University of Szeged

Doctoral School of Education

Learning and Instruction Programme

Ruth Nanjekho Wafubwa

Formative Assessment as an Instructional Approach in the Teaching and Learning of Mathematics: A Case of Secondary School Teachers and Students in Kenya

PhD Dissertation

Supervisor: Prof. Dr. Csíkos Csaba

Szeged, Hungary, 2022

Table of Contents

Acknowledgments

This dissertation has been accomplished with the support of lots of people to whom I am sincerely indebted. First and foremost, I want to thank the Hungarian Government and the Tempus Foundation for offering me the Stipendium Hungaricum scholarship which enabled me to pursue my studies. Special gratitude to Prof. Dr. Benő Csapó, the head of the Doctoral School of Education at the University of Szeged for believing in me and granting me an opportunity in this precious University.

My heartfelt appreciation goes to my supervisor, Prof. Dr. Csikos Csaba for accepting to advise and guide me throughout my doctoral studies. You influenced my line of research especially in the area of metacognition where I had little knowledge in. I aspire to continue walking in your footsteps and keep learning more from your scholarly work. I extend my appreciation to my internal defense reviewers: Dr. Klára Tarkó and Dr. Attila Rausch for taking their time to go through my work and their valuable comments that helped refine my dissertation.

Special acknowledgment to the weekly research seminar moderating professors: Dr. Molnár Edit Katalin and Prof. Dr. Vidakovich Tibor. Your feedback was a great source of encouragement and inspiration for my academic journey. I extend this acknowledgment to all the professors in the Institute of Education who taught me and those who supported me in one way or the other. My gratitude also goes to my classmates of 2018-2022. I learned a lot from you and I will forever treasure the beautiful moments we had together. Thanks to all my friends both in Kenya and those I met in Hungary for your support and encouragement.

To all my family members, thanks so much for being a pillar in my life. You provided a shoulder for me to lean on during my ups and downs. It was the hardest decision you had to make by allowing me to pursue my dreams abroad. I will forever be grateful for the patience, understanding, and support you accorded me during the long times I was away from you.

My special gratitude to the Kenya Teachers Service Commission for granting me study leave to pursue my studies. I want to assure you that I am now more knowledgeable than before. I also sincerely thank all the students and teachers who participated in my studies and the school principals who allowed me to conduct studies in their schools. Above all, I acknowledge the almighty God for His divine guidance and protection throughout my study period.

Dedication

I dedicate this dissertation to my loving mum Susan Wafubwa and the memory of my late dad Lazurus Wafubwa. Dad, as you continue resting in peace know that what you always wished for me has been fulfilled. You believed in me and always reminded me that sky was the limit. Mum, you always prayed for me and encouraged me to press on. My success is your success.

Dissertation Abstract

The main focus of this research was on formative assessment conceptualized as an instructional approach in the teaching and learning of mathematics. The participants involved in the study were the Kenyan secondary school teachers of mathematics and their students. Five studies were conducted under the theoretical framework of formative assessment as suggested by Black and Wiliam (2009). In the first study, the adapted instrument was assessed for its suitability in measuring Kenyan mathematics teachers' perceptions of formative assessment. Teachers' perceptions were rated as low to moderate on a five-point Likert scale. The second study assessed the relationship between teachers' perceived use of formative assessment strategies and their levels of metacognitive awareness. The study revealed a positive relationship between formative assessment and metacognition. The study showed that formative assessment strategies influence teachers' metacognitive regulation. Study three was an extension of study two and it examined teachers' conception and perceptions of metacognition use in mathematics classrooms. The findings revealed that teachers regarded themselves as highly metacognitive although they hardly translated the same to their teaching. The fourth study involved the development and validation of a test to measure students' proportional reasoning skills in mathematics. This test was later used as a tool to measure the impact of formative assessment on students' achievement which was conducted in study five. Analysis of items in the test showed that the test was suitable to be used for formative assessment intervention. Study five was thus experimental and aimed to find out the impact of formative assessment on students' achievement in mathematics and their levels of metacognitive awareness. The results showed that formative assessment can be used as an intervention to improve students' performance in mathematics and also improve their levels of metacognition. Theoretical and practical implications for each of the five studies are given. Suggestions for further research and recommendations are also given since formative assessment as an instructional approach is yet to be extensively studied in Kenya.

Abbreviations

ADEA	Association for the Development of Education in Africa
AMOS	Analysis of Moment Structures
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
ASEI	Activity-based, Student-centered, Experiments and Improvisation
ATCS	Assessment and teaching of 21^{st}-century skills
AVE	Average Variance Explained
BA	Bachelor of Arts
BEd	Bachelor of Education
BSc	Bachelor of Science
CBA	Competence-Based Assessment
CBC	Competency-Based Curriculum
CEMASTEA	Centre for Mathematics, Science and Technology Education in Africa
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CMIN/DF	Minimum discrepancy per Degree of Freedom
CR	Composite Reliability
EBSCO	Elton B. Stephens Company
EFA	Education for All
ERIC	Education Resources Information Center
HTMT	Heterotrait-Monotrait Ratio of Correlations
NSET	In-Service Education and Training
IST	In-Service Teacher Training
Jr. MAI	Junior Metacognitive Awareness Inventory
KCPE	Kenya Certificate of Primary Education
KCSE	Kenya Certificate of Secondary Education
KICD	Kenya Institute of Curriculum and Development
KMO	Kaiser-Meyer-Olkin
KNEC	Kenya National Examinations Council

MAIT	Metacognitive Awareness Inventory for Teachers
MEd	Master of Education
NCTM	National Council of Teachers of Mathematics,
OECD	Organization for Economic Co-operation and Development
P21	Partnership for 21st Century Skills
PCD	Perceived Classroom Discussion
PD	Professional development
PDSI	Plan, Do, See, Improve
PF	Perceived Feedback
PLI	Perceived Learning Intentions
PLS	Partial Least Squares
PPA	Perceived Peer Assessment
PRT	Proportional Reasoning Test
PSC	Perceived Success Criteria
PSA	Perceived Self-Assessment
RMSEA	Root-Mean-Square Error of Approximation
SACMEQ	Southern African Consortium for Measuring Educational Quality
SDG4	Sustainable Development Goal 4
SEM	Structural Equation Modeling
SPSS	Statistical Package for the Social Sciences
SMASSE	Strengthening of Mathematics and Science in Secondary Education
SRMR	Standardized Root Mean Square Residual
TAFL-Q	Teacher Assessment For Learning Questionnaire
TALIS	Teaching and Learning International Survey
TLI	Tucker-Lewis Index
TCPD	Teacher Continuing Professional Development
USA	United States of America
UNESCO	United Nations Educational, Scientific and Cultural Organization

List of Tables

List of Figures

Chapter 1: Introduction

1.1. Context of the Study

Improving quality and learning is one of the central goals of the post-2015 global education agenda and learning measurement. The Education for All (EFA) global monitoring report of 2013/4 indicated that quality education in sub-Saharan Africa is lagging due to factors such as shortage of trained teachers despite the high rates of students' enrollment in schools (UNESCO, 2014). Report about secondary education, for instance, shows that the majority of sub- Saharan African countries have a high student-to-teacher ratio which exceeds 30:1. The report recognized the importance of teachers' role in solving the learning crisis and therefore recommended training and equipping teachers with innovative ways of teaching and learning as one of the ways of meeting the challenges. The commitment of the United Nations Educational, Scientific and Cultural Organization (UNESCO) in meeting the Sustainable Development Goal 4 (SDG4) on quality education is reflected in the recent initiatives of the Association for the Development of Education in Africa (ADEA, 2017).

The regional assessment of some of the Sub-Saharan countries done by the Southern African Consortium for Measuring Educational Quality (SACMEQ) revealed a worrying trend of only 25.0% of students attaining a basic level in mathematics education (Hungi et al., 2010). These worrying trends require proactive measures especially relating to quality in education. Studies have suggested that formative assessment has got a lot of benefits if well implemented (Black & Wiliam, 2009; Wiliam, 2011). Some of the benefits include empowering students to learn how to learn (Swaffield, 2011), and making learning visible (Black & Wiliam, 2009; Csíkos et al., 2012; Demetriou et al., 2011). Formative assessment is thus integral to the teaching and learning process and acquisition of competencies (Pepper, 2011; Voogt & Roblin, 2012).

1.2. Problem Statement

Kenya is among the countries adopting the Competency-Based Curriculum (CBC). For the successful implementation of the CBC in Kenya, there is a need to equip teachers with the skills and tools required to carry out formative assessment in classrooms. Teachers need to embrace instructional approaches that will enable learners to engage in lifelong learning. Mathematics teachers are specifically taxed with ensuring that students make sense of mathematics concepts and relate them to authentic situations. Teachers' perceptions, competencies, and practices are therefore crucial in determining the realization of competence-based learning.

The findings by the task force (Republic of Kenya, 2012) revealed that the current summative assessment in the Kenyan education system falls short of adequately measuring learners' abilities and teachers lack sufficient training on alternative assessments. There has been a continuous decline in mathematics performance in Kenyan secondary schools over the years despite the interventions that have been put in place (Wafubwa & Obuba, 2015). Among the intervention programmes being carried out in secondary schools, none has focused on formative assessment.

The present study first examined mathematics teachers' perceptions of formative assessment and their levels of metacognitive awareness. Secondly, a formative assessment intervention study was carried out to determine the impact of formative assessment on students' achievement in mathematics and metacognition. The overall aim of the study was to assess the preparedness of secondary school mathematics teachers in implementing a competency-based curriculum and to suggest the best formative assessment practices that can be used in mathematics classrooms. It's hoped that the findings of this research will contribute knowledge towards the development of an assessment framework in mathematics instruction for secondary schools in Kenya.

1.3. Organization of the Dissertation

This dissertation is composed of five chapters. Chapter one gives an introduction of the study focusing on the context, problem statement, significance, and an outline of how the dissertation is structured.

Chapter two is a review of literature on studies related to the research topic. The main focus was on formative assessment and the teaching and learning outcomes. The review was based on the theoretical framework of formative assessment by Black and Wiliam (2009). The five strategies of formative assessment were systematically reviewed and the results showed that formative assessment can improve teaching and learning. The review also pointed out the need for more experimental studies on the impact of formative assessment.

Chapter three gives the aims, research questions, hypotheses, methods, and the synopsis of the empirical studies that are fundamental to this dissertation. Methodologies used in the studies focus on design, sample and sampling procedures, data collection procedures, instruments, and their validations.

Chapter four presents the five empirical studies which have been compiled from five separate journal articles. Study one assessed the suitability of the adapted teacher assessment for learning questionnaire in the Kenyan context. Study two examined the relationship between formative assessment and mathematics teachers' metacognitive regulation. Study three looked at teachers' conception and perceptions of their metacognitive awareness. Study four involved the construction and validation of a test to measure students' proportional reasoning skills in mathematics whereas study five examined the impact of formative assessment on students' achievement in mathematics and their metacognitive awareness. Chapter five is the final chapter which gives the conclusion and recommendations of the research followed by the list of references and appendices.

Chapter 2: Literature Review

2.1. Introduction

Educators and researchers in mathematics education have continued to address new challenges that emerge as a result of the evolving world. Currently, cognitive researchers in mathematics education are directing their attention to higher levels of mathematical thinking, complex problem solving, conceptual understanding, and sense-making (Edwards et al., 2011). Current research in the field of mathematics education is informed by sociocultural, constructivist, and social constructivist theories of learning which focus mainly on the problem-solving strategies, metacognitive processes, and beliefs (Schoenfeld, 2010). Advances in instructional practices have implications for assessment practices. There is a need of aligning assessments that will support effective mathematics instruction. The assessment practices should thus focus on the role of the learner, peer, and the teacher (Black & Wiliam, 2009; Torrance, 2012; Swaffield, 2011). Advances in instructional processes are also leading to changes in assessment practices (Mullis & Martin, 2017).

In the Kenyan education system, assessment of students which is conducted at the end of primary and secondary education is mainly summative. Examinations are administered by the national examining board known as Kenya National Examinations Council (KNEC). The results of the Kenyan Certificate of Primary Education (KCPE) are used to determine placement at secondary schools whereas the Kenya Certificate of Secondary Education (KCSE) examination is used for placement of students in universities and tertiary institutions. There are several practical issues affecting mathematics education in Kenya. The commonly cited problems are related to poor teaching methods where most classes are dominated by teachers, leaving no room for learners' creativity (Ministry of Education, 2012; Pontefract & Hardman, 2005).

Instead of assessment being seen as part of the teaching and learning process, it is taken as a sieve to determine those who can move to higher education with limited space. As a result of competing for limited slots in the local universities, the instruction process is geared towards examinations as opposed to competencies applicable to life. Prerequisite knowledge is rarely specified before the start of a new topic. Students are therefore just taught before being made aware of the reasons and procedures (Ally, 2014; Ministry of Education, 2012).

This chapter on literature review first highlights the education structure and mathematics in-service teacher training needs in Kenya which is the context of the research. The concept of metacognition and how it is related to formative assessment is briefly discussed before reviewing empirical studies related to formative assessment and the learning outcomes. The chapter then gives a summary of the literature review followed by conclusions.

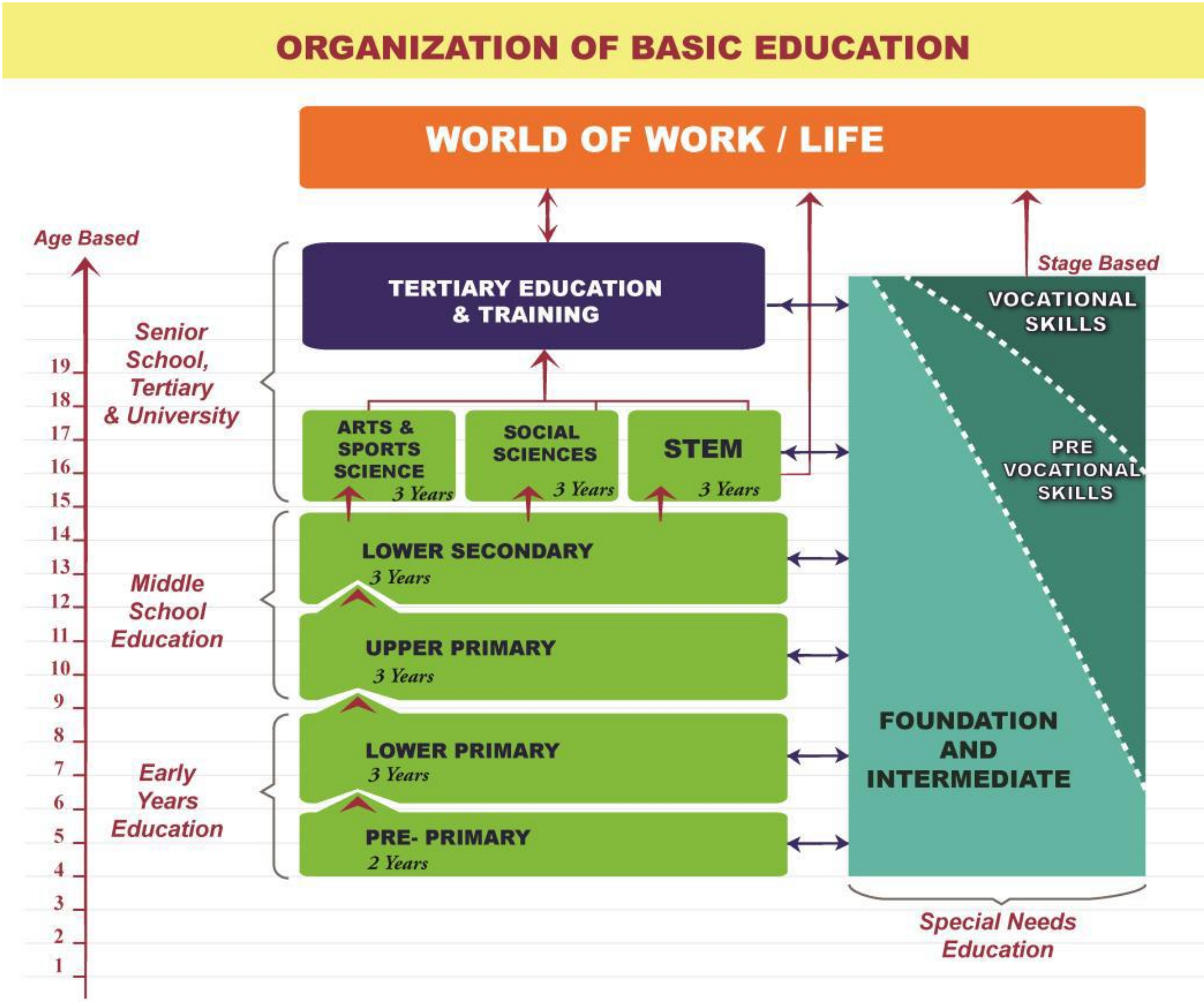
2.2. Education Structure in Kenya

There have been two education systems since Kenya gained independence in 1963. The first system established in 1963 was referred to as 7-4-2-3. This education structure was modeled after the British education system. The system was designed to provide 7 years of primary education, 4 years of lower secondary education, 2 years of upper secondary education, and a minimum of 3 years of University education. The second system known as 8-4-4 was launched in 1985 and was designed to provide 8 years of primary education, 4 years of secondary education, and a minimum of 4 years of university education. The emphasis was placed on Mathematics, English, and vocational subjects.

The country is however gradually implementing a new system that focuses on child development, skills, and competencies to be learned, and the outcome at each level from early childhood care and development to university level (see Figure 1). The new structure (2-6-6-3) referred to as the Competency-Based Curriculum (CBC) comprises two years of pre-primary, six years of primary (3 years lower and 3 years upper), six years of secondary (3 years junior and 3 years senior), two years minimum of middle-level colleges and 3 years minimum university education (KICD, 2017).

Figure 1

The New 2-6-6-3 Kenyan Education Structure



(Source: KICD, 2017, p. 28)

The new CBC is seen as the means through which the skills and knowledge needed for the economic growth and development in the country will be realized (KICD, 2017). The CBC is a learner-centered pedagogy that focuses on the learner's application of the skills learned to solve day-to-day challenges. According to KICD, competence is defined as "the ability to apply appropriate knowledge and skills to successfully perform a function" (p. 21). In the Kenyan context, every learner is expected to achieve seven core competencies which have been described as "communication and collaboration; self-efficacy; critical thinking and problem-solving;

creativity and imagination; citizenship; digital literacy; and learning to learn" (KICD, 2017, p. 21).

The assessment of competencies is expected to be achieved through a formative assessment approach which is deemed to enhance learning and learning outcomes. The implication is that teachers need a paradigm shift from the traditional teacher-centered assessment to student-centered assessment approaches. According to Clark (2012), formative assessment can be carried out by teachers and students as part of the day-to-day activity.

2.3. Mathematics Teachers' Beliefs about Teaching and Learning Mathematics

Philipp (2007) defined beliefs as psychologically held understandings about the world that are thought to be true. Beliefs are regarded as personal, steady nevertheless far beyond an individual's awareness (Cross, 2009). Studies have suggested that teachers' beliefs about teaching and learning have a significant influence on students' performance and it is needful that these beliefs are examined (Behrmann & Souvignier, 2013; Cross, 2009; Staub & Stern, 2002; Turner et al., 2009). A study by Staub & Stern (2002) for instance showed that cognitive constructivist orientation was associated with larger achievement gains in mathematical word problems as compared to a direct transmission approach. On the contrary, Behrmann & Souvignier (2013) noted that transmissive beliefs related positively to students' achievement. Most studies however support the constructivist belief approach as being positively related to learning achievement (e.g., Behrmann & Souvignier, 2015; OECD, 2009; Schunk, 2012). Adopting the right interventions can hence influence teachers' beliefs towards good classroom practices.

Most researchers on beliefs about the nature of teaching and learning have categorized beliefs based on teacher-centered and learner-centered approaches (Stipek et al., 2001; OECD, 2009). Beliefs based on teacher-centered approaches are regarded as transmission beliefs whereas those based on learner-centered approaches are regarded as constructivist beliefs (OECD, 2009). The direct transmission beliefs view teachers as the owners of knowledge and the students as the passive recipients of knowledge. The teachers' role is hence to transmit the knowledge to the students in a clear and structured way. In contrast, a constructivist requires students' active engagement in the process of acquiring knowledge. Teachers who hold the constructivist view emphasize understanding and problem solving as the context for knowledge construction (Schunk, 2012).

According to the National Council of Teachers of Mathematics (NCTM, 2015), beliefs can also be unproductive or productive. The unproductive beliefs relate to the direct transmission approach towards learning whereas the productive beliefs relate to the constructivist perspective of learning. The Teaching and Learning International Survey (TALIS) report by the Organization for Economic Co-operation and Development (OECD, 2009) showed that countries differ in the strength of teachers' endorsement of each of the two approaches. According to the TALIS report, direct transmission beliefs and constructivist beliefs about learning and instruction have been widely used in western countries.

Studies have also revealed that teachers' background factors such as gender, teaching experience, and professional development influence teachers' beliefs towards mathematics teaching and learning (Devine et al., 2013; OECD, 2009). In addition, OECD (2013) noted that teachers' beliefs vary across countries and schools. Studies by OECD (2009) revealed that female teachers are less likely than male teachers to view teaching as a direct transmission of knowledge and can easily embrace student-centered approaches to learning. A report by TALIS also postulated that teachers who take up professional development assume a variety of teaching approaches. Studies further suggest the need for schools to provide explicit training programs that target the modification of teachers' beliefs to provide an enabling learning atmosphere (OECD, 2013).

2.4. Mathematics Teachers' Continuing Professional Development and In-Service Teacher Training (IST) in Kenya

The Teachers' Continuing Professional Development (TCPD) in Kenya is mainly done through a cascade model where some teachers are trained in a particular content who in turn train their colleagues on the same (Kennedy, 2005; Ono & Ferreira, 2010). Though this cascade system is deemed to be economical since a large number of teachers can be reached within a short span, it has been less productive in Kenya. One of the programmes that have been carried out through the cascade model is the Strengthening of Mathematics and Science in Secondary Education (SMASSE) which was officially launched in Kenya in 1999 on a pilot basis and was later expanded to cover the entire country in 2004 (Ngugi & Nyakweba, 2005). Studies on the impact of SMASSE project in Kenya show that the project has failed to realize a significant positive impact on the teaching and learning of mathematics and science (Kiige, 2019; Mwangi & Mugambi, 2013; Wafubwa & Obuba, 2015). This could be because teachers passively absorb information from the facilitators who are seen as 'experts' and in turn pass the same to their learners.

The SMASSE project was born out of the need to improve the teaching, learning, and performance in mathematics and science subjects. The SMASSE In-Service Education and Training (INSET) programmes are facilitated by the Centre for Mathematics, Science and Technology Education in Africa (CEMASTEA) which was established in 2004 to provide INSET for mathematics and science teachers in Kenya and Africa. The main goal of CEMASTEA is to transform teaching by continuously developing competencies for effective curriculum delivery and improved quality of education. SMASSE programmes have since 2001 been spread to other African countries. The strategy that CEMASTEA adopted for pedagogical improvement has been the Activity-based, Student-centred, Experiments and Improvisation (ASEI) realized through the continual improvement cycle of Plan, Do, See, and Improve (PDSI) (SMASSE Project, 2008).

The CEMASTEA uses an inquiry-based approach which requires students' engagement in implementing rich problem-solving activities. This approach adopted by SMASSE project is meant to follow a Japanese classroom approach where students engage in the process of problem-solving using their strategies to solve challenging problems. This however calls for teachers who are well equipped to nurture students holistically. A study carried out by Inoue et al. (2019) revealed that the Japanese teachers modified their lesson plans on the spot to promote students' holistic development. The findings by Inoue and colleagues seem to suggest that the Japanese teachers are well versed on how to deal with eventualities in the course of lesson delivery and can therefore change their lesson plans to address any eventuality.

One of the reasons why SMASSE programmes may have failed to realize the expected impact could be attributed to the centralized nature of professional development (PD) programmes which limits teachers' active involvement in knowledge construction. One unique aspect of Japanese teachers is that they are always involved in action research through a lesson study system (Doig & Groves, 2011; Fujii, 2019). The Japanese teachers engage in PD through a lesson study system whose aim is to improve the effectiveness of the experiences that the teachers provide to their students. While engaging actively in lesson study through collaboration with one another, the Japanese teachers are always learning something new. In contrast, Kenyan teachers passively pass what they receive from the 'experts' through workshops to learners and therefore lack the opportunity for learning new things.

Professional development programmes in Kenya and Africa at large are facing challenges since most of them have not been developed from the teachers' perspective. Teachers fail to embrace the programmes because their views are left out (Mokhele & Jita, 2010). In Botswana, the failure of teachers' PD programmes has been attributed to a lack of teachers' involvement (Ramatlapana, 2009). When teachers are not involved in designing the programme, they may lack the motivation to implement it and end up doing it as a duty. A study carried out in Kenya by David and Bwisa (2013) revealed that few teachers were actively involved in continuous professional development due to a lack of support from the work environment and limited opportunities for career development. Mwangi and Mugambi (2013) also observed that SMASSE INSET programme hardly meets the needs of teachers. Research has shown that teachers have their own beliefs and perceptions which play a crucial role in the success of any new reform (Ige, 2014).

A report by the EFA global monitoring team on TCPD in Kenya suggested the use of field-based models consisting of school-based training as one of the ways of raising the quality of teaching practices (UNESCO, 2015). Face-to-face collaboration and distance learning have also been suggested as ways of cushioning the flaws of the cascade system of INSET delivery (Kafyulilo, 2013; Republic of Kenya, 2012). To effectively utilize the field-based models, it is necessary to establish the unique needs that Kenyan teachers face so that a more workable and productive in-service programme can be designed. One of the aims of this current study is to establish the perceptions of secondary schools' mathematics teachers regarding formative assessment. Understanding teachers' perceptions will help in designing interventions meant for classroom improvement and consequently teachers' professional development.

2.5. The Concept of Formative Assessment

Formative assessment came into the limelight following the seminal work of Black and Wiliam (1998) that involved the synthesis of over 250 studies, linking assessment and learning. The main features identified in these studies as forming part of formative assessment included sharing success criteria with learners; classroom talk and questioning; appropriate feedback; and peer and self-assessment. Reiterating on Black and Wiliam's work, Sadler (1998) defined formative assessment as the assessment which is intended to provide feedback on performance and accelerate learning. The Assessment reform group defined formative assessment (also referred to as assessment for learning) as involving the process of searching and clarifying evidence for use by learners and their teachers (Broadfoot et al., 2002).

During the third international conference on assessment for learning, another definition of assessment was proposed. Formative assessment was hence defined as a practice that teachers, students, and peers reflect on and respond to information emanating from it to strengthen learning (Klenowski, 2009). In line with their earlier definition, Black and Wiliam (2009) gave another definition of formative assessment as:

> Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. (Black & Wiliam, 2009, p. 9).

In all these definitions of formative assessment, the bottom line is formative assessment being seen as a daily classroom practice that is performed by the teacher, learners, and peers. Black and Wiliam's (2009) definition is however the one that is commonly cited in the literature related to formative assessment. In this definition, formative assessment is seen as a daily classroom exercise that is executed by three parties; teachers, students, and their peers. This means that formative assessment is more than teachers occasionally giving students tests or quizzes in preparation for a summative evaluation but it should be a daily classroom practice. Formative assessment in this context is supposed to show evidence of students' learning and inform the next steps of instruction. Wiliam (2011) noted that the basic idea behind formative assessment is that evidence of student learning is used to adjust instruction to better meet students' learning needs.

Whereas evidence in formative assessment is used to inform teaching and learning, in summative assessment, the evidence is used to report on where the learning is at a particular point in the learning process (Black & Wiliam, 2009; Wiliam, 2011). Unlike summative assessment which is used as a measurement instrument, formative assessment is designed to support teaching and learning continuously (Clark, 2012; Gipps, 2015). The aim of formative assessment is thus to keep track of how students are learning through continuous feedback which in turn helps in filling the learning gap. On the other hand, summative assessment evaluates how students have learned at the end of a course or unit based on some set standards.

2.5.1. Models of Formative Assessment

Formative assessment is one of the areas that have attracted researchers' attention as evidenced by several models. While there are so many models of formative assessment, this section only reviewed some of the recent models relevant to the current study.

