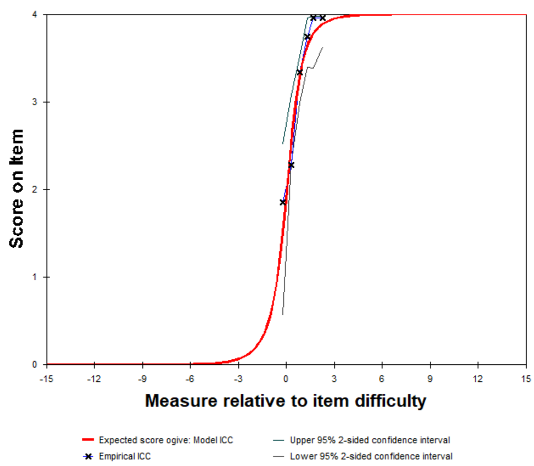




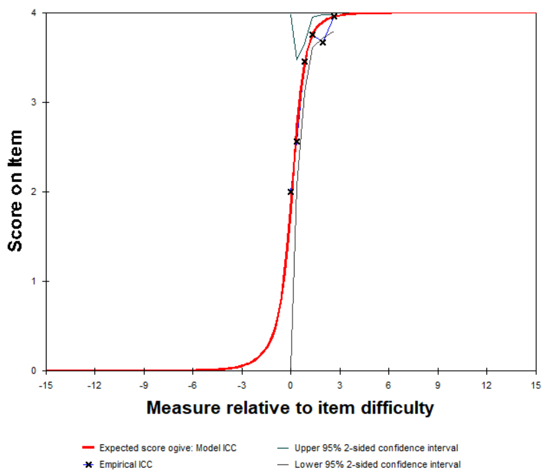
5. LGSC_E



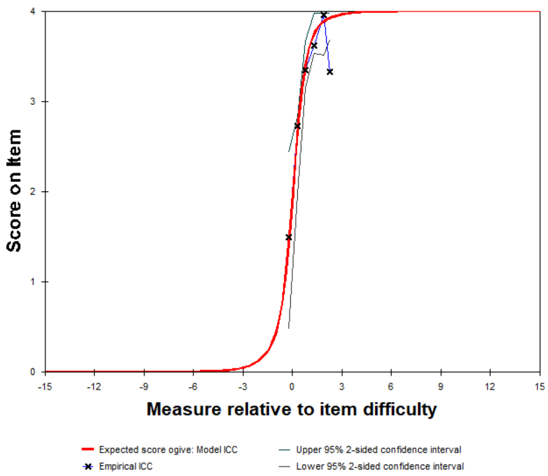
6. LGSC_F



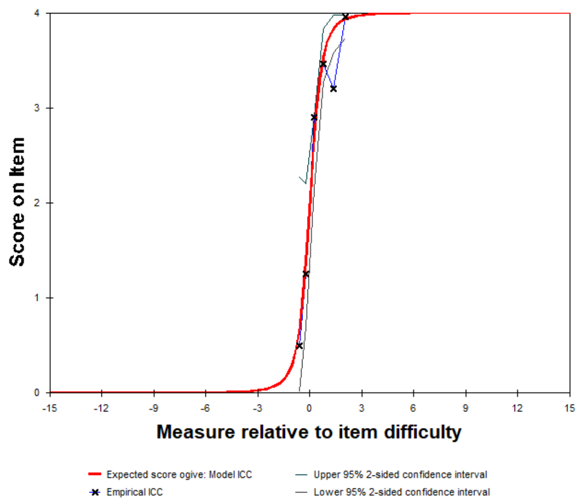
7. LGSC_G



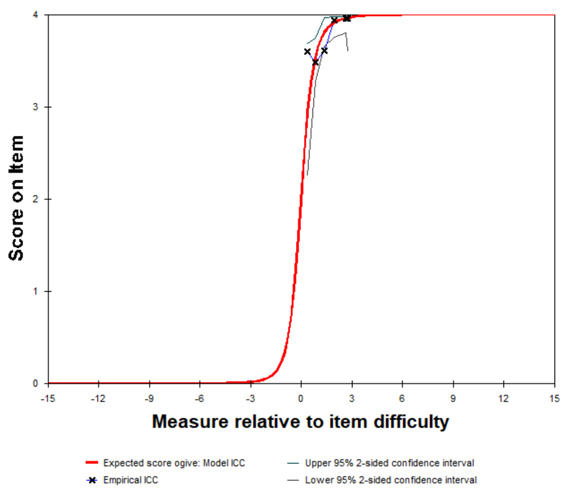
8. LGSC_H



9. LGSC_I

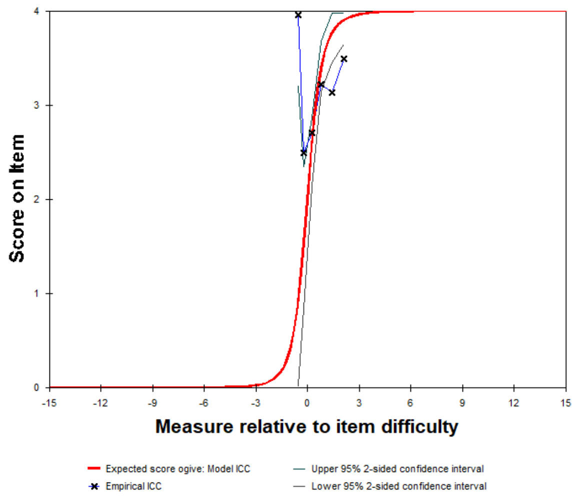


10. LGSC_J

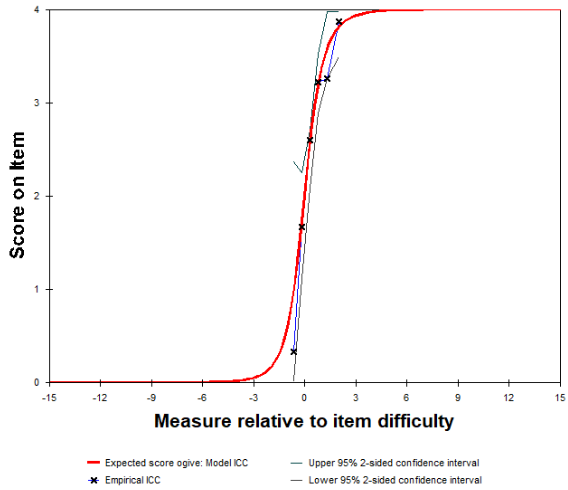


UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

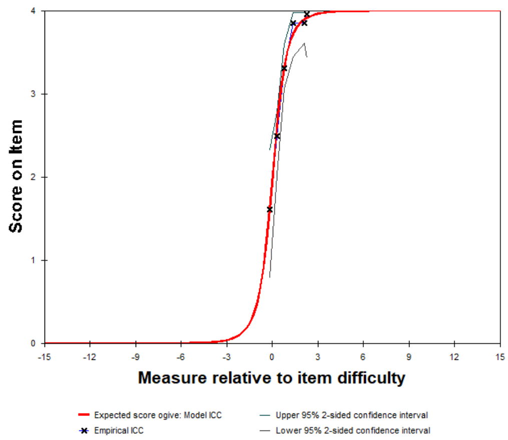
11. LGSC_K



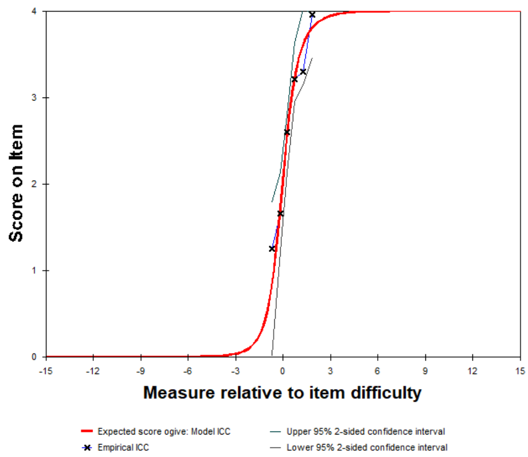
12. LGSC_L



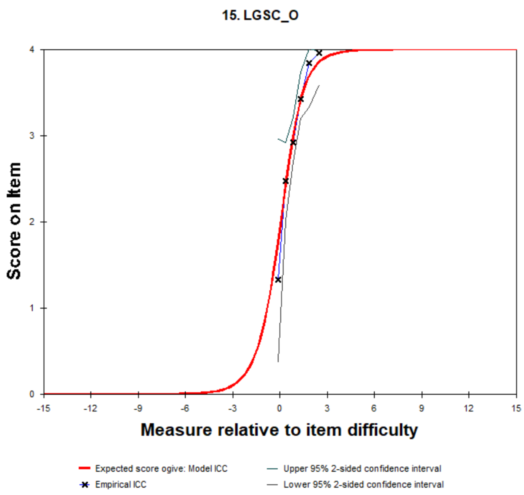
13. LGSC_M

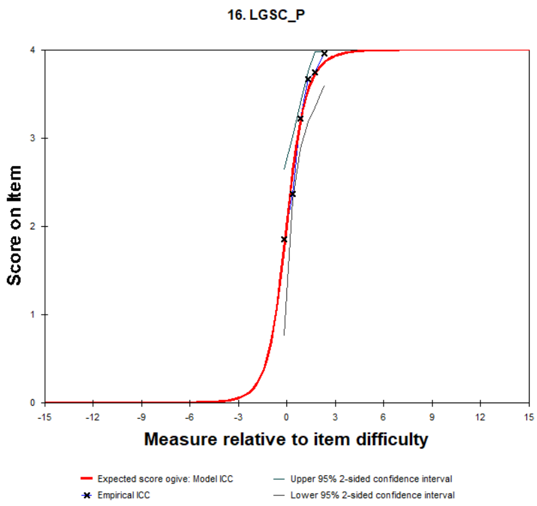


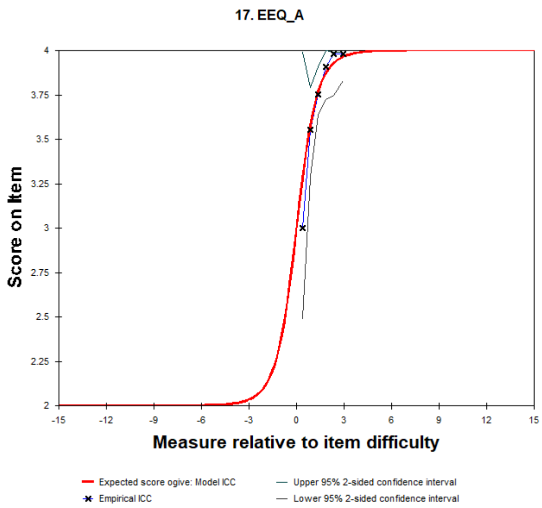
14. LGSC_N

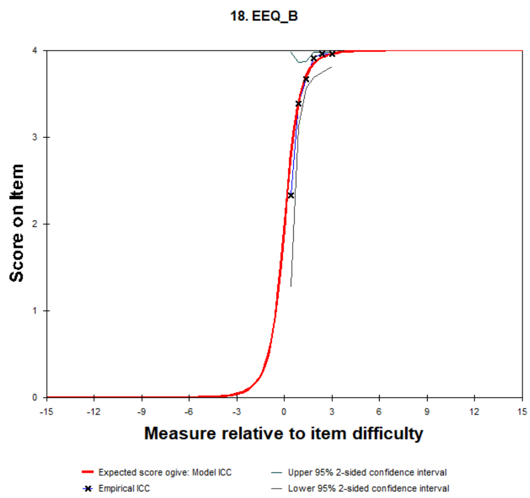


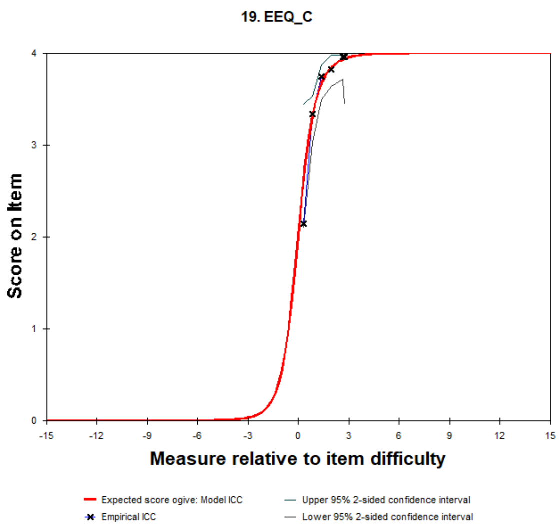
UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

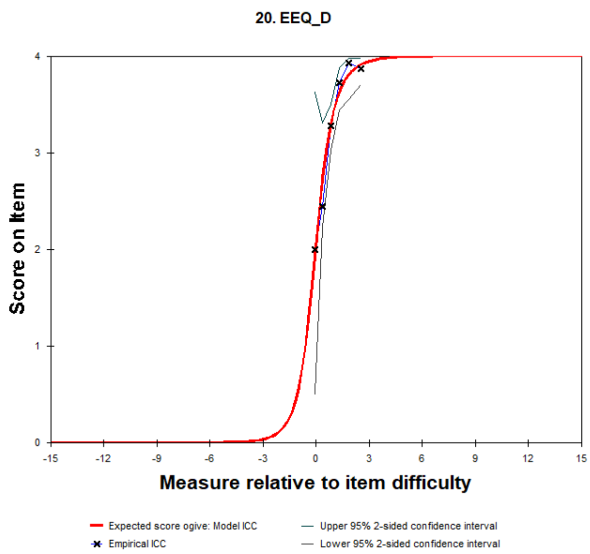




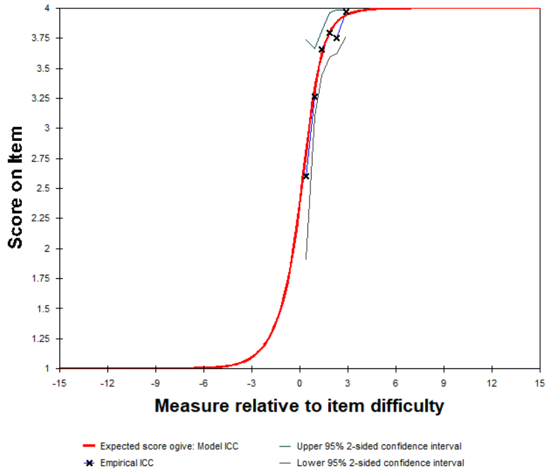




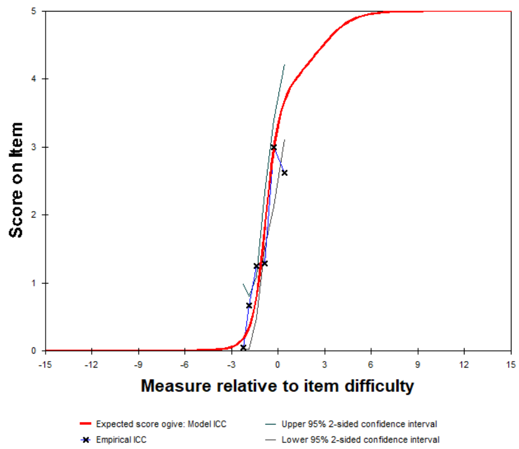




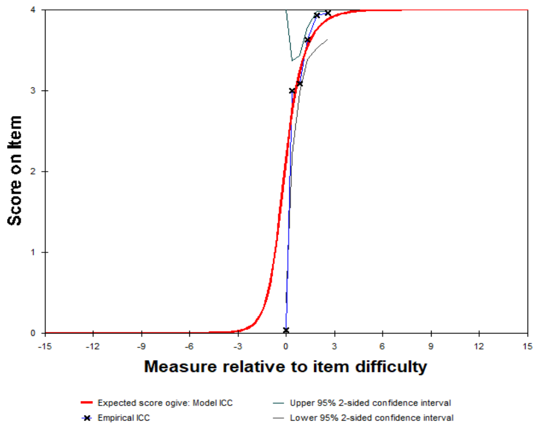
21. EEQ_E



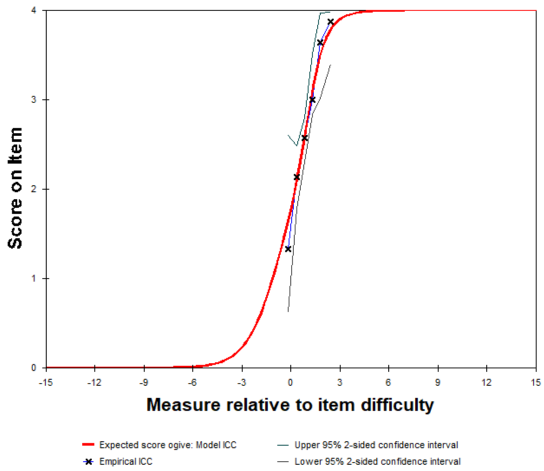
22. EEQ_F



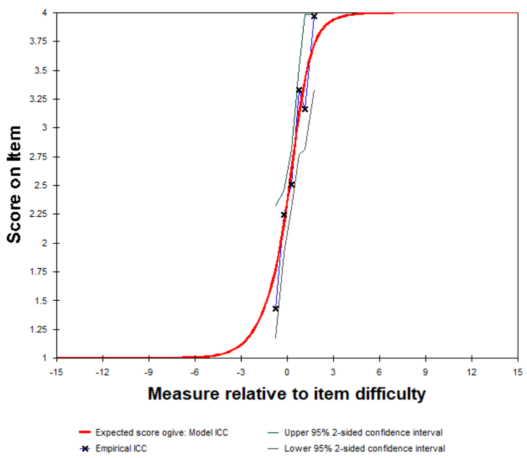
23. EEQ_G



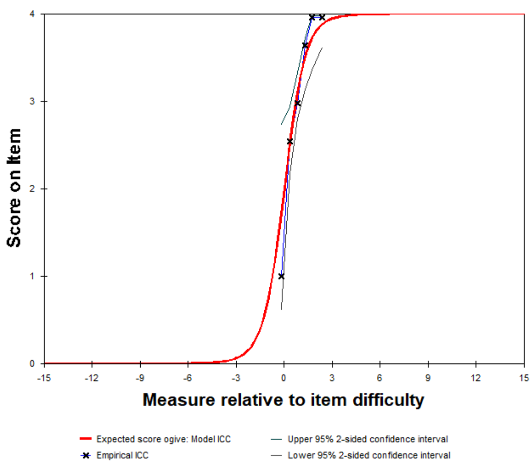
24. EEC_A

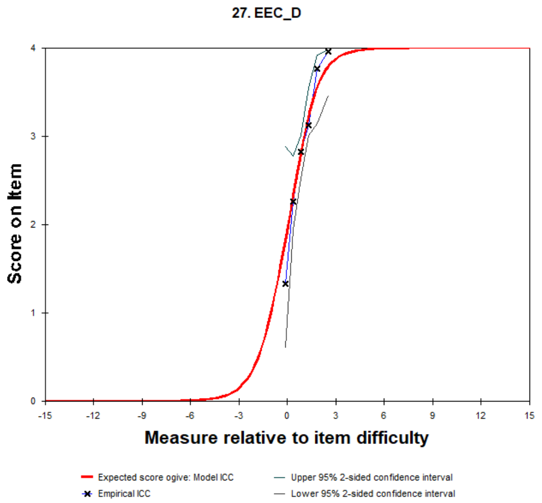


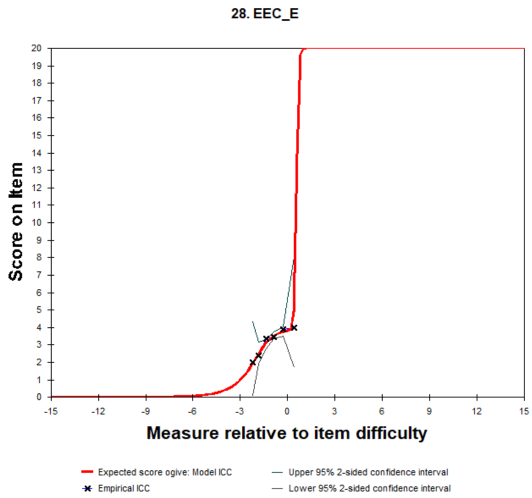
25. EEC_B

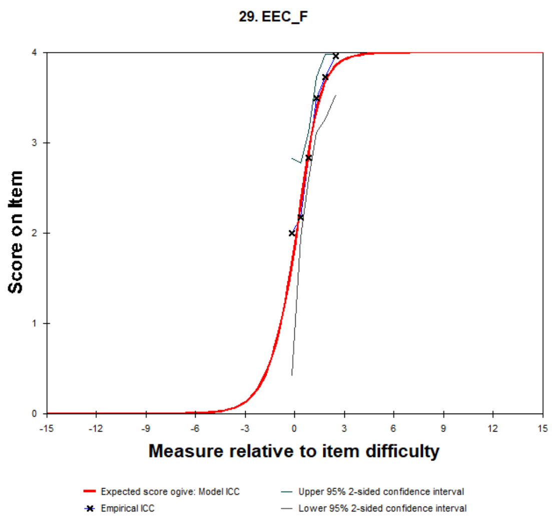


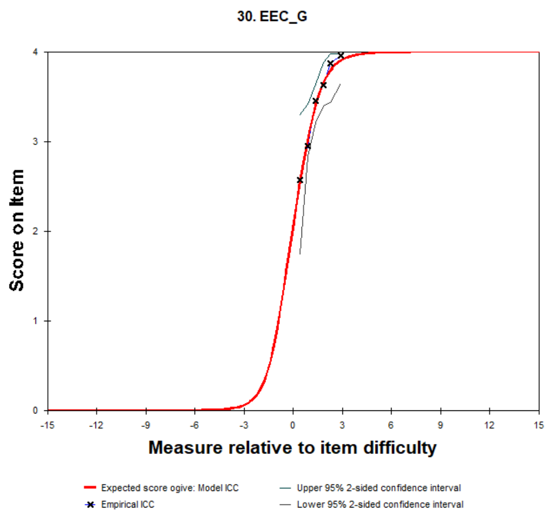
26. EEC_C

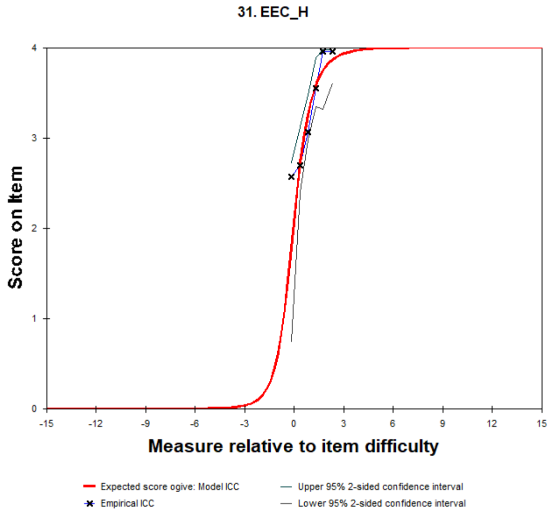


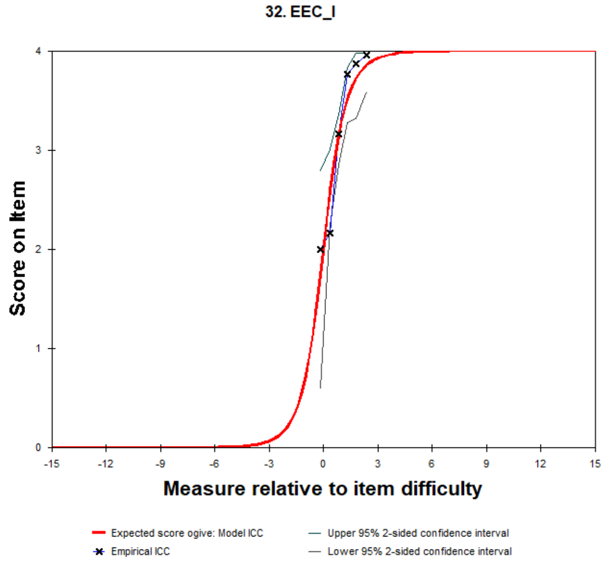

27. EEC_D
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval

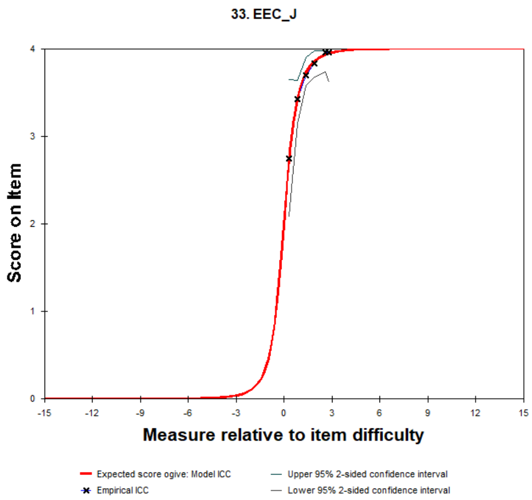

28. EEC_E
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval

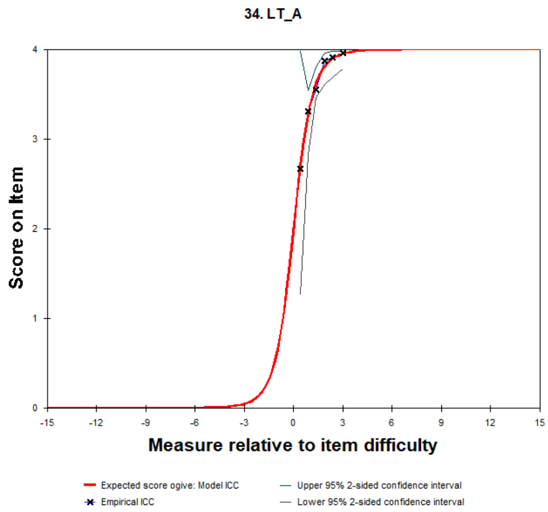

29. EEC_F
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval


30. EEC_G
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval

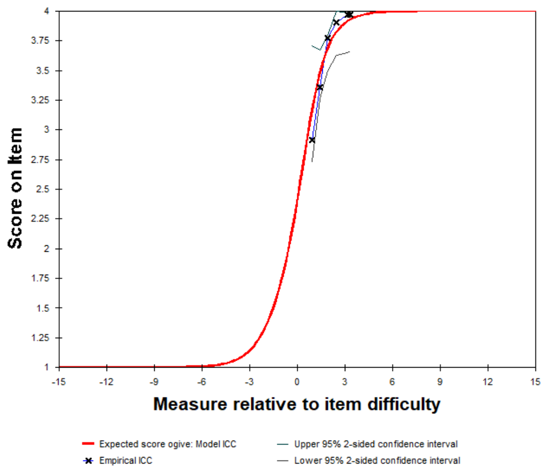

31. EEC_H
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval


32. EEC_I
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval

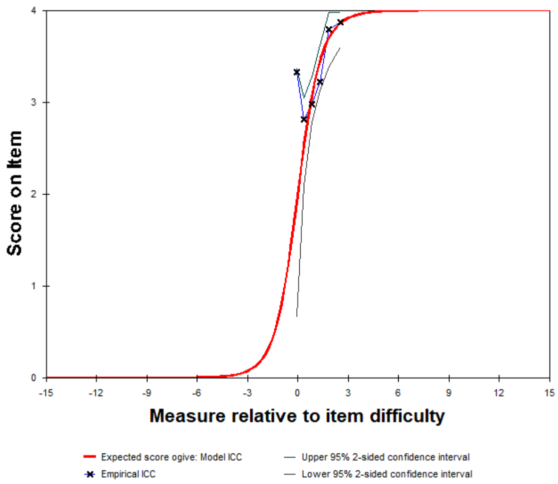

33. EEC_J
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval


34. LT_A
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval

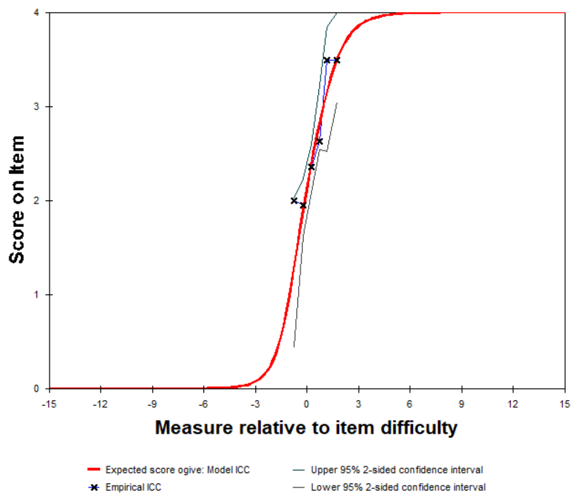
35. LT_B



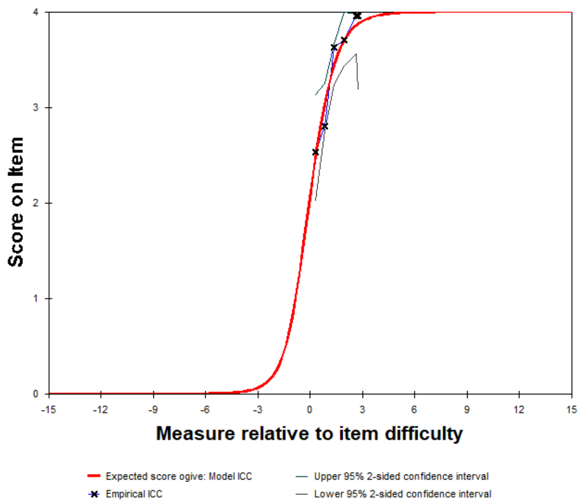
36. LT_C



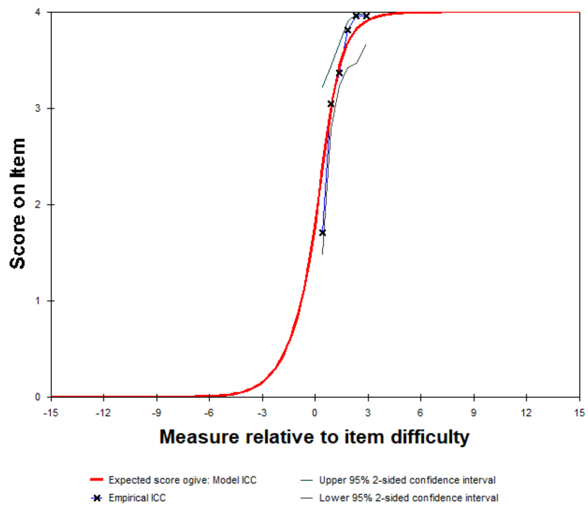
37. LT_D



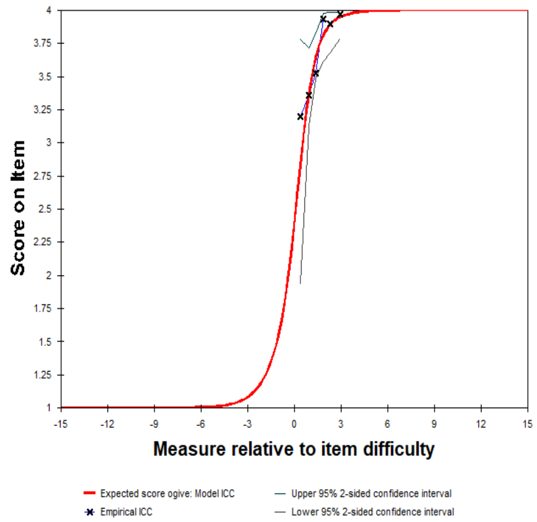
38. LT_E



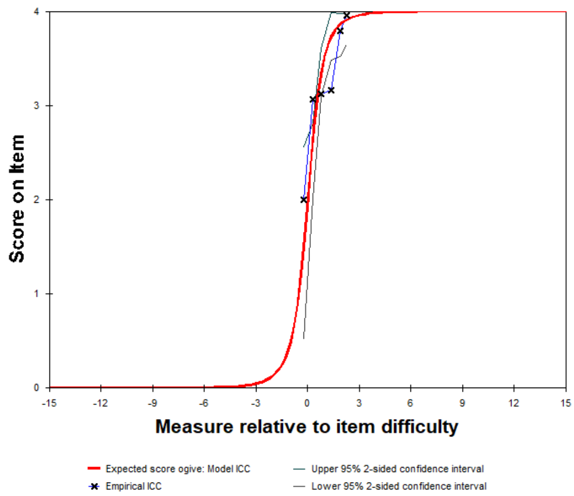
39. LT_F



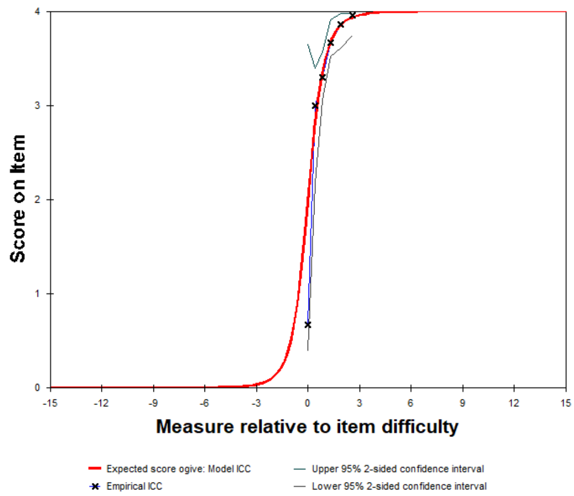
40. GPF_A



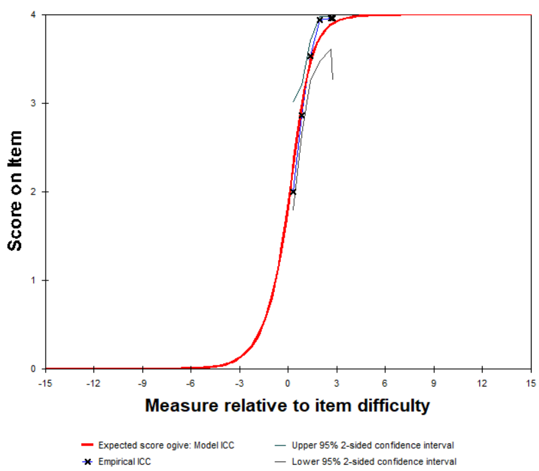
41. GPF_B



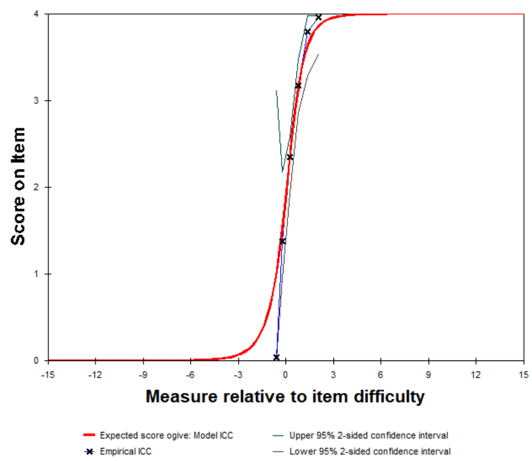
42. GPF_C



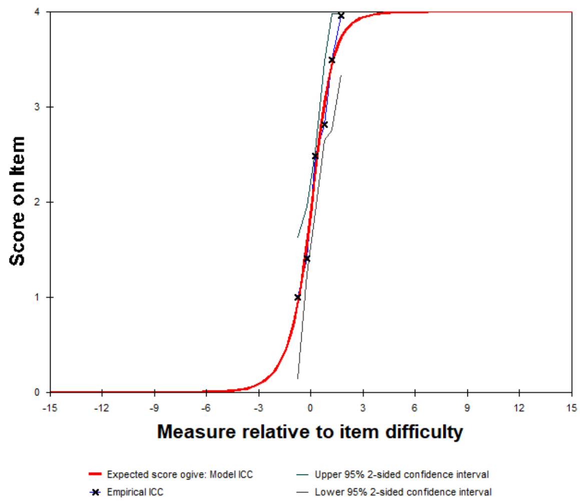
43. GPF_D



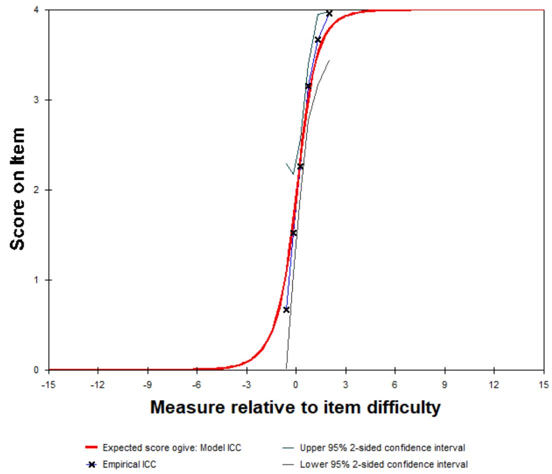
44. GPF_E



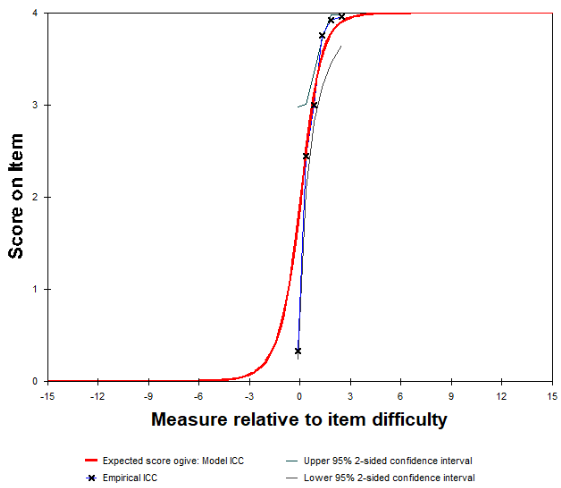
45. GPF_F



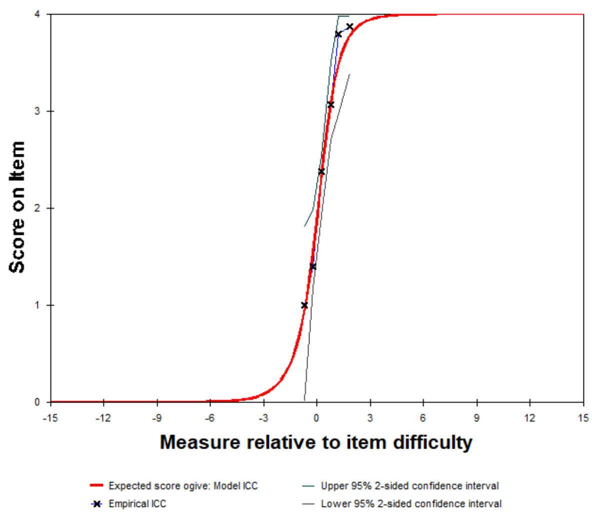
46. GPF_G



48. GPF_I



47. GPF_H



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

9.6 Appendix Civ. Chapter 4 (Phase I)- FA items category functioning after rescoring

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ITEM	CODE	VALUE	SCORE	UNWTD	UNWTD %	WTD	WTD %	AVGE MEAS	P.SD MEAS	S.E. MEAS	INFIT MNSQ	OUTFIT MNSQ	PTMA	LABEL	Order
2	1	1	1	1	55	36.6667	55	36.6667	-0.4003	1.8008	0.2451	1.4807	2.1466	-0.2841	LGSC_a	3
3	1	2	2	2	95	63.3333	95	63.3333	0.5515	1.3804	0.1424	1.1772	1.2743	0.2841	LGSC_a	4
4	2	1	1	1	64	42.6667	64	42.6667	-0.6327	1.5525	0.1956	1.0725	1.0794	-0.4463	LGSC_b	7
5	2	2	2	2	86	57.3333	86	57.3333	0.824	1.3589	0.1474	1.028	1.0052	0.4463	LGSC_b	8
6	3	1	1	1	30	20	30	20	-1.329	1.7441	0.3239	0.9615	0.7906	-0.4743	LGSC_c	11
7	3	2	2	2	120	80	120	80	0.5854	1.3283	0.1218	0.9495	0.9301	0.4743	LGSC_c	12
8	4	1	1	1	83	55.3333	83	55.3333	-0.4642	1.3547	0.1496	0.9413	0.893	-0.4597	LGSC_d	15
9	4	2	2	2	67	44.6667	67	44.6667	1.0284	1.526	0.1878	1.045	2.3764	0.4597	LGSC_d	16
10	5	1	1	1	88	58.6667	88	58.6667	-0.4226	1.4503	0.1555	1.0408	0.9892	-0.4613	LGSC_e	19
11	5	2	2	2	62	41.3333	62	41.3333	1.0898	1.4064	0.1801	1.0119	1.0203	0.4613	LGSC_e	20
12	6	1	1	1	70	46.6667	70	46.6667	-0.6552	1.3826	0.1665	0.8814	0.7623	-0.497	LGSC_f	23
13	6	2	2	2	80	53.3333	80	53.3333	0.953	1.4167	0.1594	1.057	1.0896	0.497	LGSC_f	24
14	7	1	1	1	39	26	39	26	-0.8629	1.8243	0.2959	1.1622	1.5022	-0.3912	LGSC_g	27
15	7	2	2	2	111	74	111	74	0.5768	1.3467	0.1284	1.0573	1.0537	0.3912	LGSC_g	28
16	8	1	1	1	60	40	60	40	-0.5191	1.6865	0.2196	1.2472	1.732	-0.365	LGSC_h	31
17	8	2	2	2	90	60	90	60	0.6836	1.3671	0.1449	1.1243	1.1272	0.365	LGSC_h	32
18	9	1	1	1	61	40.6667	61	40.6667	-0.4752	1.6549	0.2136	1.2905	1.4048	-0.3475	LGSC_i	35
19	9	2	2	2	89	59.3333	89	59.3333	0.667	1.4089	0.1502	1.1731	1.1946	0.3475	LGSC_i	36
20	10	1	1	1	32	21.3333	32	21.3333	-0.713	1.8583	0.3338	1.5195	1.4704	-0.2953	LGSC_j	39
21	10	2	2	2	118	78.6667	118	78.6667	0.4508	1.4448	0.1336	1.3138	1.2873	0.2953	LGSC_j	40
22	11	1	1	1	70	46.6667	70	46.6667	-0.3213	1.615	0.1944	1.295	1.673	-0.3035	LGSC_k	43
23	11	2	2	2	80	53.3333	80	53.3333	0.6608	1.4678	0.1651	1.2962	1.4519	0.3035	LGSC_k	44
24	12	1	1	1	92	61.3333	92	61.3333	-0.4029	1.4583	0.1529	1.057	0.9917	-0.4723	LGSC_l	47
25	12	2	2	2	58	38.6667	58	38.6667	1.1628	1.365	0.1808	0.9832	0.8652	0.4723	LGSC_l	48
26	13	1	1	1	69	46	69	46	-0.6183	1.5128	0.1835	1.0242	1.0388	-0.4693	LGSC_m	51
27	13	2	2	2	81	54	81	54	0.9017	1.3469	0.1506	0.9918	0.9626	0.4693	LGSC_m	52
28	14	1	1	1	94	62.6667	94	62.6667	-0.3981	1.4291	0.1482	1.019	0.9456	-0.482	LGSC_n	55
29	14	2	2	2	56	37.3333	56	37.3333	1.2107	1.3894	0.1874	0.9545	0.9327	0.482	LGSC_n	56
30	15	1	1	1	78	52	78	52	-0.637	1.4195	0.1618	0.9105	0.8405	-0.5412	LGSC_o	59
31	15	2	2	2	72	48	72	48	1.1119	1.287	0.1527	0.8532	0.7771	0.5412	LGSC_o	60
32	16	1	1	1	77	51.3333	77	51.3333	-0.6677	1.4139	0.1622	0.8988	0.8073	-0.5536	LGSC_p	63
33	16	2	2	2	73	48.6667	73	48.6667	1.1204	1.2669	0.1493	0.8298	0.7315	0.5536	LGSC_p	64
34	17	1	1	1	27	18.1208	27	18.1208	-1.5486	1.7206	0.3374	0.8387	0.6655	-0.5095	EEQ_a	67
35	17	2	2	2	122	81.8792	122	81.8792	0.5935	1.3101	0.1191	0.8945	0.8831	0.5095	EEQ_a	68
36	18	1	1	1	30	20.1342	30	20.1342	-1.4239	1.6629	0.3088	0.8716	0.6452	-0.5051	EEQ_b	71
37	18	2	2	2	119	79.8658	119	79.8658	0.616	1.3224	0.1217	0.9422	0.911	0.5051	EEQ_b	72

9.7 *Appendix Di*. Chapter 4 (Phase I)- HOTS items category functioning before rescoring