Shepard (2006) model

Shepard's (2006) model focuses on ten characteristics of formative assessment which are supposed to act as a tool to guide student learning and to also guide teachers in their instructional practice. Shepard based her model on the earlier works of Black and Wiliam (1998) and Sadler (1989). The ten characteristics border on the responsibilities of both teachers and students. In this model, teachers' responsibilities include: communicating learning goals; identifying students' learning gaps; developing plans for attaining the desired goals; encouraging students to self-monitor their learning progress; providing examples of learning goals; providing frequent assessment; providing feedback, and promoting metacognition. Students on the other hand are expected to take responsibility for their learning and engage in peer and self-assessment.

Black and Wiliam (2009) model

In trying to develop the theory of formative assessment, Black and William (2009) conceptualized formative assessment as consisting of five key strategies which can be enhanced through sharing success criteria with learners, classroom questioning, making comments, peer and self-assessment, and formative use of summative tests. The key five strategies were identified as clarifying and sharing learning intentions and criteria for success; engineering effective classroom discussions and other learning tasks; providing feedback that moves learners forward; activating students as instructional resources for one another and activating students as the owners of their learning. The five key strategies were centered on three key processes: where the learner is going, where the learner is right now, and how the learner can reach the desired end.

Margaret Heritage Formative Assessment (FA) Model

According to Heritage's model (Heritage, 2007), there are four key elements of the formative assessment process: learning progressions, teacher assessment, effective feedback, and learner involvement. The outstanding characteristic of this model is its cyclic nature with the final target of closing the learning gap. According to this model, the formative assessment process begins with identifying the learning gap, followed by teacher assessment and effective feedback through learner involvement. This process is expected to reduce the learning gap through the learning evidence observed.

2.5.2. Conceptual Framework for the Current Study

The current research was informed by the above three frameworks of formative assessment but more specifically by Black and Wiliam's (2009) framework (see Table 1). The role of the three agents (teacher, peer, and learner) and five strategies of formative assessments in the learning process in the Black and Wiliam's framework are discussed as follows:

The Three Agents

The three agents are the teacher, peer, and learner who act as partners in the learning process. Formative assessment involves the teacher as the primary actor and students as partners. Black and William (2009) described the formative assessment process as a cycle in which a teacher is continually asking a series of three questions: Where are my students heading? Where are they right now? How can I close the gap between where they are and where I want them to be? The teacher's role involves identifying learning goals for the learners and then identifying where the learners are regarding those goals. The idea of closing the gap is achieved through timely, specific, corrective feedback; adjustments to instruction; and engaging peers in the support process.

Teachers' roles emphasize setting clear goals, making aspects of success explicit, providing useful feedback, and encouraging peer and self-reflection (Boud & Falchikov, 2007; Spiller, 2012). The learner has to take responsibility in the learning process by also reflecting on a series of questions (Where am I heading? Where am I right now? How can I close the gap between where I am and where I want to be?). By so doing, learners develop self-assessment skills which are crucial for life-long learning (Taras, 2010; Leach, 2012). A peer as an agent in the learning process plays a critical role in developing judgment skills through an effective feedback mechanism.

Five Key Strategies of Formative Assessment

i) Clarifying and sharing learning intentions and criteria for success.

This is the first strategy that involves clarifying, communicating, and understanding learning intentions and criteria for success with the learners. Teachers are expected to make the lesson objectives or intended outcome clear by ensuring the learners understand them clearly. According to Wiliam (2011), teachers, learners, and their peers should jointly break down this strategy into several criteria for success.

ii) Engineering effective classroom discussions and other learning tasks

The focus of this strategy is on eliciting evidence of achievement which mostly takes the form of questioning. This strategy revolves around the role of a teacher in finding out where the learners are in their learning so that the teacher can know the kind of evidence to collect (Wiliam, 2011).

iii) Providing feedback that moves learners forward

Research has shown that good feedback is among the most powerful influences on achievement (Hattie & Timperley, 2007; Owen, 2016). The importance of feedback is reflected in the seven principles of good feedback practice in learning proposed by Nicol and MacFarlane-Dickn (2006). Apart from teachers providing feedback, learners also engage in self and peer feedback. Black and Wiliam (2009) suggest that teachers can effectively provide feedback by comment-only marking as opposed to the use of grades.

iv) Activating learners as instructional resources for each other

This strategy involves peer assessment whereby learners are involved in collaborative learning. Wiliam (2011) observed that peer assessment which is geared towards improvement rather than evaluation can be more productive than when learners interact directly with a teacher. Wiliam further remarked that learners benefit more because they work towards a common goal which enhances motivation and clarity of concepts. He suggested strategies of peer assessment like peer evaluation of homework using the rubric created by the teacher, student feedback on other learners after instruction, and peers presenting their reviews to others.

v) Activating learners as the owners of their learning

Wiliam (2011) suggested that Students owning their learning can only occur within other strategies of formative assessment which involve: Sharing learning goals; promoting the belief that ability is incremental rather than fixed; discouraging students from comparing themselves with others in terms of achievement; providing feedback that encourages learning and promoting learning autonomy among students.

Table 1

Features of Formative Assessment

Agent	The direction in which the learner is moving	Current position of the learner	How the learner can get to the desired destination
Teacher	1. Sharing intentions for learning and criteria for success.	2. Classroom discussion	3. Feedback
Peer	Understand and share learning intentions and criteria for success	4. Peer assessment	
Student	Understand intentions of learning and criteria for success	5. Self-assessment	

(Adopted from Wiliam & Thompson, 2008, p. 63)

2.6. Metacognition and Formative Assessment

Metacognition was originally defined by Flavell (1979) and Brown (1978) as the knowledge about and regulation of one's cognitive activities in the learning processes (in Veenman et al., 2006). Knowledge of cognition involves awareness of and knowledge about one's cognition whereas metacognitive skills involve planning, monitoring, and evaluating learning processes (Veenman et al., 2004; Harris et al., 2010). In the revised Bloom's taxonomy of educational objectives, metacognitive knowledge was included as the new and fourth category of knowledge dimension (Krathwohl, 2002). This underscores metacognition as crucial in the teaching and learning process.

Theoretically, formative assessment has been posited to improve metacognition as reflected in the aforementioned models of formative assessment (e.g., Black & Wiliam, 2009; Shepard, 2006). Some aspects of formative assessment such as self-assessment have also been linked to metacognition. Andrade and Du (2007) defined self-assessment as a formative assessment process that enables students to reflect, evaluate and judge the quality of their work and their learning based on the set goals or criteria. According to Grantz and Gruber (2014), self-assessment should assist students to become aware of their performance, analyze their performance, judge their performance against predetermined criteria, and plan for future performance.

Learners engaged in formative assessment are guided by three questions that point to where they are heading, where they are currently, and how they can get to their desired end (Black & William, 2009). Studies have shown that self-assessment is one of the most important skills that students require for future professional development and life-long learning (Taras, 2010; Leach, 2012). Self-assessment has been shown to have positive effects on student performance with median effect size, Cohen's d between 0.40 and 0.45 (Brown & Harris, 2013). Self-assessment is thus a valuable classroom practice where students can evaluate their work and skills required for effective learning.

While discussing the benefits of self-assessment, Leach (2012) observed that self-assessment can enhance students' metacognitive engagement. Self-assessment has been regarded by Panadero and Alonso-Tapia (2013) as two-fold. First, it can be used as an instructional strategy and as a way of helping students regulate their learning. Students engage in self-regulating processes when they engage with their work and with the work of their peers (Nicol & Macfarlane-Dick, 2006; Sharma, 2016). It will be interesting to empirically find out how formative assessment is related to metacognition and how it can influence students' achievement in mathematics. It will also be interesting to see how formative assessment can impact students' metacognitive awareness. Research on the influence of formative assessment on metacognition is rare and the current study will contribute to the limited literature and form a basis for further research in this area.

2.7. Role of Formative Assessment in Improving Students' Learning Outcomes: A Systematic Review of Literature

2.7.1. Introduction

Formative assessment is part of an instructional process that involves continuous gathering, analyzing, and reflecting on evidence to make informed judgments and improve student learning (Black & Wiliam, 2009; Wiliam, 2011). Formative assessment also referred to as assessment for learning has gained a lot of prominence in the past few decades after the seminal work of Black and Wiliam (1998). There have however been a lot of controversies regarding the efficacy of formative assessment as reflected in several meta-analyses (e.g., Briggs et al., 2012; Dunn & Mulvenon, 2009; Kingston & Nash, 2011). The benefits of formative assessment as reported by Black and Wiliam (1998) were critiqued by Dunn and Mulvenon (2009) based on methodology and the context of the study.

Some of the methodological issues cited included the use of inadequate dependent measures, uncontrolled examiner expectancy, unchecked fidelity of treatment, and the use of inappropriate statistical units of analysis. The concept of formative assessment has also been faulted by some researchers. There have been claims that formative assessment practice is limited in terms of its scope and its utilization and hence reduced to classroom tests used for monitoring students' progress (Swaffield, 2011; Torrance, 2012). Lack of consistency in the definition and application of formative assessment practices was also pointed out by Dann (2014). According to Bennett (2011), defining formative assessment as an instrument or a process is an oversimplification. Despite the issues surrounding the concept and efficacy of formative assessment, a significant number of studies have reported the benefits of formative assessment (e.g., Anderson & Palm, 2017; Ozan & Kincal, 2018).

The meta-analysis and reviews on the impact of formative assessment have mainly focused on achievement outcomes. There are other effects of formative assessment that are rarely exploited. Cauley and McMillan (2010) described formative assessment as one of the most powerful ways to enhance student motivation and achievement. By way of formative assessment, students can focus on progress through ongoing assessment, get meaningful feedback, and see concretely how they can improve. Moss et al. (2011) used the metaphor of a windmill to visualize the formative assessment process and its effects. They related the way a windmill intentionally harnesses the power of moving air to generate energy to the way the formative assessment process helps students intentionally to harness the workings of their minds to generate motivation to learn. Formative assessment should, therefore, be seen as multidimensional and a focus on only one aspect can give misleading results.

The current review focuses on how the five strategies of formative assessment (Black & Wiliam, 2009) have been applied in the studies and their effects on students' learning outcomes. No study has systematically reviewed the use of these five formative assessment strategies in the literature. The current review is done systematically so that every aspect of formative assessment is given due attention. Unlike the previous reviews which have looked at formative assessment holistically, this review seeks to find out how each formative assessment strategy has been conceptualized and used in research. The main aim of the review is to find out the aspects of formative assessment that have been under-researched and form a basis for future research. Specifically, this review was guided by the following research questions:

1. How frequently have the formative assessment strategies been studied?
2. What is the effect of formative assessment strategies on students' motivation, engagement, and achievement?

2.7.2. Review Method

Procedure

The approach taken in this review was a stepwise process that entailed formulating research questions, defining search terms, selecting databases, conducting the literature search, formulating inclusion criteria, and applying all these to selected relevant literature, and the extraction of data (Popay et al., 2006).

Databases and Search Terms

The current thematic review synthesized the studies on the "Role of formative assessment in improving students' motivation, engagement, and achievement in secondary schools". The studies in the context of high schools and middle schools were considered secondary schools. The students' age range was generally from 12 to 18 years. Three databases: ERIC, EBSCO, and ELSEVIER were used to identify the studies. Some more papers were hand-searched using the Google Scholar search engine. Search terms that were used included "formative assessment or assessment for learning or formative feedback" AND ("motivation" OR "engagement" OR "secondary schools" OR "high school" OR middle school).

Criteria for Study Inclusion

The following criteria were used to determine whether the study would be included in this review: (1) the study had to be described as formative assessment or assessment for learning (AfL); (2) the study had to include empirical data and contain at least one of the five strategies of formative assessment; (3) participants belonged to secondary education set up; (4) the study must have been published between 2015 and 2019; (5) the study had to show the effect of formative assessment; (6) the study must have undergone a peer review and published in international journals; and (7) the study must have been published in English. Unpublished work such as master thesis and conference papers were excluded from this study.

Data Extraction, Analysis, and Coding

The initial search identified 832 records which were further narrowed down after going through a quality check. Some of the considerations for the quality check were: The clarity of the research objective; clear research approach; clarity of the context of research; well-described methods and justifications; clear analysis procedure; clarity in the presentation of the results; and the relevance of the study. After assessing for eligibility, the final synthesis included 38 studies. The studies were analyzed thematically based on the strategies of formative assessment. Key features from the articles were appropriately coded and used in the analysis. These key features included (1) authors and year of publication (2) study description (3) sample characteristics (4) research design and instruments used and (5) main findings.

2.7.3. Findings and Discussion

The initial search of the databases identified 832 records and 13 more records were identified through the Google Scholar search engine. After removing duplicates, 544 studies were obtained. The second screening based on the level of the study left 105 studies. Further screening based on the quality check excluded 59 studies from the records leaving 46 articles for eligibility assessment. Further examination of the full document led to the elimination of eight articles that failed to fully focus on secondary education but included primary (elementary) and university levels. Finally, 38 articles were deemed suitable for the final analysis. Out of the 38 articles, two (5.3%) focused on learning intentions and/or success criteria; seven (18.4%) articles focused on feedback; eight (22.2%) articles focused on peer and self-assessment, while 21 (56.3%) articles focused on formative assessment as a whole. In total, the literature review resulted in 38 articles that were analyzed, and the results are presented in Tables 2 to 6.

Learning Intentions (LI) and Success Criteria (SC)

Two studies focusing on learning intentions and/or success criteria were obtained from the search process (Table 2). A study by Crichton and McDaid (2016) investigated teachers' and students' perceptions regarding the use of learning intentions and success criteria strategies within lessons. Crichton and McDaid's study revealed that both teachers and students recognized the importance of LI and SC, especially during revision. Students however felt that these strategies were rarely discussed in the classrooms and teachers also expressed concerns about the implementation challenges. Teachers, therefore, felt the need for training on how to implement learning intentions and success criteria in their lessons.

The second study by Krijgsman et al. (2019) focused on the importance of goal clarification and its relation to feedback. Goal clarification involves making explicit the learning intentions through a feedback mechanism. The study revealed that goal clarification is an important element of feedback and needs satisfaction. The two studies seem to suggest that the formative assessment strategy of sharing learning intentions and success criteria are rarely implemented in schools despite the suggested benefits.

Table 2

Studies on Learning Intentions (LI) and Success Criteria (SC)

	Author	Study description	Sample characteristics	Design and instruments	Main findings
1	Crichton and McDaid (2016)	Teachers and students perceptions	20 teachers and 20 students, Scotland	Qualitative	LI and SC strategies are rarely implemented
2	Krijgsman et al. (2019)	Teachers' lesson variability	570 students, Netherlands	Quantitative	Goal clarification affects process feedback

Appropriate Feedback

A total number of seven studies on feedback were reviewed (Table 3). All these studies were quantitative with two of them using a quasi-experimental approach (Cutumisu & Schwartz 2018; Pinger et al., 2018a); two studies used a mixed design (Kyaruzi et al., 2018; 2019) while three used only a survey approach (Jónsson et al. 2018; Van der Kleij 2019; Vattøy & Smith 2019).

Whereas feedback has been seen as having the most powerful influence on achievement (Hattie, 2008), only two intervention studies examined the impact of feedback on students' achievement. Pinger et al.'s (2018a) study on the effectiveness of feedback showed positive effects on mathematics achievement and students' interest. The intervention study by Cutumisu and Schwartz (2018) also revealed improved performance in students who engaged with critical feedback. Kyaruzi et al. (2019) focused on the impact of students' perceptions of their mathematics teachers and the study revealed that feedback use predicted students' performance to a small extent. Jónsson et al.'s (2018) and Van der Kleij's (2019) studies focused on the comparison between teachers' and students' perceptions regarding feedback. Both studies revealed differing perceptions between teachers and students with teachers having high perceptions as compared to students'. Kyaruzi et al.'s (2018) study indicated that teachers' perceptions of formative assessment positively predicted the quality of feedback while Vattøy and Smith (2019) observed that students perceive their teachers' feedback to be more useful when they are aware of learning goals.

Analysis of studies on feedback in this review has shown that effective feedback can improve students' motivation and achievement. It is however worth noting that most studies focused on perceptions of teachers and students. Only two studies, one in the USA (Cutumisu & Schwartz, 2018) and another one in Europe (Pinger et al., 2018a) used an experimental approach to investigate the effect of feedback on learning outcomes. A similar observation was realized in the meta-analysis by Van der Kleij et al. (2015). According to Van der Kleij et al., only six experimental studies done between 1968 and 2012 in secondary education settings met the criteria for inclusion in the analysis.

Table 3

Studies on Appropriate Feedback

	Author	Study description	Sample	Design	Main findings
1	Jónsson et al. (2018)	Perception of teachers and students	56 teachers & 234 students, Iceland	Quantitative	Teachers had higher perceptions than students
2	Kyaruzi et al. (2018).	Effect of teachers' perceptions	54 mathematics teachers, Tanzania	Mixed method	Teachers' perceptions predicted feedback quality
3	Van der Kleij (2019).	Teacher and student perceptions	59 teachers & 186 students, Australia	Quantitative	More positive perceptions by teachers than students
4	Vattøy and Smith (2019)	Students' perceptions of teachers' feedback	1137 students (13–16 years), Norway	Quantitative	Students perceived teachers' feedback as more useful.
5	Cutumisu and Schwartz (2018)	Impact of critical feedback choice	106 grade 8 students, California	Quasi-experimental	Improved performance in students
6	Pinger et al. (2018a)	The effectiveness of feedback	17 teachers & 426 students, Germany	Quasi-experimental	Positive effects on maths achievement and interest
7	Kyaruzi et al. (2019)	Impact of students' perceptions	2767 form three students, Tanzania	Mixed-method	Feedback predicted mathematics performance

Peer Assessment (PA) and Self-Assessment (SA)

A total number of seven studies focused on peer assessment whereas only one focused on self-assessment (Table 4). Two studies were surveys on students' perceptions of peer assessment (Rotsaert et al., 2017) and teachers' perceptions of peer assessment (Rotsaert, Panadero, & Schellens, 2018). One study used a qualitative approach to study the implementation of peer assessment (Musfirah, 2019). The remaining five studies were experimental and sought to find out the effect of peer assessment and self-assessment in different learning conditions. For instance, Nikou and Economides (2016) looked at the impact of self-assessment on student motivation and achievement. Rotsaert, Panadero, Schellens, and Raes (2018) examined the effects of peer assessment practice on peer feedback quality; Tsivitanidou et al. (2018) studied reciprocal peer assessment as a learning tool; Vanderhoven et al. (2015) investigated the effect of anonymity in peer assessment whereas Hsia et al. (2016) examined the effects of web-based PA approach. All the interventions in these experimental studies were reported to have positive effects on students' learning outcomes.

Peer assessment involves students giving feedback to peers' work which can either be verbal or in written form. According to Boud and Falchikov (2007), peer assessment involves feedback on a product or a performance, based on the criteria of excellence for that product. Self-assessment on the other hand is defined as a formative assessment process that enables students to reflect, evaluate and judge the quality of their work and their learning (Andrade & Du, 2007; Grantz & Gruber, 2014). Only one study (Nikou & Economides, 2016) in this review directly addressed the self-assessment strategy. The majority of the studies focused specifically on peer assessment. Most studies on self-assessment were excluded because they focused less on formative assessment instead most of them were geared towards self-regulated learning (Panadero & Alonso-Tapia, 2013).

The current trend on the studies on peer assessment in secondary schools is quite encouraging since most studies are increasingly focusing on intervention programs as opposed to surveys. Previously, studies were focused more on universities. For instance, in Van Gennip et al.'s (2010) literature review, only one study out of 15 studies focused on secondary education. The current study reviewed seven studies on peer-assessment out of which more than half were experimental. Although the survey studies revealed differing perceptions of teachers and students regarding peer assessment, the experimental studies showed that peer assessment can improve students' learning motivation and achievement. The majority of these studies (75.0%) were however carried out in Europe and only 25.0% in Asia. Evidently, from this analysis, peer assessment works best in online environments, and when students do it anonymously. These findings are supported by previous studies such as Tenório et al. (2016) and Fu et al. (2019).

Table 4

Studies on Peer Assessment (PA) and Self-Assessment (SA)

	Author	Description	Sample	Design	Main findings
1	Rotsaert et al. (2017)	Students' perceptions of PA	3680 students, Belgium	Quantitative	Perceptions predicted by trust
2	Nikou & Economides (2016)	SA on student motivation and achievement	66 students, Europe	Quasi-experimental	Increased learning achievement and motivation.
3	Musfirah (2019)	PA in teaching speaking skill	1 high school, Indonesia	Qualitative	The use of PA motivates students
4	Rotsaert, Panadero, & Schellens, (2018)	Use of PA among teachers	225 teachers, Belgium	Quantitative	Accuracy of PA predicts belief in educational value
5	Rotsaert, Panadero, Schellens, & Raes (2018)	Effects of PA practice	36 students, Belgium	Quasi-experimental	Peer feedback improves the quality of peer feedback
6	Tsivitanidou et al (2018)	Reciprocal PA as a learning tool in	22 students, Switzerland	Quasi-experimental	Reciprocal PA facilitates students' learning in science
7	Vanderhovn et al. (2015)	Effect of anonymity in PA	2 teachers & 69 students, Belgium	Quasi-experimental	Pupils felt more positive towards anonymity in PA
8	Hsia et al. (2016)	Effects of web-based PA	163 junior high students, Taiwan	Quasi-experimental	Improved learning performance and motivation

All the Five Strategies Combined

After analyzing studies with specific strategies as discussed above, studies that involved a combination of formative assessment strategies or at least examined the effect of formative assessment, in general, were analyzed together as shown in Table 5. In total, 21 studies focused on a combination of strategies out of which five investigated either teachers' perceptions, students' perceptions, or both (Burner, 2016; Dobish & Meyer, 2017; Saito & Inoi, 2017; Ozan & Kincal, 2018; Kippers et al., 2018; Rakoczyet al., 2019; Johnson et al., 2019). Five studies were experimental. Two quasi-experimental studies (Vogelzan & Admiraal, 2017; Pinger et al., 2018b) investigated the effect of formative assessment on chemistry achievement and instructional quality respectively.

Two longitudinal experimental studies (Wylie & Lyon, 2015; Furtak et al., 2016) determined the quality of formative assessment implementation after teachers' professional development. Rakoczy et al. (2019) however used a cluster randomized field trial with pre-tests and post-tests. One study (Yin & Buck, 2019) used collaborative action research to negotiate the understanding of formative assessment among teachers. Two qualitative studies (Van der Nest et al., 2018; Beesley et al., 2018) focused on teachers' professional development while three studies (Brink & Bartz, 2017; Cisterna & Gotwals, 2018; Lyon et al., 2019) focused on the implementation of formative assessment. Other non-experimental but quantitative studies also focused on the implementation and use of formative assessment (Bulunuz et al., 2016; Saito & Inoi, 2017; Choi et al., 2018; Xiao & Yang, 2019).

Table 5

Studies on a Combination of Formative Assessment (FA) Aspects

	Study	Description	Sample	Design	Main findings
1	Dobish and Meyer (2017)	Teachers' perceptions	305 teachers, USA	Mixed method	Improved teachers' practice and students' learning
2	Bulunuz et al. (2016)	Use of FA probes	61 students, Turkey	Quantitative	Improved students' performance.
3	Choi et al. (2018)	Automatic Item Generation (AIG)	57 students & their teachers, Korea	Mixed-method	AIG can be utilized for students and teachers
4	Burner (2016)	FA in English as a foreign language	4 teachers and 100 students, Norway	Quantitative	Differing teachers' and students' perceptions
5	Pinger et al. (2018b)	Effects of FA on instruction	35 teachers & 859 students, Germany	Quasi-experimental	Improves achievement and saves instructional time.
6	Ozan and Kincal (2018)	Effects of FA on achievement	45 students, Turkey	Mixed method	Increased achievement and better attitudes
7	Vogelzan and Admiraal (2017)	Effects of FA on achievement	69 students, the Netherlands	Quasi-experimental	Positive effect on students' achievement
8	Kippers et al. (2018)	Teachers' view on assessment for learning	479 teachers, the Netherlands	Mixed method	Lack of AfL integration into teacher practice
9	Furtak et al. (2016)	Teachers' FA abilities	9 teachers, USA	Longitudinal Experimental	Increase in teachers abilities to use FA strategies
10	Cisterna and Gotwals (2018).	Teachers enactment of FA	4 in-service science teachers, USA	Qualitative	Teachers struggled with integrating FA practices

	Study	Description	Sample	Design	Main findings
11	Yin and Buck (2019)	Collaborative action research	2 chemistry teachers, USA	Experimental	Teachers need support to use action research in FA
12	Lyon et al. (2019)	An integrated approach to FA	Six teachers, USA	Qualitative	Implementation depends on the interaction of FA practices
13	Wylie and Lyon (2015)	Breadth and quality of FA	202 teachers, USA	Experimental-longitudinal	Significant improvements in some aspects of FA
14	Xiao (2017)	Formative tests	3 classes, China	Qualitative	tests used to a certain degree
15	Xiao and Yang (2019)	FA and students' self-regulation	2 teachers, China	Qualitative	FA activities enhance students' self-regulation
16	Saito and Inoi (2017)	English Language teachers and FA	727 teachers, Japan	Quantitative	Varying degrees of FA use among teachers
17	Brink and Bartz (2017)	Teachers' use and perceptions of FA	3 teachers, USA	Mixed method	Positive impact on the use and perceptions of FA
18	Van der Nest et al. (2018)	Impact of FA activities on	Grade 9 maths teachers, S. Africa	Qualitative research	Need for multiple dimensions to understand maths concepts
19	Beesley et al. (2018)	FA professional development	7 schools, 47 teachers, USA	Qualitative	Improved teachers' practice of FA.
20	Rakoczy et al. (2019)	FA, interest, and achievement	26 teachers and 18 schools, Germany	Experimental	The indirect effect on interest
21	Johnson et al. (2019)	FA in public school districts	1,097 teachers from the USA	Mixed method	Less use of FA strategies like LI, SC, and Feedback

Although this review of literature aimed to find out how specific strategies of formative assessment were applied in research, 21 studies focused on formative assessment as a whole. The analysis of these studies has revealed that the impact of formative assessment is more pronounced when the implementation is done within the context of teachers' professional development (e.g., Wylie & Lyon, 2015; Brink & Bartz, 2017; Dobish & Meyer, 2017; Beesley et al., 2018). Generally, formative assessment is seen to have a positive effect on learning outcomes when well implemented. One outstanding fact is that most studies were focused on teachers' and students' attitudes. The sample sizes in the experimental studies were however small thus limiting the generalization of the results. A summary of all the studies based on different strategies is presented in Table 6.

Table 6

Summary of the Studies on the Five Aspects of Formative Assessment

Aspect	*n*	%	Design	Purpose	Results
Learning Intention (LI) and Success Criteria (SC)	2	5.4	Qualitative and quantitative	Teachers' and students' perceptions	Difficulties in implementing the strategies by teachers
Appropriate Feedback	7	18.9	Qualitative, quantitative, experimental	Effect of feedback use	Increase in motivation and achievement
Peer Assessment (PA) and Self-Assessment (SA)	8	18.9	Qualitative, quantitative, experimental	Teachers' and students' perceptions	Increased motivation; differing perceptions of teachers and students
All the five strategies combined	21	56.8	Qualitative, quantitative, experimental	Teachers' and students' perceptions	Enhances motivation, engagement, and achievement

2.8. Summary of the Literature Review

This chapter reviewed the literature on formative assessment and its impact on learning outcomes. The concept of formative assessment was reviewed starting with Black and Wiliam's (1998) definition. The common aspect in the various versions of the definition was the fact that formative assessment is a process of learning as opposed to summative assessment which acts as a measurement instrument. Various models were also reviewed and the striking similarity in all the models is the emphasis placed on the role of the teacher, learner, peer, and feedback in the learning process. The three agents of formative assessment (teacher, peer, and learner) were also explained briefly by looking at how they relate to formative assessment. The link between formative assessment and metacognition was briefly reviewed although empirical studies on the impact of formative assessment on metacognition were limited.

The main focus of the literature review was on peer-reviewed empirical studies that have been done within a period of five years (2015- 2019). The review was limited to international peer-reviewed studies done in English and within secondary schools' education settings. The term secondary school was used to generally include studies done in middle schools and high schools. The aim of carrying out this review was to find out the extent to which formative assessment strategies have been addressed in the studies and their influence on teaching and learning outcomes. The review was carried out systematically and analysis was done thematically based on the strategies of formative assessment described by Black and Wiliam (2009). An overview of the results revealed that most of the studies were done in Europe (17 studies), followed by the USA (11 studies), Asia (6 studies), Africa (3 studies), and Australia (1 study).

The synthesis revealed that feedback and peer assessment strategies have significantly been studied in the past five years, the majority of studies having been done in 2018 and 2019. Only two studies addressed learning intention and success criteria making it a potential strategy for future inquiry. Although there is evidence from the past studies that self-assessment can improve students' performance (Brown & Harris 2013), only one study in this review met the criteria for inclusion. The focus of most studies on self-assessment has been its relationship with students' self-regulation (Panadero et al., 2017).

There is a need for more clarity in the concept of self-assessment especially in the context of formative assessment. Seemingly, most studies have used the term self-assessment in the studies without a solid theoretical framework. These sentiments are also echoed in the study by Panadero et al., (2016) who argued that there are different conceptions of student self-assessment components yet they have been treated uniformly in the field of education. Self-assessment in this current study was conceptualized as an instructional strategy in which students engage with their work (Panadero & Alonso-Tapia, 2013) and how the strategy impacts students' achievement, engagement, and motivation.

There was no specific study focusing on the use of classroom discussion and questions which are also the main strategies of formative assessment. Although all studies were quantitative, most used questionnaires to get the teachers' and students' perceptions of formative assessment. There is a need for more focus on experimental research so that more realistic conclusions on the role of formative assessment can be obtained. In the current review, only 11 studies out of 38 were experimental. The results of these experimental studies pointed out a varying degree in the use of formative assessment strategies by teachers after undergoing professional development and a positive educational outcome for students in the treatment groups.

There is a need for more research to focus on intervention studies. The studies have suggested that the key elements of formative assessment can increase students' learning outcomes if well implemented. An important observation in this review is an increase in the use of mixed methods unlike before where the studies on formative assessment were characterized as flawed due to methodological issues (Dunn & Mulvenon, 2009; Kingston &Nash. 2011).

2.9. Conclusion and Practical Implications

The success stories of formative assessment dating back to the work of Black and William (1998) have mostly been documented in western countries. There is no documented evidence of the implementation of formative assessment in Kenya despite its empirically proven positive outcomes. One of the concerns in sub-Saharan Africa and Kenya, in particular, is on lack of quality education which is evidenced through poor employability skills in the job markets (Ally, 2014). Research has shown that learning in this 21st century has become more complex and needs a more proactive way of assessing the competencies (Boud & Falchichar, 2007). The ineffectiveness of traditional assessment practices to equip students for lifelong learning is supported by OECD (2014). The overarching need in the 21st century is equipping students with learning how to learn and this can effectively be done by embracing classroom assessment practices that enhance learning transferability.

The foregoing literature points out the need to conduct research using more efficient methodologies and design for more conclusive results on the impact of formative assessment on teaching and learning. The findings of this review have shown that formative assessment can improve learning outcomes especially when teachers undergo professional development. This implies that schools should embrace in-service training of teachers on how to effectively implement formative assessment. Secondly, teachers should be encouraged to use formative assessment strategies for better learning outcomes for their students.

Chapter 3: Research Aims and Methodology of the Empirical Studies

3.1. Introduction

This chapter discusses the overall aims of the empirical studies and the methods used in the studies. In line with the research aims, research questions and hypotheses are also outlined. While this chapter only gives a brief overview of the empirical studies, a detailed discussion of each empirical study is presented in chapter four. The five studies forming the empirical part of this dissertation are published journal articles.

3.2. Research Aims and Structure of Empirical Studies

Kenya is among the countries that have adopted the implementation framework for 21st-century competencies through Competency-Based Curriculum (CBC). So far CBC has been implemented in the lower primary schools from grades one to four. Plans are still underway to implement the curriculum in upper primary and secondary schools. The competencies are expected to be realized through learner-centered teaching approaches which can enhance learning and learning outcomes. For this to be realized, teachers need to change their teaching approaches from conventional ways to student-centered approaches.

Currently, secondary schools in Kenya are still dominated by teacher-centered teaching approaches with a summative evaluation of students. The CBC encourages Competence-Based Assessment (CBA) as opposed to summative assessment. The Kenyan curriculum framework defines CBA as being able to determine the potential of applying a set of related knowledge, skills, and abilities to successfully perform a task in a particular setting (KICD, 2017, p. 114). Studies have shown that formative assessment when used as an instructional approach can lead to many benefits which include making learning visible (Black & Wiliam, 2009; Demetriou et al., 2011. Formative assessment can therefore be used in the instructional process to enable learners to acquire competencies (Pepper, 2011; Voogt & Roblin, 2012).

In this series of studies, formative assessment has been conceptualized as a teaching approach with the overall aim of assessing its effectiveness in the teaching and learning process in Kenyan secondary schools. This dissertation thus comprises five related empirical studies. The first study was a pilot study that examined the suitability of the adapted questionnaire for measuring mathematics teachers’ perceptions of formative assessment. The study aimed to find out if the Teacher Assessment for Learning Questionnaire (TAFL-Q) which was developed in the Netherlands by Pat-El et al. (2013) could be applicable in the Kenyan context.

The second study used the validated TAFL-Q to examine the teachers' perceptions and how these perceptions are related to the teachers' metacognitive awareness. A review of the literature revealed that teachers can model students' metacognitive skills (Martinez, 2006; Tanner, 2012). Metacognition is regarded as a very important skill for enhancing 21st-century skills and competencies (Lai & Viering, 2012). In line with the context of the CBC in Kenya, teachers' metacognition is important in developing learners' competencies. The motivation for the second study was thus to find out if Kenyan mathematics teachers' perceptions of formative assessment were in any way related to their perceived metacognition.

The third study was informed by the results of the second study which showed that formative assessment predicted mathematics teachers' levels of metacognitive awareness. The aim was therefore to assess how mathematics teachers' perceived their levels of metacognition and their conception of the different facets of metacognition.

Study four involved the development of a test to measure proportional reasoning skills in mathematics. The test was later used in study five which was an intervention that entailed using formative assessment as an instructional approach to improve students' achievement in mathematics and metacognition. The proportional reasoning test and a metacognitive awareness scale were used to measure the impact of the intervention.

In summary, the studies forming part of this dissertation were carried out in the context of the Kenyan new competence-based curriculum. The overall aim of these studies is to equip teachers and curriculum developers with alternative teaching and learning approaches that can promote the key competencies in mathematics. Table 7 gives a summary of the empirical studies and the timelines.

Table 7

Timeline and Empirical Studies

Timeline	Research activities	Instruments	Samples
June to July 2019	Piloting the adapted research instruments in secondary schools	Teacher Assessment For Learning questionnaire-TAFL-Q (Pat-El et al., 2013).	Mathematics teachers; N=180
		Metacognitive Regulation Inventory for Teachers-MAIT (Balcikanli, 2011)	Mathematics teachers; N=180
July to August 2019	Collecting data using the validated instruments. Examining the relationship between formative assessment and mathematics teachers' metacognitive regulation.	TAFL-Q (Pat-El et al., 2013). MAIT (Balcikanli, 2011)	Mathematics teachers; N=213
August 2019	Interviewing teachers on metacognition	Interview Schedule (Researcher developed)	Mathematics teachers; N=10
January to March 2020	Constructing a proportional reasoning test (PRT). Piloting the test and assessing the psychometric properties	Proportional Reasoning Test (PRT).	Grade 11 students; N=45
December 2020 to February 2021	Conducting a teacher training on formative assessment and an intervention study on formative assessment	PRT (Wafubwa et al., 2020). Junior Metacognitive Awareness Inventory-Jr. MAI (Sperling et al., 2002)	Grade 11 students; N=164 Mathematics teachers; N=4

3.3. Research Questions and Hypotheses

This section presents the research questions and the hypotheses answered in the five studies that form the empirical part of this dissertation.

3.3.1. Study 1

Study one involved the validation of the adapted teacher assessment for learning questionnaire in the Kenyan context. The following two main research questions and the corresponding hypotheses were answered.

RQ1: What is the evidence of validity for the two-factor model of the Teacher Assessment For Learning questionnaire (TAFL-Q) in the Kenyan context?

H1: It is expected that the two-factor model of TAFL-Q will result in acceptable fit indices: normed chi-square (CMIN/DF) of between 1.0 and 3.0, the root-mean-square error of approximation (RMSEA) of $\leq.080$, standardized root mean square residual (SRMR) of $\leq.080$ Tucker-Lewis index (TLI) of $\geq.900$, and the comparative fit index (CFI) of $\geq.900$ (Ho, 2006; Hooper et al., 2008; Teo, 2013).

RQ2: What are the mathematics teachers' overall perceptions of formative assessment?

H2: As measured on a 5-point Likert scale (Strongly Disagree=0; Disagree=1; Undicided=2; Agree=3; Strongly Agree=4), it is expected that teachers' ratings correspond with "Agree" implying a high perception of formative assessment.

3.3.2. Study 2

In study two, the relationship between teachers' formative assessment and their metacognitive awareness was examined. The validated TAFL-Q and a metacognitive awareness inventory for teachers (MAIT) scale were used to assess the relationship. Four questions and corresponding hypotheses were answered in this study.

RQ1: What is the evidence of validity and reliability for the six-factor model of the TAFL-Q?

H3: It is expected that the scales of the six-factor model of the TAFL-Q will have Cronbach's alphas (α) of $\geq .60$ and the fit indices of the measurement model will be within the acceptable range (CMIN/DF of between 1.0 and 3.0; RMSEA of $\leq.080$; SRMR of $\leq.080$; TLI of $\geq.900$; and CFI of $\geq.900$ (Ho, 2006; Hooper et al., 2008; Teo, 2013).

RQ2: What is the evidence of validity and reliability of the MAIT scale?

H4: It is expected that the scales of the MAIT will have Cronbach's alphas (α) of $\geq .60$ and the fit indices of the model measurement will be within the acceptable range: CMIN/DF of between 1.0 and 3.0; RMSEA of $\leq.080$; SRMR of $\leq.080$; TLI of $\geq.900$; and CFI of $\geq.900$ (Ho, 2006; Hooper et al., 2008; Teo, 2013).

RQ3: How does the hypothesized structural relationship among the predictor and the observed variables fit the data?

H5: It is expected that the hypothesized structural TAFL-Q and MAIT model will have acceptable fit indices: CMIN/DF of between 1.0 and 3.0; RMSEA of $\leq.080$; SRMR of $\leq.080$; TLI of $\geq.900$; and CFI of $\geq.900$ (Ho, 2006; Hooper et al., 2008; Teo, 2013).

RQ4: How are the teachers' perceptions of formative assessment and their metacognitive skills related?

H6: It is hypothesized that there will be a positive relationship between the teachers' perceptions of formative assessment and their metacognitive skills.

3.3.3. Study 3

Study three investigated the in-service mathematics teachers' conception and perceptions of metacognition by answering three main questions and three hypotheses.

RQ1: To what extent do secondary school mathematics teachers in Kenya perceive their use of metacognitive knowledge and skills in teaching mathematics?

H7**:** Based on a 5-point Likert scale used to rate the items (Not at All=0; Rarely=1; Sometimes=2; Often=3; Always=4), it is expected that the mathematics teachers' perceptions would correspond to "Often" implying a high perception of their use of metacognitive knowledge and skills.

RQ2: What is the effect of gender, teaching experience, and level of education on the metacognitive awareness of secondary school mathematics teachers in Kenya?

H8: It is expected that the three background factors of gender, teaching experience, and level of education will have non-significant effects on mathematics teachers' metacognitive awareness.

RQ3: What conception of metacognition do Kenyan secondary school mathematics teachers report?

H9: It is expected that mathematics teachers will conceptualize metacognition in terms of knowledge about cognition and regulation of cognition.

3.3.4. Study 4

In study four, a proportional reasoning test to measure students' proportional reasoning skills on rates, ratios, and proportions was constructed and validated. In this study, two questions and two hypotheses were answered.

RQ1: What is the evidence of content-related validity of the constructed proportional reasoning test (PRT)?

H10: Based on the judgment by the subject matter experts, it is expected that the PRT will have content-related validity.

RQ2: What is the evidence of reliability and discriminant validity of the PRT?

H11: It is expected that the PRT will have Cronbach's alphas (α) of $\geq .60$, Difficulty Index (DI) range between 0.3 and 0.6; and Item Discrimination (ID) range between .20 and .50. This will imply a reliable and valid test (Ebel, 1979; Hopkins, 1998), and hence suitable for assessing students' proportional reasoning skills in mathematics.

3.3.5. Study 5

Study five assessed the impact of formative assessment as an instructional approach on students' mathematics achievement and their metacognitive awareness. The study addressed four main research questions and six hypotheses.

RQ1: How does the teaching approach influence students' performance on mathematics posttest between the intervention and control groups?

H12: It is expected that students exposed to formative assessment in the intervention group will have a higher mean score in the mathematics posttest as compared to the students in the control group. Studies show that formative assessment improves students' performance (Vogelzan & Admiraal, 2017; Andersson & Palm, 2017; Ozan & Kıncal, 2018).

RQ2: How does gender influence students' performance on mathematics posttest? What is the interaction effect between gender and the type of teaching approach, and mathematics posttest?

H13: Gender is expected to have a non-significant effect on students' posttest (Lindberg et al., 2010; Louis, & Mistele, 2012).

H14: It is expected that there will be a non-significant interaction effect between gender and the teaching approach, and mathematics posttest.

RQ3: How does the teaching approach affect the students' ratings on their levels of metacognitive awareness between the intervention and the control groups after the treatment?

H15: A significant difference in the students' ratings on their levels of metacognition in favor of the intervention group after the treatment is expected. Students in the intervention condition are exposed to monitoring, planning, and evaluation activities that promote metacognitive awareness (Kostons et al., 2012; Tanner, 2012).

RQ4: How does gender influence students' metacognitive awareness after the treatment? What is the interaction effect between gender and the type of teaching approach, and metacognitive awareness posttest?

H16: Gender is expected to have a non-significant influence on students' metacognition posttest. Studies have shown that metacognition is gender independent (Al Shabibi, & Alkharusi, 2018; Siswati & Corebima, 2017).

H17: It is expected that there will be a non-significant interaction effect between gender and the teaching approach, and metacognition posttest.

3.4. Research Methodology

3.4.1. Research design

Since this dissertation comprises five empirical studies, different research designs were employed depending on the research questions. First, a descriptive survey design was used in the first study to validate the adapted teacher for learning assessment questionnaire in the Kenyan context. The second study utilized a correlational approach to examine the relationship between formative assessment and mathematics teachers' metacognitive regulation. The third study employed a descriptive cross-sectional survey design that utilized both questionnaires and personal interviews. The third study aimed to find out how mathematics teachers perceive and understand the concept of metacognition.

The fourth study involved the construction and validation of an instrument for measuring proportional reasoning skills in mathematics. The study, therefore, employed a descriptive research design to show the psychometric properties of the test items. Study five applied a quasi-experimental pretest-posttest non-equivalent group design to assess the impact of the formative assessment intervention on learners' achievement in mathematics and their levels of metacognition. This design involved a between-subject independent variable with two levels and a within-subject independent variable with two levels. This design is considered a strong quasi-experimental approach since the researcher has control over the independent variable (Gliner et al., 2016).

3.4.2. Samples

The main research was carried out in Bungoma County which is one of the 47 counties in Kenya with an approximate population of 1.7 million people (Kenya National Bureau of Statistics, 2019). All 47 counties have similar characteristics in terms of the curriculum and distribution of teaching and learning resources. Bungoma County has nine Sub-counties with slightly more than 600 mathematics teachers. To enhance population validity, stratified random sampling was used to select mathematics teachers from the three school categories and three school types as shown in Table 8.

For the intervention study, four schools were randomly selected from 159 mixed sub-county secondary schools. After identifying the four schools, two schools were randomly assigned to the experimental group and two to a control group. Before the intervention study, the instruments were validated in a different school with a sample of 45 students. The target population for the intervention study was the form three (grade 11) students. This implied that a follow-up study could be carried out to assess the students' achievement during summative evaluation which is done at the end of secondary education (grade 12). In total, 393 teachers and 209 students formed the research sample. Table 9 shows the detailed composition of the samples and the instruments used in each study.

Table 8

Target Population

School category	School type				The approximate number of mathematics teachers
	Boys	Girls	Mixed	Total	
1 National schools	1	1	None	2	12
2 County schools	22	11	8	41	200
3 Sub county schools (district)	5	35	159	199	386
Total	28	47	167	242	598

Table 9

Study Samples

Studies	Samples	Instruments
Study 1	Mathematics teachers (N=180)	TAFL-Q (Pat-El et al., 2013)
Study 2	Mathematics teachers (N=213)	TAFL-Q (Pat-El et al., 2013) and MAIT (Balcikanli, 2011)
Study 3	Mathematics teachers (N=213)	MAIT (Balcikanli, 2011) and Interview Schedule (researcher-developed)
Study 4	Grade 11 students (N=45)	Proportional Reasoning Test (PRT)
Study 5	Grade 11 students (N=164)	PRT(Wafubwa et al., 2020) and Junior Metacognitive Awareness Inventory-Jr. MAI (Sperling et al., 2002)

3.4.3. Instruments for the Empirical Studies

Five instruments were used in this series of studies out of which three were adopted and two were researcher-made. The first instrument used in the first and second studies was a Teacher Assessment For Learning Questionnaire (TAFL-Q) which was developed by Pat-El et al. (2013). The TAFL-Q was originally developed in the Netherlands and it was necessary to test its suitability in the Kenyan context. The instrument was therefore piloted on a sample of 180 teachers and some modifications were made on it before it was used in the second study on a sample of 213 mathematics teachers.

The second instrument was the Metacognitive Awareness Inventory for Teachers (MAIT) which was used in the second and third studies. The MAIT scale was developed by Balcikanli (2011) to measure teachers' metacognitive awareness. The MAIT scale comprises two dimensions each with 12 items. The first dimension measures metacognitive knowledge and it has three subscales each with four items. The second dimension is the metacognitive regulation also comprising of three subscales. The scale was first piloted on a sample of 180 teachers before being used for the second study with a sample of 213 teachers.

The third instrument was the interview schedule which was developed by the researcher based on the two dimensions and six subscales of the MAIT scale. The schedule consisted of eight questions which were meant to get more insight into the themes covered in the MAIT scale. The questions addressed the teachers' conception and use of metacognition in teaching.

The fourth instrument was a proportional reasoning test (PRT) which was developed by the researchers (Wafubwa et al., 2020) to measure students' achievement in mathematics. The PRT consisted of 10-word problems covering five aspects of proportional reasoning. The instrument was first piloted on a sample of 45 students in the fourth study before it was used in the fifth study on a sample of 164 students.

The fifth instrument was the Junior Metacognitive Awareness Inventory (Jr. MAI) which was adopted from Sperling et al. (2002). The instrument was used to measure the young adults' metacognition and also as a tool to gauge classroom interventions. In this study, the Jr. MAI was used specifically to assess the impact of formative assessment intervention on students' metacognition. The instrument had 18 items that measured students' level of metacognition before and after the formative assessment intervention. A summary of the instruments used in the empirical studies is shown in Table 10.

Table 10

Instruments for the Empirical Studies

Instruments	Number of items
1. TAFL-Q (Pat-El et al., 2013)	28 (Appendix B)
2. MAIT (Balcikanli, 2011)	24 (Appendix C)
3. Interview Schedule (Researcher developed)	8 (Appendix D)
4. PRT (Wafubwa et al., 2020)	10 (Appendix E)
5. Jr. MAI (Sperling et al., 2002)	18 (Appendix F)

3.4.4. Procedures

Data Collection

This line of research was conducted in two phases. The first phase involved validation of the instruments which was done in 2019 July and August. The authority to carry out the research was granted by the County director in the Ministry of Education, Science and Technology of Bungoma County (Appendix G). Instruments one and two were piloted on a sample of 180 mathematics teachers. The second phase of the research was conducted in 2020 and 2021 after obtaining research approval from the Ethics and Review Committee from the University of Pwani (Appendix H) and a research permit from the National Commission for Science, Technology, and Innovation in Kenya (Appendix I).

After obtaining authority to conduct the research, the mathematics teachers in the sampled schools were approached to fill out the questionnaires. The consent of teachers was sought before administering them with the questionnaires (Appendix J). The second and third studies used questionnaires that had already been validated in the first study. In addition to the MAIT questionnaire, an interview schedule was also used. The schedule was used to interview 10 teachers from 10 randomly selected schools. The interviews were conducted in the respective schools with each taking approximately 30 minutes.

The fourth study involved the development of a test that was used in the intervention study on measuring the impact of formative assessment. The test was administered to 45 students in one mixed school and the items were analyzed based on the classical test theory. The teachers were given instructions on how to administer the test. The instrument was thus tested in a controlled classroom setup.

The fifth study was quasi-experimental and was given to four randomly selected schools. The mathematics teachers in the selected schools administered the instrument before and after the intervention. Teachers in the experimental condition were trained on how to carry out the intervention. Teachers in the control classes were left out of the training on formative assessment. They were however given a four-week scheme on the content areas which showed the amount of time required for each specific content area.

Data analysis

Quantitative data analysis involved several approaches using Statistical Package for the Social Sciences (SPSS), AMOS, and Smart PLS package. Qualitative data which was collected through interviews was analyzed thematically. Descriptive statistics involving the mean and standard deviation were computed in SPSS to determine the general characteristics of the samples. Correlation analysis was done by computing the Pearson correlation coefficient in SPSS. Independent samples T-test and one-way ANOVA were used to examine the effect of background factors on teachers' perceptions of their metacognitive awareness. The Analysis of covariance (ANCOVA) was used to examine the impact of formative assessment on students' achievement and metacognition. The Shapiro-Wilk test assessed the normality of the data whereas Levene's test assessed the homogeneity of variances.

Analysis of moment structures (AMOS) was used to conduct confirmatory factor analysis, Structural Equation Modeling (SEM), and path analysis. The fit indices that were used to assess the measurement model included the normed chi-square (CMIN/DF), the root-mean-square error of approximation (RMSEA), standardized root mean square residual (SRMR), Tucker-Lewis index (TLI), and the comparative fit index (CFI). The Smart PLS package was used to determine composite reliability (CR), the heterotrait-monotrait (HTMT) ratio, and the average variance extracted (AVE).

3.5. Summary

Chapter three discusses the main aims of the research including the timeline of the empirical studies, research questions, research hypotheses, and research methods. The research timeline (Table 7) gives an overview of when the studies were conducted, the instruments used and the sample involved. The research questions and hypotheses are presented in the order in which the studies were carried out. The answers to these research questions and hypotheses are presented in chapter four which comprises five published journal articles. Under the method section, a general overview of the different research designs, samples, instruments, and procedures used in the studies is given. The specific methodology for each study is presented in detail in chapter four.

The empirical studies carried out in chapter four are informed by the gaps in the literature as reflected in chapter two. The literature review was carried out under the theoretical framework of formative assessment as defined by Black and Wiliam (2009). From the analysis of the 38 selected studies in chapter two, it was noted that some formative assessment strategies have rarely been tested empirically. A case in point is the use of classroom discussions that help in eliciting evidence of student understanding. Other strategies that have been least used are learning intentions and success criteria, and the use of self-assessment as an instructional process.

Researchers need to utilize these underutilized strategies for the formative assessment theory to remain meaningful. Another important finding was on where the studies have been implemented. The analysis revealed that most studies have been carried out in Europe and the USA with a few in Asia and only three in Africa and one in Australia. This implies that globally formative assessment is less embraced and therefore researchers in the affected continents can take up the challenge and spearhead the implementation of formative assessment. For the future advancement of this theory, all the five key strategies defining the formative assessment process must be put into perspective.

The five studies forming the empirical part of this dissertation in chapter four were therefore carefully designed and conducted based on the gaps identified in the literature. More specifically, the studies addressed the needs of Kenyan secondary school mathematics teachers and students in the context of a competency-based curriculum. Whereas the literature review and theoretical background have already been presented in chapter two, brief literature and theoretical background for each empirical study in chapter four are also presented. Conclusions, implications, and limitations for each study are also discussed extensively.

Chapter 4: Empirical Studies

4.1. Study 1: Assessing the Suitability of the Adapted Teacher Assessment for Learning Questionnaire in the Kenyan Context

4.1.1. Introduction

Mathematical competency is one of the 21^{st}-century skills that have gained attention in the international frameworks such as the Partnership for 21^{st}-century skills (P21); En gauge; assessment and teaching of 21^{st}-century skills (ATCS); and 21^{st}-century skills and competencies for new millennium learners (Darling-Hammond, 2012; Partnership for 21st Century Skills, 2009). Studies have shown that countries worldwide have approved 21^{st-}century competencies in their curriculum but they still have challenges regarding assessment (Ananiadou & Claro, 2009; Swaffield, 2011). As observed by Ananiadou and Claro, formative assessment is fundamental to the process of teaching and learning since it enhances student learning and also improves instruction. According to Voogt and Roblin (2012), competencies can be best obtained through determined pedagogical approaches like problem-based learning, formative assessment, and cooperative learning. Formative assessment can, therefore, assist students in applying 21^{st}-century competencies in real-life situations (Pepper, 2011).

4.1.2. Theoretical Background

The present study was guided by Black and Wiliam’s (2009) formative assessment framework which was conceptualized as:

> Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited (Black & Wiliam, 2009, p. 9).

In the foregoing definition, the authors related three key processes involved in instruction and the role of the teacher, learner, and peer to the five key strategies of formative assessment. The three key processes were identified as finding out where the learner is going; establishing the current position of the learner; and establishing how the learner will arrive at the desired destination. The Teacher, the learner, and the peers were referred to as agents who work in collaboration to ensure teaching and learning take place smoothly. The five key strategies were described as follows:

> Clarifying and sharing learning intentions and criteria for success; engineering effective classroom discussions and other learning tasks; providing feedback that moves learners forward; activating students as instructional resources for each other; and activating students as the owners of their learning (Black & Wiliam, 2009, p.8).

Pat-El et al. (2013) summarized these five strategies into two broad categories which involved monitoring and scaffolding strategies. In monitoring strategies, the focus is on making learning intentions clear to the students and initiating productive classroom discussions. On the other hand, scaffolding refers to the assistance that teachers provide to students during problem-solving or task performance (Alibali et al., 2007). It is an interactive process where students construct their learning through questions and feedback from their peers and the teacher.

How Teachers Perceive Formative Assessment

Research associated with teachers' perception of formative assessment has mainly focused on teachers' presumptions on the function of formative assessment. Few studies have focused on how teachers interpret their use of formative assessment. A study carried out by Young and Jackman (2014) revealed that teachers generally had a high positive perception of the role of formative assessment. Kenney and Maeda (2016) also focused on what teachers felt was the purpose of assessment. Their findings suggested the need for using teachers' previous understanding to design relevant staff development interventions. A study by Amoako et al. (2019) revealed that a high percentage of mathematics teachers in Ghana had low knowledge of formative assessment. Alotaibi (2019) while studying the adoption of formative assessment in Saudi Arabia focused mainly on factors that may hamper the execution of formative assessment.

One of the recent instruments that have been utilized in rating the perceptions of teachers regarding formative assessment is the Teacher Assessment For Learning Questionnaire (TAFL-Q) developed by Pat-El et al. (2013). With this questionnaire, a large sample validation study was conducted in the Netherlands, and the TAFL-Q had good psychometric properties. Since its establishment, the questionnaire has been used in different contexts to measure teachers' views regarding formative assessment. For instance, the questionnaire was used by Öz (2014) in Turkey and Nasr et al. (2018) in Iran. The two studies revealed that teachers generally had high perceptions of formative assessment and there were no significant differences regarding gender, years of teaching experience, and teachers' qualifications. On the contrary, a study by Sach(2012) revealed a significant relationship between perceptions and teachers' experience.