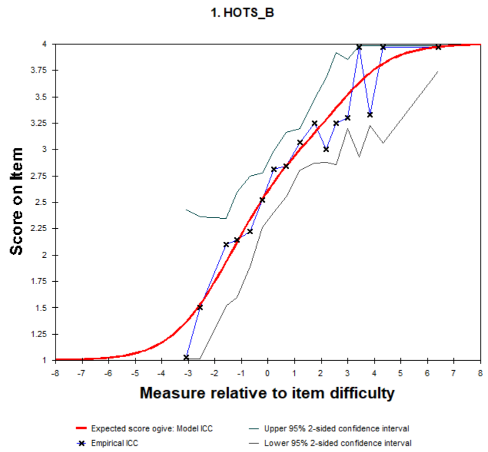
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	ITEM	CODE	VALUE	SCORE	UNWTD	UNWTD %	WTD	WTD %	AVGE MEAS	P.SD MEAS	S.E. MEAS	INFIT MNSQ	OUTFIT MNSQ	PTMA	LABEL
2	1	1	1	1	7	4.6667	7	4.6667	-1.502	0.6418	0.262	0.7157	0.7015	-0.3432	HOTS_a
3	1	2	2	2	11	7.3333	11	7.3333	-0.1778	1.194	0.3776	1.8005	1.4083	-0.2022	HOTS_a
4	1	3	3	3	87	58	87	58	0.6203	1.0286	0.1109	0.8861	0.7569	-0.2551	HOTS_a
5	1	4	4	4	45	30	45	30	2.2967	1.6462	0.2482	1.0939	0.9782	0.5477	HOTS_a
6	2	1	1	1	8	5.3333	8	5.3333	-1.4597	0.8393	0.3172	0.7548	0.7012	-0.3618	HOTS_b
7	2	2	2	2	37	24.6667	37	24.6667	0.1308	1.1543	0.1924	1.2616	1.2879	-0.3003	HOTS_b
8	2	3	3	3	80	53.3333	80	53.3333	1.0524	1.1816	0.1329	1.0964	1.085	0.0583	HOTS_b
9	2	4	4	4	25	16.6667	25	16.6667	2.6996	1.6055	0.3277	1.0763	1.0848	0.4874	HOTS_b
10	3	1	1	1	6	4	6	4	-1.6181	0.5718	0.2557	0.6849	0.5535	-0.3315	HOTS_c
11	3	2	2	2	15	10	15	10	0.0851	1.4558	0.3891	1.8143	1.6139	-0.1845	HOTS_c
12	3	3	3	3	94	62.6667	94	62.6667	0.7076	1.0782	0.1118	0.8471	0.7997	-0.2101	HOTS_c
13	3	4	4	4	35	23.3333	35	23.3333	2.479	1.6781	0.2878	1.0872	0.991	0.5248	HOTS_c
14	4	1	1	1	13	8.6667	13	8.6667	-1.0655	1.0345	0.2986	0.9798	1.1572	-0.3933	HOTS_d
15	4	2	2	2	19	12.6667	19	12.6667	-0.333	0.9795	0.2309	0.8543	0.8985	-0.3109	HOTS_d
16	4	3	3	3	80	53.3333	80	53.3333	0.8194	0.9446	0.1063	0.7392	0.6786	-0.0983	HOTS_d
17	4	4	4	4	38	25.3333	38	25.3333	2.6177	1.4962	0.246	0.9082	0.8581	0.6049	HOTS_d
18	5	1	1	1	6	4	6	4	-1.3721	0.8085	0.3616	0.8908	0.9292	-0.2999	HOTS_e
19	5	2	2	2	18	12	18	12	-0.6131	0.9884	0.2397	0.8509	0.8192	-0.3664	HOTS_e
20	5	3	3	3	77	51.3333	77	51.3333	0.731	1.1451	0.1314	1.0221	1.1344	-0.1515	HOTS_e
21	5	4	4	4	49	32.6667	49	32.6667	2.2006	1.4327	0.2068	0.9418	0.9916	0.5407	HOTS_e
22	6	1	1	1	18	12	18	12	-0.8977	1.2614	0.3059	1.1201	1.1144	-0.4325	HOTS_f
23	6	2	2	2	53	35.3333	53	35.3333	0.2797	0.8602	0.1193	0.8244	0.7869	-0.3187	HOTS_f
24	6	3	3	3	53	35.3333	53	35.3333	1.5575	1.2323	0.1709	1.0856	1.1792	0.275	HOTS_f
25	6	4	4	4	26	17.3333	26	17.3333	2.4475	1.6202	0.324	1.3282	1.4344	0.4265	HOTS_f
26	7	1	1	1	10	6.6667	10	6.6667	-1.3721	0.6886	0.2295	0.684	0.6806	-0.3927	HOTS_g
27	7	2	2	2	48	32	48	32	0.0244	0.9495	0.1385	0.8757	0.8732	-0.4059	HOTS_g
28	7	3	3	3	70	46.6667	70	46.6667	1.2573	0.9921	0.1194	0.7889	0.7765	0.1715	HOTS_g
29	7	4	4	4	22	14.6667	22	14.6667	3.1539	1.5711	0.3428	0.9078	0.8804	0.5702	HOTS_g
30	8	1	1	1	28	18.7919	28	18.7919	-0.305	1.2448	0.2396	1.4487	1.4475	-0.3827	HOTS_h
31	8	2	2	2	52	34.8993	52	34.8993	0.3311	1.0901	0.1527	1.094	1.0884	-0.2908	HOTS_h
32	8	3	3	3	48	32.2148	48	32.2148	1.3678	0.9557	0.1394	0.9888	0.9189	0.1739	HOTS_h
33	8	4	4	4	21	14.094	21	14.094	3.3077	1.4067	0.3145	0.7864	0.7749	0.5944	HOTS_h
34	9	1	1	1	28	18.7919	28	18.7919	-0.3617	1.417	0.2727	1.3634	1.2445	-0.3998	HOTS_i
35	9	2	2	2	71	47.651	71	47.651	0.5435	0.9736	0.1164	0.8083	0.761	-0.252	HOTS_i
36	9	3	3	3	36	24.1611	36	24.1611	1.8162	0.9764	0.165	0.8951	0.862	0.3009	HOTS_i
37	9	4	4	4	14	9.396	14	9.396	3.5681	1.6685	0.4628	1.0735	1.0832	0.5251	HOTS_i

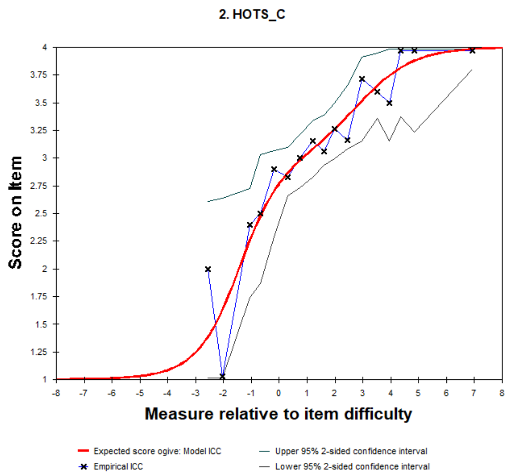
9.8 Appendix Dii. Chapter 4 (Phase I)-HOTS Items Maps with Thurstonian Thresholds before Rescoring

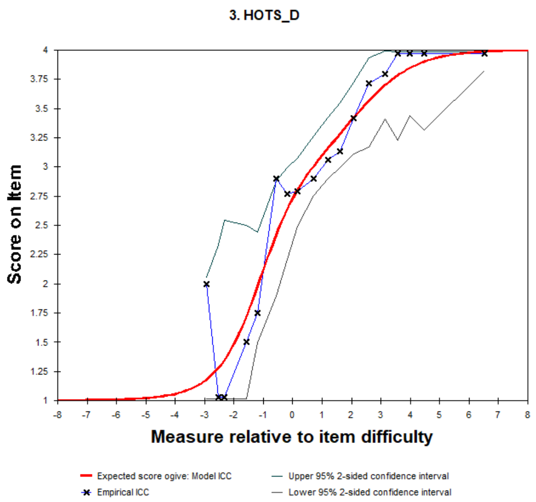
```
INPUT: 150 PERSON  15 ITEM REPORTED: 150 PERSON  15 ITEM  57 CATS WINSTEPS 4.7.1.0
----------------------------------------------------------------------------------

MEASURE PERSON - MAP - ITEM - 50% Cumulative probabilities (Rasch-Thurstonian thresholds)
             <more>|<rare>  1         2         3         4
    5          XX  +
                   |
                X  |
                   |
                   |
              XXX  |                      HOTS_J.4
    4             T+
               XX  |                                HOTS_I.4
                   |
           XXXXXX  |
                   |                      HOTS_G.4  HOTS_H.4
             XXXX  |                      HOTS_B.4
    3              +                      HOTS_F.4
             XXXX  |
              XXX  |                      HOTS_C.4
                                          HOTS_O.4
                  S|                      HOTS_D.4
             XXXX  |                      HOTS_K.4
         XXXXXXXX  |
                   |                      HOTS_E.4
    2        XXXX  +                      HOTS_I.3
      XXXXXXXXXXX  |
         XXXXXXXX  |
                   |T           HOTS_N.4  HOTS_L.4
        XXXXXXXXX  |            HOTS_J.3  HOTS_H.3
         XXXXXXXX M|
    1       XXXXX  +            HOTS_F.3  HOTS_M.4
                   |S
          XXXXXXX  |
        XXXXXXXXX  |            HOTS_G.3
       XXXXXXXXXX  |
          XXXXXXX  |
    0      XXXXXX  +M           HOTS_B.3
                   |            HOTS_O.3
           XXXXXX S|
              XXX  |            HOTS_D.3
```

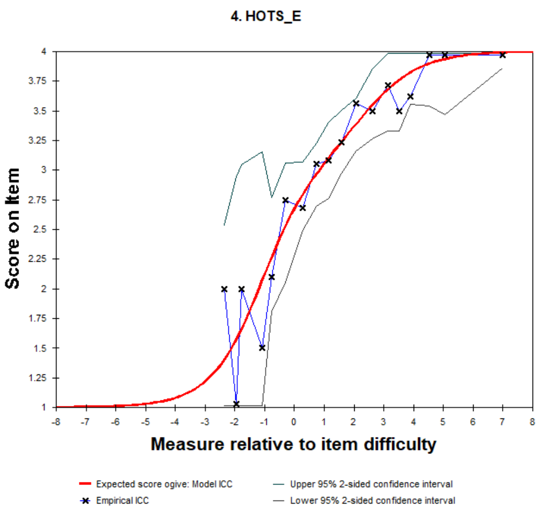
```
               X   |              HOTS_H.2
                                  HOTS_I.2
            XXXX   |S
 -1           XX   +   HOTS_J.2   HOTS_E.3
          XXXXXX   |              HOTS_C.3
              XX   |   HOTS_F.2   HOTS_K.3
              XX   |T  HOTS_D.2
                   |
                  T|   HOTS_O.2   HOTS_L.3
 -2                +   HOTS_G.2
               X   |   HOTS_C.2
                       HOTS_B.2
                       HOTS_N.3
                       HOTS_E.2
               X   |              HOTS_M.3
                   |
                   |
               X   |   HOTS_K.2
                   |
 -3                +   HOTS_L.2
                   |
                   |   HOTS_M.2
                   |                   |
 -4                +
                   |   HOTS_I.1
                       HOTS_H.1
                   |
                   |

 -5                +
            <less>|<freq>   0         1         2         3
```

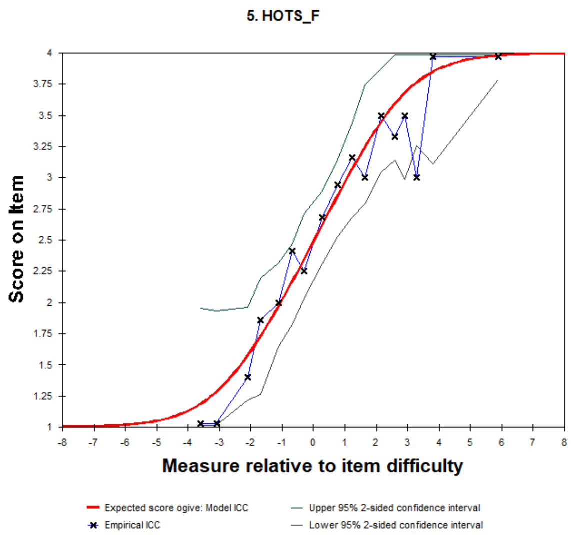
9.9 *Appendix Diii*. Chapter 4 (Phase I)-HOTS items multiple ICCs before rescoring

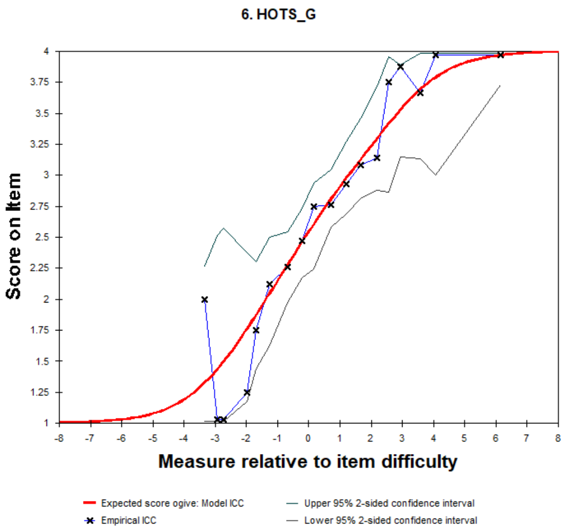


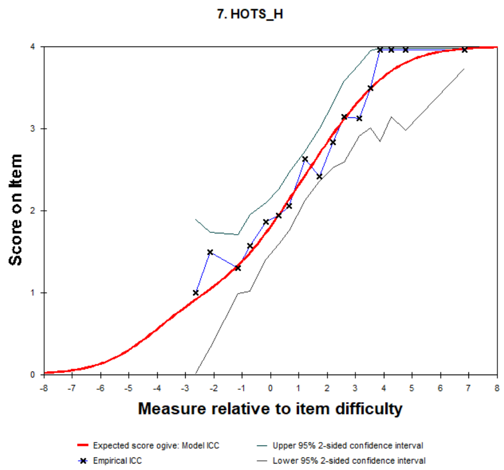


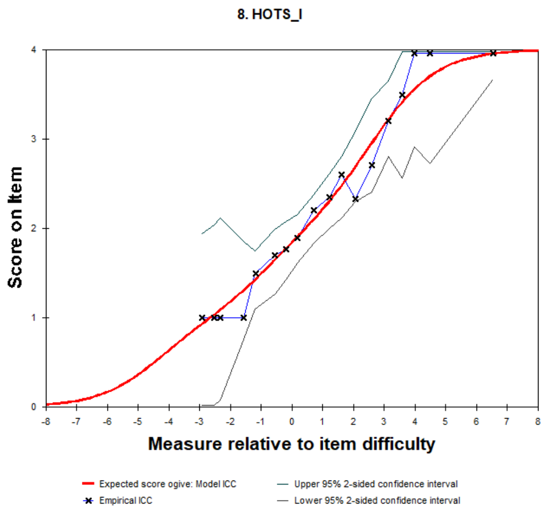


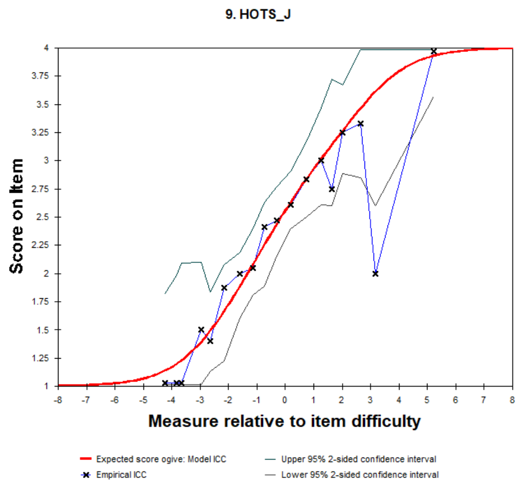


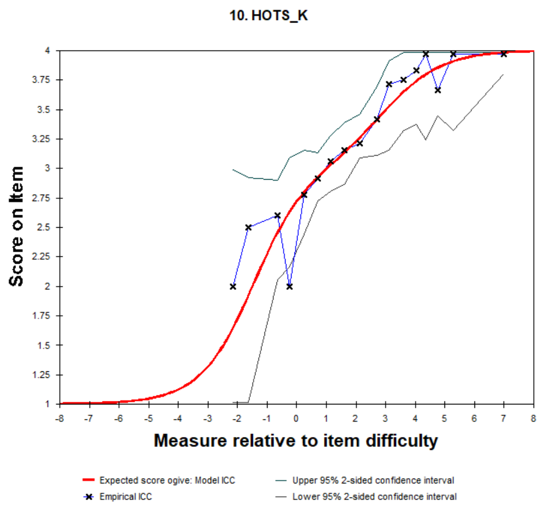




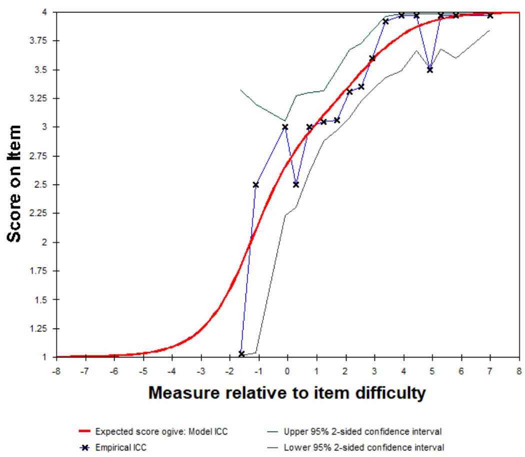

7. HOTS_H
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval


8. HOTS_I
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval

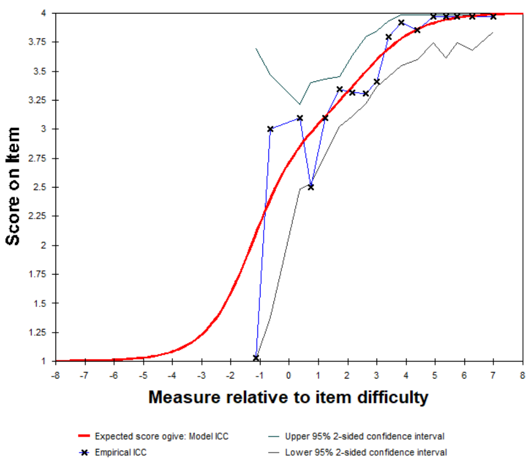

9. HOTS_J
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval


10. HOTS_K
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval

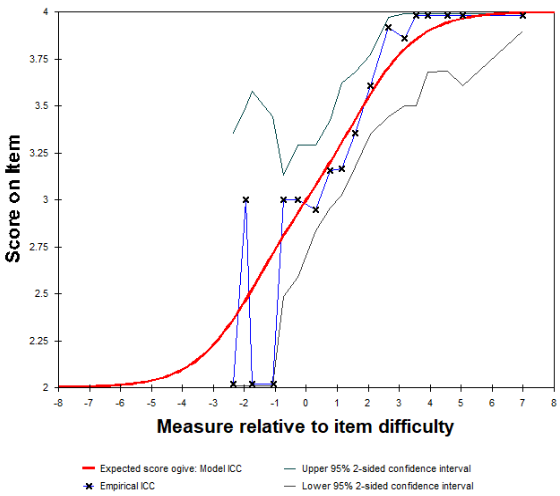
11. HOTS_L



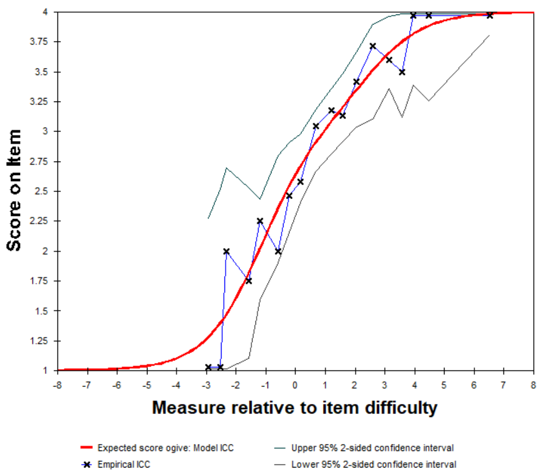
12. HOTS_M



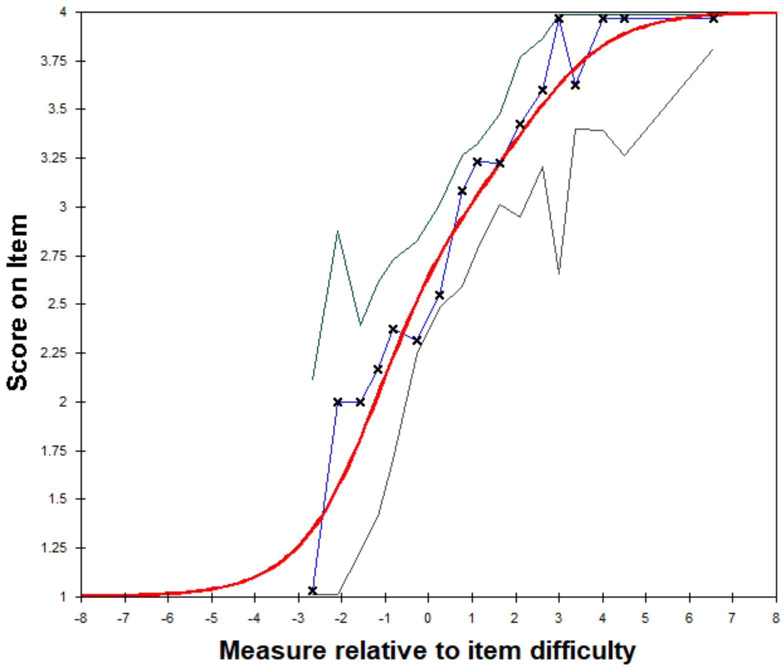
13. HOTS_N



14. HOTS_O



15. HOTS_O


Score on Item
Measure relative to item difficulty
4
3.75
3.5
3.25
3
2.75
2.5
2.25
2
1.75
1.5
1.25
1
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8

9.10 Appendix Div. Chapter 4 (Phase I)- HOTS items category functioning after rescoring

	ITEM	CODE	VALUE	SCORE	UNWTD	UNWTD %	WTD	WTD %	AVGE MEAS	P.SD MEAS	S.E. MEAS	INFIT MNSQ	OUTFIT MNSQ	PTMA	LABEL
2	1	1	1	1	18	12	18	12	-1.9942	1.6241	0.3939	1.1631	1.1818	-0.4006	HOT1_FormativeAssessUnderstand
3	1	2	2	2	87	58	87	58	-0.3452	1.303	0.1405	0.9698	0.8991	-0.23	HOT1_FormativeAssessUnderstand
4	1	3	3	3	45	30	45	30	1.524	1.7271	0.2604	1.1465	1.101	0.5318	HOT1_FormativeAssessUnderstand
5	2	1	1	1	45	30	45	30	-1.3635	1.6484	0.2485	1.2319	1.2405	-0.4876	HOT2_PlanFormativeAssess
6	2	2	2	2	80	53.3333	80	53.3333	0.1859	1.339	0.1506	1.068	1.2441	0.097	HOT2_PlanFormativeAssess
7	2	3	3	3	25	16.6667	25	16.6667	1.9654	1.6394	0.3346	1.1217	1.1396	0.4697	HOT2_PlanFormativeAssess
8	3	1	1	1	21	14	21	14	-1.831	1.9862	0.4441	1.2828	1.2605	-0.4022	HOT3_CurriculumGuidesAssess
9	3	2	2	2	94	62.6667	94	62.6667	-0.2273	1.2703	0.1317	0.8885	0.8521	-0.1712	HOT3_CurriculumGuidesAssess
10	3	3	3	3	35	23.3333	35	23.3333	1.7848	1.6441	0.282	1.0079	0.9845	0.5257	HOT3_CurriculumGuidesAssess
11	4	1	1	1	32	21.3333	32	21.3333	-1.8567	1.577	0.2832	1.0347	1.019	-0.5264	HOT4_CurriculumGuidesIndiv
12	4	2	2	2	80	53.3333	80	53.3333	-0.1324	1.129	0.127	0.7767	0.721	-0.0865	HOT4_CurriculumGuidesIndiv
13	4	3	3	3	38	25.3333	38	25.3333	1.9121	1.484	0.244	0.8853	0.8691	0.595	HOT4_CurriculumGuidesIndiv
14	5	1	1	1	24	16	24	16	-2.2675	1.4643	0.3053	0.8646	0.8545	-0.5378	HOT5_InteractAllStudents
15	5	2	2	2	77	51.3333	77	51.3333	-0.1796	1.2964	0.1487	1.0575	1.104	-0.1093	HOT5_InteractAllStudents
16	5	3	3	3	49	32.6667	49	32.6667	1.447	1.4746	0.2128	1.0003	1.017	0.5368	HOT5_InteractAllStudents
17	6	1	1	1	71	47.3333	71	47.3333	-1.1247	1.484	0.1774	1.0514	1.0406	-0.584	HOT6_FormativeAssessHOT
18	6	2	2	2	53	35.3333	53	35.3333	0.7132	1.333	0.1849	1.1601	1.0415	0.2773	HOT6_FormativeAssessHOT
19	6	3	3	3	26	17.3333	26	17.3333	1.7194	1.6608	0.3322	1.3467	1.922	0.4202	HOT6_FormativeAssessHOT
20	7	1	1	1	58	38.6667	58	38.6667	-1.3824	1.461	0.1935	0.9608	0.9462	-0.5995	HOT7_AdminSupportHOT
21	7	2	2	2	70	46.6667	70	46.6667	0.4246	1.0826	0.1303	0.8057	0.684	0.2052	HOT7_AdminSupportHOT
22	7	3	3	3	22	14.6667	22	14.6667	2.4142	1.6289	0.3554	0.9277	1.1225	0.5358	HOT7_AdminSupportHOT
23	8	1	1	1	80	53.6913	80	53.6913	-0.9278	1.5663	0.1762	1.1903	1.1675	-0.5446	HOT8_TrainingHOT
24	8	2	2	2	48	32.2148	48	32.2148	0.4688	1.1207	0.1635	1.1451	0.969	0.169	HOT8_TrainingHOT
25	8	3	3	3	21	14.094	21	14.094	2.5543	1.4206	0.3177	0.8336	0.7512	0.5535	HOT8_TrainingHOT
26	9	1	1	1	99	66.443	99	66.443	-0.7292	1.5348	0.155	1.0011	0.994	-0.5615	HOT9_TrainingFA
27	9	2	2	2	36	24.1611	36	24.1611	0.9612	1.1184	0.189	1.0319	0.9153	0.2878	HOT9_TrainingFA
28	9	3	3	3	14	9.396	14	9.396	2.822	1.6845	0.4672	1.0251	1.4683	0.4864	HOT9_TrainingFA
29	10	1	1	1	83	55.3333	83	55.3333	-0.8238	1.6552	0.1828	1.317	1.3334	-0.5051	HOT10_InstructMaterialsHOT
30	10	2	2	2	55	36.6667	55	36.6667	0.7492	1.2704	0.1729	1.1541	1.0498	0.3002	HOT10_InstructMaterialsHOT
31	10	3	3	3	12	8	12	8	2.4851	1.7896	0.5396	1.2502	1.2409	0.3923	HOT10_InstructMaterialsHOT
32	11	1	1	1	18	12	18	12	-2.5808	1.4176	0.3438	0.7929	0.7956	-0.5174	HOT11_ProblemSolvingHOT
33	11	2	2	2	91	60.6667	91	60.6667	-0.2281	1.2123	0.1278	0.8452	0.8535	-0.1646	HOT11_ProblemSolvingHOT
34	11	3	3	3	41	27.3333	41	27.3333	1.7039	1.5798	0.2498	0.9523	0.9164	0.5577	HOT11_ProblemSolvingHOT
35	12	1	1	1	12	8	12	8	-2.9797	1.501	0.4526	0.7765	0.7696	-0.4766	HOT12_VaryLevelsQuestioning
36	12	2	2	2	74	49.3333	74	49.3333	-0.6401	1.1226	0.1314	0.784	0.7655	-0.35	HOT12_VaryLevelsQuestioning
37	12	3	3	3	64	42.6667	64	42.6667	1.3402	1.499	0.1889	0.9482	0.9308	0.6152	HOT12_VaryLevelsQuestioning

9.11 *Appendix E*. Chapter 4 (Phase I)- FA items- Measure order

TABLE 14.1 Teacher_Questionnaire_Sello_02.07.201 ZOU555WS. TXT May 7 2021 8:21
INPUT: 150 PERSON 48 ITEM REPORTED: 150 PERSON 48 ITEM 96 CATS WINSTEPS 4.7.1.0

PERSON: REAL SEP.: 3.07 REL.: .90 ... ITEM: REAL SEP.: 4.48 REL.: .95

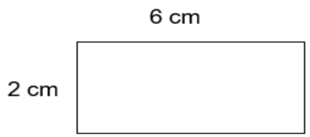
ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM	G
1	245	150	-.52	.20	1.30	3.35	1.84	3.38	.28	.47	63.0	72.6	LGSC_a	0
2	236	150	-.18	.19	1.05	.66	1.05	.33	.45	.47	70.5	70.9	LGSC_b	0
3	270	150	-1.63	.24	.96	-.22	.82	-.47	.47	.45	84.2	83.0	LGSC_c	0
4	217	150	.50	.19	.99	-.08	1.71	3.74	.46	.48	72.6	71.4	LGSC_d	0
5	212	150	.68	.19	1.03	.34	1.01	.10	.46	.47	69.9	72.2	LGSC_e	0
6	230	150	.03	.19	.97	-.44	.91	-.49	.50	.48	74.7	70.7	LGSC_f	0
7	261	150	-1.18	.21	1.09	.87	1.39	1.31	.39	.46	76.7	78.5	LGSC_g	0
8	240	150	-.33	.19	1.18	2.17	1.50	2.37	.36	.47	64.4	71.5	LGSC_h	0
9	239	150	-.29	.19	1.22	2.71	1.32	1.65	.35	.47	64.4	71.3	LGSC_i	0
10	268	150	-1.53	.23	1.31	2.27	1.43	1.26	.30	.45	75.3	81.9	LGSC_j	0
11	230	150	.03	.19	1.30	3.58	1.57	3.02	.30	.48	61.0	70.7	LGSC_k	0
12	208	150	.83	.19	1.02	.25	.91	-.46	.47	.47	71.2	73.1	LGSC_l	0
13	231	150	.00	.19	1.01	.13	1.00	.08	.47	.48	74.0	70.6	LGSC_m	0
14	206	150	.91	.20	.99	-.13	.94	-.31	.48	.47	74.0	73.6	LGSC_n	0
15	222	150	.32	.19	.88	-1.58	.81	-1.25	.54	.48	77.4	70.9	LGSC_o	0
16	223	150	.28	.19	.86	-1.84	.77	-1.54	.55	.48	74.7	70.9	LGSC_p	0
17	271	149	-1.80	.25	.90	-.69	.70	-.82	.51	.45	85.5	84.7	EEQ_a	0
18	268	149	-1.62	.24	.92	-.54	.69	-.90	.51	.45	84.1	82.9	EEQ_b	0
19	250	149	-.78	.20	.91	-.95	.80	-.82	.52	.47	78.6	74.8	EEQ_c	0
20	242	149	-.47	.20	.81	-2.51	.78	-1.11	.57	.47	80.0	72.4	EEQ_d	0
21	254	148	-1.03	.21	.94	-.62	.97	-.04	.49	.46	80.6	76.9	EEQ_e	0
22	171	148	2.50	.26	1.21	1.22	1.41	1.19	.32	.43	84.7	87.1	EEQ_f	0
23	234	149	-.17	.19	.89	-1.51	.79	-1.25	.54	.48	75.2	71.0	EEQ_g	0
24	201	149	1.06	.20	1.13	1.40	1.20	1.09	.40	.47	69.0	74.4	EEC_a	0
25	192	149	1.43	.21	1.01	.14	1.00	.07	.46	.46	74.5	77.0	EEC_b	0
26	224	149	.20	.19	.98	-.23	.93	-.38	.49	.48	71.7	70.9	EEC_c	0
27	206	149	.86	.20	1.06	.67	1.01	.11	.45	.47	67.6	73.3	EEC_d	0
28	231	148	-.10	.19	.99	-.04	.90	-.55	.49	.47	68.8	70.8	EEC_e	0
29	216	149	.49	.19	1.08	.96	1.01	.11	.44	.48	68.3	71.4	EEC_f	0
30	237	149	-.28	.19	1.09	1.08	1.04	.30	.43	.48	68.3	71.4	EEC_g	0
31	224	149	.20	.19	.82	-2.50	.73	-1.77	.58	.48	80.0	70.9	EEC_h	0
32	221	149	.31	.19	.83	-2.38	.75	-1.66	.57	.48	78.6	71.0	EEC_i	0
33	256	149	-1.03	.21	1.05	.57	1.21	.83	.41	.46	79.3	77.0	EEC_j	0
34	262	150	-1.23	.22	.94	-.55	.76	-.82	.50	.46	80.1	78.9	LT_a	0
35	254	150	-.88	.20	.81	-2.11	.68	-1.42	.57	.46	82.9	75.6	LT_b	0
36	225	150	.21	.19	1.07	.97	1.08	.55	.44	.48	65.1	70.8	LT_c	0
37	176	150	2.31	.25	.91	-.57	.80	-.59	.48	.43	86.3	85.5	LT_d	0
38	227	150	.14	.19	.91	-1.22	.86	-.87	.53	.48	73.3	70.7	LT_e	0
39	240	150	-.33	.19	.93	-.87	.87	-.67	.51	.47	74.0	71.5	LT_f	0
40	261	150	-1.18	.21	1.06	.58	.97	-.02	.44	.46	76.7	78.5	GPF_a	0
41	236	150	-.18	.19	1.25	3.05	1.42	2.18	.33	.47	62.3	70.9	GPF_b	0
42	250	149	-.78	.20	.99	-.03	.90	-.35	.48	.47	73.1	74.8	GPF_c	0
43	230	149	-.02	.19	.89	-1.52	.81	-1.18	.54	.48	74.5	70.7	GPF_d	0
44	215	150	.57	.19	.92	-.94	.85	-.96	.52	.48	72.6	71.7	GPF_e	0
45	191	150	1.53	.21	.96	-.37	.84	-.65	.49	.46	78.8	77.8	GPF_f	0
46	203	150	1.03	.20	.85	-1.72	.74	-1.49	.55	.47	79.5	74.3	GPF_g	0
47	198	150	1.23	.20	.92	-.88	.81	-.93	.52	.47	75.3	75.5	GPF_h	0
48	234	150	-.11	.19	.84	-2.24	.72	-1.77	.57	.47	75.3	70.7	GPF_i	0
MEAN	230.0	149.5	.00	.20	1.00	.0	1.00	.0			74.4	74.4		
P.SD	23.9	.6	.96	.02	.13	1.5	.28	1.3			6.3	4.5		

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

9.12 *Appendix F*. Chapter 5 (Phase II)- Mathematics HOTS items validation by reviewers

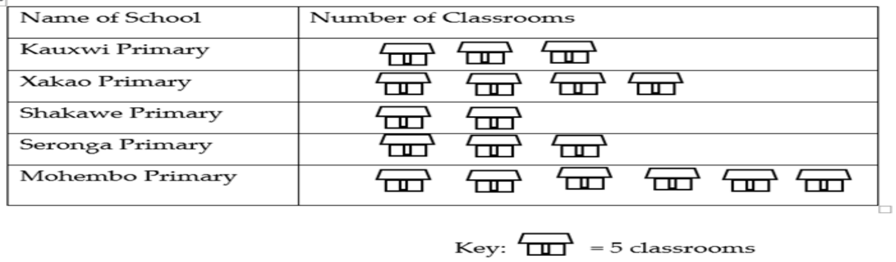
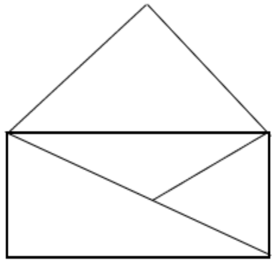
Item ID	Item description	Content	Processes of Comprehension/ Blooms	Level of agreement rate from 0 to 1.0			
				Rater 1	Rater 2	Rater 3	Mean
Q1 a &b	Use the list of numbers below to answer question 1 (a) and (b) 13, 15, 17, 20, 23, 30 (a) Which of the numbers are odd? Answer……………………………. (2) (b) Which of the numbers are multiples of 5? Answer……………..……………….. (2)	Number and operations	Synthesis	1.00	0.80	1.00	0.93
Q2	The statement given below shows a pattern. Use it to answer question 2. 14 1 + 4 = 5 25 2 + 5 = 7 2. What do the letters X and Y represent if 36 X + 6 = Y? Answer X =……………………………… (2) Answer Y = ………...…………………… (2)	Number and operations	application	0.30	0.40	0.30	0.33

Q3	**3.** What is the distance around the rectangle shown below  ***Working:*** Answer........................ cm (2)	Measure	Application	0.80	0.70	1.00	0.83
Q4	**4.** A mother received 65 kg of sorghum from her son. She also received 86 kg of sorghum from her daughter. How many kg of sorghum did she receive altogether? ***Working:*** *Answer* .... (2)	Measure	Application	1.00	0.80	0.80	0.87

Q5	**5.** A builder had 812 bags of cement. He used 271 bags to build a house. How many bags are left? ***Working:*** *Answer*............................. (2)	Number and operations	Application	0.60	0.70	0.80	0.70
Q6	**6.** A litre of petrol costs P5.40. Moreri buys 3 litres of petrol. How much money does Moreri have to pay? *Working:* *Answer P*........................... (2)	Number and operations	Application	0.60	0.50	0.75	0.62

Q7	7. Lolo ate $\frac{3}{9}$ of her cake in the morning. She later ate a piece that make $\frac{5}{9}$ of the whole cake What is the total fraction eaten? ***Working:*** *Answer*.............................. (2)	Number and operations	Application	0.20	0.30	0.50	0.33
Q8	8. A father divides P15 among his 2 children equally. How much does each child get? ***Working:*** *Answer*...................... (2)	Number and operations	Synthesis	0.70	0.80	0.70	0.73

<table>
<tr>
<td>Q9</td>
<td>
The information below shows part of a bill of items.
Use it to answer question 10 (a) and (b)
<table>
<tr><th>Item</th><th>Number of items</th><th>Cost per item</th><th>Total cost</th></tr>
<tr><td>Pencil</td><td>2</td><td>P3.50</td><td>...</td></tr>
<tr><td>Pen</td><td>3</td><td>P4.50</td><td>P13.50</td></tr>
<tr><td>Ruler</td><td>1</td><td>P7.75</td><td>P7.75</td></tr>
<tr><td colspan="3">Total</td><td>...</td></tr>
</table>
er
:ions
9.
(a) What is the total cost of 2 pencils?
Working Answer P......................... (2)
(b) What is the total cost of all the items?
Working: Answer P...................... (2)
</td>
<td>Synthesis</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
<td>0.40</td>
</tr>
</table>

Q10	**Use the information below to answer question 10**  **10.** (a) Which two schools have the same number of the classroom? *Answer*...................... (1) (b) How many classrooms are in Mohembo Primary School? ***Working:*** *Answer*......................... (2) (c) One primary school in Gaborone has 45 classrooms. How many pictures will be used to show this information? ***Working:*** *Answer*.......................... (2)	statistics	Analysis	0.65	0.60	0.65	0.63
Q11	**11.** How many triangles are in the shape below?  *Answer*.................... (1)	Problem Solving	Application	0.50	0.60	0.70	0.60

<table>
<tr>
<td>Q12</td>
<td>
12. The time table below shows the journey of a bus from Lobatse to Gaborone.
Use it to answer question 12 (a) to (c)
<table>
<tr><th></th><th>Arrival Time</th><th>Department Time</th></tr>
<tr><td>Lobatse</td><td></td><td>6:30 am</td></tr>
<tr><td>Boatle</td><td>7: 00 am</td><td>7:15 am</td></tr>
<tr><td>Otse</td><td>8:15 am</td><td>8:30 am</td></tr>
<tr><td>Gaborone</td><td>10:00 am</td><td></td></tr>
</table>
(a) How many hours does the bus take to travel from Lobatse to Gaborone?
Working Answer .... (1)
(b) At what time does the bus leave Boatle?
Working: Answer .... (2)
(c) What is the total stopping time between Boatle and Otse?
Working Answer..................... (2)
</td>
<td>Measure</td>
<td>Application</td>
<td>0.30</td>
<td>0.30</td>
<td>0.45</td>
<td>0.35</td>
</tr>
</table>

Q13	**Use the pictures below to answer question 13 (a) and (b).**  (a) Sello uses a P20 note to pay for a packet of tea. How much chance does she get? *Working* *Answer* P…………………….. (1) (b) Goitse paid a total of P56.90 for two food items. Which **two** items did he buy? *Working:* *Answer* ………………….. (2) (c) Tumi buys one box of Jungle Oats She pays using pula and thebe coins only. How many pula and thebe coins does she use? *Working:* *Answer* ………………….. (2)	Number and operations	Application	0.40	0.35	0.35	0.37
	Total						7.70
	Average						0.59

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

9.13 *Appendix Gi*. Chapter 5 (Phase II)- Mathematics HOTS items for pre-assessment

MATHEMATICS	MAY 2019
Marks: 40	Time: 75 Minutes

PRE-ASSESSMENT TEST

SURNAME: ____________________ **FIRST NAME:** ______________________

Sex: Male ________ **Female** _________ **Date of Birth:** ____________________

School: ____

Sub-Region: ____

INSTRUCTIONS

1. Read all questions carefully before you answer.
2. Answer all questions in the spaces provided.
3. Write in pencil so that you can erase easily if you need to.
4. Where necessary, show your working

THIS QUESTION PAPER CONTAINS 7 PAGES.

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

Use the list of numbers below to answer question 1 (a) and (b)

13, 15, 17, 20, 23, 30

1.

(a) Which of the numbers are odd?

Answer……………………………. (2)

(b) Which of the numbers are multiples of 5?

Answer…………….……………….. (2)

The statement given below shows a pattern. Use it to answer question 2.

14 ⟶ 1 + 4 = 5

25 ⟶ 2 + 5 = 7

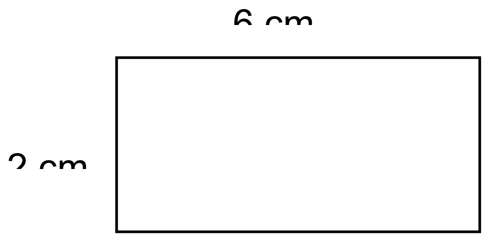
2. What do the letters **X** and **Y** represent if 36 ⟶ **X** + 6 = **Y**?

Answer **X** =………………………………. (2)

Answer **Y** = ………...……………………. (2)

---

3. What is the distance around the rectangle shown below



Working:

Answer…………………… cm (2)

---

4. A mother received 65 kg of sorghum from her son.

She also received 86 kg of sorghum from her daughter.

How many kg of sorghum did she receive altogether?

Working:

Answer ………………………… (2)

---

5. A builder had 812 bags of cement.

He used 271 bags to build a house.

How many bags are left?

Working:

Answer………………………… (2)

---

6. A litre of petrol costs P5.40. Moreri buys 3 litres of petrol. How much money does Moreri have to pay?

Working:

Answer P……...…………………… (2)

---

7. Lolo ate $\frac{3}{9}$ of her cake in the morning. She later ate a piece that makes $\frac{5}{9}$ of the whole cake

What is the total fraction eaten?

Working:

Answer……………………….. (2)

---

8. A father divides P15 among his 2 children equally.

How much does each child get?

Working:

Answer……………………… (2)

The information below shows part of a bill of items.

Use it to answer question 10 (a) and (b)

Item	Number of items	Cost per item	Total cost
Pencil	2	P3.50	…
Pen	3	P4.50	P13.50
Ruler	1	P7.75	P7.75
Total			…

9.

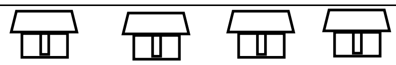
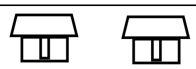
(a) What is the total cost of 2 pencils?

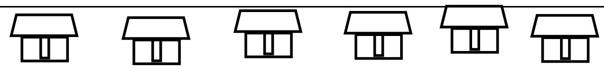
Working *Answer* P…………………… (2)

(b) What is the total cost of all the items?

Working: *Answer* P………………… (2)

Use the information below to answer question 10

Name of School	Number of Classrooms
Kauxwi Primary	
Xakao Primary	
Shakawe Primary	

Seronga Primary	
Mohembo Primary	

Key:  = 5 classrooms

10.

(a) Which two schools have the same number in the classroom?

Answer…………… (1)

(b) How many classrooms are in Mohembo Primary School?

Working:

Answer…………………… (2)

(c) One primary school in Gaborone has 45 classrooms.

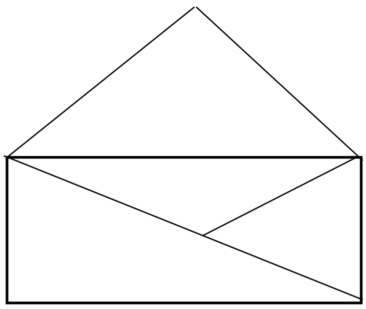
How many pictures will be used to show this information?

Working:

Answer……………………… (2)

---

11. How many triangles are in the shape below?



Answer………………… (1)

---

12. The timetable below shows the journey of a bus from Lobatse to Gaborone.

Use it to answer question 12 (a) to (c)

	Arrival Time	Department Time
Lobatse		6:30 am
Boatle	7: 00 am	7:15 am
Otse	8:15 am	8:30 am
Gaborone	10:00 am	

(a) How many hours does the bus take to travel from Lobatse to Gaborone?

Working *Answer* ………………… (1)

(b) At what time does the bus leave Boatle?

Working: *Answer* ………………… (2)

(c) What is the total stopping time between Boatle and Otse?

Working *Answer*………………… (2)

Use the pictures below to answer question 13 (a) and (b).

(a) Sello uses a P20 note to pay for a packet of tea.

How much chance does she get?

Working *Answer* P……………………. (1)

(b) Goitse paid a total of P56.90 for two food items.

Which **two** items did he buy?

Working: *Answer* …………………. (2)

(c) Tumi buys one box of Jungle Oats

She pays using pula and thebe coins only.

How many pula and thebe coins does she use?

Working:

Answer …………………. (2)

"End"

9.14 *Appendix Gii*. Chapter 5 (Phase II)- Mathematics HOTS items for post-assessment

MATHEMATICS	October 2019
Marks: 40	Time: 75 Minutes

POST-ASSESSMENT TEST

SURNAME: ____________________ **FIRST NAME:** ____________________

Sex: Male ________ **Female** _________ **Date of Birth:** ______________

School: ____

Sub-Region: ____

INSTRUCTIONS

5. Read all questions carefully before you answer.
6. Answer all questions in the spaces provided.
7. Write in pencil so that you can erase easily if you need to.
8. Where necessary, show your working

THIS QUESTION PAPER CONTAINS 7 PAGES.

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO

Use the list of numbers below to answer question 1 (a) and (b)

13, 15, 17, 20, 23, 30

1.

(c) Which of the numbers are even?

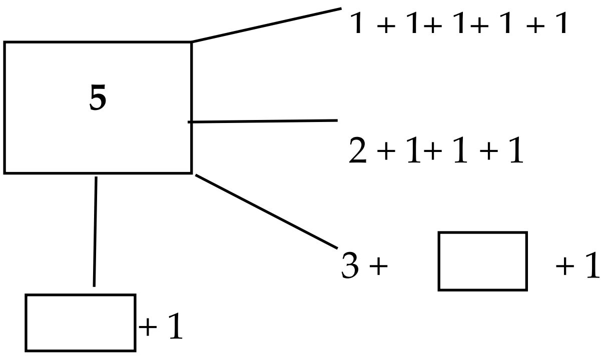
Answer……………………………. (2)

(d) Which of the numbers are multiples of 5?

Answer……….….………………. (2)

---

2. Complete the pattern below by filling in the missing in the boxes.

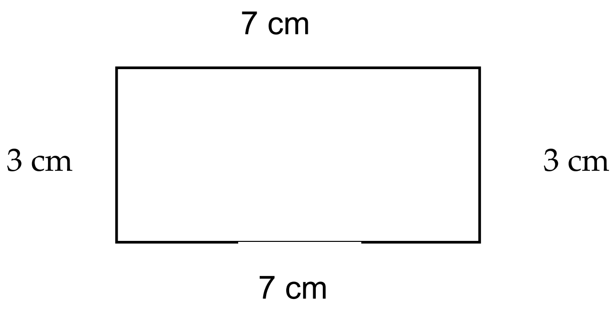


(2)

(2)

---

3. What is the area of the rectangle shown below



Working:

Answer…………………… cm (2)

---

4. There are 235 parents and 172 children in a church.

How many people are there altogether in the church?

Working:

Answer ………………………… (2)

---

5. A farmer had 539 bags of sorghum. He sold 478 bags. How many bags were left?

Working:

Answer………………………… (2)

---

6. A litre of petrol costs P5.40. Moreri buys 3 litres of petrol.

How much money does Moreri have to pay?

Working:

Answer P……..…………………… (2)

---

7. Lolo ate $\frac{3}{4}$ of bread loaf in the morning. She later ate a piece that makes $\frac{1}{4}$ of the whole bread

What is the total fraction eaten?

Working:

Answer………………………….. (2)

---

8. A father divides P15 among his three children equally.

How much does each child get?

Working:

Answer…………………. (2)

The information below shows part of a bill of items.