Focusing on mathematics teachers, studies have revealed a remarkable correlation between mathematics teachers' perceptions and students' problem-solving achievement (Handal & Herrington, 2003). A study carried out by Morrissette (2011) revealed that formative assessment practices depended on teachers' professional culture. According to Kyaruzi et al. (2018), mathematics teachers' feedback practices are positively predicted by their perceptions of formative assessment. Generally, studies have shown that teachers' perceptions of teaching and learning strongly influence their way of teaching. It is therefore paramount that teachers' perceptions are gauged before being introduced to new reforms in education settings (Dekker & Feijs, 2005).

Kenyan Context

Students' assessment in Kenya has predominantly been summative with minimal use of formative assessment (Milligan, 2017; Ngware et al., 2014). The current summative assessment is however short of measuring learners' abilities effectively. Furthermore, teachers lack adequate training in assessment processes (Republic of Kenya, 2012). The persistent dismal achievement in mathematics and science subjects over the years has been reflected in several reports (e.g., KNEC, 2014; KNEC, 2017). The concern in sub-Saharan Africa and Kenya, in particular, is on the lack of quality education which is evidenced through poor employability skills in the job markets (Ally, 2014). The inability of traditional assessment practices in empowering students for lifelong learning is supported by OECD (2014). The overarching need of the 21st century is thus equipping learners with relevant competencies which can effectively be done through classroom assessment practices that enhance learning transferability.

Poor teaching methods are among the commonly cited problems affecting students' performance in Kenya (Ngware et al., 2014; Wafubwa & Obuba, 2015). Teacher-centered teaching approaches have dominated the classrooms leaving no room for learners' creativity. The pressure from the high stake examinations has made teachers teach what is tested (Sayed & Kanjee, 2013). The assessment has thus been conceived as a filter to determine those who can join universities with limited capacity (Ally, 2014). There is currently a lack of in-service training of secondary school teachers regarding the implementation of formative assessment. The teachers conceive formative assessment through formal tests, quizzes, or homework given to students in the course of the instruction process and it is not clear how data from these tests, quizzes, or homework is used.

Despite the current challenges facing secondary education in Kenya, the country is committed to the adoption of a competency-based curriculum that is in line with 21st-century skills. The competency-based curriculum implies new learning, instruction, and assessment approaches. Teachers will be required to carry out classroom assessment practices that will be in line with formative assessment strategies. The success of the competency-based curriculum will, therefore, require teachers who are armed with the vital skills and tools to carry out formative assessment in the classrooms. Teachers' beliefs, attitudes, competencies, and practices are therefore crucial in determining the realization of competence-based learning. The present study sought to assess the suitability of the adapted TAFL-Q in measuring mathematics teachers' perceptions of formative assessment in Kenyan secondary schools.

4.1.3. Method

Participants

The sample consisted of 180 (138 males) secondary school mathematics teachers from 50 selected public secondary schools in Bungoma County. The sample was obtained through stratified and simple random sampling techniques to get the best representation across the entire county. Out of 180 participants, 143 (79.4%) possessed a Bachelor of Education (BEd) degree; 14 (7.8%) had either a Bachelor of Arts or Bachelor of Science (BA/BSc) degree; 21 (11.7%) had a master (MA) degree in fields different from education while two (1.1%) had a master's degree in education (MEd). Regarding teaching experience, 109 (60.6%) participants had up to 5 years; 31 (17.2%) between 6 to 10 years; 22 (12.2%) between 11 to 15 years and 18 (10.0%) above 15 years.

Adaptation of the TAFL-Q Questionnaire

The original 28 items questionnaire had two scales; "perceived monitoring" with 16 items and "perceived scaffolding" with 12 items (Pat-El et al., 2013). All items were closed-ended and measured on a 5 point scale which ranged from strongly disagree (SD = 1) to strongly agree (SA = 5). Since the original and the adapted versions were both in English, no translation was needed, instead, some phrases which seemed to be wordy were shortened while retaining the original meaning. For instance, the phrase “I give my students opportunity” was shortened to “I allow my students”. Since the questionnaire was adapted to the mathematics teachers, the introduction part of the questionnaire had specific instructions allowing mathematics teachers to focus on their mathematics classrooms.

Procedure and Data Analysis

The researchers physically distributed the questionnaires to the sampled schools’ principals who later gave the questionnaires to mathematics teachers. The first task was to assess the model fit in the Kenyan context by performing a confirmatory factor analysis in Amos Graphics 23. Statistical data analyses were conducted using IBM SPSS Statistics 25. Descriptive statistics were computed to show the characteristics of the sample. Factor analysis was done to determine the validity of the questionnaire in the Kenyan context. The Smart PLS package was used to determine the composite reliability (CR), the heterotrait-monotrait (HTMT) ratio, and the average variance extracted (AVE). The CR and AVE were used to examine the construct validity of the emergent scales whereas the HTMT ratio of correlations was used to assess the discriminant validity.

4.1.4. Results

Confirmatory factor analysis of the 28 items TAFL-Q

We conducted a confirmatory factor analysis (CFA) on the two-factor model of the TAFL-Q (see Appendix B) to test the model fit on the Kenyan sample. Using maximum likelihood (ML) estimation and the standardization method, we assessed the goodness of fit of the measurement model using different goodness of fit indices. The indices included the normed chi-square (CMIN/DF), the root-mean-square error of approximation (RMSEA), standardized root mean square residual (SRMR), Tucker-Lewis index (TLI), and the comparative fit index (CFI). We also evaluated the convergent validity of the constructs by considering the composite reliability (CR) and average variance extracted (AVE). The following results were obtained: CMIN/DF = 2.643, RMSEA = .096, SRMR = .085, TLI = .653, CFI = .680 (See Appendix B1)

According to Glynn et al. (2011), the recommended value for the normed chi-square (CMIN/DF) should be within the range of 1.0–3.0. The value of 2.643 is therefore within the recommended range. The RMSEA and the SRMR values should be below .080 and the two incremental fit indices (TLI and CFI) above .900 (Hoope et al., 2008; Teo, 2013). The results of RMSEA and SRMR as well as of the TLI and CFI for this model are all indicative of a poor model fit. The composite reliability for the monitoring and scaffolding scales was .88. The AVE for monitoring and scaffolding scales was .32 and .39 respectively. Whereas CR for the two scales was above the recommended cut-off of .70 (Fornell & Larcker, 1981), the values of AVE were way below the cutoff point of .50.

To improve the model fit, we used modification indices to covary error terms that had values more than 10 and fell within the same construct (Hermida, 2015). The resultant model improved a little but it could still not fit the data well (CMIN/DF = 1.884, RMSEA = .073, SRMR =.069, TLI = .810, CFI = .840). Further analysis of the factor loadings revealed that seven out of 16 factors on the monitoring scale and one factor out of 12 factors on the scaffolding scale had low loadings of less than .50. It was therefore imperative to reexamine the factor structure for the Kenyan sample.

Exploratory factor analysis

We used the precepts recommended by Williams et al. (2010) to conduct factor analysis on our Kenyan sample. We first checked if the data was suitable for factor analysis by computing a correlation matrix for the 28 items. The items correlated at least at .30 with at least one other item and no item correlated above .50 with other items. This verified that items shared some common variance. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) was .840 and Bartlett's test of sphericity was significant ($\chi 2$ (378) = 2046.028, $p < .001$), suggesting that the data were appropriate for factor analysis. The communalities were all above .40 except for item 1 which had a communality of .29 and was therefore not considered for factor analysis. Principal components analysis was used to analyze the factors since the main aim was to identify how factors would load in a different cultural-educational context. Factors were thus extracted by the use of Kaiser's criteria (eigenvalue > 1 rule), the Scree test, and the cumulative percent of variance extracted.

The first six factors extracted based on the initial eigenvalues explained 11.0%, 9.4%, 9.3%, 8.9%, 8.4% and 8.0% of the variance respectively. The seventh factor had an eigenvalue just slightly above one with an explained variance of 6.9%. Solutions for the three, four, five, six, and seven factors were each examined using varimax rotations of the factor loading matrix. We preferred the six-factor solution which explained 54.9% of the total variance.

Nine items were eliminated based on failing to meet the minimum criteria of not loading above .30 on any factor, loading less than .40 on any factor, and no cross-loading of .30 or above (Williams et al., 2010). Item 1 "*I encourage my students to reflect upon how they can improve their assignments"* was eliminated because it failed to load above .30 on any factor. Item 3 "*While working on their assignments, I ask my students how they think they are doing"* had loadings of .46 and .47 on factors 2 and 5. Items 6 "*I ask my students to indicate what went well and what went badly concerning their assignments",* 7 "*I encourage students to reflect upon their learning processes and how to improve their learning"* and 19 "*During my class, students are allowed to show what they have learned"* also loaded on more than two factors with similar factor loadings between .40 and .50. Items 11 "*I give students guidance and assistance in their learning"*, 15 *"I discuss with my students how to utilize their strengths to improve on their assignment",* 25 "*I give my students opportunities to ask questions"* and 28 "*I can recognize when my students reach their learning goals"* had low factor loadings of below .40 and were therefore eliminated.

A final principal components factor analysis of the remaining 19 items was conducted using varimax rotations, with six factors together explaining 66.0% of the total variance. All items in this analysis had primary loadings of over .50. The factor loading matrix for this final solution is presented in Table 11. Further exploration of the factors using oblimin rotation yielded similar outcomes. The factors were labeled based on the formative assessment strategies in the model proposed by Black and Wiliam (2009). Reliability analysis of the 19 items resulted in a high Cronbach's Alpha of .87. For more details on item-total correlation analysis, see Appendix B2.

Table 11

Factor Loadings and Communalities for 19 Items from the Original TAFL-Q

	Factors and Communalities						
Items	PSC	PPA	PSA	PLI	PF	PCD	Com
2. After a test, I discuss the answer given with each student				.72			.61
4. I involve my students in thinking about how they want to learn at school				.79			.73
5. I allow my students to decide on their learning objectives				.70			.64
8. I inform my students of their strong points concerning learning					.74		.70
9. I inform my students of their weak points concerning learning					.66		.53
10. I encourage my students to improve on their learning processes					.71		.69
12. I discuss assignments with my students to help them understand the content better						.73	.71
13. I discuss with my students the progress they have made						.75	.69
14..After an assessment, I inform my students on how to improve their weak points						.68	.70
16. Together with my students, I consider ways on how to improve on their weak points			.55				.48
17. I adjust my instructions whenever I notice that my students do not understand a topic			.74				.70
18. I provide my students with guidance to help them gain an understanding of the content taught			.80				.78
20. I ask questions in a way my students understand		.79					.72
21. By asking questions during class, I help my students gain an understanding of the content taught		.82					.77
22. I am open to student contribution in my class		.61					.68
23. I allow my students to ask each other questions during class	.57						.48
24. I ensure that my students know what areas they need to work on to improve their results	.61						.55
26. My students know what the evaluation criteria for their work are	.75						.69
27. I ensure that my students know what they can learn from their assignments	.72						.53

Note: N=180. Factor loadings below .40 are suppressed: PSC=Perceived Success Criteria; PPA=Perceived Peer Assessment; PSA=Perceived Self-Assessment; PLI=Perceived Learning Intentions PF=Perceived Feedback; PCD=Perceived Classroom Discussion; Com= Communalities.

Descriptive analyses (mean, standard deviations, and Pearson correlations) were calculated to ascertain teachers' perceptions of formative assessment (Table 12). Results indicated that teachers' level of agreement ranged from low (a mean score of between 2.0 to 2.5) to moderate (mean score between 2.6 and 3.5). The construct of Perceived Self-Assessment had the highest score (M=3.31, SD=0.66) which was rated as moderate whereas the Perceived Learning Intentions had the lowest score (M=2.45, SD=0.95) which was rated as low. Teachers also had moderate perceptions on the constructs of Perceived Feedback (M=3.07, SD=0.73), Perceived Classroom Discussion (M=3.14, SD=0.67), Perceived Peer Assessment (M=3.01, SD=0.77), and Perceived Success Criteria (M=2.81, SD=0.78).

Bivariate correlations (see Table 12) among the emergent factors revealed low to moderate positive correlations between the variables. There was a weak positive correlation between the Perceived Learning Intentions scale and the Perceived Self-Assessment scale. This weak relationship was expected and it proved that the two scales measure different constructs. Generally the low to moderate positive correlations show that the six scales measure different constructs that are somehow related.

Table 12

Descriptive Statistics and Correlations for Emergent Study Variables

	n	*M*	*SD*	1	2	3	4	5	6
1. Perceived Learning Intentions(PLI)	3	2.45	0.95	-					
2. Perceived Feedback(PF)	3	3.07	0.73	.31**	-				
3. Perceived Classroom Discussion (PCD)	3	3.14	0.67	.35**	.41**	-			
4. Perceived Peer Assessment (PPA)	3	3.01	0.77	.35**	.49**	.46**	-		
5. Perceived Self-Assessment(PSA)	3	3.31	0.66	.09	.36**	.39**	.47**	-	
6. Perceived Success Criteria (PSC)	4	2.81	0.78	.38**	.34**	.38**	.47**	.43**	-

**. Correlation is significant at the 0.01 level (2-tailed).

We further examined the psychometric properties of the emergent factors by looking at convergent validity and discriminant validity. We determined convergent validity using the composite reliability (CR) and the average variance extracted (AVE) as suggested by Ab Hamid et al. (2017). The AVE and CR values were computed using Smart PLS 3 software which is one of the most recent packages for structural equation modeling (Ringle et al., 2015). The AVE shows the average amount of variance explained by a construct to the overall variance of its indicators (Ringle et al., 2015). As reflected in Table 13, the values of AVE and CR for all the constructs were above the recommended threshold of .50 and .70 respectively (Fornell & Larcker, 1981).

We assessed discriminant validity using the Heterotrait-Monotrait (HTMT) ratio of correlations which is a new method of examining discriminant validity in variance-based structural equation modeling. This method is deemed to be a superior alternative to the Fornell-Larcker criterion and examination of cross-loadings approaches for determining discriminant validity (Henseler et al., 2015). To ascertain the discriminant validity, the value of the HTMT should be less than .85, and values above .85 imply the absence of discriminant validity (Henseler et al., 2015; Kline 2011). The values of the HTMT for this analysis ranged between .16 and .69, hence confirming the discriminant validity. Table 13 shows the results of the CR, AVE, and HTMT ratio of correlations computed using the bootstrap method.

Table 13

Six-Factor Model CR, AVE, and HTMT Ratio of Correlations

Latent factors	*CR*	*AVE*	*1*	*2*	*3*	*4*	*5*	*6*
1. Perceived Learning Intentions(PLI)	.83	.63	-					
2. Perceived Feedback(PF)	.82	.60	.60	-				
3. Perceived Classroom Discussion (PCD)	.83	.62	.51	.45	-			
4. Perceived Peer Assessment (PPA)	.85	.66	.63	.69	.50	-		
5. Perceived Self-Assessment(PSA)	.87	.68	.54	.51	.16	.63	-	
6. Perceived Success Criteria (PSC)	.83	.55	.54	.48	.53	.65	.58	-

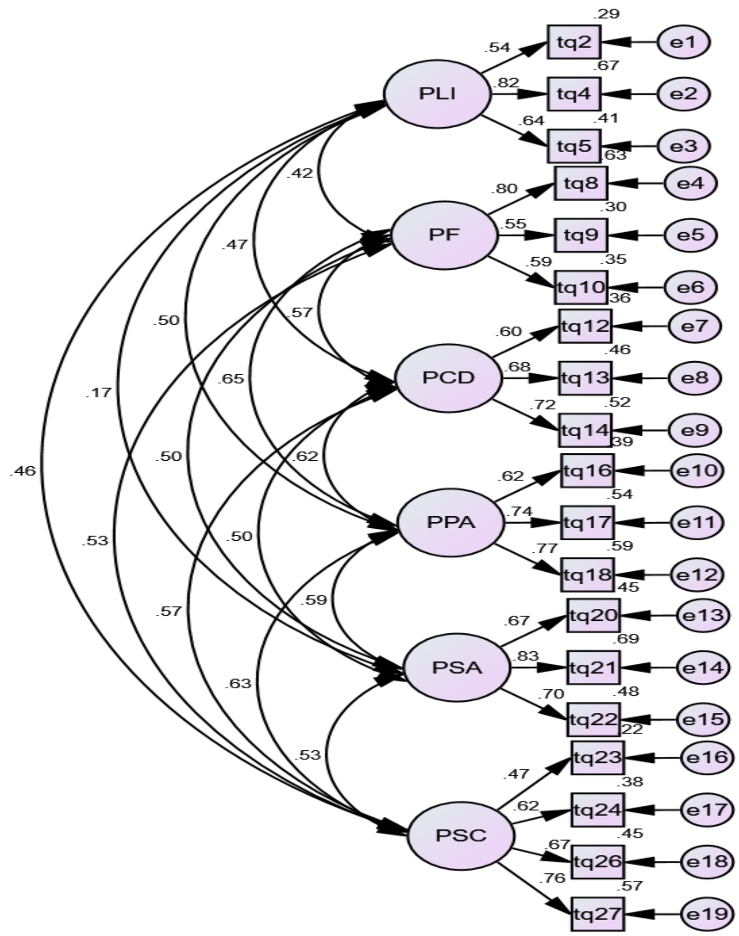
Note: CR=Composite Reliability; *AVE*=Average Variance Extracted; *HTMT*= Heterotrait-Monotrait

The measurement model (Figure 2) was estimated using maximum likelihood and had the fit indices as follows: $\chi 2(51) = 116.002$; $p<.001$; CMIN/DF = 1.92, RMSEA = .070, SRMR = .060, TLI = .880, CFI= .850. The value of the CMIN/DF was within the recommended range of 1.0-3.0. The RMSEA and SRMR were also within the acceptable range of .080 and below (Hooper et al., 2008; Teo, 2013). Although the incremental fit indices fell marginally below .900

which is the generally accepted threshold, according to Ho (2006), these values are still within the recommended range.

Figure 2

Confirmatory Factor Analysis: Standardized Loadings and Correlations of the Latent Factors and Error Terms



Note: PSC=Perceived Success Criteria; PPA=Perceived Peer Assessment; PSA=Perceived Self-Assessment; PLI=Perceived Learning Intentions PF=Perceived Feedback; PCD=Perceived Classroom Discussion: all items (tq) are numbered in the order presented in the text; *e*= Error terms

4.1.5. Discussion

This study sought to examine whether the adapted TAFL-Q can be appropriate in measuring mathematics teachers' perceptions of formative assessment in Kenyan secondary schools. Since the adapted questionnaire was used in a different cultural setup from the original questionnaire, we first did a confirmatory factor analysis to assess if the questionnaire fits the Kenyan sample. The results of the two-factor model of the TAFL-Q failed to fit well with the Kenyan data. It was therefore necessary to carry out an exploratory factor analysis to explore the structural stability of the Kenyan sample. The results suggested a six-factor structure for 19 out of 28 items based on a principal exploratory factor analysis with varimax rotation.

Although no six-factor solution exists in the literature regarding TAFL-Q, the six factors extracted in our present study are underpinned by the formative assessment theoretical framework of Black and Wiliam (2009). These factors can however be reinforced through revising or rephrasing the items with poor primary loadings and perhaps by including other items. For instance, item 1 which never loaded on any factors can be replaced with a new item. Item 3 may probably have been misunderstood by teachers and the loadings may be higher if it can be revised. In the same vein, all other items (6, 7, 11, 15, 19, 25, and 28) which had low loadings can be revised and retained in the questionnaire.

Descriptive statistics indicated that mathematics teachers' level of agreement ranged from low to moderate. The results suggested that teachers had average perceptions of their use of self-assessment, feedback, classroom discussion, peer assessment, and success criteria strategies. They however had a low agreement on the use of learning intentions strategies. Learning intentions and success criteria define where the learner is going (Black & Wiliam, 2009). Even though the low perception of the use of learning intention strategies may seem to suggest that mathematics teachers rarely involve students in their academic journey, the low perception could have been influenced by factors such as large class sizes and heavy workload for teachers.

One of the aspects of learning intentions criteria is involving learners in coming up with learning objectives. This would mean that every learner's opinion is considered. This may however be less practical in Kenyan classrooms which are characterized by large class sizes and a high workload for teachers. Teachers could be willing to involve students in the use of learning intentions but this may be practically impossible due to the aforementioned reasons. The moderate perception of the other five strategies should also be interpreted cautiously since this is an instrument that needs to be developed further.

The lack of in-service training of secondary school teachers on formative assessment and the phrasing of some items may have likely influenced the teachers' ratings of their perceptions. Teachers may lack a clear understanding of formative assessment given the fact that their interpretation was based on their pre-service training understanding. The findings by Crichton and McDaid (2016) on teachers' perceptions of learning intentions and success criteria strategies also showed that teachers agreed on the usefulness of these strategies but rarely practiced them in classrooms.

Limitation of the present study

The first limitation regards the sample size. Although our sample size was generally within the accepted range for factor analysis (de Winter et al., 2009), we still felt a sample of more than 200 would have been the best. This study relied only on data from a self-report questionnaire which may affect the validity of the results. While the data provide evidence of validity, future studies should consider incorporating qualitative approaches like observation and interviewing to give room for greater objectivity and accuracy of results. Nevertheless, the present results are a pointer to formative assessment practices that should be considered for future studies.

Conclusion and implications

From our study, we conclude that the adapted TAFL-Q questionnaire needs to be further developed before it can be used to rate the mathematics teachers' views on formative assessment in Kenya. The emerging six-factor structure can be further cross-validated with a different sample to confirm the stability of the scales. The items that were dropped from the original questionnaire can be revised and added back to the scale to increase the internal consistency. Since the present study is part of a larger study that involves an intervention programme, the results of the descriptive statistics have pointed out the need for clearly defining the concept of formative assessment to the Kenyan mathematics teachers. Our results have also pointed out the need for further development of the TAFL-Q as an instrument for measuring mathematics teachers' perceptions of formative assessment in the Kenyan context.

4.2. Study 2: Formative Assessment as a Predictor of Mathematics Teachers' Levels of Metacognitive Regulation

4.2.1. Introduction

The emergence of 21st-century competencies has come along with the challenges of teaching and assessment (Lai & Viering, 2012). One of the ways that have been deemed suitable for teaching and assessing these competencies is by the use of formative assessment strategies (Griffin & Care, 2013; Shute & Becker, 2010). Studies have shown that formative assessment benefits both teachers and learners. Unlike summative assessment which is used as a measurement instrument, formative assessment is designed to support teaching and can, therefore, be used as a teaching tool (Clark, 2012; Gipps et al., 2015). Teachers who use formative assessment strategies such as classroom discussions, questioning, effective feedback, self-assessment, and peer assessment (Black & Wiliam, 2009) enhance student achievement.

Formative assessment also acts as a valuable professional development opportunity for teachers (OECD, 2005) since teachers need to optimize their content knowledge to apply quality and effective formative assessment strategies (Heritage, 2007; Sadler, 2009). The main goal of formative assessment has been seen as promoting students' learning to learn skills (OECD, 2005). Formative assessment also builds students' skills at peer-assessment and self-assessment and helps students to develop a range of effective learning strategies (Chan, 2010). When students actively build their understanding, they develop invaluable skills for lifelong learning.

Formative assessments, therefore, enable students to become autonomous and self-regulating learners. According to Vrugt and Oort (2008), self-regulation (metacognitive skills) is an important aspect of learning in academic performance because learners are actively engaged in the learning process. Shepard (2006) as well noted that formative assessment encourages students' metacognition and reflection in their learning. As teachers clarify learning intentions and criteria for success through feedback, students can regulate their learning and become partners in filling the learning gaps (Heritage, 2007).

It is therefore important to teach students metacognitive strategies so that they construct their understanding through deep learning. Teachers can only teach students metacognitive strategies if they are metacognitive in their teaching. In other words, teachers must also become learners for learning to be visible (Hattie & Yates, 2013). Metacognition enables teachers to be aware of their strengths and weaknesses and can, therefore, be more effective in their teaching (Ben-David & Orion, 2013).

4.2.2. Theoretical Background

Formative Assessment

Formative assessment as a classroom practice came into the limelight following the seminal work of Black and Wiliam (1998) which involved the synthesis of over 250 studies, linking assessment and learning. Black and Wiliam (2009) after considering the main features of teaching and learning defined formative assessment as a classroom practice that is executed by both teachers and learners. According to Wiliam (2011), the basic idea behind formative assessment is that evidence of student learning is used to adjust instruction to better meet students' learning needs. Formative assessment is thus designed to support teaching and learning continuously (Clark, 2012; Gipps, et al., 2015).

To meet students' learning needs, Black and Wiliam (2009) identified five key strategies that could be used in classrooms as involving learning intentions and criteria for success, classroom discussions, feedback, peer assessment, and self-assessment. Research has shown that formative assessment practices such as self and peer assessment have a positive impact on self-regulated learning (Panadero & Alonso-Tapia, 2014; Panadero & Broadbent, 2018; Zimmerman & Schunk 2011). Generally, studies agree that formative assessment plays a role in self-regulated learning. Notwithstanding, some aspects of formative assessment have rarely been studied empirically.

Borrowing from Black and Wiliam's (2009) framework of formative assessment, Pat-El et al. (2013) conceptualized formative assessment (assessment of learning) as comprising of monitoring and scaffolding dimensions. They viewed the monitoring dimension as consisting of strategies that deal with feedback and self-monitoring whereas the scaffolding dimension deals with instruction-related processes such as classroom questioning. As noted by Lee and Mark (2014), monitoring strategies entail students examining their learning progress through self-monitoring to identify learning strengths and weaknesses. Scaffolding, on the other hand, involves a classroom interaction through sharing learning intentions, success criteria, and how success is evaluated (Pat-El et al., 2015).

Metacognition

Metacognition was originally defined by Flavell (1979) and Brown (1978) as the knowledge about and regulation of one's cognitive activities in the learning processes (in Veenman et al., 2006). Since then, metacognition has been substantively studied in the field of educational psychology and has been often seen as a form of executive control involving monitoring and self-regulation (Demetriou et al., 2011). Based on Flavell's definition, most researchers have conceptualized metacognition as consisting of two broad dimensions: knowledge about cognition and regulation of cognition (e.g., Lai & Viering, 2012; Veenman, 2011; Williams & Atkins, 2009).

Knowledge of cognition also referred to as knowledge and awareness of one's cognition is composed of declarative, procedural, and conditional knowledge (Harris et al., 2010). Declarative knowledge is the kind of knowledge required to accomplish a task. Procedural knowledge deals with how to apply learning strategies. Conditional knowledge on the other hand relates to knowledge of when, where, and why in applying particular procedures or strategies (Harris et al., 2010; Mahdavi, 2014).

Regulation of cognition or metacognitive skills is described as an acquired repertoire of procedural knowledge for monitoring, guiding, and controlling one's learning and problem-solving behavior (Veenman, 2011). Planning involves choosing relevant strategies and providing the required resources to attain the learning goals. Monitoring refers to skills necessary to regulate one's learning like self-assessment skills while evaluation refers to the process of judging the achievements made (Harris et al., 2010; Mahdavi, 2014).

Metacognitive skills play a great role in guiding and controlling the execution of tasks (Veenman, 2011). Studies have further shown that metacognitive skills training greatly improve the performance of students (Kramarski & Mevarech, 2003; Mevarech & Fridkin, 2006; Veenman et al., 2006). Through metacognitive skills, learners are therefore able to often carry out self-evaluation of task performance, self-monitoring, and planning. An intervention study by Csíkos and Steklács (2010) focusing on planning, monitoring, and evaluation skills resulted in a positive achievement among the Hungarian students. Other intervention studies with similar outcomes include Naseri et al. (2017) and Roll et al. (2011).

Formative Assessment and Metacognition

Whereas a substantial number of studies have shown that student metacognitive skills training leads to a positive outcome, little is known about what influences teachers' metacognition. As much as studies strongly advocate that students should be made aware of the importance of metacognition through ways such as teacher's modeling (Martinez, 2006; Tanner, 2012), it has also been noted that teachers lack adequate knowledge about metacognition and they, therefore, need to be trained on metacognitive instruction (Veenman et al., 2006). Enhancing awareness, improving self-knowledge, and ensuring enabling learning environments have been described by Schraw (1998) as some of the instructional strategies for promoting metacognitive awareness.