Use it to answer question 9 (a) and (b)

Item	Number of items	Cost per item	Total cost
Pencil	2	P3.50	…
Pen	3	P4.50	P13.50
Ruler	1	P7.75	P7.75
Total			…

9.

(c) What is the total cost of 2 pencils?

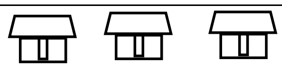
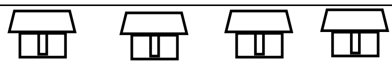
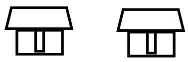
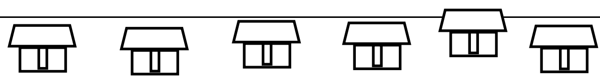
Working

Answer P…………………….. (2)

(d) What is the total cost of all the items?

Working: *Answer* P………………….. (2)

Use the information below to answer question 10

Name of School	Number of Classrooms
Kauxwi Primary	
Xakao Primary	
Shakawe Primary	
Seronga Primary	
Mohembo Primary	

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Key:  = 5 classrooms

10.

(d) Which two schools have the same number in the classroom?

Answer…………………. (1)

(e) How many classrooms are in Mohembo Primary School?

Working: *Answer*……………………. (2)

(f) One primary school in Gaborone has 45 classrooms.

How many pictures will be used to show this information?

Working *Answer*……………………… (2)

Write <u>True</u> if the statement is correct and <u>False</u> if the statement is not correct from 11a and 11b

11.

a. Mpho bought 8 sweets and Neo bought 5 sweets. Together they have bought 13 sweets

Answer………………… (1)

b. Mr Pule had 15 cattle, a lion ate 8 of them, and he was left with 6 cattle.

Answer………………… (1)

12.

(a) Three children visited their grandmother. They arrived at the following times:
They arrived at the following times:

Kabo: 5 am

Tebatso: 2 pm

Lebanang: 9 am

Who arrived at first?

Answer………………… (2)

(b) If Tebatso left her grandmother 's place at 2.30 pm and arrived home at 4.30 pm

How many hours does Tebatso take to travel from the grandmother's place to her home

Working: *Answer* ………………… (2)

Use the pictures below to answer question 13 (a) and (b).

(d) Sello uses a P20 note to pay for a packet of tea.

How much change does she get?

Working: *Answer* P……………………. (1)

(e) Goitse paid a total of P56.90 for two food items. Which **two** items did he buy?

Working: *Answer* P ………………... (2)

(f) Tempo buys one box of instant porridge at P12.95

He pays using **pula** and **thebe** coins only.

How many pula and thebe coins does he use?

Working:

Answer P………………. (2)

9.15 *Appendix Hi*. Chapter 5 (Phase II)- Pre-test Maths category functioning before rescoring

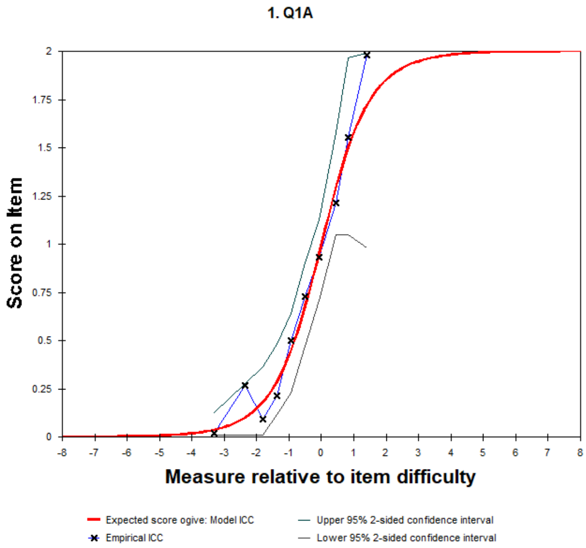
1	ITEM	CODE	VALUE	SCORE	UNWTD	UNWTD %	WTD	WTD %	AVGE MEAS	P.SD MEAS	S.E. MEAS	INFIT MNSQ	IT MNS	PTMA	LABEL
2	1	0	0	0	160	56.14	160	56.1404	-2.2007	1.0321	0.0819	0.9661	0.9632	-0.501	Q1A
3	1	1	1	1	68	23.86	68	23.8596	-1.2694	0.7595	0.0928	0.8733	1.1356	0.2444	Q1A
4	1	2	2	2	57	20	57	20	-0.9669	0.6862	0.0917	1.1269	1.1612	0.3611	Q1A
5	1	3	3	-1	0	0	0	0	0	0	0	0	0	0	Q1A
6	1	4	4	-1	0	0	0	0	0	0	0	0	0	0	Q1A
7	2	0	0	0	150	52.632	150	52.6316	-2.3018	1.0135	0.083	0.9549	0.914	-0.567	Q1B
8	2	1	1	1	35	12.281	35	12.2807	-1.5498	0.647	0.111	0.8637	0.9178	0.0643	Q1B
9	2	2	2	2	100	35.088	100	35.0877	-0.9403	0.619	0.0622	0.9372	0.8271	0.5495	Q1B
10	2	3	3	-1	0	0	0	0	0	0	0	0	0	0	Q1B
11	2	4	4	-1	0	0	0	0	0	0	0	0	0	0	Q1B
12	3	0	0	0	259	90.877	259	90.8772	-1.8291	1.0428	0.0649	1.4536	1.9228	-0.29	Q2
13	3	1	1	1	2	0.7018	2	0.7018	-1.4823	0.4504	0.4504	4.0448	0.9081	0.0198	Q2
14	3	2	2	2	6	2.1053	6	2.1053	-1.3747	0.4972	0.2224	5.7204	3.0795	0.0494	Q2
15	3	3	3	3	4	1.4035	4	1.4035	-0.5133	0.5489	0.3169	0.7728	1.291	0.1373	Q2
16	3	4	4	4	14	4.9123	14	4.9123	-0.4672	0.4769	0.1323	1.045	4.0929	0.2714	Q2
17	4	0	0	0	247	86.667	247	86.6667	-1.8512	1.0515	0.067	1.0927	1.1211	-0.288	Q3
18	4	1	1	1	6	2.1053	6	2.1053	-1.7301	0.5717	0.2557	3.435	2.2716	0.0002	Q3
19	4	2	2	2	32	11.228	32	11.2281	-0.8097	0.656	0.1178	1.1597	1.5073	0.3097	Q3
20	4	3	3	-1	0	0	0	0	0	0	0	0	0	0	Q3
21	4	4	4	-1	0	0	0	0	0	0	0	0	0	0	Q3
22	5	0	0	0	82	28.772	82	28.7719	-2.7571	1.0516	0.1168	0.9828	1.1415	-0.615	Q4
23	5	1	1	1	43	15.088	43	15.0877	-1.6126	0.8313	0.1283	1.3907	1.4035	0.0474	Q4
24	5	2	2	2	160	56.14	160	56.1404	-1.2383	0.6773	0.0537	0.9677	0.9507	0.5272	Q4
25	5	3	3	-1	0	0	0	0	0	0	0	0	0	0	Q4
26	5	4	4	-1	0	0	0	0	0	0	0	0	0	0	Q4
27	6	0	0	0	132	46.316	132	46.3158	-2.4518	0.9841	0.086	0.8678	0.8162	-0.632	Q5
28	6	1	1	1	40	14.035	40	14.0351	-1.4721	0.6803	0.1089	0.908	0.8971	0.0991	Q5
29	6	2	2	2	113	39.649	113	39.6491	-0.9825	0.5848	0.0553	0.8417	0.768	0.5735	Q5
30	6	3	3	-1	0	0	0	0	0	0	0	0	0	0	Q5
31	6	4	4	-1	0	0	0	0	0	0	0	0	0	0	Q5
32	7	0	0	0	192	67.368	192	67.3684	-2.1185	1.0059	0.0728	0.9275	0.8994	-0.525	Q6
33	7	1	1	1	26	9.1228	26	9.1228	-1.1445	0.7111	0.1422	0.8325	0.9705	0.1757	Q6
34	7	2	2	2	67	23.509	67	23.5088	-0.8512	0.5797	0.0714	0.9725	0.834	0.461	Q6
35	7	3	3	-1	0	0	0	0	0	0	0	0	0	0	Q6
36	7	4	4	-1	0	0	0	0	0	0	0	0	0	0	Q6
37	8	0	0	0	118	41.404	118	41.4035	-2.4066	1.0951	0.1012	1.2652	1.4962	-0.536	Q7
38	8	1	1	1	8	2.807	8	2.807	-1.6119	0.4967	0.1877	0.5765	0.463	0.0192	Q7
39	8	2	2	2	159	55.79	159	55.7895	-1.2369	0.7238	0.0576	1.2464	1.2247	0.5249	Q7
40	8	3	3	-1	0	0	0	0	0	0	0	0	0	0	Q7
41	8	4	4	-1	0	0	0	0	0	0	0	0	0	0	Q7
42	9	0	0	0	231	81.053	231	81.0526	-1.9676	1.0103	0.0666	0.9026	0.8878	-0.461	Q8
43	9	1	1	1	12	4.2105	12	4.2105	-0.7605	0.3967	0.1196	0.3556	0.1392	0.1923	Q8
44	9	2	2	2	42	14.737	42	14.7368	-0.7118	0.578	0.0903	0.9864	0.7969	0.4004	Q8
45	9	3	3	-1	0	0	0	0	0	0	0	0	0	0	Q8
46	9	4	4	-1	0	0	0	0	0	0	0	0	0	0	Q8
47	10	0	0	0	207	72.632	207	72.6316	-2.079	0.9971	0.0695	0.9073	0.8798	-0.534	Q9A
48	10	1	1	1	12	4.2105	12	4.2105	-1.0553	0.4003	0.1207	0.45	0.1903	0.1339	Q9A

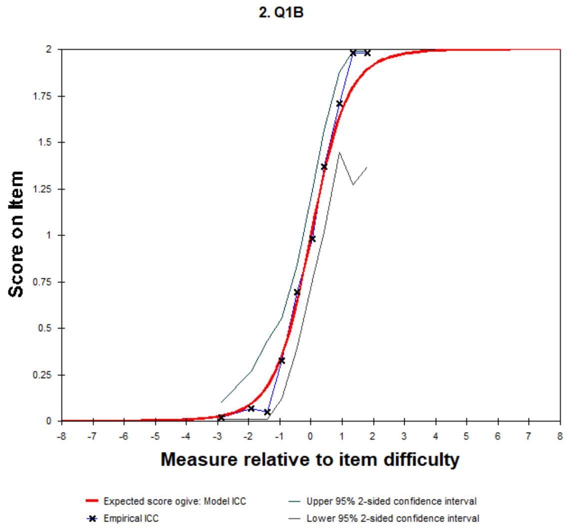
9.16 *Appendix Hii*. Chapter 5 (Phase II)-Pre-test Maths Items Maps with Thurstonian thresholds before rescoring

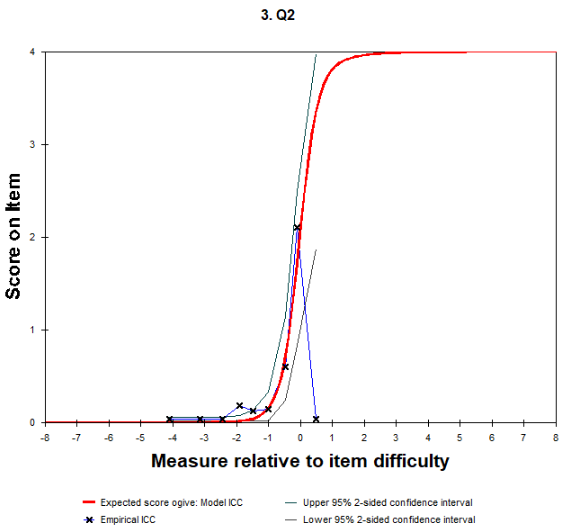
```
INPUT: 285 PERSON  21 ITEM REPORTED: 285 PERSON  21 ITEM  61 CATS WINSTEPS 4.7.1.0
--------------------------------------------------------------------------------


MEASURE     PERSON - MAP - ITEM - 50% Cumulative probabilities (Rasch-Thurstonian
thresholds)
                <more>|<rare> 1        2        3        4
    4                 +
                      |
                      |
                      |
                      |
                      |
                      |   Q12A.1
                      |
    3                 +
                      |
                      |T
                      |
                      |
                      |
                      |
                      |
    2                 +    Q11.1
                      |
                      |
                      |   Q13A.1
                      |   Q12C.1   Q12C.2
                      |S
                      |   Q13B.1   Q13B.2
    1                 +
                      |
                      |   Q13C.1   Q13C.2
                      |   Q10C.1   Q10C.2
                   .  |
                   .  |
                      |
                   . T|   Q10B.1    Q9B.2
                                   Q10B.2
    0              #  +M   Q9B.1                      Q2.4
                 ###  |   Q12B.1     Q3.2     Q2.3
                                   Q12B.2
                 ###  |     Q3.1     Q2.2
                            Q2.1     Q8.2
             .######  |             Q1A.2
```

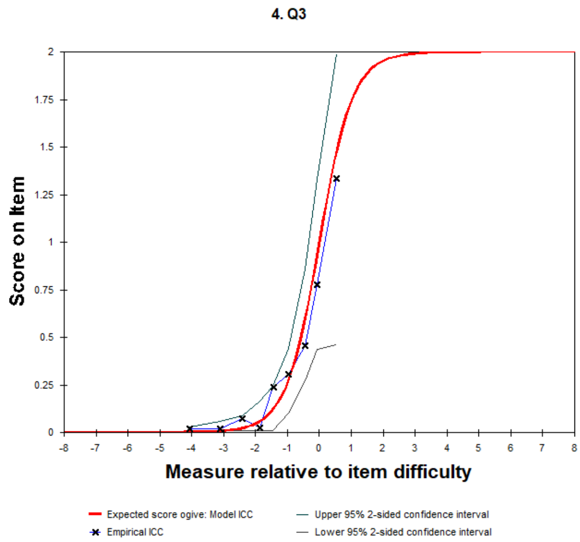
```
                 ##  |     Q8.1
            .####### S|             Q6.2
                                   Q9A.2
                .###  |   Q10A.1
                           Q9A.1
               #####  |
   -1       ########  +     Q6.1
          ##########  |             Q1B.2
            ########  |S            Q5.2
           #########  |    Q1A.1
             ####### M|    Q1B.1
                      |
         ###########  |     Q5.1     Q7.2
          ##########  |     Q7.1     Q4.2
   -2                 +
           .########  |
                     S|
          ##########  |     Q4.1
                      |
         .##########  |T
                      |
   -3                 +
                      |
            .####### T|
                      |
                      |
                      |
                      |
                      |
   -4     ##########  +
                <less>|<freq> 0        1        2        3
EACH "#" IS 2: EACH "." IS 1
```