In enhancing general awareness, Schraw (1998) pointed out the important role played by the teacher and other students in modeling cognitive and metacognitive skills. This implies that both teachers and learners work together in designing the learning intentions and success criteria. In the formative assessment framework by Black and Wiliam (2009), clarifying learning intentions and criteria for success is the first strategy that points to where the leaner is going. This is jointly done by the teacher, the learner, and the peer. Schraw further noted that students should be given a chance to regularly reflect on their drawbacks and achievements. This involves self and peer assessment through discussions and teacher feedback.

According to the formative assessment framework by Black and Wiliam (2009), the learning gap can be filled through strategies like effective classroom discussions, feedback, peer assessment, and self-assessment (Braund & DeLuca, 2018). Formative assessment is hence a learning process that can enhance students' metacognitive knowledge. Theoretically, metacognition is seen as a multidimensional set of general skills that are crucial for developing 21st-century skills and competencies (Lai & Viering, 2012). However, few empirical studies have investigated the relationship between formative assessment and metacognition. Baas et al. (2014) investigated the relationship between assessment for learning (formative assessment) and metacognition among elementary school students. The results showed that formative assessment strategies involving monitoring and scaffolding predicted the students' use of cognitive and metacognitive strategies.

The Current Study

The preceding literature has illustrated how the aspects of formative assessment relate to metacognition especially the regulation of cognition. Limited empirical research has however been done to show a clear relationship between formative assessment and metacognition. This study aims to fill the gap regarding the limited literature by empirically examining the relationship between teachers' perceptions of formative assessment and their levels of metacognitive skills. We aim to show how the use of formative assessment strategies affects the teachers' metacognitive regulation in terms of planning, monitoring, and evaluating skills (Balcikanli, 2011). Based on the literature review, we hypothesized that the use of formative assessment strategies will have a positive effect on teachers' levels of metacognitive skills. This study, therefore, sought to answer the research question: "what is the relationship between mathematics teachers' formative assessment strategy use and their levels of metacognitive regulation?"

4.2.3. Methods

Sample

There were two sets of samples consisting of 180 and 213 secondary school mathematics teachers from secondary schools in Kenya. The two samples were collected in two different regions. Stratified and simple random sampling techniques were employed to obtain a representative sample from different school categories as shown in Table 14.

Table 14

Sociodemographic Characteristics of Participants

		Sample 1(N=180)		Sample 2(N=213)	
Characteristic	Description	*n*	%	*n*	%
Gender	Male	138	76.7	157	73.7
	Female	42	23.3	56	26.3
Teacher qualification	BEd	143	79.4	166	77.9
	BA/BSc	14	7.8	19	8.9
	Diploma	21	11.7	25	11.7
	MEd	2	1.1	3	1.4
Teaching experience	Up to 5 years	109	60.6	125	58.7
	6 to 10 years	21	17.2	41	17.2
	11 to 15 years	32	12.2	27	12.7
	Above 15 years	18	10.0	20	9.4

Measures

Questionnaires were used to measure the perceptions of mathematics teachers' use of formative assessment strategies and their levels of metacognitive regulation.

1. The teacher Assessment For Learning Questionnaire

The teacher Assessment For Learning Questionnaire (TAFL-Q) was used to measure the teachers' perceptions of formative assessment. The TAFL-Q (see Appendix B). was constructed by Pat-El et al. (2013) using a sample of secondary school teachers from the Netherlands. The questionnaire consisted of 28 closed-ended items divided into two scales: perceived monitoring (16 items) and perceived scaffolding (12 items). The items were measured on a five-point Likert scale.

2. Metacognitive Awareness Inventory for Teachers

The Metacognitive Awareness Inventory for Teachers (MAIT) scale was used to measure the teachers' level of metacognitive awareness (see Appendix C). The MAIT scale was constructed by Balcikanli (2011) who considered the two components of metacognition: metacognitive knowledge and metacognitive skills, with three scales under each component. The scales under metacognitive knowledge included declarative knowledge (DK), procedural knowledge (PK), and conditional knowledge (CK). On the other hand, the scales under metacognitive regulation included planning (P), monitoring (M), and evaluating (E). Each of the six scales was composed of four items which were measured on a five-point Likert scale. The present study only considered the component of metacognitive skills which had the scales of planning, monitoring, and evaluating.

Data collection procedure and analysis

After obtaining clearance from the Ministry of education, the researchers visited the sampled schools and physically delivered questionnaires to the teachers through the heads of mathematics departments. After being assured of anonymity of participation, each teacher took approximately 25 minutes to fill in the questionnaires. Confirmatory factor analysis was conducted using Amos Graphics 23 to test the model fit to the first sample (n=180). Since the TAFL-Q had a poor fit, exploratory factor analysis was done using IBM SPSS Statistics 25 to obtain a new factor structure.

Descriptive and inferential statistics were computed to analyze the data collected. Descriptive statistics were computed to obtain the participants' levels of agreement regarding formative assessment and metacognitive skills use. The relationship between the variables was measured using structural equation modeling (SEM) and analysis of moment structures (AMOS).

4.2.4. Findings

Measurement Model Development

The assumptions of multivariate normality and linearity of the data from the TAFL-Q and MAIT were evaluated through IBM SPSS Statistics 25 based on Kline's (2011) guidelines. Using Cook's distance and box plots, no significant univariate or multivariate outliers were observed. The data were normally distributed without any missing data. Maximum likelihood estimation was therefore used in the analysis.

1. The TAFL-Q analysis

A confirmatory factor analysis (CFA) was conducted on TAFL-Q using a sample of 180 mathematics teachers to see whether the model fits the questionnaire data of the Kenyan sample. The results showed a poor fit with the following fit indices: CMIN/DF = 2.643, RMSEA = .096, SRMR = .085, TLI = .653, CFI= .680. It was, therefore, necessary to establish a new factor structure for the sample. Factor analysis by principal components analysis and Varimax Kaiser Normalization rotation resulted in a six-factor structure consisting of 19 items from the original 28 items. The nine items were eliminated after a careful analysis based on the guidelines suggested by Williams et al., 2010).

The emergent six factors were: Perceived Learning Intentions (PLI); Perceived Feedback (PF); Perceived Classroom Discussion (PCD); Perceived Peer Assessment (PPA); Perceived Self-Assessment (PSA) and Perceived Success Criteria (PSC). These emergent factors are well supported by the formative assessment theoretical framework and were therefore labeled based on the strategies of formative assessment (Black & Wiliam, 2009).

Using a different sample of 213 secondary school mathematics teachers, the new version of the TAFL-Q resulted in an improved acceptable model with the following fit indices: CMIN/DF = 2.009, RMSEA = .069, SRMR = .054, TLI = .862, CFI= .889. Although the values of the two incremental indices; TLI (.862) and CFI (.889) were slightly below the recommended threshold of .900, they were still within the acceptable range (Ho, 2006).

2. The MAIT analysis

The MAIT consisted of six scales but CFA was conducted for three scales which represented the metacognitive regulation dimension because that was the area of our focus. The three scales of metacognitive regulation had a total of 12 items that measured planning, monitoring, and evaluating skills. The CFA resulted in an acceptable model on both samples that were used in this study. The fit indices for the first sample of 180 were: CMIN/DF = 2.275, RMSEA = .084, SRMR = .058, TLI = .913, CFI= .933 (See Appendix C1) while the second sample of 213 resulted to the following fit indices: CMIN/DF = 2.2411, RMSEA = .082, SRMR = .053, TLI = .917, CFI= .936. No item was therefore eliminated from the MAIT scale. The scales and reliabilities of the two instruments are shown in table 15.

Table 15

Scales, Cronbach's α, and Sample Items for the TAFL-Q and MAIT

	Scale	*n*	Cronbach's α	Sample item
TAFL-Q	Perceived Learning Intentions (PLI)	3	.70	After a test, I discuss the answer given with each student
	Perceived Feedback(PF)	3	.66	I inform my students of their strong points concerning learning
	Perceived Classroom Discussion (PCD)	3	.68	I discuss with my students the progress they have made
	Perceived Peer Assessment (PPA	3	.76	I adjust my instructions whenever I notice that my students do not understand a topic
	Perceived Self-Assessment (PSA)	3	.74	I ask questions in a way my students understand
	Perceived Success Criteria (PSC)	4	.72	I allow my students to ask each other questions during class
MAIT Scale	Planning (P)	4	.77	I know what I am expected to teach
	Monitoring (M)	4	.83	I try to use teaching techniques that worked in the past
	Evaluating (E)	4	.87	I have a specific reason for choosing each teaching technique I use in class

Note: TAFL-Q= Teacher Assessment for Learning Questionnaire; *MAIT*= Metacognitive Regulation Inventory for Teachers

Final Model

All the latent variables in the TAFL-Q and MAIT were considered in the final measurement model with a sample of 213 mathematics teachers. The model fit for the nine latent variables in the measurement model resulted to an adequate model fit for the data: CMIN/DF = 1.789, RMSEA = .061, SRMR = .055, CFI= .879. Table 16 shows the mean, standard deviation, and correlations among the latent variables. Analysis of the standardized residual matrix for the measurement model revealed no statistically significant residual (all absolute values were less than two).

Table 16

Descriptive Statistics and Correlations for Latent Variables

	Scale	*M*	*SD*	1	2	3	4	5	6	7	8
1	PLI	2.45	.95	-							
2	PF	3.07	.73	.31**	-						
3	PCD	3.14	.67	.35**	.41**	-					
4	PPA	3.01	.77	.35**	.49**	.46**	-				
5	PSA	3.31	.66	.09	.36**	.39**	.47**	-			
6	PSC	2.81	.78	.38**	.34**	.38**	.47**	.43**	-		
7	Planning	2.90	.81	.32**	.28**	.20**	.41**	.23**	.49**	-	
8	Monitoring	3.05	.83	.26**	.34**	.38**	.46**	.34**	.47**	.54**	-
9	Evaluation	2.67	1.05	.35**	.26**	.32**	.52**	.30**	.57**	.62**	.66**

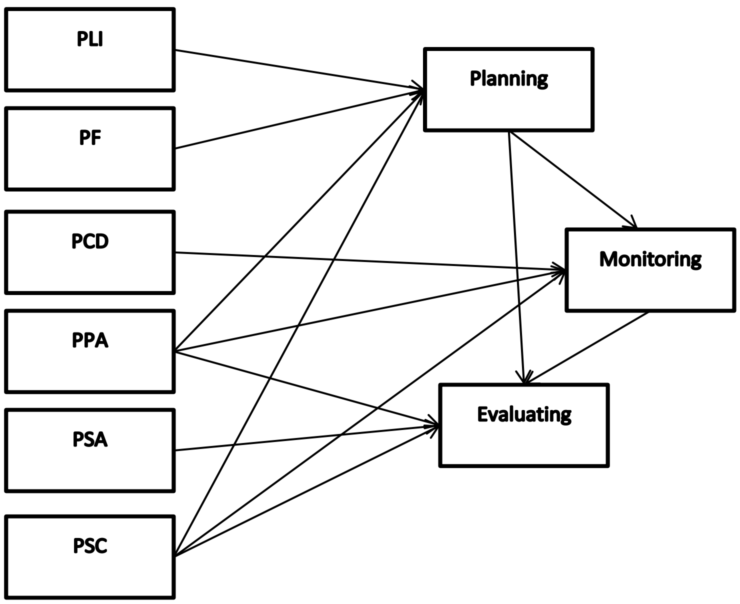
Note: **. Correlation is significant at the 0.01 level (2-tailed); PSC=Perceived Success Criteria; PPA=Perceived Peer Assessment; PSA=Perceived Self-Assessment; PLI=Perceived Learning Intentions PF=Perceived Feedback; PCD=Perceived Classroom Discussion

The Structural Model

The final analysis involved testing the hypothesized structural relationships among the predictor and observed variables through bootstrapping analysis. Since the TAFL-Q failed to fit the Kenyan data well, we had to obtain a new factor structure that slightly changed our original hypothetical relationship but without affecting the relationship among the emergent variables. The Kenyan sample resulted in a six-factor structure from the original two-factor structure. The first three scales (PLI, PF, and PCD) fitted so well under the monitoring dimension on the original scale whereas the PSA, PPA, and PSC scales fitted under the scaffolding dimension. The new hypothetical relationship is reflected in figure 3. The predictors of planning were hypothesized to be PLI, PF, PPA, and PSC; the predictors of monitoring were PLI and PF through planning, PCD, PPA, and PSC. The predictors of evaluating were PCD through monitoring, PPA, PSA, and PSC.

Figure 3

Hypothetical Relationship between Formative Assessment Strategies and Metacognitive Skills



Note: Formative assessment strategies are: PLI=Perceived Learning Intentions; PF=Perceived Feedback; PCD=Perceived Classroom Discussion; PPA=Perceived Peer Assessment; PSA=Perceived Self-Assessment; PSC=Perceived Success Criteria, while Metacognitive skills are: Planning, Monitoring, and Evaluating.

The hypothesized structural model fitted the data (n=213) well with the following fit indices: $\chi 2(13) = 9.839$; $p = .277$; CMIN/DF = 1.218, RMSEA = .032, SRMR = .012, TLI = .989, CFI= .998. There were no post-hoc modifications from the analysis because the indices indicated a good fit between the model and the observed data. Furthermore, there was no problem regarding the residual analysis. Table 17 and Figure 4 show the results of the estimates for the parameters. Regression analysis (see Table 17) revealed significant path relations in almost all the predictors of metacognitive regulation (planning, monitoring, and evaluating). Only two paths: PSA to evaluating and PF to planning failed to show a significant effect. Results of squared multiple correlations (R^2) showed that the predictors of planning explained 33.0% of its variance; predictors of monitoring explained 41.0% of its variance and the predictors of evaluating explained 55.0% of its variance.

Direct Effects

The results show a direct positive effect of success criteria (β=.17) and peer assessment (β=.15) on evaluating skills; classroom discussion (β=.16) and peer assessment (β=.13) on monitoring skills; learning intentions (β=.13), feedback (β=.03), peer assessment (β=.19) and success criteria (β=.36) on planning skills. Self-assessment (β=-.06) is however negatively related to evaluating skills. The effects of feedback (β=.03) on planning and self-assessment (β=-.06) on evaluating were non-significant (p>.05) as reflected in Table 17.

Indirect Effects

Significant mediated effects were found between PPA and monitoring through planning (β= .027, p< .05); PCD and evaluating through monitoring (β=.033, p<.05); PPA and evaluating through monitoring (β=.009 p<.01); PSC and monitoring through planning (β=.011, p<.01); PSC and evaluating through planning (β=.008, p<.01).

Table 17

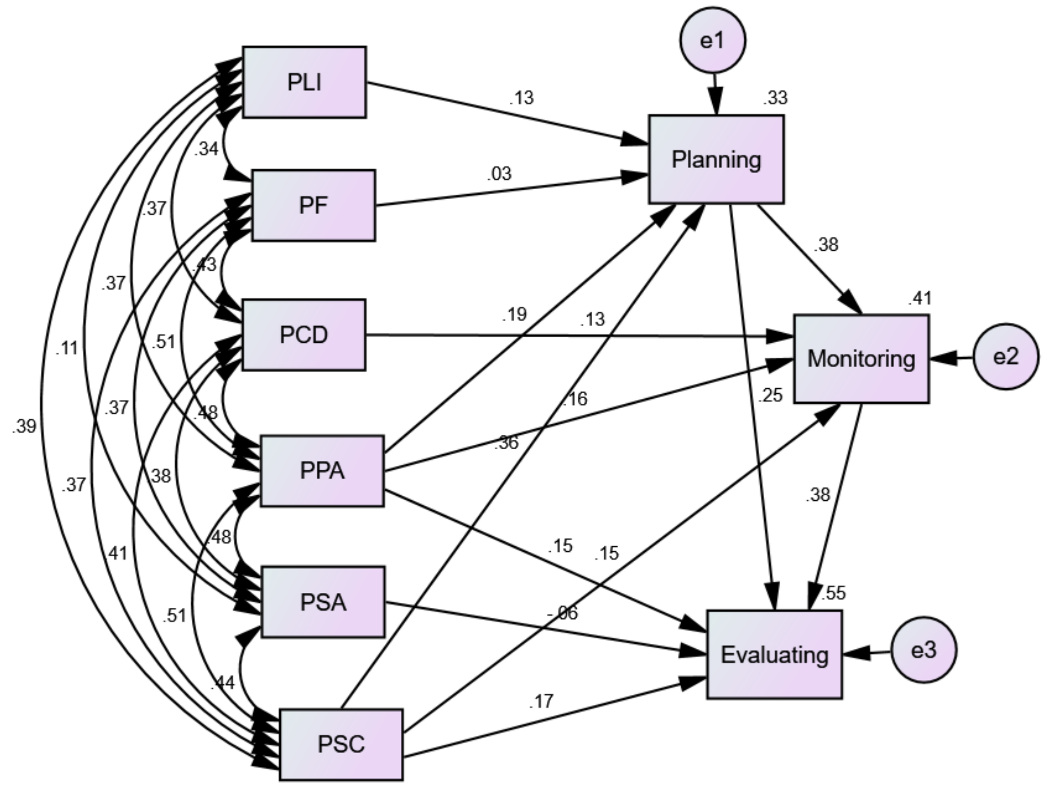
Standardized and Unstandardized Coefficients of the Structural Model

Outcome		Predictor	*B*	*SE B*	*β*	*t*	*P*
Planning	<---	PPA	0.17	0.06	.19	2.576	.010
Planning	<---	PLI	0.09	0.05	.13	2.034	.042
Planning	<---	PF	0.03	0.06	.03	0.504	.614
Planning	<---	PSC	0.32	0.06	.36	5.316	.000
Monitoring	<---	Planning	0.44	0.07	.38	5.964	.000
Monitoring	<---	PCD	0.16	0.07	.14	2.191	.028
Monitoring	<---	PPA	0.16	0.07	.16	2.348	.019
Monitoring	<---	PSC	0.16	0.70	.15	2.247	.025
Evaluating	<---	PSC	0.28	0.08	.17	2.755	.006
Evaluating	<---	Monitoring	0.49	0.08	.39	6.453	.000
Evaluating	<---	PSA	-0.09	0.09	-.06	-1.061	.289
Evaluating	<---	Planning	0.36	0.09	.25	4.090	.000
Evaluating	<---	PPA	0.19	0.08	.15	2.467	.014

Note. *B*=unstandardised beta; *SE B*=standard error for unstandardised beta; *β* =standardised beta; t=t-test statistic; p=probability value.

Figure 4

Standardized Solutions of the Structural Model



Note: Formative assessment strategies are: PLI=Perceived Learning Intentions; PF=Perceived Feedback; PCD=Perceived Classroom Discussion; PPA=Perceived Peer Assessment; PSA=Perceived Self-Assessment; PSC=Perceived Success Criteria, while Metacognitive skills are: Planning, Monitoring, and Evaluating.

4.2.5. Discussion

This study aimed to establish the relationship between mathematics teachers' perceptions of formative assessment and metacognitive regulation using the TAFL-Q and MAIT. First, the validity of the two questionnaires was examined in the Kenyan context. A confirmatory factor analysis on a sample of 180 mathematics teachers showed that the TAFL-Q had a poor fit whereas the MAIT had a good fit for the sample. An exploratory factor analysis of the TAFL-Q using the same sample resulted in a six-factor structure consisting of 19 items from the original 28 items. The six-factor solution explained 66.0% of the total variance. The new structure was deemed suitable since every scale had at least three items with good reliabilities. Furthermore, the new factors were still in line with the theoretical framework of formative assessment (Black & Wiliam, 2009).

The new version of the TAFL-Q was subjected to a confirmatory factor analysis using a different sample of 213 mathematics teachers. The new structure resulted in an acceptable model with good fit indices. The confirmatory analysis of the three scales representing metacognitive regulation on the MAIT scale had good fit indices on both the first and the second samples. There was, therefore, no adjustment of the items on the MAIT scale.

Although the TAFL-Q has gained popularity in different cultural contexts, some studies that have used this questionnaire failed to examine the contextual suitability of the instrument. For instance, Öz (2014) used the questionnaire to measure the perceptions of Turkish English teachers but failed to report on confirmatory factor analysis. Similarly, the same questionnaire was used in Tanzania before a confirmatory analysis of the factors (Kyaruzi et al., 2018). However, Nasr et al. (2018) found the questionnaire fit to measure the perceptions of Iranian English teachers. Due to cultural differences and different educational practices, questionnaires may fail to elicit similar structures across different samples (Brown et al., 2017). It is therefore important to examine the structure of an existing scale when dealing with a different cultural context.

Several inventories have been used to conceptualize metacognition for the past four decades (e.g., Balcikanli, 2011; Miholic, 1994; Paris & Jacobs, 1984; Schraw & Dennison, 1994). The most widely used and cited in the literature is the Metacognitive Awareness Inventory (MAI) which was developed by Schraw and Dennison (1994) to measure metacognitive awareness for adults. Since the focus was generally on adults, Balcikanli (2011) found it necessary to develop a metacognitive inventory that specifically measures teachers' metacognitive awareness. The MAIT was therefore simply a modification of the MAI to fit the teaching context. Although the MAIT has been less widely used, the original version (MAI) has proved its validity across different cultural contexts. The MAIT was found suitable for measuring the level of metacognitive skills among secondary school mathematics teachers in Kenya.

The second part of the study involved assessing the relationship between teachers' perceptions of formative assessment and metacognitive regulation. The results showed a significant positive relationship between most of the factors. For instance, learning intentions

(PLI), success criteria (PSC), and peer assessment (PPA) significantly predicted teachers evaluating skills. This implies that mathematics teachers develop evaluation skills when they use formative assessment strategies like sharing learning intentions and success criteria with students and engineering them as instructional resources of one another. Self-assessment, however, had no significant effect on teachers evaluating skills.

Monitoring strategies were significantly predicted by classroom discussion (PCD) and peer assessment (PPA). Through classroom discussion and peer assessment, students can reflect and monitor their learning process. Planning strategies were significantly predicted by learning intentions (PLI), peer assessment (PPA), and success criteria (PSC). This shows that mathematics teachers' planning strategies are enhanced when they share learning intentions and success criteria with students. When the planning strategies are in place, it becomes easier for teachers to monitor and evaluate the learning.

It's worth noting that among the predictors, peer assessment and success criteria had a significant effect on the three outcome variables of metacognitive skills. This underscores the importance of these formative assessment strategies in metacognitive regulation. Overall, the results were in line with our hypothesized relationship except for the relationship between self-assessment and evaluation which turned out to be negative although insignificant.

4.2.6. Conclusion

The current study builds on the work of Black and Wiliam (2009), Pat-El et al. (2013), and Balcikanli (2011) to try and conceptualize the relationship between formative assessment and metacognitive regulation. Unlike metacognitive regulation which has been consistently conceptualized and widely measured, formative assessment (assessment for learning) was first measured instrumentally by Pat-El et al. (2013). Furthermore, only a few studies have assessed the relationship between formative assessment and metacognitive regulation. The findings of our study show that monitoring and scaffolding dimensions as described by Pat-El et al. (2013) can be disintegrated into subcomponents of Learning Intentions, Feedback, Classroom Discussion, Peer Assessment, Self-Assessment, and Success Criteria. Our results have shown that formative assessment strategies predict teachers' levels of metacognitive regulation.

There is a need for subsequent work to consider other variables of formative assessment and how they relate to metacognitive regulation. The work should also involve multiple approaches to measuring the constructs of formative assessment and metacognitive regulation. More research is needed to test the relationship exhibited in the present study with other samples, especially in different cultural contexts. The findings of this study will contribute to future theory development and designing effective intervention programs for classroom instruction.

4.3. Study 3: In-Service Mathematics Teachers' Conception and Perceptions of Metacognition in their Teaching Experience

4.3.1. Introduction

Metacognition refers to being aware of one's thought processes (Merriam-Webster, 2012). According to Flavell (1979), metacognition is the knowledge that helps one to regulate his/her cognitive activities in the learning processes. Metacognition has been regarded as consisting of two major parts namely, metacognitive knowledge and metacognitive skills (Lai & Viering, 2012; Veenman, 2011; Veenman et al., 2006; Williams & Atkins, 2009). Over the past four decades, research on metacognition and its implications to teaching and learning is increasingly becoming an area of focus among educational researchers. This is due to the proven benefits of metacognition in children's learning (e.g., Csíkos & Steklács, 2010; Naseri et al., 2017; Roll et al., 2011).

Metacognitive knowledge has widely been categorized into declarative, procedural, and conditional knowledge (Harris et al., 2010). Metacognitive skills or regulation of cognition have also been conceptualized under three components: planning, monitoring, and evaluating (Harris et al., 2010). Metacognitive skills generally encompass the acquired repertoire of strategic knowledge necessary for monitoring, directing, and managing learning activities (Veenman, 2011). Self-monitoring, planning, and self-evaluation are therefore part of the metacognitive skills that can be practiced in a learning situation. According to Veenman (2011), the indicators of metacognitive skillfulness include planning, monitoring, note-taking, and time and resource management.

Various studies have shown that the training of students in metacognitive skills greatly enhances their performance (e.g., Donker et al., 2014; Kramarski & Mevarech, 2003). A study by Veenman & Spaans (2005) on solving mathematical word problems by secondary school students revealed that students perform better when learning is supported by a series of metacognitive cues. Other studies have also shown the need for metacognitive skills training in schools (e.g., Mevarech & Fridkin, 2006; Muijs et al., 2014).

Teachers' Metacognition and Academic Achievement

Teachers are expected to prepare students to construct their knowledge. This implies that teachers must model their teaching to the extent that students can engage in monitoring and self-regulating learning behaviours. Significant research on how metacognitive skills training impacts learners' academic achievement has been done (e.g., Baas et al., 2014; Hattie, 2013; Stel & Veenman, 2010). Some studies have also focused on pre-service teachers' metacognition and its impact on teachers' academic achievement.

Whereas Hashmi et al. (2019) found a negative correlation between metacognition and academic achievement among the prospective teachers, other studies yielded a positive correlation (Abdellah, 2015; Young & Fry, 2008). Studies related to teachers' metacognition have mostly focused on pre-service teachers' metacognition. However, little is known about in-service teachers' metacognition. It is necessary to also assess how practicing teachers are utilizing their metacognitive knowledge and skills in the process of teaching and learning.

Metacognition and Background Factors (Teaching Experience, Gender, and Qualification)

According to Young and Fry (2008), the more experienced (graduate) education students had higher scores of their metacognitive skills as compared to the less experienced (undergraduate) education students. There was however no difference regarding their scores on knowledge of cognition. A recent study by Kallio et al. (2020) revealed that more experienced teachers (more than 10 years) had a higher perception of support of their learners' metacognitive awareness than the less experienced teachers did. Stewart et al. (2007) and Jiang et al. (2016) also found similar results.

Regarding gender and teachers' qualifications, Kallio et al. (2020) observed that women's perception of the support of their learners' metacognition was higher than men's perceptions, and teachers with masters' degrees were seen to provide more assistance to their learners as compared to bachelors' degree holders. However, Usher (2019) found no significant gender differences among the pre-service teachers. Likewise, in the study carried out by Ekici et al. (2019), the pre-service male and female teachers had similar perceptions of their metacognition. Ibrahim and Watts (2016) too observed that both men and women had similar perceptions of their metacognitive skills. These studies however mainly focused on prospective teachers. It will be interesting to find out how gender, teachers' qualifications, and teaching experience affect the Kenyan in-service mathematics teachers' perceptions of metacognitive awareness.

The Current Study

Research on metacognition has mainly focused on the learners and pre-service teachers. Few recent studies on in-service teachers' metacognition have focused on preschool, primary and special education teachers (Kallio et al., 2020; Sulaiman et al., 2021; Thienngam et al., 2020). No study has however focused on mathematics teachers' metacognition in secondary schools. Studies have shown the need for training the learners in metacognition. The implication is that teachers must be metacognitive in their teaching so that they teach or model the same to their learners (Martinez, 2006; Tanner, 2012; Wilson & Bai, 2010). Teachers' understanding of their levels of metacognition can be a head start in the training of learners' metacognitive skills. In-service teachers' metacognition is however one of the under-researched areas.