9.17 *Appendix Hiii* - Chapter 5 (Phase II)-FA- Multiple ICCS before Rescoring

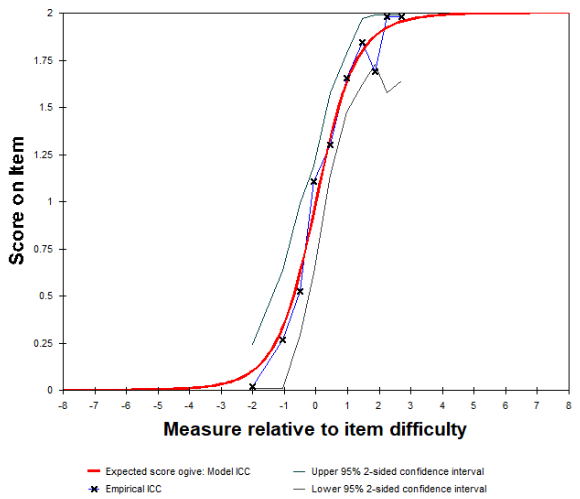




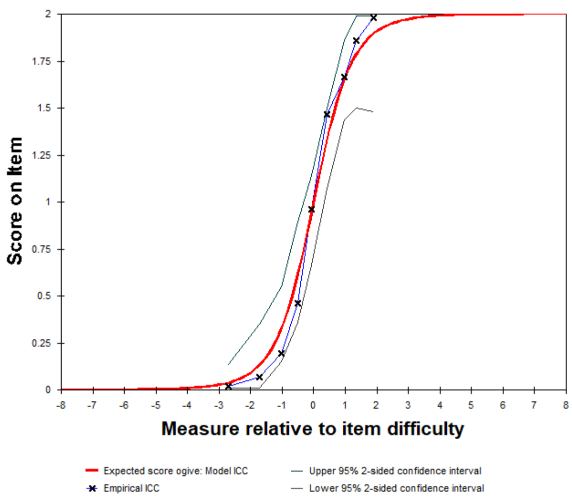




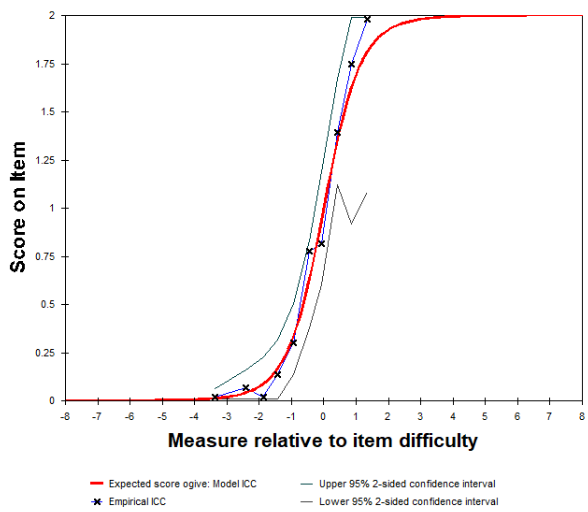
5. Q4



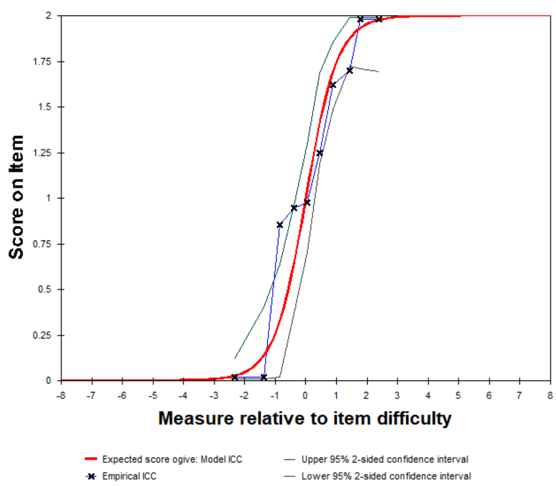
6. Q5



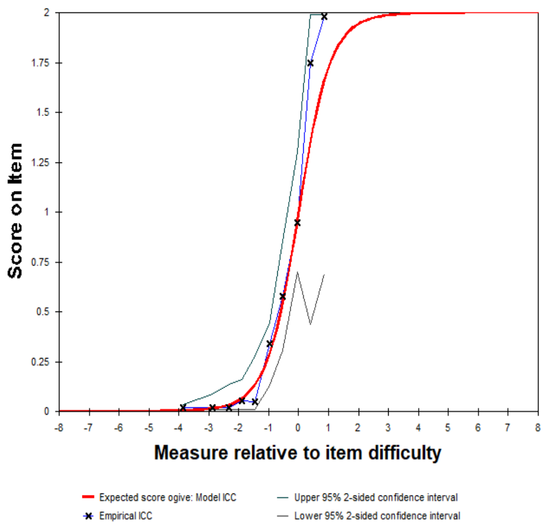
7. Q6



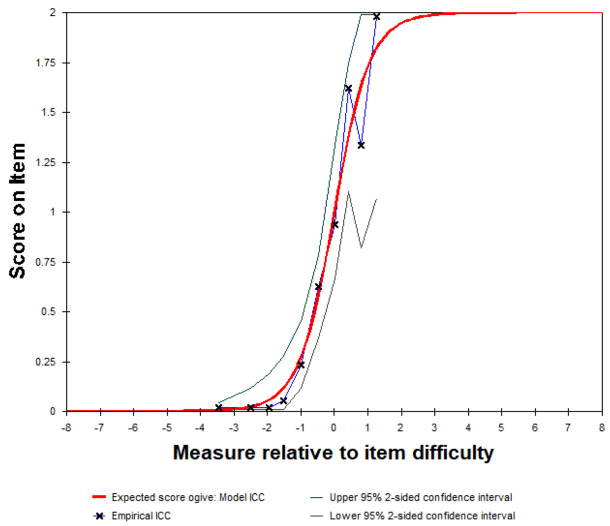
8. Q7



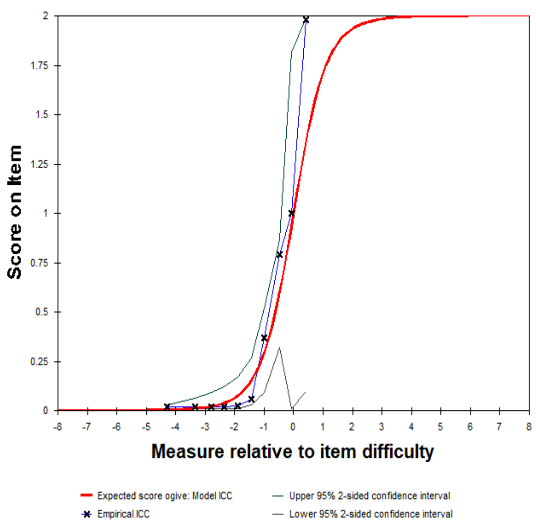
9. Q8



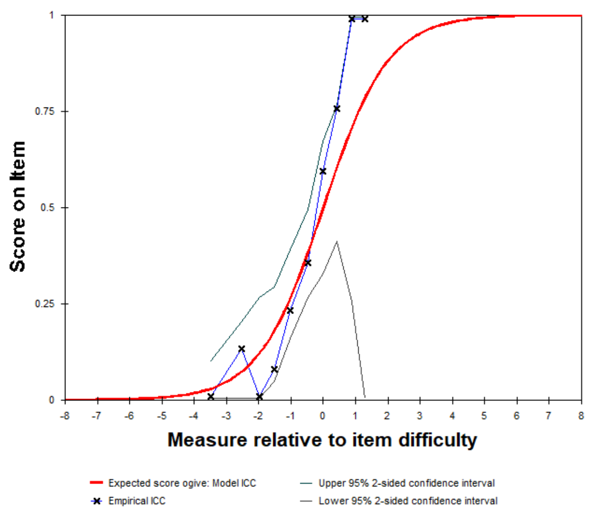
10. Q9A

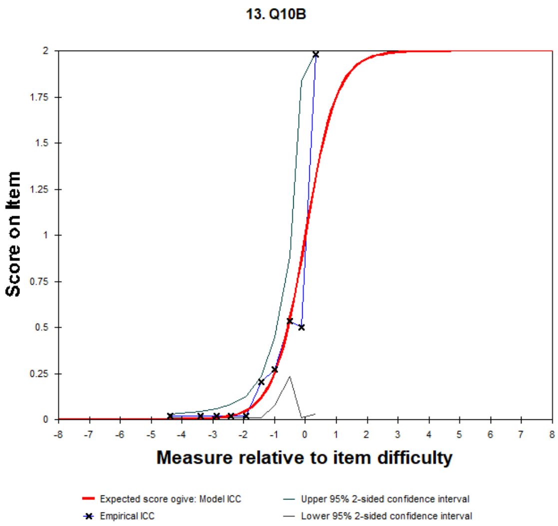


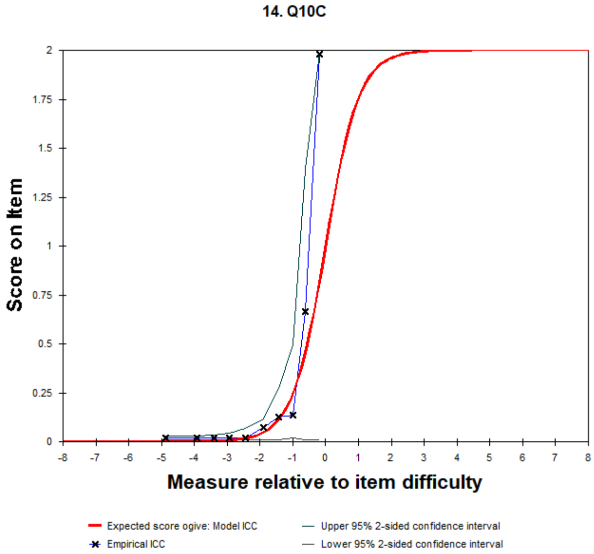
11. Q9B

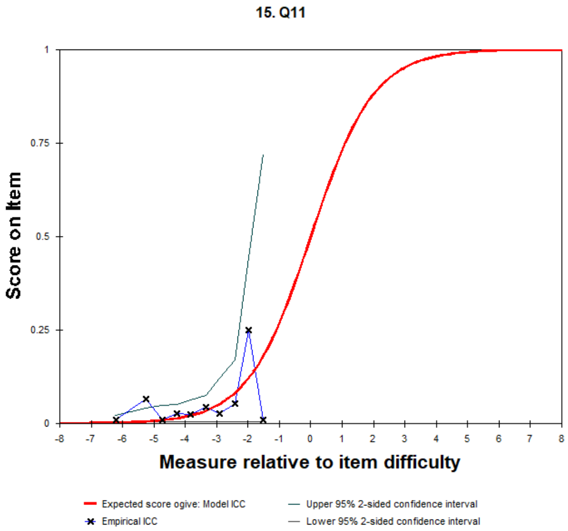


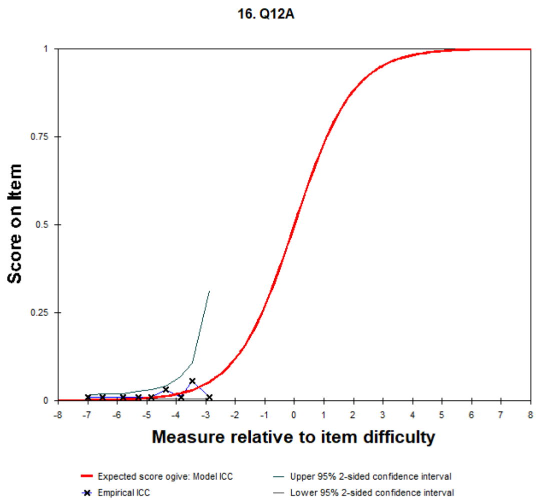
12. Q10A



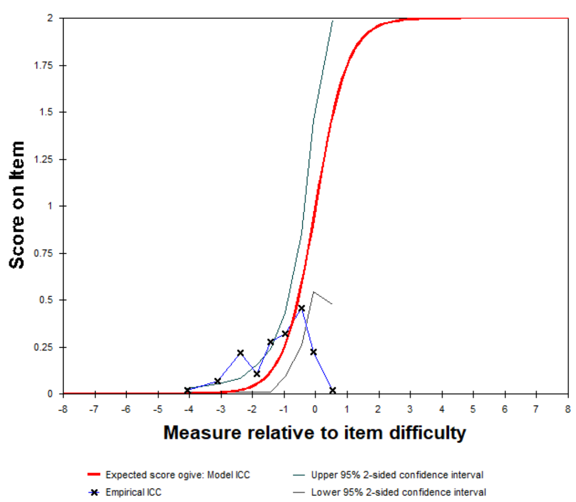

13. Q10B
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval


14. Q10C
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval

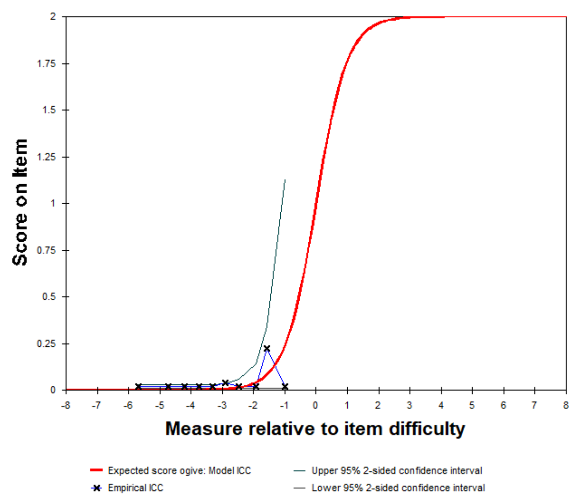

15. Q11
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval


16. Q12A
Score on item
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval

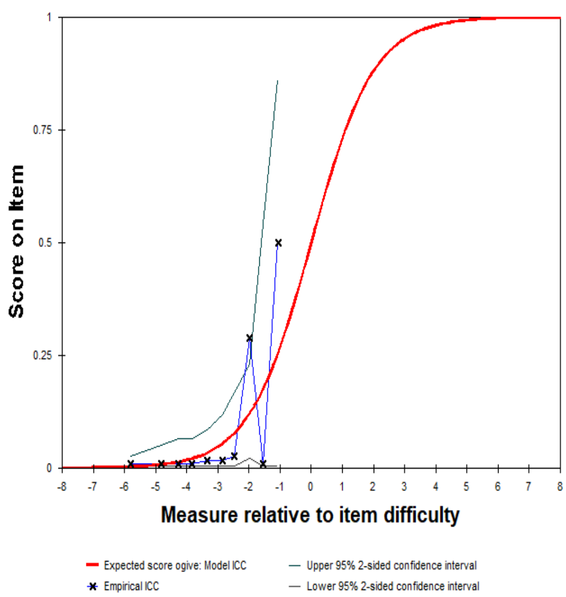
17. Q12B



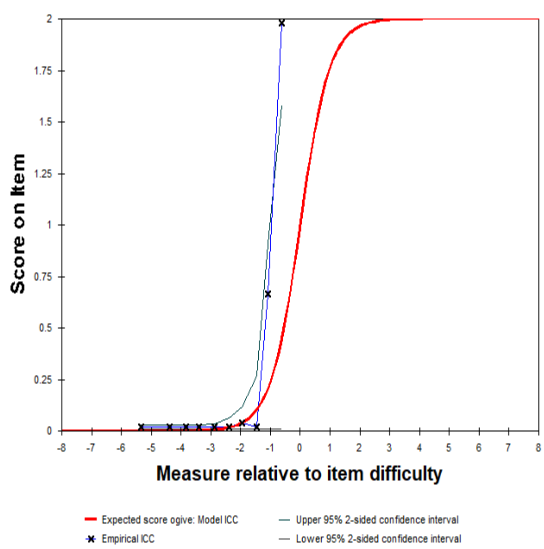
18. Q12C

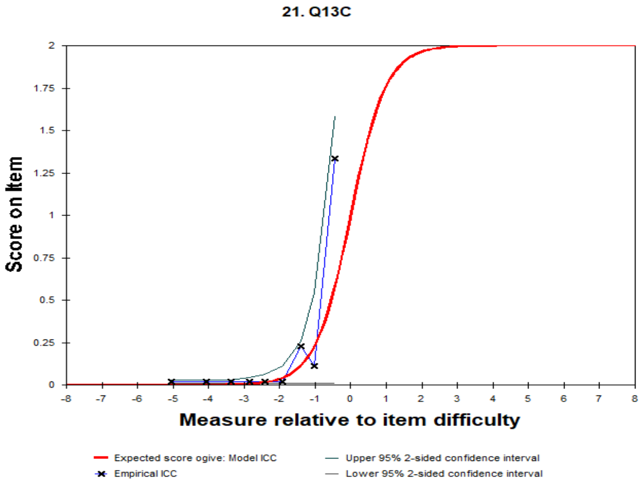


19. Q13A



20. Q13B




21. Q13C
Score on Item
2
1.75
1.5
1.25
1
0.75
0.5
0.25
0
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
Measure relative to item difficulty
Expected score ogive: Model ICC
Empirical ICC
Upper 95% 2-sided confidence interval
Lower 95% 2-sided confidence interval

9.18 *Appendix Hiv*. Chapter 5 (Phase II)- Pre-test Maths category functioning after rescoring

ITEM	CODE	VALUE	SCORE	UNWTD	UNWTD %	WTD	WTD %	AVGE MEAS	P.SD MEAS	S.E. MEAS	INFIT MNSQ	OUTFIT MNSQ	PTMA	LABEL
1	0	0	0	294	51.9435	294	51.944	-2.0549	0.9339	0.0546	0.8839	0.867	-0.571	Q1A
1	1	1	1	134	23.6749	134	23.675	-1.0293	0.6805	0.059	0.8207	0.928	0.257	Q1A
1	2	2	2	138	24.3816	138	24.382	-0.7626	0.7127	0.0609	1.0702	1.071	0.411	Q1A
2	0	0	0	242	42.7562	242	42.756	-2.2215	0.9133	0.0588	0.8943	0.789	-0.617	Q1B
2	1	1	1	60	10.6007	60	10.601	-1.4132	0.6229	0.0811	0.9106	0.791	0.028	Q1B
2	2	2	2	264	46.6431	264	46.643	-0.8519	0.6779	0.0418	0.936	0.888	0.594	Q1B
3	0	0	-1	473	100	473	100	-1.6329	0.9752	0.0449	0	0	0	RQ2
3	1	1	-1	0	0	0	0	0	0	0	0	0	0	RQ2
3	2	2	-1	0	0	0	0	0	0	0	0	0	0	RQ2
4	0	0	0	453	82.8154	453	82.815	-1.6878	0.9664	0.0455	1.0452	1.066	-0.42	RQ3
4	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ3
4	2	2	2	94	17.1846	94	17.185	-0.5549	0.6834	0.0709	1.0301	1.237	0.42	RQ3
5	0	0	0	243	43.0089	243	43.009	-2.1253	1.042	0.067	1.3008	1.923	-0.542	RQ4
5	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ4
5	2	2	2	322	56.9911	322	56.991	-1.0164	0.6717	0.0375	1.0031	0.992	0.542	RQ4
6	0	0	0	331	58.5841	331	58.584	-1.952	0.9788	0.0539	1.1632	1.429	-0.539	RQ5
6	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ5
6	2	2	2	234	41.4159	234	41.416	-0.8443	0.6336	0.0415	0.9902	0.9	0.539	RQ5
7	0	0	0	421	74.5133	421	74.513	-1.8109	0.9247	0.0451	0.9121	0.861	-0.527	RQ6
7	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ6
7	2	2	2	144	25.4867	144	25.487	-0.582	0.6541	0.0547	0.9126	0.927	0.527	RQ6
8	0	0	0	249	44.306	249	44.306	-2.073	1.0279	0.0653	1.3114	1.612	-0.508	RQ7
8	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ7
8	2	2	2	313	55.694	313	55.694	-1.0356	0.729	0.0413	1.2099	1.244	0.508	RQ7
9	0	0	0	466	82.4779	466	82.478	-1.7128	0.9476	0.0439	0.9335	0.894	-0.461	RQ8
9	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ8
9	2	2	2	99	17.5221	99	17.522	-0.4792	0.644	0.0651	0.9757	0.723	0.461	RQ8
10	0	0	0	432	76.8683	432	76.868	-1.7939	0.9035	0.0435	0.7965	0.753	-0.548	RQ9A
10	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ9A
10	2	2	2	130	23.1317	130	23.132	-0.4792	0.618	0.0544	0.8128	0.687	0.548	RQ9A
11	0	0	0	508	90.7143	508	90.714	-1.6241	0.9379	0.0417	0.8237	0.838	-0.431	RQ9B
11	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ9B
11	2	2	2	52	9.2857	52	9.2857	-0.1276	0.5485	0.0768	0.685	0.389	0.431	RQ9B
12	0	0	0	351	62.5668	351	62.567	-1.8968	0.9448	0.0505	0.8765	0.86	-0.516	Q10A
12	1	1	1	210	37.4332	210	37.433	-0.8213	0.7116	0.0492	0.8743	0.858	0.516	Q10A

12	2	2	-1	0	0	0	0	0	0	0	0	0	0	Q10A
13	0	0	0	481	85.4352	481	85.435	-1.6708	0.947	0.0432	1.0117	0.974	-0.444	RQ10B
13	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ10B
13	2	2	2	82	14.5648	82	14.565	-0.4035	0.5789	0.0643	0.8926	0.587	0.444	RQ10B
14	0	0	0	503	89.6613	503	89.661	-1.6203	0.9587	0.0428	0.993	0.959	-0.39	RQ10C
14	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ10C
14	2	2	2	58	10.3387	58	10.339	-0.3292	0.6125	0.0811	0.9436	0.629	0.39	RQ10C
15	0	0	0	505	90.3399	505	90.34	-1.5658	0.9901	0.0441	0.9885	0.991	-0.24	Q11
15	1	1	1	54	9.6601	54	9.6601	-0.7445	0.8816	0.1211	0.9792	1.168	0.24	Q11
15	2	2	-1	0	0	0	0	0	0	0	0	0	0	Q11
16	0	0	0	550	98.9209	550	98.921	-1.4904	0.9886	0.0422	0.8619	0.973	-0.147	Q12A
16	1	1	1	6	1.0791	6	1.0791	-0.0645	1.0185	0.4555	0.5038	0.607	0.147	Q12A
16	2	2	-1	0	0	0	0	0	0	0	0	0	0	Q12A
17	0	0	0	478	86.1261	478	86.126	-1.5754	0.9994	0.0458	1.2928	1.367	-0.222	RQ12B
17	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ12B
17	2	2	2	77	13.8739	77	13.874	-0.9336	0.7871	0.0903	1.9116	2.687	0.222	RQ12B
18	0	0	0	548	99.4555	548	99.456	-1.4878	0.9985	0.0427	1.3509	1.115	-0.028	RQ12C
18	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ12C
18	2	2	2	3	0.5445	3	0.5445	-1.113	0.6407	0.453	9.99	4.426	0.028	RQ12C
19	0	0	0	492	88.3303	492	88.33	-1.6245	0.9593	0.0433	0.899	0.925	-0.359	Q13A
19	1	1	1	65	11.6697	65	11.67	-0.4968	0.7807	0.0976	0.778	0.799	0.359	Q13A
19	2	2	-1	0	0	0	0	0	0	0	0	0	0	Q13A
20	0	0	0	528	94.964	528	94.964	-1.5638	0.9756	0.0425	0.9674	0.955	-0.301	RQ13B
20	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ13B
20	2	2	2	28	5.036	28	5.036	-0.1744	0.6253	0.1203	0.9689	0.488	0.301	RQ13B
21	0	0	0	544	98.1949	544	98.195	-1.5154	0.9952	0.0427	1.4041	1.172	-0.179	RQ13C
21	1	1	1	0	0	0	0	0	0	0	0	0	0	RQ13C
21	2	2	2	10	1.8051	10	1.8051	-0.1666	0.5675	0.1892	1.1616	0.554	0.179	RQ13C

9.19 *Appendix I.* Chapter 5 (Phase II)- Mathematics Pre- and post-test Assessment framework

Pre- Assessment Items Used and all blue highlighted items were retained as "ANCHOR items" Model fit & Difficult analysis										
Item ID	**Item No. Pre-test**	**Item No. Post-test**	**Anchor item (Y/N)**	**ItemType (Constructed /TrueorFalse)**	**Maximum Score**	**Maximum Categories**	**Item description**	**Missing Value**	**Processes of Comprehension/ Blooms**	**Comments**
Q1a	Q1a		N	CR	2	3	Which of the numbers are odd	9	Synthesis	Categories were not collapsed
Q1b	Q1b	1	Y	CR	2	3	Which of the numbers are multiples of 5	9	Synthesis	Categories were not collapsed
Q2	Q2		N	CR	4	5	Complete the pattern	9	Application	Categories collapsed from 4 to 2
Q3	Q3		N	CR	2	3	What is the distance around the rectangle	9	Application	Categories collapsed from 3 to 2
Q4	Q4		N	CR	2	3	How many kg of sorghum did she receive altogether	9	Application	Categories collapsed from 3 to 2
Q5	Q5		N	CR	2	3	How many bags are left	9	Application	Categories collapsed from 3 to 2
Q6	Q6	2	Y	CR	2	3	How much money does Moreri have to pay	9	Application	Categories collapsed from 3 to 2
Q7	Q7		N	CR	2	3	What is the total fraction eaten	9	Application	Categories collapsed from 3 to 2

Q8	Q8	3	Y	CR	2	3	How much does each child get	9	Synthesis	Categories collapsed from 3 to 2
Q9a	Q9a	4	Y	CR	2	3	What is the total cost of 2 pencil	9	Synthesis	Categories collapsed from 3 to 2
Q9b	Q9b	5	Y	CR	2	3	What is the total cost of all the items	9	Synthesis	Categories collapsed from 3 to 2
Q10a	Q10a	6	Y	CR	1	2	Which two schools have the same number of classrooms	9	Analysis	Categories were not collapsed
Q10b	Q10b	7	Y	CR	2	3	How many classrooms are in Mohembo Primary School	9	Analysis	Categories collapsed from 3 to 2
Q10c	Q10c	8	Y	CR	2	3	How many pictures will be used to show this information	9	Analysis	Categories collapsed from 3 to 2
Q11	Q11		N	CR	1	2	How many triangles are in the shape	9	Application	Item deleted from pre-test analysis
Q12a	Q12a		N	CR	1	2	How many hours does the bus take to travel from Lobatse to Gaborone	9	Application	Item deleted from pre-test analysis
Q12b	Q12b		N	CR	2	3	At what time does the bus leave Boatle	9	Application	Item deleted from pre-test analysis
Q12c	Q12c		N	CR	2	3	What is the total stopping time between Boatle and Otse	9	Application	Item deleted from pre-test analysis
Q13a	Q13a	9	Y	CR	1	2	How much change does she get	9	Application	Categories were not collapsed
Q13b	Q13b	10	Y	CR	2	3	Which two items did he buy	9	Application	Categories collapsed from 3 to 2

Q13c	Q13c		N	CR	2	3	How many pula and thebe coins does she use	9	Application	Categories collapsed from 3 to 2

New Items Used the Post-assessment – some revision made to the items (See the comments accordingly)

Item ID	Item No. Post-test	Item No. Post-test	New item	ItemType (Constructed)	Maximum Score	Maximum Categories	Item description	Missing Value	Processes of Comprehension/ Blooms	Comments
NQ1a	NQ1a		Yes	CR	2	3	Which of the numbers are even	9	Synthesis	Replaced the odd number with even
NQ2	NQ2		Yes	CR	4	5	Complete the pattern	9	Application	Set a new pattern
NQ3	NQ3		Yes	CR	2	3	What is the area of the rectangle	9	Application	Replaced the perimeter with Area
NQ4	NQ4		Yes	CR	2	3	How many people are there altogether in the church	9	Application	Changed the scenario
NQ5	NQ5		Yes	CR	2	3	How many bags were left	9	Application	Changed the scenario
NQ7	NQ7		Yes	CR	2	3	What is the total fraction eaten	9	Application	Changed the scenario
NQ11a	NQ11a		Yes	True/False	1	2	Mpho bought 8 sweets and Neo bought 5 sweets. Together they have bought 13 sweets	9	Application	Changes the scenario
NQ11b	NQ11b		Yes	True/False	1	2	Mr Pule had 15 cattle, a lion ate 8 of them, and he was left with 6 cattle	9	Application	Changed the scenario

NQ12a	NQ12a		Yes	CR	2	3	Who arrived first	9	Application	Changed the scenario for the time
NQ12b	NQ12b		Yes	CR	2	3	How many hours does Tebatso take to travel from grandmother's place to her home?	9	Application	Changed the scenario for the time
NQ13c	NQ13c		Yes	CR	2	3	How many pula and thebe coins does he use?	9	Application	Changed the item which cheaper product to accord easy breaking of Notes and coins