The current study explored the levels of metacognitive awareness among practicing mathematics teachers in Kenya. The motivation behind this study is the fact that no study has been done in Kenya regarding teachers' metacognition despite the role it plays in students' achievement. This study will therefore form a basis for further research on metacognitive awareness among teachers in Kenya and how it can be utilized for the benefit of students' learning. The present study responded to the following research questions:

1. To what extent do secondary school mathematics teachers in Kenya perceive their use of metacognitive knowledge and skills in teaching mathematics?
2. Is there any effect of gender, teaching experience, and level of education on the metacognitive awareness of secondary school mathematics teachers in Kenya?
3. What conception of metacognition do Kenyan secondary school mathematics teachers report?

4.3.2. Methodology

Sample

The sample consisted of 213 (157 males) secondary school mathematics teachers from 50 selected public secondary schools in Kenya. . The 50 schools were obtained through a stratified random sampling technique to include the different categories of schools (National, County, or

Sub-county) and different school types (boys', girls' or mixed). The sample of 213 teachers was then purposefully obtained by considering all mathematics teachers in the selected schools who taught a grade 11 class. Out of 213 participants, 166 (77.9%) had a bachelor of education (BEd) degree; 19 (8.9%) had either a Bachelor of Arts or Bachelor of Science (BA/BSc) degree; 25 (11.7%) had a diploma in education while 3 (1.4%) had a Master's degree in education (MEd). Regarding teaching experience, 125 (58.7%) had an experience of up to 5 years; 41 (17.2%) an experience of between 6 to 10 years; 27 (12.7%) an experience between 11 to 15 years and 20 (9.4%) with an experience above 15 years.

Design

The study adopted a descriptive cross-sectional survey design that utilized both questionnaires and personal interviews (Creswell, 2009). This method was deemed viable for the current study since the aim was to relate the quantitative and the qualitative findings and hence gain more insight into the mathematics teachers' perception and conception of metacognition. The questionnaire was used to collect the quantitative data whereas structured interviews acquired qualitative data.

Instruments

1. Questionnaire

The self-report paper and pencil questionnaire comprised of two parts. The first part contained questions related to demographic characteristics and the second part consisted of 24 items (questions) adopted from Balcikanli (2011) as shown in Table 18. The 24 items were used to rate the mathematics teachers' metacognitive awareness. The Metacognitive Awareness Inventory for Teachers (MAIT) scale (Balcikanli, 2011, p. 1323) was developed for measuring teachers' metacognitive awareness. In the current study, the scale was adopted for mathematics teachers. The teachers were clearly instructed to relate the responses to their mathematics classes since some mathematics teachers also taught a different second subject.

The MAIT scale is composed of two dimensions each with 12 items. The metacognitive knowledge dimension has three subscales each with four items: declarative knowledge (DK), procedural knowledge (PK), and conditional knowledge (CK). The metacognitive regulation dimension included planning (P), monitoring (M), and evaluating (E) subscales each with four items (Balcikanli, 2011, p. 1323). Items were rated on a 5-point Likert scale ranging from 'Not at All' (NA=0) to 'Always' (A=4).

Table 18

A 24 Item Questionnaire and the Subscales

Scale	Items	Questions
Declarative Knowledge (DK)	1	I am aware of the strengths and weaknesses in my teaching
	2	I know what skills are most important to be a good teacher
	3	I know what I am expected to teach
	4	I have control over how well I teach
Procedural Knowledge (PK)	5	I try to use teaching techniques that worked in the past
	6	I have a specific reason for choosing each teaching technique I use in the class
	7	I am aware of what teaching techniques I use while I am teaching
	8	I use helpful teaching techniques automatically
Conditional Knowledge (CK)	9	I use my strengths to compensate for my weaknesses in my teaching
	10	I can motivate myself to teach when I need to teach
	11	I use different teaching techniques depending on the situation
	12	I know when each teaching technique I use will be most effective
Planning (P)	13	I pace myself while I am teaching to have enough time
	14	I teach my specific goals before I start teaching
	15	I ask myself questions about the teaching materials I am going to use
	16	I organize my time to best accomplish my teaching goals
Monitoring (M)	17	I ask myself periodically if I meet my teaching goals while I am teaching
	18	I find myself assessing how useful my teaching techniques are while I am teaching
	19	I check regularly to what extent my students comprehend the topic while I am teaching
	20	I ask myself questions about how well I am doing while I am teaching
Evaluating (E)	21	I ask myself how well I have accomplished my teaching goals once I have finished
	22	I ask myself if I could have used different techniques after each teaching experience
	23	After teaching a point, I ask myself if I would use it more effectively next time
	24	I ask myself if I have considered all possible techniques after teaching a point

Note. Adopted from, Metacognitive awareness inventory for teachers (MAIT), by C. Balcikanli, 2011, *Electronic Journal of Research in Educational Psychology*, *9* (3), 1309–1332, adopted with permission.

2. Interview Schedule

Eight interview questions were developed by the researcher to get more insight into the themes covered by the subscales on the MAIT. The interview sample consisted of 10 teachers (men=7) who were selected from ten randomly selected schools. The structured questions focused on the teachers' understanding and use of metacognitive knowledge and skills in their teaching. The interview questions were as follows:

1. What would you consider as your strengths and weaknesses in teaching mathematics?
2. What do you consider before deciding on a particular method to use in your teaching?
3. How do you make use of your strengths and weaknesses?
4. How often do you prepare your lesson plans?
5. How often do you evaluate your teaching goals at the end of each lesson?
6. How often do you question whether you are meeting your teaching objectives while teaching?
7. To what extent did your teaching approach predict students' achievement at the end of term one exams?
8. Do you consider yourself a metacognitive teacher? Briefly explain.

Data Collection Procedure

The research was conducted after obtaining ethical approval from Pwani University of Kenya and after getting a license from the Nation Commission for Science, Technology, and Innovation. The researchers visited the schools and administered the questionnaires to the mathematics teachers with the assistance of the heads of mathematics departments. The filling of each questionnaire took approximately 25 minutes to complete. The interview was conducted on different dates after seeking the teachers' consent. The lead researcher interviewed teachers separately after assuring them of the confidentiality of the recorded scripts. The interview took between 30 to 35 minutes for every teacher.

Data Analysis

1. Questionnaire Analysis

The adapted MAIT scale was validated by performing a confirmatory factor analysis on a separate sample of 180 mathematics teachers. Different fit indices, which included CMIN/DF, RMSEA, SRMR, TLI, and the CFI, were used to test the model fit to the Kenyan sample. Confirmatory factor analysis (N = 180) of the six-factor model resulted to acceptable fit indices as follows: CMIN/DF = 2.275, RMSEA = .084, SRMR = .058, TLI = .913, CFI= .933 (Ho, 2006; Schumacker & Lomax, 2004). Reliability analysis of the questionnaire was determined using Cronbach's Alpha for the different subscales and the overall scale (DK=.65; PK=.67; CK=.73; P=.64; M=.70; E=.86; Overall scale=.92). All the alphas were within the acceptable range hence confirming that the MAIT measured precisely the underlying constructs. The analysis of the item-total correlations for the overall MAIT scale is presented in Appendix C2.

2. Interview Analysis

The thematic content analysis approach was used to scrutinize interview data by first identifying the common themes across the data set (Creswell, 1994). The analysis involved six steps as suggested by Miles and Huberman (1994): (1) Listening to the tape and transcription of the interview (2) Becoming familiar with the transcripts by reading them over and over (3) Coding (4) Summarizing the coded data (5) Data interpretation (6) confirming the findings. Confirmation of the findings was done by taking the transcript and the interpretation back to the respondents and they were asked if they agreed with the interpretation. All the participants agreed with the interpretation.

4.3.3. Findings

The findings of the present study were categorized into quantitative (descriptive, t-test, and ANOVA statistics) and qualitative (interview analysis).

Descriptive Statistics

Descriptive statistics were used to examine the perceptions of teachers regarding their metacognitive awareness. The analysis indicated a mean score of 3.31 for items under the metacognitive knowledge dimension and a mean score of 2.91 for items under the metacognitive skills dimension. All the items under the metacognitive knowledge dimension had a mean score of between 3.11 and 3.53 which corresponded to “often” on the Likert scale implying mathematics teachers had a high perception of their metacognitive knowledge. Regarding the metacognitive skills dimension, the mean score range was between 2.63 and 3.13 with an average of 2.91 corresponding to “often”. Teachers had the lowest perception of Items in the subscale of evaluative skills (21, 22, 23, and 24) as compared to items in other subscales. Generally, the perceptions of mathematics teachers regarding their level of metacognition can be rated as positive although the perceptions regarding metacognitive knowledge were higher than perceptions regarding metacognitive skills. Tables 19 and 20 give a summary of the descriptive statistics regarding the teachers’ perceptions of their levels of metacognitive awareness.

Table 19

Descriptive Statistics of Teachers' Perceptions on Metacognitive Knowledge (N=213)

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Overall mean
M	3.32	3.36	3.46	3.53	3.20	3.37	3.44	3.24	3.27	3.27	3.20	3.11	3.31
SD	0.79	0.74	0.68	.75	0.83	0.78	0.74	0.85	0.78	0.88	0.92	0.83	0.45

Table 20

Descriptive Statistics of Teachers' Perceptions on Metacognitive Skills (N=213)

	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20	Item 21	Item 22	Item 23	Item 24	Overall mean
M	3.00	2.59	3.08	3.13	3.00	3.05	3.11	3.07	2.90	2.63	2.64	2.70	2.91
SD	0.89	1.22	0.95	0.85	0.98	0.96	0.91	1.02	1.11	1.15	1.24	1.21	0.71

Further analysis of the three subscales of metacognitive knowledge revealed that teachers had a higher perception of their declarative knowledge as compared to procedural knowledge and conditional knowledge. Regarding the subscales of metacognitive skills, teachers had a higher perception of their monitoring skills as compared to planning and evaluating. Table 21 shows the descriptive results of the six subscales of the metacognitive scale.

Table 21

Descriptive Statistics of Perceptions on Metacognitive Awareness Sub-Skills (N=213)

	Declarative knowledge	Procedural knowledge	Conditional knowledge	Planning	Monitoring	Evaluating
M	3.41	3.31	3.32	2.95	3.05	2.72
SD	0.52	0.53	0.63	0.67	0.79	1.00

T-Test and ANONA Results

Independent samples T-test and one-way ANOVA were used to examine the effect of background factors (gender, teaching experience, and qualification) on teachers' perceptions of their metacognitive awareness. The t-test results exhibited non-significant differences in the means of male (M=3.31, SD=0.46) and female teachers (M=3.32, SD=0.40); t (211) =-.13, p=.90

for metacognitive knowledge. Similarly, no significant difference in the means of male (M=2.88, SD=0.73) and female teachers (M=2.98, SD=0.60); t (211) =-1.08, p=.28 for metacognitive skills was found.

There was no effect of teacher qualification on teachers' perceptions at p=.05 for the four categories of teacher qualification [F (3, 209) =.56, p=.64] for metacognitive knowledge and metacognitive skills [F (3, 209) =.59, p=.62]. Similarly, a one-way ANOVA conducted on teaching experience resulted in a non-significant difference [F (3, 209) =.16, p=.92] for the four categories of teaching experience regarding metacognitive knowledge. There was also no significant difference [F (3,209) =.53, p=.66] for the four categories of teaching experience regarding metacognitive skills. Table 22 gives a summary of the mean and standard deviations of metacognitive knowledge and skills for the background factors.

Table 22

Descriptive statistics of Teachers' Perceptions Based on Background Factors

Background factors		Metacognitive Knowledge		Metacognitive Skills	
		M	*SD*	*M*	*SD*
Gender	Male	3.31	0.46	2.88	0.73
	Female	3.32	0.40	2.98	0.63
Teacher qualification	BEd	3.32	0.45	2.92	0.72
	BA/BSc	3.20	0.47	2.90	0.72
	Diploma	3.30	0.41	2.70	0.70
	MEd	3.53	0.53	3.05	0.55
Teaching experience	Up to 5 years	3.30	0.44	2.93	0.74
	6 to 10 years	3.31	0.47	2.85	0.69
	11 to 15 years	3.35	0.47	2.96	0.68
	Above 15 years	3.27	0.46	2.75	0.62

Interview Analysis

Ten teachers participated in the interview. The interview consisted of eight semi-structured questions with associated probes. The interview aimed to help triangulate the research and hence strengthen the findings from the quantitative data. The analysis centred on the two major themes of metacognition and the six sub-themes that had already been established during the quantitative phase.

Theme 1: Metacognitive Knowledge

i) Declarative knowledge

The interview guide for questions on declarative knowledge sought to find out more information on teachers' awareness of person, task, and strategy variables. The main question that was asked related to teachers' declarative knowledge about their strengths and weaknesses: *what would you consider as your strengths and weaknesses in teaching mathematics?* This question aimed to find out if teachers were aware of their strengths and weaknesses. All teachers interviewed responded that they were aware of their strengths related to content knowledge, class control, giving feedback to students, and the use of different student-centred approaches in teaching. Regarding knowledge of their weaknesses, eight teachers admitted they have some weaknesses and were aware of them. Some of the weaknesses mentioned included poor time management and a negative attitude towards some topics.

> Teacher 1: I always have a problem with time management in almost all of my classes however much I try. To me, 40 minutes are never enough and I even hardly notice how time passes when I am teaching. I think I do much of the talking. I may need to work on this weakness.

Two teachers were however skeptical about mentioning their weaknesses. One of the teachers (teacher 4) said:

> I enjoy teaching mathematics and I don't have issues with any topic. I have taught mathematics for over 15 years and the major challenge I have always encountered is on dealing with students' negative attitudes towards some topics.

ii) Procedural Knowledge

The interview guide for questions on procedural knowledge sought to gain an understanding of teachers' utilization of their procedural knowledge. The main question asked was: *what do you consider before deciding on a particular method to use in your teaching?* This question elicited varied responses based on teachers' teaching experience. One more experienced teacher reported that he used various methods depending on the topic and the class size.

> Teacher 2: My choice of the teaching method depends on the topic and the class size. If for example am teaching a topic like statistics, I just divide students into groups and give

them a task to work on. Some topics will require me to do a demonstration on the chalkboard and ask students to work individually especially if the class is large.

One teacher (less than 5 years of teaching experience) demonstrated the use of procedural knowledge although he seemed to be slightly limited as compared to the more experienced teacher.

> Teacher 7: Mathematics is a hands-on activity and so I always prefer working with students in small groups. I first do a demonstration on the chalkboard then give them activities in groups. I don't want to teach mathematics the way I was taught (teacher-centered approach) because it makes students passive participants.

iii) Conditional Knowledge

Questions on conditional knowledge investigated teachers' application of their declarative and procedural knowledge. The main question concerned how teachers use their strengths and weaknesses: *how do you make use of your strengths and weaknesses?* Most teachers said that they maximize their strengths and seek assistance in their weak areas.

> Teacher 9: Whenever I come across a difficult problem, I usually ask my colleagues to assist or sometimes I give the challenge to my students. My students may know some concepts better than I do. Team teaching has worked well in my school since no one has a monopoly on knowledge.

Theme 2: Metacognitive skills

i) Planning

The questions on planning focused on how teachers plan their teaching to attain lesson goals. The main question asked was: *how often do you prepare your lesson plans?* All the 10 teachers admitted that they rarely prepared lesson plans although they always had an idea of what to teach. Regarding lesson organization, the ten teachers mentioned that they divided their lesson into the introduction, body, and conclusion although in most cases they hardly followed the plan.

> Teacher 6: I prepare the lesson plan once in a while because it's part of the requirements. It's rarely practical for every lesson. Like on average I have 28 lessons per week so it's

almost impossible to have an elaborate lesson plan. Besides, I may prepare and end up failing to follow it.

ii) Monitoring

The questions on monitoring focused on teachers' self-testing skills needed to regulate learning. The main question asked was: *how often do you question yourself whether you are meeting your teaching objectives while teaching?* Questions on monitoring elicited different responses, which included the following:

> Teacher 5: Based on my experience, with time I have known what works for me and what fails to work. In most cases, I assess my teaching techniques after the lesson. I can be able to tell if students understood the concept or if they need a remedial lesson.

> Teacher 10: Sometimes it's hard to check the understanding of the students during the lesson but after giving them some tasks, I can be able to check the understanding based on their responses to the tasks.

The analysis showed that teachers generally self-test their skills after the lessons. Some of the reasons they pointed out as affecting their self-monitoring processes were the heavy workload and large class sizes. It seemed from the responses that teachers engage in self-monitoring processes only after assessing the students' performance on continuous assessment tests.

iii) Evaluating

The questions on evaluating related to how teachers examine their progress as they strive to achieve the set goals. An example of the question asked was: *how often do you evaluate your teaching goals at the end of each lesson?* Responses to the questions relating to evaluation showed that this was one of the skills that teachers rarely thought about. All teachers said they rarely evaluate their teaching goals because in most cases they hardly prepare the lesson plan.

> Teacher 9: Once I finish teaching the first class, I am always in a hurry to attend the next class. I don't even have time to evaluate my teaching as much as I know the importance of doing so. During my free time, I am busy marking students' books and rarely get time for reflection.

Related to evaluation, teachers were requested to rate how their teaching approach predicted their students' achievement based on the previous end-of-term performance (*Do what extent did your teaching approach predict students' achievement at the end of term one exams?*). Out of the ten teachers, seven teachers reported that they tried to do their best to employ the student-centered approaches in their teaching but their efforts were rarely reflected in students' performance. When they were requested to rate on a five-point scale, seven rated themselves on a scale of '1' which corresponded to rarely, and three rated themselves on a scale of '2' which corresponded to sometimes.

The final question asked teachers whether they considered themselves as metacognitive teachers (*Do you consider yourself a metacognitive teacher? Briefly explain!*). This seemed to be a challenging question for the teachers as most of them exhibited little understanding of what metacognition was all about. The first two teachers that were interviewed considered themselves as metacognitive teachers but their reasons showed a lack of clear understanding of what metacognition involved.

> Teacher 1: I think I am a metacognitive teacher because I involve my students in active learning through discussions and group activities that are mainly hands-on.
>
> Teacher 2: I think I am a metacognitive teacher because I prepare my lessons and reflect on what I want to teach even though I don't always have a lesson plan with me in class.

Based on the responses of the first two teachers, there was a need to explain the meaning of metacognition and its components to the remaining eight teachers before they were asked whether they considered themselves as metacognitive teachers or otherwise. All eight teachers said they were aware of their metacognitive knowledge (declarative, procedural, and conditional knowledge), and used it largely while teaching students. Regarding the regulation of cognition, five of them reported that they monitor and reflect on their teaching but in most cases fail to plan and evaluate their teaching. Three teachers felt they were less effective in utilizing their metacognitive skills:

> Teacher 5: I think to be a metacognitive teacher needs more than just writing down a lesson plan. One must be intentional and have enough time to plan and reflect on the lesson to be taught. As much as I would wish to plan and evaluate my teaching, I have little time for that. On a scale of "1 to 5", I can give myself a 3.

Teacher 7: I sometimes reflect on the topic but sometimes fail to prepare adequately due to other factors like large class sizes and greater workload without any extra pay.

4.3.4. Discussion

Previous research findings have suggested that metacognitive skills training improve students' achievement (Baas et al. 2015; Hattie 2013; Muijs et al. 2014; Stel & Veenman, 2010). These findings are however based on the assumption that teachers are themselves metacognitive and can train students on how to use metacognitive strategies in their learning. Perhaps the question that researchers need to ask themselves is whether the teachers are aware of their metacognition. The argument paused in the present study is that one can only give out what he/she has. In other words, teachers can only train students on how to be metacognitive if they are also metacognitive. The present study thus explored the mathematics teachers' perceptions of their metacognition, the effect of background factors on metacognition, and teachers' conception of metacognition.

Teachers' Perceptions

Research question one established the extent to which secondary school mathematics teachers in Kenya perceived their level of metacognitive knowledge and skills in teaching mathematics. Descriptive statistics analysis showed that mathematics teachers rated themselves highly regarding their level of metacognitive awareness although their rating for metacognitive knowledge was higher than the rating for metacognitive skills.

The analysis thus implies that teachers have positive perceptions of their metacognitive awareness. Descriptive statistics results were also supported by the qualitative analysis of the interview where teachers reported that they were aware of their metacognitive knowledge and skills. Özsoy and Günindi (2011) similarly found a medium to high-level metacognitive awareness of the pre-service teachers. Koc and Kuvac (2016) as well found positive perceptions of metacognitive awareness by prospective science teachers.

Background Factors and Teachers' Perceptions

The second research question examined if there was any effect of gender, teaching experience, and level of education on the metacognitive awareness of secondary school mathematics teachers in Kenya. This question was answered by conducting an independent samples t-test and ANOVA statistics. The outcome of the analysis showed a non-significant effect of gender, teaching experience, and level of education on both the metacognitive knowledge and metacognitive skills of the participants.

Previous studies although dealing with pre-service teachers also found statistically non-significant gender differences in teachers' metacognitive awareness (Alci & Karatas, 2011; Ekici et al., 2019). These findings show that background factors of gender, experience and level of education have little influence on the in-service teachers' perceptions of metacognition.

Teachers' Conception of Metacognition

The third research question sought to get a deeper understanding of the Kenyan secondary school mathematics teachers' conception of metacognition as reported through interviews. The analysis was based on the two major themes and the subthemes identified during the questionnaire analysis. The first theme, metacognitive knowledge which relates to person, task, and strategy variables (Mahdav, 2014) is divided into declarative knowledge, procedural knowledge, and conditional knowledge. The interview of teachers regarding their metacognitive knowledge revealed that teachers made use of their metacognitive knowledge. For instance, they were able to identify their strengths and weaknesses (declarative knowledge); used varied teaching approaches (procedural knowledge); and made good use of their strengths and weaknesses (conditional knowledge).

The second theme analyzed teachers' metacognitive skills which encompass planning, monitoring, and evaluation. Regulation of cognition enables one to know what to do (task orientation), what to achieve (goal setting) and how to achieve the goal (planning). Teachers get involved in the regulation of cognition when they plan their teaching, monitor themselves during lesson delivery, and evaluate the outcome of their teaching. Analysis of teachers' metacognitive skills revealed that teachers generally had an understanding of the skills involved in the regulation of cognition but rarely put them to use. These results show that more effort is needed for teachers to translate their awareness of metacognitive regulation to their teaching.

Teachers mentioned large class sizes, greater workload, and lack of motivation as some of the reasons that hindered their use of metacognitive skills. The interview analysis generally revealed that teachers failed to utilize their metacognitive skills fully. Earlier studies on teachers' promotion of self-regulatory strategies in classes have shown that teachers rarely supported the use of metacognitive skills in classrooms but instead promoted cognitive strategies (Dignath & Büttner, 2018; Dignath-van Ewijk et al., 2013; Spruce & Bol, 2015).

One interesting finding in the interview analysis was the misapprehension that some teachers had about the concept of metacognition. This was revealed through question eight when teachers were asked whether they considered themselves as metacognitive teachers and to give reasons why they thought so. Most of the teachers took metacognition to mean students being actively involved in learning and therefore the assumption was that as long as students were engaged in discussion and group work activities, metacognitive strategies were being utilized. The responses from the teachers revealed that the understanding of active learning is also misunderstood.

Active learning is defined as “anything that engages students in doing things and thinking about the things they are doing” (Bonwell & Eison, 1991, p. 19). Metacognition on the other hand involves students thinking about their learning process by engaging in self-regulatory processes like self-monitoring and self-evaluation. In other words, active learning appeals to cognitive processes instead of metacognitive processes. Active learning can be treated as a prerequisite for metacognition but it doesn’t imply that active engagement of students in learning activities is being metacognitive. Teachers in most cases apply active learning pedagogies like engaging students in hands-on activities but little do they consider minds-on activities.

Limitations

The present study was conducted in the western part of Kenya and the results can therefore be generalized only to the schools in the western region of Kenya. Future research can consider replicating the same study to other regions in the country. Although triangulation using questionnaires and interviews was employed in the study, future studies can also incorporate the observation of teachers in classes to increase the validity of the results.

4.3.5. Conclusions and Suggestions

Research has shown that metacognition is crucial to the learning process and it determines learning performance to a large extent. Both quantitative and qualitative analysis of data in this study has revealed that mathematics teachers perceive themselves as highly metacognitive. These perceptions are however hardly translated into the teachers’ actual classrooms. The report from the teachers’ interview showed that teachers face challenges that hinder them from utilizing their metacognitive skills. There is, therefore, a need for addressing the challenges so that teachers can consciously utilize their metacognitive skills and model the same to their students.

This study has revealed the need for explicit training and instructing students on metacognitive strategies. Teachers must be made aware of the benefits of metacognition in students' learning. There is, therefore, a need of introducing metacognition as a course in the teachers' training colleges and in-service training so that teachers are equipped on how to integrate metacognition in their teaching. It is important to integrate metacognition in the in-service training of teachers so that teachers will be able to model the same to the students. Teachers can also make use of other metacognitive strategies like mind maps and concept maps that will help students to consciously solve new and challenging problems.

4.4. Study 4: Proportional Reasoning in Mathematics: Construction and Validation of a Test to Measure Students' Proportional Reasoning on Rates, Ratios, and Proportions

4.4.1. Introduction

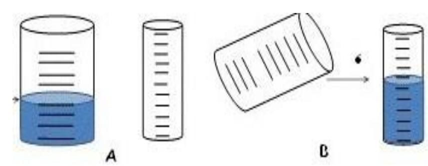
Proportional reasoning is generally regarded as the ability to compare objects using multiplicative reasoning instead of additive reasoning. According to Van de Walle and Lovin (2006), proportional reasoning involves a comparison of multiplicative relationships between quantities. Studies have revealed that most students have problems with proportional reasoning because they fail to differentiate between situations that call for additive reasoning and those that require multiplicative reasoning (Nunes & Bryant, 2009; Gläser & Riegler, 2015). Whereas additive reasoning originates from actions such as putting together and separating sets, multiplicative reasoning develops from actions such as one-to-many correspondence and sharing (Nunes & Bryant, 2015).

The importance of proportional reasoning cannot be overemphasized. This kind of reasoning is applied across all grades and subjects. Children as early as age five already have some intuitions about intensive quantities which form the basis of proportional reasoning (Nunes et al, 2012). Although the perception of proportional reasoning is mainly on ratios, rates, and rational numbers, proportionality is generally applied in other areas involving measurement (Ayan & Isiksal-Bostan, 2019). Proportional reasoning has also been regarded as a life skill that is crucial for daily decision-making (Howe & Bryant, 2011).

Literature Review

Proportional reasoning is a very important tool that children learn from early grades until high school. Although a significant number of studies have focused on promoting proportional reasoning in students, these studies have acknowledged the fact that proportional reasoning is challenging for most children (Al-Wattban, 2001; Charalambous & Pitta-Pantazi, 2007; Singh, 2000). Research has also revealed that proportional reasoning, particularly on rational numbers, affects students of all ages. The following two paraphrased examples which were given to a sample of German university students from the science, technology, engineering, and mathematics (STEM) faculty on a pretest (Gläser & Riegler, 2015) prove this assertion.

1. The diagrams below show two tins of different sizes but marked with the same scale on each of them. Oil is poured into the broad tin until it reaches the fourth mark. When the same oil is poured into the small tin as demonstrated in diagram B, it rises to the sixth mark. If both tins are emptied and oil is poured into the broad tin until it reaches the sixth mark, to what level can the oil rise if it is poured into the small cylinder?



2. Mary and Anna started riding bikes at different times and then rode at the same constant rate. When Mary covered six kilometers, Anna had covered eight kilometers. How far will Mary be when Anna covers 12 kilometres?

Question one tested students on proportional reasoning whereas question two on additive reasoning. With a sample of 446 students, only 47.7% answered the first question correctly, 45.5% answered the second question correctly, and only 22.6% answered both questions correctly. Gläser & Riegler (2015) noted that a considerable number of students applied proportional reasoning to the bike problem even when they were expected to use additive reasoning. In this case, the students had difficulties differentiating between scalar and functional relations.

Apart from Germany, these problems have also been used in other universities in the United States of America to identify difficulties with proportional reasoning. Nunes and Bryant

(2009) observed that students in primary and secondary schools use additive procedures to solve multiplicative reasoning problems and vice versa. When similar items were given to students on a post-test after an intervention, the researchers reported no change in the pattern of reasoning. Those students who used the additive reasoning in the pre-test also used the same reasoning in the post-test. The results suggested that most of the students were unable to reason proportionally.

Van Dooren, et al. (2005) presented similar items to the bike ride which tested on additive strategy but students ended up using proportional strategies. Van De Bock and Verschaffel (2010) observed that students also use scalar relations in solving ratio problems. According to Nunes and Bryant (2015), children reason more successfully about the problem when they can identify two quantities related by a fixed ratio. Children also tend to use intuitive strategies through experimentation without necessarily being aware of the proportional relationships (MacDonald & Wilkins, 2016). To develop proportional reasoning, students must, therefore, understand functional relations which are essential for mathematical modeling in science. Since scalar reasoning can develop without schooling, teachers should concentrate more on developing functional relations in students.

Theoretical Background

Different frameworks have been used by researchers in studies related to proportional reasoning. Lamon (1993) conceptualized proportional reasoning under four semantic problem types as follows: (a) “Well-Chunked Measures” which compares two extensive measures yielding to an intensive measure; (b) “Part-Part-Whole” which involves expressing an extensive measure of a single subset of a whole in two or more subsets; (c) “Associated Sets” where two sets lack a common connection and (d) “Stretchers and Shrinkers” which Involve problems requiring scaling up (stretching) or scaling down (shrinking).

Building on this framework, Allain (2000) developed an instrument testing on seven areas of proportional reasoning among the middle (secondary) school students which involved comparison, missing value, associated sets, part-part-whole mixture problems, comparisons, graphical interpretation, and Stretcher. Tjoe and de la Torre (2014) focused on the attributes of proportional reasoning that are relevant to the eighth-grade students (13-14 years). The attributes focused on skills and concepts, fractions, ratios and proportions, and algorithms for proportional reasoning problems.

Conceptual Framework for the Present Study

The present study builds on the Lamon (1993) and Allain (2000) frameworks and conceptualizes proportional reasoning under five aspects of proportional reasoning (Figure 5). It is worth mentioning that these five areas are in line with the instructional objectives as stated in the Kenya Institute of Curriculum and instruction (KICD) for secondary mathematics. As a way of improving education quality and learning outcomes, the Kenyan government introduced the competency-based curriculum (CBC) for basic education in 2018. The CBC pays attention to how learners can display the ability to transfer knowledge and skills learned in the classroom (KICD, 2017). To measure the competencies, teachers are expected to carry out the competency-based assessment (CBA).

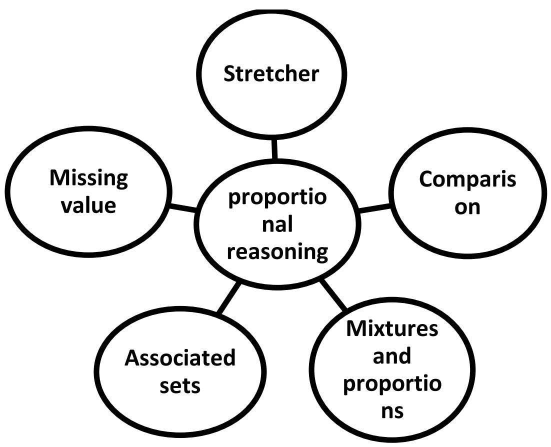
The mathematics curriculum in Kenya requires that learners are assessed on their ability to think critically, solve problems, and their responsiveness to authentic tasks. The test should, therefore, contain authentic tasks that require critical thinking. The CBA generally focuses more on the assessment of higher-order skills than lower-order skills. This study is part of a paradigm shift that moves away from assessing students for the sake of passing exams to authentic assessment.

The overarching aim of this study is therefore to develop mathematics tasks on the topic of rates, ratios, and proportions that will assess students' proportional reasoning skills. The review of literature has shown that students at all levels have difficulties in solving problems that require proportional reasoning and this affects their overall achievement in mathematics. One of the reasons for the poor performance is that students are unable to differentiate between scalar and functional relations in solving rational problems. It is therefore important that classroom teachers design interventions that can help students develop proportional reasoning.

The previous studies have majorly concentrated on identifying misconceptions that hinder students from developing proportional reasoning strategies. The studies thus intentionally included items that needed additive reasoning as well as multiplicative reasoning. Since the present instrument intends to measure the effectiveness of an intervention program, the focus is on items that measure multiplicative reasoning. This instrument is designed to measure the impact of formative assessment on students' achievement on a proportional reasoning test. This kind of instrument is more than a conventional test meant to test students' knowledge. The intention is to tap into students' cognitive ability and knowledge transfer to authentic situations. The aim of this study is therefore to develop an instrument that will measure students' level of proportional reasoning before and after an intervention. Specifically, the study answered the following question: *Is the mathematics achievement test on rates, ratios, and proportions a reliable and valid measure of students' proportional reasoning skills?*

Figure 5

A Conceptual Framework for the Proportional Reasoning Skills



4.4.2. Methodology: Development Framework

The items in the test were developed based on the phases described in the Standards for Psychological and Educational Testing (AERA et al., 1999) which in the case of this study were compressed into three phases. The first phase involved describing the intention of the test and the extent of the construct. The second phase involved development and evaluation, selection of the items, and scoring guide. The third phase involved piloting of the items, discussion of the pilot findings, assembly, and evaluation of the test.

Phase One: Describing the Purpose of the Test and the Scope of the Construct

The purpose of the test was to assess students' abilities on proportional reasoning related to the topic on rates, ratios, and proportions. The test comprised of word problems that relate to real-life situations covering five aspects of proportional reasoning: missing value, associated sets, mixtures and proportions, comparison problems, and stretcher. The overall aim of the test was to measure the impact of the formative assessment instruction at the pre and posttest. To achieve the intended aim of the test, we, therefore, expected the test to be reliable, valid, and manageable.

Phase Two: Development and Evaluation of the Test Specifications

This phase involved designing the format of the items, specifying the psychometric properties, considering the test duration, population composition, and the procedures for the test administration. The items together with a scoring guide were developed by the researchers based on: the sample problems in the literature (Allan, 2000; Gläser & Riegler, 2015; Tjoe & de la Torre, 2014); instructional objectives in the curriculum; and sample questions from the standardized national tests. The psychometric properties were determined by checking the validity and reliability of the test. The content validity was determined by a team of subject experts and the Item level analysis was done by computing the difficulty and discrimination indices. A total number of 10 items were carefully designed to reflect the instructional objectives as described in the Kenyan form three (grade 11) secondary mathematics coursebook and the different categories of proportional reasoning skills.

The focus was on rates, ratio, and proportions which are taught in form three (Grade 11) according to the Kenya Institute of Curriculum Development (KICD). Since the interest of the test was on higher thinking skills, only questions based on application, analysis, evaluation, and synthesis (higher thinking skills) were considered (Haladyna & Rodriguez, 2013). A rigorous revision and consideration of the syllabus and previous questions were done before deciding on the questions for the pilot study. Table 23 shows the areas tested and the problems from each area.

Table 23

Selected Problems for the Proportional Reasoning Test

Type of problem	Problem	Source
Comparison	1 Last week, Mary answered 24 out of 30 questions correctly in an exam. This week, she answered 20 out of 24 questions correctly in another exam. For which exam did Mary have better results? Explain your answer.	Researcher made
	2. Nafula bought 3 lollypops at 12 shillings and Atieno bought five lollypops at 20 shillings. Who used less amount of money? Explain your answer.	Adapted from (Allain, 2000) and modified
Missing value	3. How many glasses of orange juice can you make with 12 cups of water if it takes 8 cups of water to make 14 glasses of orange juice? Show your work.	Researcher made
	4. The diagrams below (see Appendix E) show two tins of different sizes but are marked with the same scale on each of them. Oil is poured into the broad tin until it reaches the fourth mark. When the same oil is poured into the small tin as demonstrated in diagram B, it rises to the sixth mark. If both tins are emptied and oil is poured into the broad tin until it reaches the sixth mark, to what level can this oil rise if it is poured into the small cylinder?	Adapted from Gläser & Riegler (2015)
Associated sets	5. A group of 7 girls shares 3 chapatis equally and another group of 9 boys shares 4 chapatis equally. Who gets a bigger size of chapati, a girl or a boy? Explain your answer.	Adapted from (Allain, 2000) and modified
	6. Mary has the option of working in Mombasa or Nairobi. She discovered that the workers in Mombasa work 8 hours per day and receive Ksh 24000 every 15 days and those in Nairobi work 6 hours per day and receive Ksh 20000 every 12 days. If she decides to work for 20 days, which job option will be best for her? Explain your answer.	Researcher made
Mixtures and proportions	7. Your father decides to give a piece of land as an inheritance to your three brothers Joe, Alex, and Peter in the ratio 4:5:3. Peter being the firstborn feels he has already accumulated enough wealth and therefore decides to share his portion equally with Joe and Alex. What will be the ratio of Joe's share to Alex's share? Show your work.	Researcher made
	8. In a mixture of 60 litres, the ratio of orange concentrate to water is 7:5. If the principal of a school wants to make orange juice for the students in the ratio of 3:2, how many liters of water should he add to the mixture? Show your work.	Researcher made
Stretcher	9. The figures below show two similar rectangles (see Appendix E). The height of the first rectangle is 6cm and the width is 8cm. The width of the second rectangle is 12cm. Explain how you would find the height of the larger rectangle.	Adapted from (Lamon, 1993) and modified
	10. The heights of two trees taken three years ago were eight feet for the tree (I) and ten feet for the tree (II). When the heights were taken today, tree (I) was 14 feet and tree (II) was 16 feet. Which of the trees increased most over the past three years?	Adapted from Allain (2000)

Scoring guide

The scoring guide was created concurrently with item construction. Each item had a scoring guide based on the instructional objective and the response expected. Since the questions were open-ended, varied strategies were expected from students. The scoring guide was therefore flexible and accommodative of different possible strategies that students could use to solve a given task. As noted by Csíkos (2016), children have personal preferences as to which strategy to apply on a given task.

Ratios and proportions are taught progressively since the early years of primary school. The students could therefore have several approaches to solving these tasks. The scoring guide hence gives room for consideration of a strategy that may make sense in solving a given task. All items had the same weight with a maximum of three points. The scoring guide (Table 24) was thus based on a 4 point scale ranging from 0 to 3 which was similar to the one created by Allain (2000). Table 25 shows a sample problem and the corresponding scoring guide.

Table 24

A scoring Guide

Score	Description
3	Shows understanding of the concept (1 point) Applies a strategy to solve the problem (1 point) Obtains the correct answer or explains the answer (1 point)
2	Shows understanding of the concept (1 point) Applies a strategy to solve the problem (1 point) Obtains an incorrect answer possibly due to a math error (0 point)
1	Possesses some understanding of the concept (1 point) Fails to apply a strategy to solve the problem or shows no work (0 points) Incomplete answer or obtained the correct answer probably by guessing(0 points)
0	Possesses a misconception (0 points) Applies an incorrect strategy to solve the problem or shows no work (0 points) Obtains an incorrect answer (0 points)

Table 25

Sample Problem and the Corresponding Scoring Guide

Sample problem	Problem objective and expected students response	Scoring guide
Last week, Mary answered 24 out of 30 questions correctly in an exam. This week, she answered 20 out of 24 questions in another exam. On which exam did Mary have better results? Explain your answer	This is a problem that tests students on comparison of ratios Expected response: Last week This week 24:30 20: 24 or 4: 5 5: 6 Mary had better results on this week's exam because the ratio is higher Alternatively, some students may just express ratios as fractions and compare which fraction is bigger.	• Can the student express the problem as a ratio form (1 point) • Can the student deduce from the ratios which performance was better (1 point) • Can the student explain the answer? This will help in deducing if the student has an understanding of the concept or could be having some misconceptions (1point)

Expert Review

After the test items were constructed, a team of experts consisting of three mathematics graduate students, and a professor assessed the item's content validity. During the review process, the experts focused on three issues. They first compared the items with various sample problems found in the relevant literature and established whether they were aligned with the Kenyan secondary mathematics curriculum. In the second step, they checked the difficulty level of the items and the third step involved checking whether the questions were correctly worded and appropriate to the sample and study context.

The same test was also given to two experienced mathematics teachers from two different secondary schools in Kenya for review. Based on a long-term teaching experience, they pointed out some weaknesses related to the phrasing of two questions and suggested how the questions could be rephrased to fit the cultural context of students. This was done jointly so the two experts were both in agreement regarding the test content and its alignment to the curriculum and the expected skills.

Item Revision

Based on the reviews from the expert judgment and the two experienced secondary school mathematics teachers from Kenya, two questions that were both adopted from the previous studies were revised to fit the cultural context of the current study. One of the questions was a pizza question which was testing on comparison skills (A group of 7 girls shares 3 pizzas equally and another group of 9 boys shares 4 pizzas equally. Who gets a bigger size of the pizza, a girl or a boy? Explain your answer)? Since students in Kenya rarely eat pizza, we replaced the pizza with 'chapati' which is common in Kenya and normally refers to flatbread.

The other question which was slightly modified was a question on missing value and required students to identify the number of cups of coffee that can be made using 12 cups of water given that eight cups of water can make 14 cups of coffee. We replaced coffee with orange juice since coffee is a less popular beverage in Kenya. The rest of the items were also thoroughly revised and some slight rephrasing was done on them. The final test consisted of 10 items together with the scoring rubric. Each question was given the same weight on a scale of 0 to 3.

Phase Three: Item Administration (Piloting)

After a thorough revision of the questions, the final version of the test was administered to 45 students from one mixed school. The participants consisted of 21 male students (M=1.38, SD=0.35) and 24 female students (M=1.27, SD=0.32). Before administering the test, students were provided with both oral and written instructions relating to the purpose of the test, how they should respond to questions, and to take the test seriously. The test was done in a controlled classroom under the supervision of a mathematics classroom teacher. The test duration was 60 minutes and all students finished within the stipulated time.

4.4.3. Pilot Findings and Discussions

The results are presented based on the three analyses that were done: reliability, item difficulty, and item discrimination.

Reliability

Cronbach's alpha was used to determine the internal consistency of the test. The overall test reliability for the 10 items was .83 with a mean of 1.30 and a standard deviation of 0.38 which was within the acceptable range (Cohen et al., 2007). All items had item-total correlation

(ITC) values ranging from .32 to .67 (see Table 27) indicating that the items fitted well to the whole test.

Item Difficulty

Item difficulty is a measure of the percentage of students answering a test item correctly and it helps in determining how easy the item is (Hopkins, 1998). Item difficulty index (p-value) can also be used to determine the validity of test items. The difficulty index ranges from .0 to 1.0 where the higher the p-values the greater the percentage of students answering the item correctly. The difficulty of each item was computed using the formula for open-ended items (Tjoe & de la Torre, 2014) as illustrated below:

$$p = \frac{\sim fX - nX_{min}}{n(X_{max} - X_{min}}$$

Where: $\sim fX$ -the total number of points earned by all students on an item,

n -the number of students,

X_{min} - the smallest item score possible, and

X_{max} - the highest item score possible.

The difficulty index of the test items ranged from .39 to .50 which implied moderately difficult items (Table 27).

Item Discrimination

The item discrimination index was used to measure each test item to distinguish the performance of students. This was done by calculating the difference in the percentage of high achieving students who got an item correct and the percentage of low achieving students who got the item correct. The discrimination index ranges from -1 to +1 where positive numbers above 0.2 show that an item is positively discriminating. A discriminating index (D) values less than zero shows a negatively discriminating item which is regarded as a poor item for the test.

Generally, based on the classical test theory item analysis (Ebel, 1979; Hopkins, 1998), items with an index value less than .20 are regarded as poor items and should, therefore, be discarded or completely revised. Items with an index value of between .20 and .29 are marginal items and need some revision. Items ranging between .30 and .39 are reasonably good items whereas those with an index above .40 are regarded as very good items. Table 26 shows the discriminant level of items for the test in this study. The discrimination index was calculated using the upper (U) 27.0% and the lower (L) 27.0% of the test scores. The possible score for every question ranged from 0 to 3 with a maximum possible total test score of 30 and a minimum of 0. The following formula (Tjoe & de la Torre, 2014) was used to calculate the item discrimination index for every item.

Item discrimination (D) $=P_U - P_L$

Where: P_U is the difficulty indices for the Upper (U) group and
P_L is the difficulty indices for the Lower (L) group.

The item DI ranged from .17 to .44 (see Table 26). Item q8 had the lowest DI of .17 with a moderate level of difficulty. It is, however, important to consider that the DI should be interpreted based on the purpose of the test. According to Mehrens and Lehman (1991), there are various reasons why items can have a low DI which may unnecessarily imply a poor item. For instance, a typical classroom test may have a low DI simply because it is measuring a variety of instructional objectives. Haladyna and Rodriguez (2013) noted that item difficulty should be interpreted based on how the students were prepared for the test or their previous cognitive experiences. In the present study, the items were measuring proportional reasoning and most likely some students hardly understand the concept and therefore may have used a wrong strategy.

Table 26

Summary Statistics for the PRT Items (N=45)

Item	M	SD	ITC (r)	DI (p)	ID (D)
q1	1.24	0.98	.32	.41	.28
q2	1.36	0.53	.60	.45	.22
q3	1.29	0.63	.45	.43	.36
q4	1.27	0.54	.57	.42	.33
q5	1.29	0.51	.65	.43	.25
q6	1.44	0.62	.50	.48	.22
q7	1.20	0.46	.61	.40	.25
q8	1.20	0.46	.46	.40	.17
q9	1.18	0.39	.66	.39	.25
q10	1.51	0.76	.67	.50	.44

Note: M=mean; SD=Standard deviation; ITC=Item-total correlation; DI=Difficulty index; ID=Item discrimination

4.4.4. Conclusion and Future Steps

This study aimed to construct and validate a proportional reasoning test (PRT) instrument. The target domain to be measured was students' proportional reasoning skills on rates, ratios, and proportions in mathematics. The items were organized hierarchically based on the three areas of competency which represented the proportional reasoning skills. Content-related validity evidence was determined by a team of subject matter experts (SMEs) who reached a consensus regarding the items' level of difficulty and the accuracy of the scoring guide.

All items had a higher cognitive demand since they tested on the application of skills learned to real-life experiences and therefore received the same weight. The item analysis revealed that all the items had a moderate difficulty level ranging from .39 to .61. The discrimination index for most items ranged from .22 to .44. Based on experts' review and item-level analysis, PRT is feasible in measuring proportional reasoning skills among the form three (grade 11) students in Kenya.

This instrument was however only tested with a small sample. The future step is to test the instrument with a larger student population of different abilities. This will improve the reliability of the instrument and make it generalizable to a larger population. There is also a need to analyze the responses of students on each item. This will help in formulating distractors that can cover common misconceptions. From the common misconceptions, it will be possible to make closed items that can be easily computerized hence enabling automatic scoring.

4.5. Study 5: Impact of Formative Assessment Instructional Approach on Students' Mathematics Achievement and their Metacognitive Awareness

4.5.1. Introduction

Formative assessment is deemed to be a beneficial approach to instruction (Wiliam, 2011). The ground-breaking investigation on the impact of formative assessment (Black & Wiliam, 1998) has led to more research related to formative assessment. However, various concerns regarding both the concept and efficacy of formative assessment have been raised by researchers (Bennett, 2011; Dann, 2014). Although inconsistencies in defining formative assessment and how it is applied have been notable concerns, increasing research on formative assessment is being conducted. This is reflected in the most recent review on how formative assessment strategies have been employed as well as their impact on students' learning outcomes (Wafubwa, 2020). According to the review, most scholars conceptualize formative assessment in terms of specific strategies. Krijgsman et al. (2019) for instance focused on goal clarification and feedback, which constitute two strategies of formative assessment.

Studies on feedback have revealed the powerful impact that feedback has on achievement (Hattie & Timperly, 2007). After conceptualizing formative assessment as formative feedback, Pinger et al. (2018) found that feedback embedded in instruction can enhance students' performance. Although other studies have also explored the impact of formative feedback (e.g., Cutumisu & Schwartz, 2018; Kyaruzi et al., 2019), only a paucity of experimental studies on the influence of feedback has been carried out, particularly in secondary schools (Van der Kleij et al., 2015).

In most studies, formative assessment has been conceptualized as peer assessment (Hsia et al., 2016; Rotsaert, Panadero, & Schellens, 2018; Tsivitanidou et al., 2018; Vanderhovn et al., 2015). In Black and Wiliam's framework, peer assessment strategy requires learners to be resourceful with each other in instructional processes. Peer assessment of given criteria can either be written or imparted verbally (Boud & Falchikov, 2007). According to Wiliam (2011), peer assessment is more productive when the focus is on improvement as opposed to evaluation. Empirical studies that have assessed the influence of peer assessment in different learning conditions have demonstrated that it enhances learning outcomes. Rotsaert, Panadero, & Schellens (2018) in their study involving peer assessment and feedback, showed that reciprocal peer assessment ensured immediate feedback. Tsivitanidoue et al. (2018) also utilized peer assessment as a learning tool.

Although there are a limited number of studies conceptualizing formative assessment as self-assessment, especially in the secondary schools' context, peer assessment is closely associated with self-assessment. Nikou and Economides (2016) studied the effect of self-assessment on students' motivation and achievement. In other studies, self-assessment has been conceptualized as part of self-regulation (e.g., Panadero et al., 2016; Panadero et al., 2017). These studies seem to suggest that the learning benefits of self-assessment can only be realized if students' are trained in self-regulation skills because it is innately difficult to acquire accurate self-knowledge (Dunning et al., 2004; McDonald & Boud, 2003).

Although most studies have examined teachers and students' perceptions, scholars have conceptualized formative assessment as a combination of five strategies (e.g., Burner, 2016; Dobish & Meyer, 2017; Johnson et al., 2019; Kippers et al., 2018; Ozan & Kincal, 2018; Saito & Inoi, 2017; Widiastuti et al., 2020). Few experimental studies related to the impact of formative assessment have demonstrated its positive influence on students' learning outcomes (e,g., Andersson & Palm, 2017; Vogelzan & Admiraal, 2017). A study by Pinger et al. (2018) however revealed no improvement in the quality of instruction. Formative assessment has also been conceptualized in terms of tests and/or questions given to students at regular intervals to assess their learning (Heritage & Heritage, 2013). In the latter study, formative assessment was visualized as continuous assessment tests.

Self-assessment as a formative assessment strategy has been posited to enhance students' metacognitive awareness (Andrade, 2010; Taras, 2010) because students who participate in self-assessment monitor their thinking processes and can assess their learning process in general. Metacognition has been commonly described as the knowledge related to one's thought processes as well as the regulation of cognitive activities (Flavell, 1979). Thus, metacognition comprises two facets: metacognitive knowledge and metacognitive regulation. According to Schraw and Moshman (1995), the acquisition of metacognitive awareness can be promoted by instructional strategies that activate students' self-knowledge and regulatory skills. Studies have shown that training students' metacognition influenced learning outcomes positively (Csíkos & Steklács, 2010; Naful et al., 2021; Naseri et al., 2017; Roll et al., 2011). One may conclude that self-assessment emphasizes high levels of metacognition, which influence learning styles and consequently learning achievement.

The Present Study

Formative assessment in the present study was employed as an instructional approach to enhance students' mathematics achievement on proportional reasoning skills and improve their metacognitive awareness. Black and Wiliam's (2009) framework, which envisions formative assessment as a classroom practice comprising five strategies, was employed as the theoretical framework of the study. The five strategies, which are subsequently described, are supposed to be utilized by the teacher, learner, and peer to identify and address learning gaps.

The first strategy may be described as sharing learning goals and criteria for attaining these goals. This strategy requires learners to know in which direction they are heading.

Furthermore, the teacher must involve learners in understanding what success looks like. The second strategy is effective classroom discussions. Discussions that primarily involve questioning are meant to reveal students' comprehension. Through discussion and questioning, the teacher can collect evidence of students' learning. The third strategy involves the provision of the feedback given by the teacher but also involved learners and their peers. Through feedback, the learner can discover whether the learning goals are being attained. The teacher can also adjust instructional approaches to attend to students' needs.

Peer assessment in which students act as each other's instructional resources is the fourth strategy. Peer assessment is beneficial because learners work in collaboration toward a common goal. The fifth strategy comprises self-assessment which involves activating students to own their learning. Self-assessment has to be incorporated in other formative assessment strategies (Wiliam, 2011). Black and Wiliam's (2009) framework postulate that these five strategies should be guided by three learning processes.

The first process in Black and Wiliam's framework involves identifying the direction in which the learners are heading. The second process encompasses establishing their current position while the third process comprises knowing how the learners will reach their final destination. In accordance with this framework, the present study conceptualized formative assessment as an instructional approach that encompasses five strategies, three processes, and three agents, namely: teacher, student, and peer (Table 27).

Table 27

Features of Formative Assessment

Agent	Where the learner is going	Where the learner is right now	How to get there
Teacher	1. Intentions for learning and criteria for success	2. Classroom discussion	3. Feedback
Peer	Understanding and sharing learning intentions and criteria for success	4. Peer assessment	
Learner	Understanding learning intentions and criteria for success	5. Self-assessment	

Note: Adopted from Wiliam and Thompson (2008)

Although research on formative assessment and its impact on students' achievement has extensively been done, a few studies have focused on formative assessment as integration of five strategies. As previously observed, most studies have focused on specific strategies and in particular, on feedback and peer assessment. Furthermore, research has rarely examined specific skills in mathematics. However, this study is novel in that proportional reasoning skills in mathematics and students' metacognition were explored. Proportional reasoning is a crucial life skill utilized in day-to-day decision-making (Howe & Bryant, 2011).

Proportional reasoning in mathematics is among the areas in which students perform dismally in the Kenyan mathematics curriculum. It was considered that if students' proportional reasoning improved, their overall performance in mathematics would also improve and their metacognitive awareness would increase as well. The present study thus contributes knowledge to the verifiable impact of formative assessment conceptualized as an instructional approach on students' mathematics achievement and their metacognitive awareness.

Research Questions

In this study, the following four research questions were answered:

1. Is there a significant difference in the students' performance on mathematics posttest between the intervention and control groups?
2. Does gender influence students' performance on mathematics posttest? Is there a significant interaction between gender and the type of teaching approach, and mathematics posttest scores?
3. Is there a significant difference in the students' ratings on their levels of metacognitive awareness between the intervention and the control groups after the treatment?
4. Does gender influence students' metacognitive awareness after the treatment? Is there a significant interaction between gender and the type of teaching approach, and metacognitive awareness posttest scores?

4.5.2. Method

Participants

The participants included 164 grade 11 students (84 boys) from four low achieving rural secondary schools in the western part of Kenya. While two of the schools were randomly assigned to the experimental group, the remaining two formed the control group. There were 84 participants in the intervention group and 80 in the control group. Whereas the four schools had two classes each, only one class was selected randomly to take part in the study. Table 28 shows the demographics of the sample. Four teachers, one from each school voluntarily participated in this research. The participants had similar socio-economic and cultural backgrounds. In addition, the four schools were classified as sub-county schools according to Kenya’s classification of schools. Therefore, one may deduce that the participants had similar characteristics about their socio-economic background and academic performance.

Table 28

Sample Demographics

Group	School/Class	Gender	*n*
Control	A	Boys	40
	B	Girls	40
Experimental	C	Boys	44
	D	Girls	40

Design

This study adopted a quasi-experimental pretest-posttest non-equivalent group design where schools were randomly assigned to either intervention group or control group. The intervention group consisted of 84 students whereas the control group comprised 80 students as reflected in Table 28. While the intervention schools were subjected to a formative assessment instructional approach, the control schools were taught by employing a conventional approach. The teachers taught the same content and matching tasks were given to the schools in both groups. Furthermore, both groups were given identical pretests and posttests on rates, ratios, and proportions. The participants also completed the Junior Metacognitive Awareness Inventory (Jr. MAI), which measured their levels of metacognitive awareness.