9.20 *Appendix Ji*. Chapter 5 (Phase II)- Pupils'pretest-test maths-DIF Analysis

PERSON CLASS/	Obs-Exp Average	DIF MEASURE	DIF S.E.	PERSON CLASS/	Obs-Exp Average	DIF MEASURE	DIF S.E.	DIF CONTRAST	JOINT S.E.	Rasch-Welch t	d.f.	Prob.	Mantel Chi-squ	Prob.	Size CUMLOR	Active Slices	ITEM Number	Name
1.0	.01	-.91	.12	.	.36	-1.83	.80	.91	.81	1.13	3	.3422	.7537	.3853	1.43	4	1	Q1A
1.0	.01	-.91	.12	2.0	-.02	-.87	.14	-.04	.18	-.22	249	.8296	.0512	.8210	-.06	20	1	Q1A
1.0	-.02	-1.31	.11	.	.44	-2.41	.78	1.10	.79	1.39	3	.2600	1.8947	.1687		4	2	Q1B
1.0	-.02	-1.31	.11	2.0	.00	-1.34	.12	.03	.16	.16	251	.8767	.0103	.9193	.03	20	2	Q1B
1.0	-.02	-.09	.10	.	-.08	-.07	.39	-.02	.40	-.05	3	.9659	.5167	.4723	-1.31	4	3	Q2
1.0	-.02	-.09	.10	2.0	.03	-.18	.12	.09	.16	.57	245	.5717	1.3488	.2455	.67	20	3	Q2
1.0	.00	-.15	.14	.	-.55	1.25<	1.40	-1.41	1.41	-1.00	3	.3911	2.0000	.1573		4	4	Q3
1.0	.00	-.15	.14	2.0	.02	-.23	.16	.07	.21	.34	247	.7305	.5775	.4473	.33	20	4	Q3
1.0	.01	-2.23	.12	.	-.91	.00	.76	-2.24	.77	-2.90	3	.0624	3.5008	.0613		4	5	Q4
1.0	.01	-2.23	.12	2.0	.02	-2.25	.13	.01	.18	.06	258	.9497	.0027	.9585	-.02	20	5	Q4
1.0	.00	-1.53	.11	.	-.14	-1.17	.80	-.36	.81	-.45	3	.6853	.1508	.6978	-.62	4	6	Q5
1.0	.00	-1.53	.11	2.0	.00	-1.53	.12	.00	.16	.00	253	1.000	.0303	.8618	.05	20	6	Q5
1.0	.00	-.85	.11	.	.38	-1.80	.78	.96	.79	1.22	3	.3107	2.8219	.0930	1.59	4	7	Q6
1.0	.00	-.85	.11	2.0	-.01	-.83	.13	-.02	.17	-.12	248	.9007	.0086	.9260	-.03	20	7	Q6
1.0	.05	-1.98	.11	.	.21	-2.36	.75	.39	.76	.51	3	.6474	.1056	.7453	-.50	4	8	Q7
1.0	.05	-1.98	.11	2.0	-.07	-1.78	.12	-.20	.16	-1.23	256	.2199	2.2053	.1375	-.44	20	8	Q7
1.0	-.03	-.33	.13	.	.34	-1.16	.81	.84	.82	1.02	3	.3818	1.0000	.3173		4	9	Q8
1.0	-.03	-.33	.13	2.0	.02	-.43	.15	.10	.19	.54	249	.5914	.3296	.5659	.20	20	9	Q8
1.0	.00	-.77	.11	.	-.34	-.05	.71	-.71	.72	-.99	3	.3941	.1429	.7055	-.69	4	10	Q9A
1.0	.00	-.77	.11	2.0	.01	-.77	.13	.00	.17	.00	247	1.000	.0361	.8493	.07	20	10	Q9A
1.0	-.01	.11	.15	.	.05	-.05	.72	.15	.74	.21	3	.8502	.1429	.7055	.41	4	11	Q9B
1.0	-.01	.11	.15	2.0	.01	.02	.18	.09	.24	.36	247	.7221	.0192	.8898	-.08	20	11	Q9B
1.0	.03	-.91	.18	.	.08	-1.19	1.16	.28	1.18	.24	3	.8257	.4414	.5064	.92	4	12	Q10A
1.0	.03	-.91	.18	2.0	-.04	-.52	.22	-.39	.29	-1.36	246	.1747	1.8270	.1765	-.47	20	12	Q10A
1.0	-.04	.30	.18	.	.60	-1.16	.81	1.46	.83	1.76	3	.1768	3.5714	.0588		4	13	Q10B
1.0	-.04	.30	.18	2.0	.03	.02	.18	.28	.25	1.11	256	.2699	.8137	.3670	.49	20	13	Q10B
1.0	.01	.58	.21	.	.31	-.06	.69	.64	.73	.88	3	.4433	.1429	.7055	.69	4	14	Q10C
1.0	.01	.58	.21	2.0	-.02	.97	.37	-.39	.43	-.91	197	.3621	1.5687	.2104	-1.36	20	14	Q10C
1.0	-.01	2.38	.51	.	-.06	1.78<	1.95	.60	2.01	.30	3	.7853				4	15	Q11
1.0	-.01	2.38	.51	2.0	.02	1.55	.43	.83	.67	1.24	254	.2143	.2342	.6285	.65	20	15	Q11
1.0	.01	2.68	.59	.	-.02	1.80<	1.96	.88	2.05	.43	3	.6964	.3333	.5637		4	16	Q12A
1.0	.01	2.68	.59	2.0	-.01	4.63<	1.83	-1.95	1.92	-1.02	148	.3111	1.0368	.3086		20	16	Q12A
1.0	-.05	-.03	.14	.	-.55	1.21<	1.37	-1.24	1.38	-.90	3	.4340	.5319	.4658		4	17	Q12B
1.0	-.05	-.03	.14	2.0	.07	-.39	.15	.36	.21	1.73	257	.0843	.8725	.3503	.35	20	17	Q12B
1.0	.01	1.20	.36	.	-.04	1.17<	1.34	.03	1.39	.02	3	.9839	.3333	.5637		4	18	Q12C
1.0	.01	1.20	.36	2.0	-.01	2.27<	1.30	-1.07	1.35	-.79	142	.4305	.6057	.4364		20	18	Q12C
1.0	.01	1.40	.33	.	.16	.22	1.26	1.18	1.31	.90	3	.4327	.0294	.8638	1.61	4	19	Q13A
1.0	.01	1.40	.33	2.0	-.02	1.99	.51	-.59	.61	-.97	214	.3343	.2984	.5849	-.81	20	19	Q13A
1.0	.01	.99	.30	.	-.08	1.17<	1.34	-.18	1.38	-.13	3	.9018				4	20	Q13B
1.0	.01	.99	.30	2.0	-.01	1.30	.51	-.31	.59	-.53	201	.5986	1.0000	.3173		20	20	Q13B
1.0	.03	.60	.22	.	-.14	1.17<	1.34	-.57	1.36	-.42	3	.7023				4	21	Q13C
1.0	.03	.60	.22	2.0	-.02	1.30	.51	-.70	.56	-1.26	167	.2103	.1418	.7065	-.40	20	21	Q13C
2.0	-.02	-.87	.14	.	.36	-1.83	.80	.95	.81	1.17	3	.3266	.5054	.4771	1.23	3	1	Q1A
2.0	-.02	-.87	.14	1.0	.01	-.91	.12	.04	.18	.22	249	.8296	.0512	.8210	.06	20	1	Q1A
2.0	.00	-1.34	.12	.	.44	-2.41	.78	1.07	.79	1.35	3	.2698	2.0264	.1546	3.08	3	2	Q1B
2.0	.00	-1.34	.12	1.0	-.02	-1.31	.11	-.03	.16	-.16	251	.8767	.0103	.9193	-.03	20	2	Q1B
2.0	.03	-.18	.12	.	-.08	-.07	.39	-.11	.41	-.27	3	.8071	1.0497	.3056		3	3	Q2
2.0	.03	-.18	.12	1.0	-.02	-.09	.10	-.09	.16	-.57	245	.5717	1.3488	.2455	-.67	20	3	Q2
2.0	.02	-.23	.16	.	-.55	1.25<	1.40	-1.48	1.41	-1.05	3	.3710	1.1040	.2934		3	4	Q3
2.0	.02	-.23	.16	1.0	.00	-.15	.14	-.07	.21	-.34	247	.7305	.5775	.4473	-.33	20	4	Q3
2.0	.02	-2.25	.13	.	-.91	.00	.76	-2.25	.77	-2.92	3	.0617	3.2489	.0715		3	5	Q4
2.0	.02	-2.25	.13	1.0	.01	-2.23	.12	-.01	.18	-.06	258	.9497	.0027	.9585	.02	20	5	Q4
2.0	.00	-1.53	.12	.	-.14	-1.17	.80	-.36	.81	-.45	3	.6858	.7472	.3874		3	6	Q5
2.0	.00	-1.53	.12	1.0	.00	-1.53	.11	.00	.16	.00	253	1.000	.0303	.8618	-.05	20	6	Q5
2.0	-.01	-.83	.13	.	.38	-1.80	.78	.98	.79	1.24	3	.3033	7.8478	.0051		3	7	Q6
2.0	-.01	-.83	.13	1.0	.00	-.85	.11	.02	.17	.12	248	.9007	.0086	.9260	.03	20	7	Q6
2.0	-.07	-1.78	.12	.	.21	-2.36	.75	.58	.76	.76	3	.5001	.4085	.5227	1.01	3	8	Q7
2.0	-.07	-1.78	.12	1.0	.05	-1.98	.11	.20	.16	1.23	256	.2199	2.2053	.1375	.44	20	8	Q7
2.0	.02	-.43	.15	.	.34	-1.16	.81	.73	.82	.89	3	.4385	1.0000	.3173		3	9	Q8
2.0	.02	-.43	.15	1.0	-.03	-.33	.13	-.10	.19	-.54	249	.5914	.3296	.5659	-.20	20	9	Q8
2.0	.01	-.77	.13	.	-.34	-.05	.71	-.71	.72	-.99	3	.3960	.1000	.7518		3	10	Q9A
2.0	.01	-.77	.13	1.0	.00	-.77	.11	.00	.17	.00	247	1.000	.0361	.8493	-.07	20	10	Q9A
2.0	.01	.02	.18	.	.05	-.05	.72	.07	.74	.09	3	.9344	1.0000	.3173		3	11	Q9B
2.0	.01	.02	.18	1.0	-.01	.11	.15	-.09	.24	-.36	247	.7221	.0192	.8898	.08	20	11	Q9B
2.0	-.04	-.52	.22	.	.08	-1.19	1.16	.67	1.18	.57	3	.6097	.4149	.5195	.10	3	12	Q10A
2.0	-.04	-.52	.22	1.0	.03	-.91	.18	.39	.29	1.36	246	.1747	1.8270	.1765	.47	20	12	Q10A
2.0	.03	.02	.18	.	.60	-1.16	.81	1.18	.83	1.42	3	.2516	1.0000	.3173		3	13	Q10B
2.0	.03	.02	.18	1.0	-.04	.30	.18	-.28	.25	-1.11	256	.2699	.8137	.3670	-.49	20	13	Q10B
2.0	-.02	.97	.37	.	.31	-.06	.69	1.03	.79	1.31	4	.2602				3	14	Q10C
2.0	-.02	.97	.37	1.0	.01	.58	.21	.39	.43	.91	197	.3621	1.5687	.2104	1.36	20	14	Q10C
2.0	.02	1.55	.43	.	-.06	1.78<	1.95	-.23	1.99	-.11	3	.9158				3	15	Q11
2.0	.02	1.55	.43	1.0	-.01	2.38	.51	-.83	.67	-1.24	254	.2143	.2342	.6285	-.65	20	15	Q11
2.0	-.01	4.63<	1.83	.	-.02	1.80<	1.96	2.83	2.68	1.06	10	.3157				3	16	Q12A
2.0	-.01	4.63<	1.83	1.0	.01	2.68	.59	1.95	1.92	1.02	148	.3111	1.0368	.3086		20	16	Q12A
2.0	.07	-.39	.15	.	-.55	1.21<	1.37	-1.60	1.38	-1.16	3	.3295	.1666	.6831		3	17	Q12B
2.0	.07	-.39	.15	1.0	-.05	-.03	.14	-.36	.21	-1.73	257	.0843	.8725	.3503	-.35	20	17	Q12B
2.0	-.01	2.27<	1.30	.	-.04	1.17<	1.34	1.10	1.87	.59	11	.5689				3	18	Q12C
2.0	-.01	2.27<	1.30	1.0	.01	1.20	.36	1.07	1.35	.79	142	.4305	.6057	.4364		20	18	Q12C
2.0	-.02	1.99	.51	.	.16	.22	1.26	1.77	1.36	1.30	4	.2629				3	19	Q13A
2.0	-.02	1.99	.51	1.0	.01	1.40	.33	.59	.61	.97	214	.3343	.2984	.5849	.81	20	19	Q13A
2.0	-.01	1.30	.51	.	-.08	1.17<	1.34	.13	1.44	.09	3	.9346				3	20	Q13B
2.0	-.01	1.30	.51	1.0	.01	.99	.30	.31	.59	.53	201	.5986	1.0000	.3173		20	20	Q13B
2.0	-.02	1.30	.51	.	-.14	1.17<	1.34	.13	1.44	.09	3	.9355				3	21	Q13C
2.0	-.02	1.30	.51	1.0	.03	.60	.22	.70	.56	1.26	167	.2103	.1418	.7065	.40	20	21	Q13C

9.21 *Appendix Jii*. Chapter 5 (Phase II)- Pupils' posttest maths-DIF Analysis

TABLE 30.1 Post Maths. sav.sav   ZOU645WS.TXT May 9 2021 12:27
INPUT: 272 PERSON 21 ITEM REPORTED: 272 PERSON 21 ITEM 43 CATS WINSTEPS 4.7.1.0

DIF class/group specification is: DIF= @GENDER

PERSON CLASS/	Obs-Exp Average	DIF MEASURE	DIF S.E.	PERSON CLASS/	Obs-Exp Average	DIF MEASURE	DIF S.E.	DIF CONTRAST	JOINT S.E.	Rasch-Welch t	d.f.	Prob.	Mantel Chi-squ	Prob.	Size CUMLOR	Active Slices	ITEM Number	Name
1	.03	-1.69	.22	2	-.03	-1.31	.20	-.39	.30	-1.30	269	.1938	.5502	.4582	-.27	17	1	Q1B_ANCH
1	.01	-.93	.20	2	-.01	-.81	.20	-.12	.28	-.42	269	.6747	.0031	.9554	-.04	17	2	Q6_ANCHO
1	-.01	.97	.21	2	.01	.82	.24	.15	.32	.46	262	.6487	.0014	.9703	.07	17	3	Q8_ANCHO
1	.01	-.86	.20	2	-.02	-.69	.20	-.17	.28	-.59	268	.5574	.0058	.9395	-.03	17	4	Q9A_ANCH
1	.01	.63	.20	2	-.02	.82	.24	-.20	.32	-.62	258	.5356	.0005	.9826	-.05	17	5	Q9B_ANCH
1	.06	-1.11	.20	2	-.07	-.40	.21	-.72	.29	-2.49	268	.0133	3.4662	.0626	-.65	17	6	Q10A_ANC
1	-.02	.96	.21	2	.02	.65	.24	.31	.32	.98	264	.3258	.2329	.6294	.21	17	7	Q10B_ANC
1	-.01	2.31	.29	2	.01	2.12	.35	.19	.45	.41	256	.6821	.0164	.8981	-.05	17	8	Q10C_ANC
1	.01	1.81	.25	2	-.01	2.00	.34	-.19	.42	-.46	245	.6428	.0068	.9344	.08	17	9	Q13A_ANC
1	.01	3.50	.43	2	-.01	4.63	1.01	-1.14	1.10	-1.03	177	.3047	.3847	.5351	-1.03	17	10	Q13B_ANC
1	.00	.00	.19	2	.00	-.04	.21	.05	.29	.17	265	.8657	.0044	.9468	-.06	17	11	Q1A_POST
1	-.07	.28	.15	2	.07	-.17	.16	.45	.22	2.07	267	.0396	3.4838	.0620	.53	17	12	Q2_POST
1	-.02	1.06	.22	2	.02	.82	.24	.24	.33	.73	263	.4635	.0172	.8957	.01	17	13	Q3_POST
1	.02	-3.04	.31	2	-.02	-2.65	.24	-.39	.39	-.99	256	.3242	.0232	.8790	-.04	17	14	Q4_POST
1	.06	-2.00	.23	2	-.06	-1.26	.20	-.73	.31	-2.38	267	.0182	2.0958	.1477	-.59	17	15	Q5_POST
1	-.11	-.82	.20	2	.11	-2.09	.22	1.27	.29	4.36	265	.0000	7.0817	.0778	.83	17	16	Q7_POST
1	.02	-2.86	.29	2	-.02	-2.54	.23	-.31	.37	-.84	259	.4006	2.6105	.1062	-.68	17	17	Q11A_POS
1	.03	-2.69	.28	2	-.03	-2.19	.22	-.51	.35	-1.43	259	.1529	3.3752	.0662	-.75	17	18	Q11B_POS
1	.01	.67	.20	2	-.01	.82	.24	-.16	.32	-.49	258	.6253	.2799	.5968	-.21	17	19	Q12A_POS
1	-.07	.88	.21	2	.07	.05	.22	.84	.30	2.78	269	.0058	6.7344	.0795	.89	17	20	Q12B_POS
1	.01	2.08	.27	2	-.02	2.54	.41	-.47	.49	-.96	229	.3400	.0021	.9638	-.16	17	21	Q13C_POS
2	-.03	-1.31	.20	1	.03	-1.69	.22	.39	.30	1.30	269	.1938	.5502	.4582	.27	17	1	Q1B_ANCH
2	-.01	-.81	.20	1	.01	-.93	.20	.12	.28	.42	269	.6747	.0031	.9554	.04	17	2	Q6_ANCHO
2	.01	.82	.24	1	-.01	.97	.21	-.15	.32	-.46	262	.6487	.0014	.9703	-.07	17	3	Q8_ANCHO
2	-.02	-.69	.20	1	.01	-.86	.20	.17	.28	.59	268	.5574	.0058	.9395	.03	17	4	Q9A_ANCH
2	-.02	.82	.24	1	.01	.63	.20	.20	.32	.62	258	.5356	.0005	.9826	.05	17	5	Q9B_ANCH
2	-.07	-.40	.21	1	.06	-1.11	.20	.72	.29	2.49	268	.0133	3.4662	.0626	.65	17	6	Q10A_ANC
2	.02	.65	.24	1	-.02	.96	.21	-.31	.32	-.98	264	.3258	.2329	.6294	-.21	17	7	Q10B_ANC
2	.01	2.12	.35	1	-.01	2.31	.29	-.19	.45	-.41	256	.6821	.0164	.8981	.05	17	8	Q10C_ANC
2	-.01	2.00	.34	1	.01	1.81	.25	.19	.42	.46	245	.6428	.0068	.9344	-.08	17	9	Q13A_ANC
2	-.01	4.63	1.01	1	.01	3.50	.43	1.14	1.10	1.03	177	.3047	.3847	.5351	1.03	17	10	Q13B_ANC
2	.00	-.04	.21	1	.00	.00	.19	-.05	.29	-.17	265	.8657	.0044	.9468	.06	17	11	Q1A_POST
2	.07	-.17	.16	1	-.07	.28	.15	-.45	.22	-2.07	267	.0396	3.4838	.0620	-.53	17	12	Q2_POST
2	.02	.82	.24	1	-.02	1.06	.22	-.24	.33	-.73	263	.4635	.0172	.8957	-.01	17	13	Q3_POST
2	-.02	-2.65	.24	1	.02	-3.04	.31	.39	.39	.99	256	.3242	.0232	.8790	.04	17	14	Q4_POST
2	-.06	-1.26	.20	1	.06	-2.00	.23	.73	.31	2.38	267	.0182	2.0958	.1477	.59	17	15	Q5_POST
2	.11	-2.09	.22	1	-.11	-.82	.20	-1.27	.29	-4.36	265	.0000	7.0817	.0778	-.83	17	16	Q7_POST
2	-.02	-2.54	.23	1	.02	-2.86	.29	.31	.37	.84	259	.4006	2.6105	.1062	.68	17	17	Q11A_POS
2	-.03	-2.19	.22	1	.03	-2.69	.28	.51	.35	1.43	259	.1529	3.3752	.0662	.75	17	18	Q11B_POS
2	-.01	.82	.24	1	.01	.67	.20	.16	.32	.49	258	.6253	.2799	.5968	.21	17	19	Q12A_POS
2	.07	.05	.22	1	-.07	.88	.21	-.84	.30	-2.78	269	.0058	6.7344	.0795	-.89	17	20	Q12B_POS
2	-.02	2.54	.41	1	.01	2.08	.27	.47	.49	.96	229	.3400	.0021	.9638	.16	17	21	Q13C_POS

Width of Mantel slice: MHSLICE = .010 logits, Zero cell adjustment: MHZERO = .0000

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

9.22 *Appendix K*. Chapter 6 (Phase III)- Teacher interview guide questions

INTERVIEW SCHEDULE ON FORMATIVE ASSESSMENT TO ENHANCE PUPILS' HOTS IN MATHEMATICS

Introduction

Following the workshop training and implementation of the formative assessment strategies, you are kindly invited to participate in the last Phase of this PhD study. The purpose of this interview is to determine your experience and reflection following the FA strategies intervention when teaching HOTS in mathematics. The results from this interview will advance our understanding and teachers' experience for formative assessment in the classroom.

All responses are anonymous and your participation is completely voluntary, there are no immediate personal benefits or disadvantages for your participation, and you are also free to discontinue at any time. However, your participation is very much appreciated and will assist in the education process for Botswana.

Name of researcher	Mr S. E. Moyo
Names of interviewee	
Place of interview	Kanye Sub-Region; Kanye Intervention Schools
Topic	Formative Assessment and HOTS
Date of interview	
Time of interview	
Where do you teach? Name of the primary school	
Your teaching experience (in years	

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Research Sub-Question 5;

1. ***What are teachers' experiences following the FA Strategies intervention and Mathematics teaching on the pupils learning outcomes***
   i. From your experience, what formative assessment strategies do you frequently use in your classroom? (at least 2 strategies)
   ii. Describe the extent to which formative assessment strategies (*you mentioned in question 1*) provide the opportunity for pupils to develop confidence in HOTS items? Any impact?
   iii. Explain how you use formative assessment strategies in your class? Which strategies did you like most and why?
      a. Learning goals and success criteria
      b. Effective discussion (Questioning & Collaboration)
      c. Learning tasks (HOTS)
      d. Feedback on instruction (including peer-and self-assessment)
   iv. Do you consider your experience with formative assessment intervention as professional development? Why?
   v. To what extent did the workshop training help you to implement formative assessment?
   vi. What are the pupils' reactions when you implement formative assessment strategies in your classroom?
   vii. Any improvement in pupils' achievement in mathematics following the formative assessment intervention?
   viii. What encouragement did you receive from the head of the department/principal or teachers to implement formative assessment?
   ix. How do the large class sizes and heavy workloads impede the implementation of formative assessment in your class?
   x. What are the main problems you faced regarding the implementation of formative assessment?

Research Sub-Question 6;

2. ***What are teachers' reflections following the FA strategies intervention and mathematics teaching pupils' HOTS***
   i. In your opinion, how are the main barriers/problems to implement formative assessment in your class can be addressed?
   ii. What is your recommendation to other teachers towards the implementation of formative assessment in their classroom?
   iii. What other points would you like to add regarding the implementation of formative assessment to improve HOTS in Mathematics?

9.23 *Appendix L.* Chapter 6 (Phase III)- Summary statistics for stacked analysis

TABLE 3.1 Pre AND Post-Test Data (stacked data) _12.09.2019_DICH ZOU254WS.TXT May 9, 2021, 10:58
INPUT: 272 PERSON 27 ITEM REPORTED: 278 PERSON 21 ITEM 43 CATS WINSTEPS 4.7.1.0

SUMMARY OF 272 MEASURED PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	9.1	21.0	-.63	.58	1.00	.00	.99	.08
SEM	.2	.0	.08	.01	.02	.07	.05	.05
P.SD	4.1	.2	1.34	.10	.33	1.09	.85	.81
S.SD	4.1	.2	1.34	.10	.33	1.09	.85	.82
MAX.	19.0	21.0	2.69	1.07	2.27	3.45	9.90	3.45
MIN.	1.0	18.0	-4.07	.53	.44	-2.40	.21	-1.46

REAL RMSE .63 TRUE SD 1.18 SEPARATION 1.88 PERSON RELIABILITY .78
MODEL RMSE .59 TRUE SD 1.20 SEPARATION 2.02 PERSON RELIABILITY .80
S.E. OF PERSON MEAN = .08

PERSON RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .80 SEM = 1.85
STANDARDIZED (50 ITEM) RELIABILITY = .91

SUMMARY OF 21 MEASURED ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	120.8	277.9	.00	.17	.99	-.18	1.01	-.14
SEM	16.1	.1	.38	.01	.04	.48	.10	.44
P.SD	72.2	.3	1.69	.06	.16	2.14	.44	1.96
S.SD	74.0	.4	1.73	.06	.16	2.20	.45	2.01
MAX.	234.0	278.0	3.78	.39	1.40	5.59	2.38	4.00
MIN.	7.0	277.0	-2.80	.11	.79	-3.92	.43	-2.84

REAL RMSE .18 TRUE SD 1.68 SEPARATION 9.11 ITEM RELIABILITY .99
MODEL RMSE .18 TRUE SD 1.68 SEPARATION 9.35 ITEM RELIABILITY .99
S.E. OF ITEM MEAN = .38

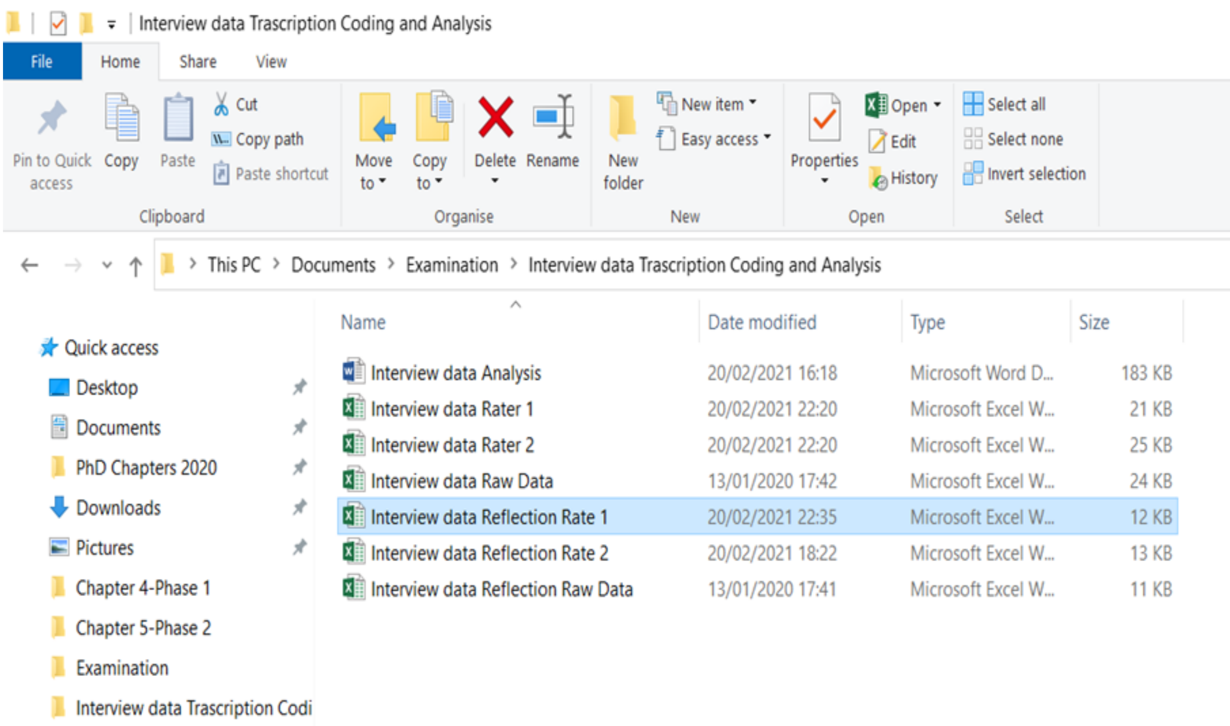
LACKING RESPONSES: 6 ITEM
ITEM RAW SCORE-TO-MEASURE CORRELATION = -.94
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

9.24 *Appendix M*. Chapter 6 (Phase III)- Teacher interviews coding scheme

School/ Teacher Pseudonym	Sex	Qualifications	Position	Specialization	No of Pupil in class	No. of year teaching Std4	Experience (in Year)
School: PS1 Teacher: T1-PS1	Female	Primary Teaching Certificate	Teacher	Generals	32	2	23
School: PS2 Teacher: T2-PS2	Female	Diploma in Primary Education	Senior 1 Teacher	Languages	27	2	24
School: PS3 Teacher: T3-PS3	Female	Diploma in Primary Education	Senior 1 Teacher	Languages	35	5	27
School: PS4 Teacher: T4-PS4	Female	Primary Teaching Certificate	Teacher	Generals	29	3	25
School: PS5 Teacher: T5-PS5	Female	Diploma in Primary Education	Teacher	Social Studies and Religious & Moral Education	38	2	20
School: PS6 Teacher: T6-PS6	Female	Diploma in Primary Education	Teacher	English and Setswana	34	2	19

School: PS7 Teacher: T7-PS7	Female	Bachelor of Education	Teacher	Physical Education	39	2	12



Main category	**Code**	**Sub-Code**	**Examples**
SRQ6. (i). From your experience, what formative assessment strategies do you frequently use in your classroom? (at least 2 strategies) (FAS)	FAS	FAS_LGSC (6)	I have used all of them because there is no how one can separate them in a lesson. Because they complement each other during teaching and learning
		FAS_EEQ (5)	I am an old teacher and I used the old model of asking question whereby children will be raising their hands. But now that since you came up with the idea of putting their name in a container and randomly picking one, so it means that every learner is ready, that I might pick his or her name.
		FAS_EEC (3)	….I then put pupils into a group for discussion
		FA_LT(2)	I engage all pupils by learning task, peer teaching where a pupil writing something on the chalkboard, giving pupils time to act and giving feedback
		FA_GPF (2)	I more often write the Learning goal on the manila and display it on the board, even success criteria, I write them on the manila and share them on the

			chalkboard. I also use pupils' self-assessment during teaching and after the lesson, I give them exercise.
Main category	**Code**	**Sub-Code**	**Examples**
SRQ6. (ii). Describe the extent to which formative assessment strategies (you mentioned in question 1) provide the opportunity for pupils to develop confidence in HOTS items? Any impact?		concentra (4)	During questioning, I always discourage my learner to raise their hands, so I use the pupils' name card to pick the learner randomly. This technique made my learners were very alert, prepared and showed a high level of concentration during our class discussion, maybe would be his or her turn to answer a question.
		direct learning (3)	. **SC-** yes because you have to follow it, and know-how to assess the learner. I was explaining the SC together with the LG so that it directs both the teacher and learners. You have to know what is expected of them and or what they should do. **Questioning**- I Liked the way we ask a question we shouldn't or pupil should be picked randomly, this indeed helped me to engage all the learner. Planning for higher-order questions in class.
	Impact FAS DL&EE		With exit car, it helps a kid to work independently when given a task and likely to improve their thinking capacity. When pupils are made to write on the exit card. You can pick the mistake from every pupil ....
		Impr thinking and problem solver (5)	from here I give them a short exercise to do alone on their EXIT CARD individually, so I can be able to see whether they have understood or something that is lacking and help them as I reviewed their EXIT card by asking them to show me their answers. .... Like I said earlier, I was also giving pupils some group tasks which involved HOTS, this was used to help them think deeply before answering the question and practice thinking independently as problem solvers.
			...it improves independent learning. Because, during working, I didn't have to go around to help them, they just look at the SC, what is it, that we are going to do next, after these, we are doing that, and they end up an understanding thing or what we were trying to discuss with easy.
Main category	**Code**	**Sub-Code**	**Example**
SRQ6. (iii)Do you consider your experience with formative assessment	Impact Training WS_PD	Planning and time (3)	It is very professional, and ... so the workshop has assisted me a lot, because I now spent time on planning good learning activities for learners, and I have enough time to assist pupils based on the FA

intervention as professional development? Why? (ImpactofTraining)			strategies and techniques. ....but nowadays, I plan and write it for a learner to read it together.
		Value addition (3)	it helps to guide the teacher because it helps you to know the direction. It added value to my teaching style.
			Yes, that's why I even went to tell, my colleagues about the strategies which I learnt to tell my colleagues about the strategies which I learnt from the workshop.
		FA apply across the level(4)	FA is very professional, this one is helpful, it's very good, it rather should be from all the classes across all levels.
Main category	**Code**	**Sub-Code**	**Examples**
To what extent did the workshop training help you to implement formative assessment? (impacttrainingWshop)	Impact TrainingWshop	improve Planning (4)	…. assisted to re-focus teaching and learning. It assisted me with planning, in particular, asking HOTS questions
		engaging all pupils (6)	Those strategies are common no one in this school using them particularly random pick of pupils and exit card. It is very helpful in engaging all learners
		Motivate learning (4)	. And the way we were taught at the workshop, yeah! It brought arousal to the learners, now I seem highly motivated and learn easy and my teaching is also easy in the classroom
Main category	**Code**	**Sub-Code**	**Examples**
What are the pupils' reactions when you implement formative assessment strategies in your classroom? (pupil react)	Pupilreact	excitement (2)	My pupils were very excited and they were ready to answer. That since ever, I started using the FA, particularly the flask card, every pupil paired attention in class discussion because the teacher will not rely on those pupils who used to raise their hand always if the teacher asks a question. Rather they were aware that they may be select too randomly using a name card
		confident (3)	Pupil likes the use of exit card and random selection of pupils when answering questions. The strategies and techniques assisted me in building pupils confidence and pupils were always alert.
		enjoyment (3)	They enjoyed it. Those strategies I was using for example when I used whiteboard, they enjoyed it and felt free, they became open to me, to one another, they assisted or help each other.

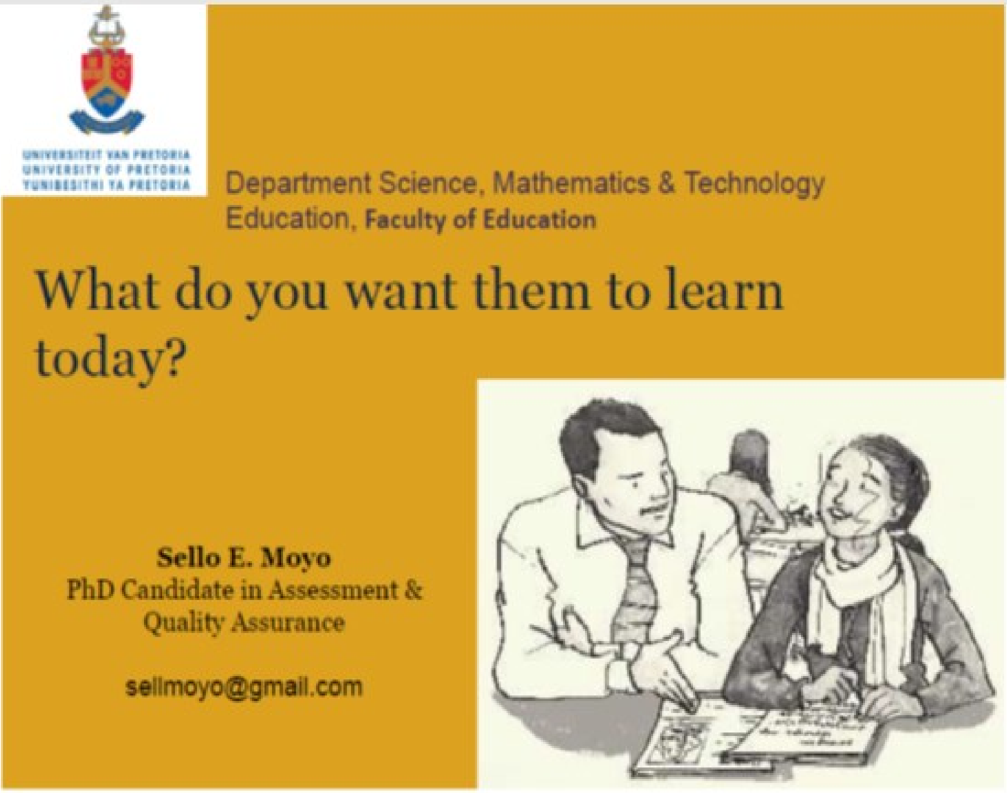
		Learning interest & excitement (4)	There was too much excitement because most of the time, they were doing the work for themselves not me as the teacher doing it for them that they are much more involved in doing the work. It encourages more highly competition and learning was so interesting for them.
			My pupils are so interested in the FA
			I think my learners were so interested and as you know young people are so interested in new things. My learners are no interested in doing the thing on their own, interested in hands-on activities, whereby everyone is given time to do a particular task as I said with exit card and that every learner is so engaged in the classroom and then show their solution to the teacher, it makes the learner so interested.
Main category	**Code**	**Sub-Code**	**Examples**
Any improvement in pupils' achievement in mathematics following the formative assessment intervention? (PL&IL)	PL&IL		There is a lot of improvement, as I speak if you take my result analysis, my pupils have made with an average of 90% in Maths.
			General the improved as compared with other classes
			They got 90%. To be honest, the pupils like it even when I leave them alone for meeting on arrival, they show me what they did in groups
			At the end of first, their average in maths was 29% and then at end of the second term they moved to 95.1%, this is a tremendous improvement,
			Even those pupils who got "D", when I paged their scripts, I observed that they were just lazy to add up thing, but in general you can see that the method for solving they know it, though the answer was wrong.
Main Category	**Code**	**Sub-Code**	**Examples**
What encouragement did you receive from the head of the department/principal or teachers to implement formative assessment? (ExpSuppt	ExpSuppt	General support (2)	The school, in general, knows that I'm doing something which is adding value to the pupils
		HOD&ScHD (5)	I do receive support from my colleague at standard 4 level who happens to be also my HOD
			I think the management gave me support, firstly by allowing me to go to the workshop for two days, and secondly by allowing you to assess me. My HOD had time to go through what I was doing in class. Aah, she was very happy

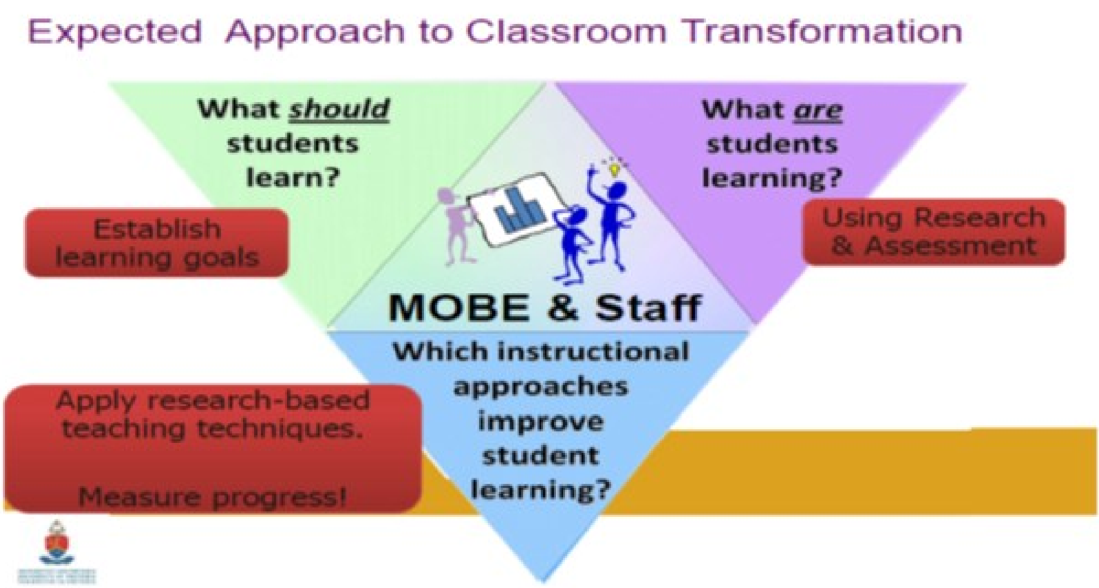
			My School head came to my class and she said the school has realized that there was a difference in my class performance particularly in maths and her interest in the strategies which would also assist other classes. And in general, my HOD is supportive.
Main category	**Code**	**Sub-Code**	**Examples**
How do the large class sizes and heavy workloads impede the implementation of formative assessment in your class? (ImpactFASWKL)	Impact FAS WKL	Large class sizes (5)	The use of FA is helpful in large class like mine because when I started implementing FA, I find it to be a pupil's centre approach, hence they, the pupils spend time engaging in activities, helping one another and make learning very easy. I don't even feel a workload
		Small class sizes(2)	My class is small and with FA is everything went well. And the workload is not bad particularly with the implementation of FA.
			With FA, it helps me to manage the large class size, unlike me do it. Nowadays I gave pupils' clear instructions and expectation, then they work. Unlike me talking the whole day
		hectic to plan (1)	... after I have been introduced to FA, my workload has reduced as most tasks are done by the pupils and they understand quickly. .... initially, the preparation of materials was a bit hectic and however... When I catch up, the work was so easy and helpful to assist all learners in the class.
Main category	**Code**	**Sub-Code**	**Examples**
What are the main problems you faced regarding the implementation of formative assessment? (ChalleIMFAS)	Challel MFAS	Resources (4)	.... lack of technology (computers, laptops) has hampered my innovation to explore some good activities.
			Lack of resources
			...no materials, I had to improvise, using manila for whiteboard, and again I even asked pupils to ask their parents to buy for them those whiteboard markers.
			.... the pupils were used to traditional teaching as well as way of answering the question by raising hands, and others.
		ChalImp Plan (2)	*Learners are a bit slower, some go could not read, hence I consumed my teaching time.*
Main category	**Code**	**Sub-Code**	**Examples**
	Reflecti	Resource (4)	The schools should be equipped with teaching technology like, pupils' tablets, computers,

SRQ(i)In your opinion, how are the main barriers/problems to implement formative assessment in your class can be addressed? (Reflection)			classroom projectors, these will enhance teaching and learning through FA
			If we have a resource like tablets will give the pupils much exposure to a learning activity.
			When FA, is implemented, at least there should be some material in place to assist the teacher to implement those strategies…I even used my money to buy the whiteboard markers.
		Reflec_Mix-able clas (3)	….to assist the mixed ability class since it is effective particularly read problem and pair attention to address the problem
Main category	**Particip ant**	**Sub-Code**	**Examples**
(SRQ(ii) What is your recommendation to other teachers towards the implementation of formative assessment in their classroom? (RecIM)	RecIM	across the level (5)	I think it should do or used in all schools and starting at the foundation level/phase and going up the level so that pupils would acquire the key skills for HOTS or Innovative skill at an early age and become independent thinker as they grow up. And with FA is very possible.
			If we start from the lower class, the pupils from those class, they will go on familiar with strategies to the high level/upper primary. And again pupils will get used to being engaged across all levels.
		Tr & Subj(2)	we can apply this in maths and other subjects, not Maths only. It also should be implemented across the level.
			……so that they may use it even in other subjects because it shows that it worked very much.
			applied to all classes across the level.
		pupil_centered (3)	…. it added value to my work to reduce the workload. I do recommend it to other teachers because it a pupil-centred approach and not teacher-centred…….In this FA the teacher does less and learners do a lot of work.

9.25 *Appendix N*. Chapter 5 (Phase II)- Sample of work slides

DAY ONE





At the end of this workshop

You will be able to…

DAY 1

1. Develop and communicate your learning goals (LG) and Success Criteria (SC) for a given topic using the Bloom's cognitive behaviours (espec HOTS)
2. Recognise the value of aligning assessments with goals.
3. Improve pupils' participation and engagement in the classroom

DAY 2

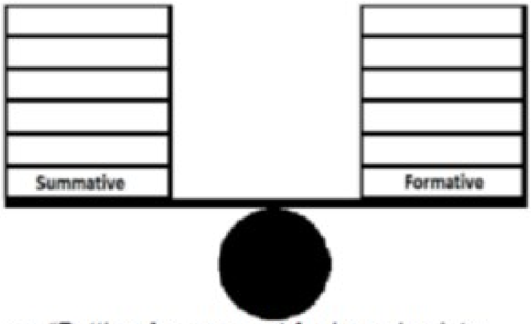
4. Develop higher order thinking skill (HOTS) questions.
5. Determine ways of providing effective feedback as well as pupils peer and self-assessment.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Activity 1: Reflect on your Current Practice

Assessment of and for learning

4 minutes



Source: "Putting Assessment for Learning into Practice," David Spendlove, 2009)

DAY 2

Department Science, Mathematics & Technology Education, **Faculty of Education**

What do you want them to learn today?

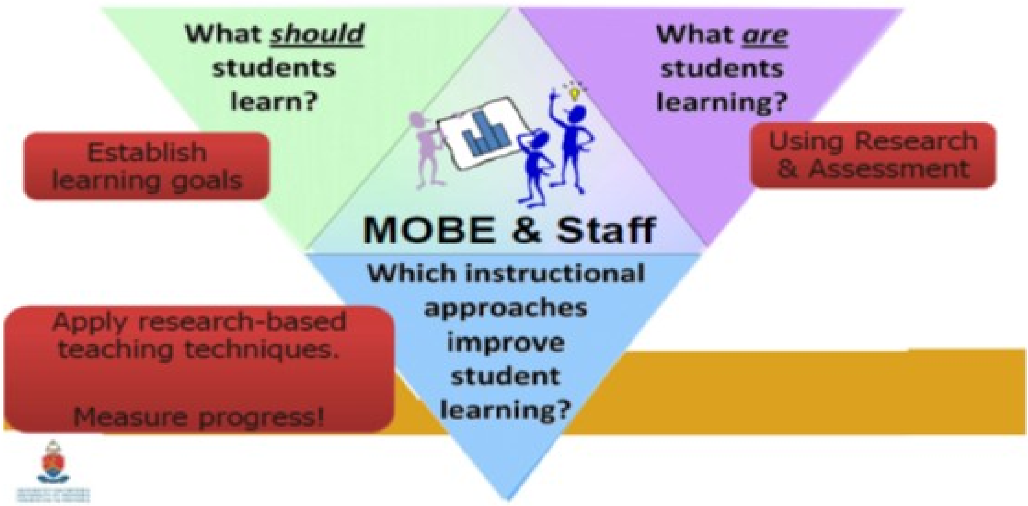
Sello E. Moyo
PhD Candidate in Assessment & Quality Assurance

sellmoyo@gmail.com

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Expected Approach to Classroom Transformation



At the end of this workshop

You will be able to…

DAY 1

1. Develop and communicate your learning goals (LG) and Success Criteria (SC) for a given topic using the Bloom's cognitive behaviours (espec HOTS)
2. Recognise the value of aligning assessments with goals.
3. Improve pupils' participation and engagement in the classroom

DAY 2

4. Develop higher order thinking skill (HOTS) questions.
5. Determine ways of providing effective feedback as well as pupils peer and self-assessment.

Formative(Assessment for Learning): Key Elements

1. Learning Intention and Success Criteria
2. Effective Feedback
3. Effective Questioning
4. Self assessment - Pupils as owners of their own learning
5. Peer Assessment - Pupils as Instructional resources for each other

Formative Assessment strategies need to be emphasised into practice over time.

AfL Assessment for Learning – A Practical Guide 2010 p38

WORKSHOP TOPIC 4:

Develop higher order thinking skill (HOTS) questions

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Problem Statement (HOTS)

The pupils' performance in Trends in International Mathematics and Science Study (TIMSS) in 2015 has revealed that Botswana pupils have not acquired substantial mathematics skills when ranked along their international cohort. It was evident through TIMSS 2015 in which Botswana's higher grade pupils (standard 6 instead of 4) took part, they ranked third from the bottom of the fifty-nine (59) countries and nations and 425 000 students participated in TIMSS in 2015.TIMSS in 2015 described Botswana pupils who participated in the study as unable to apply HOTS such as critical thinking and problem solving in mathematics (Masole, Gabalebatse, Guga, Pharithi & BEC, 2016).

What are higher order skills?

Higher Order Skills

Higher order skills are skills involving **analysis**, **evaluation** and **synthesis** (creation of new knowledge). These are thought to be of a 'higher order', requiring different learning and teaching methods than the learning of facts and concepts.

Higher Order Thinking

Higher order thinking involves the learning of complex judgmental skills such as critical thinking and problem solving.
Higher order thinking is more difficult to learn or teach but also more valuable because such skills are more likely to be usable in new and unfamiliar situations.

Higher Order Questioning

Higher order questions require answers that go beyond simple information and as such both the language and thinking behind them is more complex. They take learners into abstract language functions, such as giving and justifying opinions, speculation and hypothesising.

Are you already doing some of these?

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

At what level do we want that knowledge

Lower Order Thinking
Lower Order Skills
Lower Order Questioning

Higher Order Thinking
Higher Order Skills
Higher Order Questioning

"Slide Sample End"

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

9.26 *Appendix O*. Chapter 3 – Approval Letter from UP Ethics Committee

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Education

Ethics Committee

29 November 2018

Mr Sello Moyo

Dear Mr Moyo

REFERENCE: SM 18/10/02

This letter serves to confirm that your application was carefully considered by the Faculty of Education Ethics Committee. The final decision of the Ethics Committee is that your application has been **approved** and you may now start with your data collection. The decision covers the entire research process and not only the days that data will be collected. The approval is valid for two years for a Masters and three for Doctorate.

The approval by the Ethics Committee is subject to the following conditions being met:

1. The research will be conducted as stipulated on the application form submitted to the Ethics Committee with the supporting documents.
2. Proof of how you adhered to the Department of Basic Education (DBE) policy for research must be submitted where relevant.
3. In the event that the research protocol changed for whatever reason the Ethics Committee must be notified thereof by submitting an amendment to the application (Section E), together with all the supporting documentation that will be used for data collection namely; questionnaires, interview schedules and observation schedules, for further approval before data can be collected. Non-compliance implies that the Committee's approval is null and void. The changes may include the following but are not limited to:
    - Change of investigator,
    - Research methods any other aspect therefore and,
    - Participants
    - Sites

The Ethics Committee of the Faculty of Education does not accept any liability for research misconduct, of whatsoever nature, committed by the researcher(s) in the implementation of the approved protocol.

Upon completion of your research you will need to submit the following documentations to the Ethics Committee for your Clearance Certificate:

- Integrated Declaration Form (Form D08),
- Initial Ethics Approval letter and,
- Approval of Title.

Please quote the reference number **SM 18/10/02** in any communication with the Ethics Committee.

Best wishes

Prof Liesel Ebersöhn
Chair: Ethics Committee
Faculty of Education