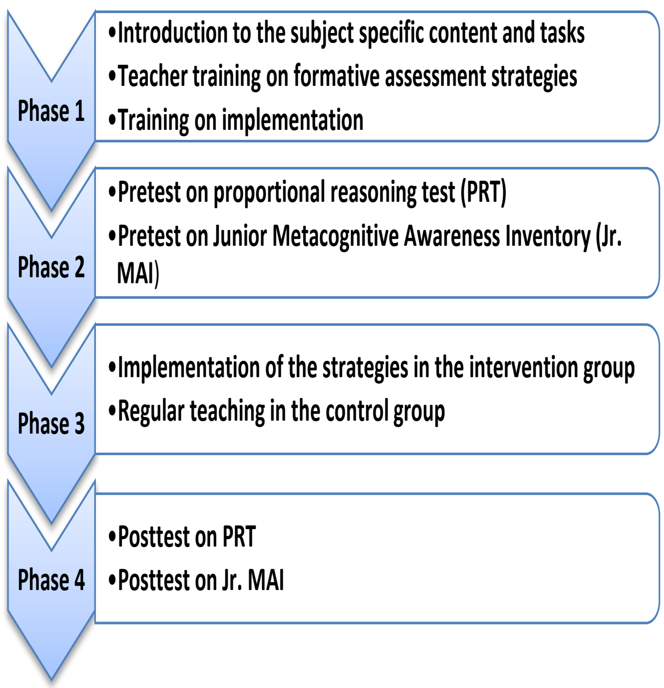
Procedure

This study comprised four phases (Figure 6). During the first phase, the four participating teachers were exposed to the proportional reasoning topic and the subtopics that they were expected to teach. Training sessions on formative assessment strategies and the implementation, therefore, were conducted with the teachers in the intervention group. On the other hand, teachers in the control group never received any training on formative assessment. In the second phase, students in both groups completed the pretest on the Proportional Reasoning Test (PRT) and their levels of metacognitive awareness were measured using the Jr. MAI. The teachers were given instructions on how to administer the test and questionnaire.

Phase three was an implementation phase where teaching and learning took place in both conditions. Fourteen 40-minute lessons were conducted to complete the intervention in four weeks. Whilst the teachers in the experimental group utilized the materials and tasks given during the training session, those in the control group employed a conventional approach in their teaching. The five strategies of formative assessment were implemented in the 14 lessons in the intervention group. In the final phase, participating students did the PRT and completed the Jr. MAI.

Figure 6

Intervention Phases



Teacher Training and Experimental Conditions

The four participating teachers had already been informed about the project during the survey study a year before and had decided to participate willingly. The training, which comprised workshops, was divided into two parts. During the first part, which lasted for a day, proportional reasoning and specific areas that had to be covered under each topic were introduced. The teachers also discussed how they could handle the five areas of proportional reasoning in 14 lessons over four weeks. They agreed on a flexible four-week scheme that could be employed to guide them teach the topic. The time allocation was in line with the recommended time stipulated in the Kenyan mathematics curriculum (KICD, 2017). The second part of the training was conducted over two days and involved only the teachers in the intervention group who received training on how to implement formative assessment strategies.

The teacher training in the intervention group was centered on five strategies of formative assessment as well as how the strategies could be implemented. During the training, implementation challenges were addressed, and the teachers agreed on a three-step flexible guideline to implement the strategies (Appendix A1). During the training session, teachers were also provided with supplementary materials on formative assessment strategies. The training program and specific areas that the training focused on are displayed in Appendix A2.

In the experimental classes, the teacher first stated the rationale for each particular lesson and briefly engaged students in designing their success through questioning and discussion. Second, the teacher gave tasks to students individually which was followed by a group discussion. The discussion enabled students to obtain feedback from each other. Finally, the teacher gave feedback to individual groups to generate and share ideas in groups. These ideas were subsequently discussed with the whole class. The groups, which were also referred to as study groups, comprised four to five students with mixed abilities. The whole process involved the students and teachers jointly identifying and communicating the learning and performance goals. The participants' current levels of understanding were assessed and strategies and skills to reach goals were generated.

Measures

1. Proportional Reasoning Test (PRT)

Previously, the researchers and two mathematics teachers who had considerable experience in teaching high school mathematics and were also national examiners developed a written mathematics test on proportional reasoning skills. The 10 items of PRT cover all the content areas on rates, ratios, and proportions, a topic covered in the Kenya secondary schools mathematics curriculum. The test was constructed from word problems relating to real-life situations and examines five aspects of proportional reasoning: missing values, associated sets, mixtures and proportions, comparison problems, and stretcher (Appendix E). Two items assess each aspect and there is an equal distribution of marks across the 10 items.

The content validity was determined by a team of mathematics subject experts. Furthermore, item-level analysis was performed by examining the difficulty and discrimination indices. The PRT test was piloted on a sample of 45 students and analysis of items showed the difficulty level ranged between .39 and .61, thus implying a moderately difficult test. All the items have a higher cognitive demand and therefore, require students to employ self-regulatory skills to solve them. Formative assessment which involves self-regulation strategies was hence deemed to be the best approach to enhance students' proportional reasoning. For the present sample, Cronbach's alpha coefficient was .72, suggesting acceptable reliability. More details on how the PRT was developed and validated can be found in Wafubwa et al. (2020).

2. Junior Metacognitive Awareness Inventory (Jr. MAI)

This inventory was constructed by Sperling et al. (2002) to assess young adults' metacognition as well as a tool to gauge classroom interventions. In this study, the adapted Jr. MAI (Appendix F) was employed to measure students' level of metacognitive awareness before and after a formative assessment intervention. The inventory comprised 18 items, which measured the knowledge and regulation dimensions of metacognition. Examples of items in the knowledge dimension were: "*I know when I understand something*" and "*I can make myself learn when I need to*". On the regulation dimension, examples of items were: "*I think about what I need to learn before I start working*" and "*I pay attention to important information*" (Sperling et al., 2002, p. 76). Items were evaluated on a 5 point scale ranging from 1 (not at all) to 5 (always).

The pretest and posttest questionnaires were partially completed by five and nine participants, respectively. Thus, the data of the 14 incomplete questionnaires were excluded from the analysis. The two-factor solution of principal component factor analysis was performed to ascertain validity. The Kaiser-Meyer-Olkin (KMO) was .652 and Bartlett's test value was significant (χ^2 (351) = 940.316, $p < .001$). While 10 factors loaded on the knowledge dimension, eight factors loaded on the regulation dimension. Although some of the factors intended for the knowledge dimension following the original scale loaded on the regulation dimension and vice versa, our concern was to measure students' metacognition in general so we considered the entire scale. Cronbach's alpha coefficient revealed that the reliability for the whole scale was .78, which is considered to be acceptable internal consistency.

Study Variables

There were two active independent variables in this study. Formative assessment was employed as a between-groups independent variable (IV) with two levels: formative assessment and no formative assessment. The second active IV was, change over time, which was a within-subjects IV with two levels: pretest and posttest. Gender was utilized as an attribute IV with two levels: boy and girl. Pretest scores on both achievement and metacognitive awareness were employed as a covariate in the Analysis of covariance (ANCOVA). Two dependent variables were measured in this study: students' mathematics achievement, which was measured using PRT, and metacognition, which was assessed by employing the Jr. MAI.

The Implementation Process

The pretest was administered to both the experimental and control classes a week after the teacher training workshops. The teachers in the experimental group implemented formative assessment through a process of reciprocal classroom interaction that involved the teachers, students, and learning resources. The implementation involved a three-step guide that specified strategies to be used during each step (Appendix A1). In the first step, strategies that described learning intentions and success criteria were employed. In the second step, self-assessment, peer assessment, and discussion strategies were utilized. The third step involved peer assessment, self-assessment, and feedback strategies. The guide included all five strategies blended in the lessons.

In the course of the implementation process, the lead researcher conducted a follow-up twice a week to ensure the implementation was being conducted as planned. However, we were unable to observe the teaching because of COVID 19 pandemic-related restrictions. The posttest was administered one month after the pretest. During the administration of both the pretest and posttest, the teachers were given a set of instructions to ensure test fidelity. The instructions included the time required to complete the test, authorized instruments, and spacing of students when doing the test. The scoring of the tests was conducted externally by the lead researcher and other two experienced teachers from a different region to that of the study. Scoring was blinded in that the examiners were unaware of which group the students belonged to. Once the scoring was complete, the teachers were given back the students' scripts to provide feedback that inform instruction.

Data Analysis.

In the study, students were the unit of analysis. Both the pre and posttest data obtained using the two research instruments were examined for parametric tests assumptions. While the Shapiro-Wilk test was employed to examine whether the data were normally distributed, Levene's test examined homogeneity of variance. Analysis of covariance (ANCOVA) was utilized in testing for significant variations in the posttest scores on achievement and metacognition, thus assessing the impact of formative assessment. The effect sizes were determined by partial Eta Squared values.

4.5.3. Findings

Mathematics Achievement

The impact of formative assessment was assessed through ANCOVA. The group variable was formative assessment. Whereas the independent variable was the students' pretest mean score, the dependent variable was the students' posttest mean score. The results of the Shapiro-Wilk test for both groups were insignificant ($p > .05$), thus implying a normal distribution for the covariate and dependent variable. Furthermore, an examination of boxplots didn't show any extreme outliers. Levene's test found non-significant results ($p = .200$) for equality of error variance. The homogeneity of regression slopes assumption was also examined and the effect was found to be non-significant ($p = .335$) hence, the assumption was met.

After using pretest scores as the covariate, the ANCOVA results showed a significant variation in the posttest scores between the students in the intervention group and the control group, $F(1, 159) = 6.227$, $p = .014$, $\eta 2 = .38$. The results imply that students who were taught by employing formative assessment strategies improved in their achievement on PRT in comparison to those who were taught by using conventional approaches. The adjusted means, standard errors (SE), means (M), and standard deviations (SD) for the posttest groups are presented in Table 29.

Table 29

Posttest Means Scores

Group	*n*	Adjusted *M*	*SE*	*M*	*SD*
Experimental	84	1.69	0.056	1.70	0.58
Control	80	1.49	0.057	1.48	0.52

A two-way ANCOVA was conducted to show how gender influenced the posttest scores. The assumptions of ANCOVA were all met. Both the Shapiro-Wilk test and Levene's test were non-significant. The results showed that gender had no significant influence on the posttest scores, $F(1, 159) = 0.322$, $p = .571$, $\eta 2 = .002$. Furthermore, no significant interaction between gender and the type of teaching approach, and the mathematics posttest scores was exhibited, $F(1, 159) = 0.347$, $p = .557$, $\eta 2 = .002$. The adjusted means, SE, means (M) and SD for the posttest mean scores based on gender are displayed in Table 30. The pretest mean scores were used as the covariate.

Table 30

Posttest Mean Scores Based on Gender

Gender	Group	*n*	Adjusted *M*	*SE*	*M*	*SD*
Boys	Control	40	1.56	0.072	1.53	0.47
	Experimental	44	1.68	0.066	1.70	0.57
Girls	Control	40	1.45	0.073	1.42	0.56
	Experimental	40	1.66	0.064	1.71	0.60

Metacognition

The influence of formative assessment on students' metacognitive awareness was determined by conducting ANCOVA. After checking all the assumptions of ANCOVA, the results of the analysis indicated a remarkable difference in the metacognition posttest scores between the formative assessment and control groups, $F(1, 145)= 128.260$, $p < .001$, $\eta 2 = .469$. The ANCOVA results revealed that metacognitive awareness rating was higher among the students who received instruction using formative assessment than those who were taught using conventional methods. In Table 31, the adjusted means, SE, means (M), and SD for the posttest scores while using the pretest mean scores as the covariate are shown.

Table 31

Posttest Means Scores

Group	*n*	Adjusted *M*	*SE*	*M*	*SD*
Intervention	77	4.30	0.26	4.26	0.29
Control	73	3.88	0.26	3.91	0.49

A two-way ANCOVA was used to estimate the effect of gender on the posttest scores of metacognition. The results showed non-significant effect of gender on the posttest scores, F(1, 145) = 1.142, p = .287, η2 = .008. In addition, no significant interaction between gender and the type of teaching approaches on the posttest scores was exhibited, F(1, 145) = .088, p = .767. η2 = .001. These results suggest that both boys and girls responded in a similar way to the teaching method. The adjusted means, SE, means (M), and SD for the posttest mean scores based on gender with pretest mean scores as the covariate are displayed in Table 32.

Table 32

Posttest Mean Scores Based on Gender

Gender	Group	*n*	Adjusted *M*	*SE*	*M*	*SD*
Boys	Control	35	3.89	0.038	3.98	0.46
	Experimental	41	4.32	0.035	4.27	0.28
Girls	Control	38	3.86	0.037	3.85	0.51
	Experimental	36	4.27	0.038	4.25	0.31

4.5.4. Discussion

In this research, the influence of formative assessment on students' mathematics achievement and their metacognitive awareness was explored. Four research questions guided the study. Research question one sought to establish if there was a significant difference in the posttest scores between intervention and control groups after the treatment. Results exhibited a significant difference between the posttest scores of the two groups with a medium effect size (η2 = .38) after controlling for the pretest scores. This implies that students who were exposed to formative assessment strategies performed better than those who were taught conventionally.

Other studies have also demonstrated that the utilization of formative assessment strategies improves students' performance (Ozan & Kıncal, 2018; Pinger et al., 2018; Vogelzan & Admiraal, 2017). However, most of these studies focused only on one or two strategies of formative assessment. In the current study, formative assessment was conceptualized as an instructional approach encompassing five strategies that are embedded in instruction. Anderson and Palm (2017) who also conceptualized formative assessment as a combination of strategies found a significant effect on students' achievement (Cohen's d = .66). The effect size they found was larger than that of this study. While we employed students as the unit of analysis, Anderson and Palm used teachers as the unit of analysis. This could be possibly the reason for the difference in the effect sizes.

However, some considerations should be taken when interpreting the effect sizes related to formative assessment. The first consideration is on how formative assessment is conceptualized and the second is on how it is implemented (Bennett, 2011). Research has indicated that the efficacy of formative assessment has been hampered by poor implementation processes (Randel et al., 2016). Therefore, teachers' preparation and support are crucial for the execution of formative assessment. Although the duration for teacher training for this study was short, professional development can hardly be the sole determining factor for the successful implementation of an intervention (Johnson et al., 2019; Randel et al., 2016; Yin & Buck, 2019). Studies have further shown that apart from inadequate professional development, teachers fail to implement formative assessment because of their heavy workload and lack of motivation (Crichton & McDaid, 2016; Jacoby et al., 2014). Because the teachers in this study received external support and participated willingly in the study, we can deduce that the improvement in the experimental group was due to formative assessment strategies.

The second research question was concerned with whether gender influenced mathematics achievement scores after the intervention and whether a significant interaction between gender and the type of teaching approach and the posttest scores were exhibited. The results revealed that after controlling for pretest scores, gender had a non-significant influence on the posttest scores. There was also no evident interplay between gender and the type of teaching approaches, and the posttest scores. The results suggest that formative assessment had a similar influence on the learning of both male and female students. Therefore, one may infer that the improvement in achievement in the intervention group was associated with the formative assessment instructional approach and gender had no effect on this approach. Recent studies on gender and achievement have also shown that gender has a non-significant influence on mathematics achievement (Lindberg et al., 2010; Louis & Mistele, 2012; Scheiber et al., 2015).

The third research question focused on whether the teaching approach had an influence on the students' posttest scores on metacognitive awareness. The results revealed that students who received instruction using formative assessment had a higher metacognitive rating than those who were taught using conventional methods. The items on Jr. MAI assessed students' metacognitive awareness, which comprised knowledge and skills dimensions. Knowledge of cognition involves awareness of and knowledge about one's cognition (Harris et al., 2010). On the other hand, metacognitive skills involve planning, monitoring, and evaluating learning processes (Veenman & Beishuizen, 2004).

Although our literature search hardly yielded studies related to the influence of formative assessment on students' metacognition, scholars have demonstrated that formative assessment and metacognitive skills are related (Baas et al., 2014; Wafubwa & Csíkos, 2021). Empirical studies have also shown the benefits of metacognition on students' achievement, especially when students are trained to be metacognitive (Csíkos & Steklács, 2010; Roll et al., 2011; Naseri et al., 2017; Dafik & Rohim, 2019; Naful et al., 2021). Being metacognitive implies that one is conscious of his or her thought processes and can regulate cognition through processes such as monitoring, planning, and evaluating.

Formative assessment strategies, in particular, self-assessment and feedback, also involve the self-regulation processes of monitoring, planning, and evaluating. Therefore, it is convincible that formative assessment strategies improved students' metacognition, which is reflected in the higher ratings on the posttest scores. Although we acknowledge that it is rare for students to have a true knowledge of themselves, based on their performance on the PRT, one may conclude that the ratings on metacognitive awareness inventory could be a reflection of what students feel about their metacognition. However, proportional reasoning is only one skill in mathematics. Therefore, the results should be elucidated only within the proportional reasoning skills context.

The fourth research question was concerned with the influence of gender on metacognitive awareness posttest scores and whether there was an interaction between gender and the type of teaching approach and metacognitive awareness posttest scores. The results revealed that gender had an insignificant influence on metacognitive awareness posttest scores. Furthermore, no significant interaction between gender and teaching approach, and the posttest scores were found. Although research on gender and students' metacognition has been less extensively studied, some studies have suggested that students' metacognition is gender independent (Al Shabibi & Alkharusi, 2018; Siswati & Corebima, 2017).

4.5.5. Conclusion

The findings of the present study indicate that formative assessment strategies based on students' needs can lead to improved learning outcomes when employed carefully. The formative assessment approach used in this study was planned following the challenges that students encountered with proportional reasoning in mathematics. Students had challenges in solving mathematics questions that needed the use of metacognitive strategies such as planning, monitoring, and evaluating. However, we may have reservations in assuming that the same results can be realized with other mathematics topics. Rather, it is dependent on how teachers conceptualize formative assessment.

When formative assessment is conceptualized as continuous assessment tests, only some aspects of formative assessment strategies may be employed, which may result in an insignificant impact. On the contrary, if teachers view formative assessment as a classroom practice that can show evidence of student learning and enable them to make decisions on how to improve instruction, all of the five strategies will be utilized for better learning outcomes. This study has revealed that formative assessment strategies can improve the performance of low achieving students and also improve their metacognitive awareness. The results of this study can benefit teachers and curriculum developers in designing formative assessment intervention programs that can boost students' achievement in mathematics and improve their metacognitive awareness. It is recommended that future studies explore other topics in mathematics using the same formative assessment approaches.

Chapter 5: Conclusions, Implications, Recommendations, and Summary

5.1. Conclusions and Implications

This dissertation entails different studies related to formative assessment and the teaching and learning of mathematics in secondary schools in Kenya. The motivation behind this project was based on the current changes in the Kenyan education curriculum. Kenya is currently transitioning from a traditional education system to a competency-based system which requires new teaching and learning approaches. Formative assessment is one of the approaches that can be used to enhance students' competencies. The review of literature has however shown that the use of formative assessment in the teaching and learning process is rare in Africa and more specifically in Kenyan secondary schools.

The current research project hence sought to fill the gap in the literature by conducting a series of studies using different research designs. The project was thus carried out in two phases and involved five studies. The first phase involved exploring the perceptions of teachers regarding formative assessment and their levels of metacognition. The second phase involved an intervention study on how formative assessment can be used to improve students' mathematics achievement and metacognition. In this experimental phase, a test for measuring achievement in mathematics was developed, validated, and then used in the intervention study. Based on the individual studies, the following conclusions can be drawn.

5.1.1. Study 1

The first study involved piloting the adapted TAFL-Q to assess whether it was suitable in measuring Kenyan mathematics teachers' perceptions of formative assessment. In this study, two research questions and two hypotheses were answered. RQ1: "What is the evidence of validity for the two-factor model of the TAFL-Q in the Kenyan context?" As stated in H1, it was hypothesized that the two-factor model of TAFL-Q will result in acceptable fit indices (CMIN/DF of between 1.0 and 3.0; RMSEA of $\leq .08$; SRMR of $\leq .08$; TLI of $\geq .900$; and CFI of $\geq .900$). Confirmatory factor analysis was used to confirm the hypothesized two-factor model of the TAFL-Q. Contrary to the expectations (H1), the analysis resulted to a poor fit to the Kenyan sample (CMIN/DF = 2.643, RMSEA = .096, SRMR = .085, TLI = .653, CFI= .680).

The factor structure was then explored using the principal components analysis with Varimax Kaiser Normalization rotation. Nine items were eliminated and a final analysis of the remaining 19 items resulted in six factors with a total explained variance of 66%. The emergent factors were further assessed through convergent and discriminant validity and the results were valid. The six-factor model fitted well to the questionnaire data of the Kenyan sample (CMIN/DF = 1.92, RMSEA = .070, SRMR = .060, TLI = .880, CFI= .850).

The second research question (RQ2) assessed teachers' overall perceptions of formative assessment. It was expected that the teachers' ratings of their perceptions would correspond to "Agree" implying a high perception (H2) as measured on a 5-point Likert scale (Strongly Disagree=0; Disagree=1; Undecided=2; Agree=3; Strongly Agree=4). Descriptive statistics analysis however showed that teachers' perceptions ranged from low to moderate hence not confirming H2. Though most studies show teachers having a high perception of formative assessment, low to moderate perceptions in this study may be an indication that Kenyan mathematics teachers are less familiar with formative assessment strategies.

Study one revealed that the adapted two-factor TAFL-Q was unfit for measuring Kenyan secondary school teachers' perceptions of formative assessment. The improved six-factor TAFL-Q model was however used to gauge teachers' perceptions of formative assessment. The perceptions were expected to be high (H2) but the analysis showed low to moderate perceptions. Some of the items that were dropped from the original two-factor model of the TAFL-Q were not replaced in the new version of the six-factor TAFL-Q. This may have affected the internal consistency of the new version. The results imply the need to further develop the TAFL-Q as an instrument for measuring mathematics teachers' perceptions of formative assessment in the Kenyan context.

5.1.2. Study 2

The second study investigated the relationship between mathematics teachers' perceived use of formative assessment strategies and their levels of metacognition. The main aim of this study was to show how the use of formative assessment strategies affects the teachers' metacognitive regulation in terms of planning, monitoring, and evaluating skills. This study answered four research questions with the corresponding hypotheses. The first question (RQ1) was concerned with establishing the evidence of validity and reliability for the six-factor model of the TAFL-Q. It was expected (H3) that the scales of the six-factor model of the TAFL-Q will have Cronbach's alphas (α) of $\geq .60$ and that the fit indices of the model measurement will be within the acceptable range (CMIN/DF of between 1.0 and 3.0; RMSEA of $\leq .080$; SRMR of

$\leq .080$; TLI of $\geq .900$; and CFI of $\geq .900$). As per the expectations, CFA of the six-factor model of the TAFL-Q indicated that all the values of the fit indices were within the recommended range hence confirming H3.

The second research question (RQ2) assessed the evidence for validity and reliability of the MAIT scale. It was hypothesized that the scales of the MAIT will have Cronbach's alphas (α) of $\geq .60$ and the fit indices of the model measurement will be within the acceptable range: CMIN/DF of between 1.0 and 3.0; RMSEA of $\leq .08$; SRMR of $\leq .08$; TLI of $\geq .900$; and CFI of $\geq .900$ (H4). The structure of the MAIT was first confirmed on a sample of 180 teachers and then validated in the second study on a sample of 213 teachers. The results of Cronbach's alphas for all the three scales of the MAIT were all above .70 indicating high reliability. Confirmatory factor analysis showed the fit indices with values that were within the recommended range. Hypothesis two (H4) was thus confirmed.

In the third research question (RQ3): "How does the hypothesized structural relationship among the predictor and the observed variables fit the data?" it was expected that the hypothesized TAFL-Q and MAIT structural model will have acceptable fit indices: CMIN/DF of between 1.0 and 3.0; RMSEA of $\leq .080$; SRMR of $\leq .080$; TLI of $\geq .900$; and CFI of $\geq .900$ (H5). As expected, the fit indices for the measurement model were all within the recommended range (CMIN/DF = 1.218, RMSEA = .032, SRMR = .012, TLI = .989, CFI= .998), hence confirming H5.

The fourth research question (RQ4) assessed the relationship between teachers' perceptions of formative assessment and their metacognitive skills. It was hypothesized that there will be a positive relationship between the teachers' perceptions of formative assessment and their metacognitive skills (H6). The relationship between the TAFL-Q and MAIT was modeled using structural equation modeling and path analysis in AMOS graphics software. The results of the model showed that teachers evaluating skills are positively predicted by learning intentions, success criteria, and peer assessment. Monitoring skills are positively predicted by classroom discussion and peer assessment. Planning skills, on the other hand, are positively predicted by feedback, peer assessment, and success criteria. As per the hypothesis (H6), there was a positive relationship between the teachers' perceptions of formative assessment and metacognition.

Understanding the relationship between formative assessment and metacognition can help teachers in modeling learning strategies and skills in the learners. More research is needed to test the relationship exhibited in the present study with other samples, especially in different cultural contexts. The findings of this study contribute to future theory development and designing effective intervention programs for classroom instruction.

5.1.3. Study 3

The third study investigated the in-service mathematics teachers' conception and perceptions of metacognition. Three research questions and corresponding hypotheses were answered. In the first research question (RQ1): "To what extent do secondary school mathematics teachers in Kenya perceive their use of metacognitive knowledge and skills in teaching mathematics?", it was hypothesized that the mathematics teachers' perceptions will correspond to "Often" on a 5-point Likert scale (Not at All=0; Rarely=1; Sometimes=2; Often=3; Always=4), implying a high perception of their use of metacognitive knowledge and skills (H7). Results from descriptive statistics indicated a mean of 3.31 for items under the dimension of metacognitive knowledge and a mean of 2.91 for items under the metacognitive skills dimension. The mean for the two dimensions corresponds to "Often" on the Likert scale implying that teachers had high perceptions of both their metacognitive knowledge and skills. The H7 was thus confirmed although the teachers' ratings of their perception of metacognitive knowledge were slightly higher than metacognitive skills.

Research question two (RQ2) assessed the effect of gender, teaching experience, and level of education on the metacognitive awareness of secondary school mathematics teachers in Kenya. It was hypothesized that the three background factors of gender, teaching experience, and level of education will have non-significant effects on mathematics teachers' metacognitive awareness (H8). The independent samples T-test and one-way ANOVA were used to examine the effect of the background factors. The t-test results exhibited non-significant differences in the means of male and female teachers for both metacognitive knowledge and metacognitive skills. The ANOVA results similarly showed non-significant effects of teachers' academic qualifications and years of teaching experience on their metacognitive knowledge and skills. The T-test and one-way ANOVA results thus confirmed H8.

Research question three (RQ3), concerned the conception of metacognition that the Kenyan secondary school mathematics teachers reported. It was expected that mathematics teachers will conceptualize metacognition in terms of knowledge about cognition and regulation of cognition (H9). The interview results were thematically analyzed based on the two dimensions of metacognition (knowledge and skills). Regarding metacognitive knowledge, the analysis revealed that teachers generally made use of their metacognitive knowledge. Teachers were able to express their declarative, procedural and conditional knowledge, which are the three facets of metacognitive knowledge. Analysis of teachers' metacognitive skills revealed that teachers generally had an understanding of the skills involved in the regulation of cognition but rarely put them to use. These results suggest that teachers should be encouraged to apply their metacognitive skills to classroom teaching.

Both quantitative and qualitative analyses reveal that Kenyan secondary school mathematics teachers perceive themselves as metacognitive. The teachers however hardly implement the knowledge and skills in classrooms. This could be partly due to the challenges like large class sizes and lack of motivation that teachers reported. There is a need for the Kenyan Ministry of Education to address these challenges so that teachers can comfortably utilize their metacognitive skills and model the same to the learners.

5.1.4. Study 4

Study four involved the construction and validation of a proportional reasoning test (PRT) to measure students' proportional reasoning skills on rates, ratios, and proportions in mathematics. In this study, two questions and two hypotheses were answered. In RQ1, the evidence of content-related validity of the constructed PRT was assessed. Based on the judgment by the subject matter experts, it was hypothesized that the PRT will have content-related validity (H10). Research question two (RQ2), assessed the evidence of reliability and discriminant validity of the PRT. It was expected that the PRT will have Cronbach's alphas (α) of $\geq .60$, Difficulty Index (DI) range between 0.3 and 0.6; and Item Discrimination (ID) range between .20 and .50. This will imply a reliable and valid test (H11) which is of a moderate level of difficulty and can discriminate the students' performance.

The test items were carefully designed to reflect the aspects of proportional reasoning identified in the existing literature and aligned them to the instructional objectives as stated in the Kenya secondary mathematics curriculum. The test underwent different developmental processes to establish content-related validity before it was piloted. Content-related validity evidence was determined by a team of subject matter experts who reached a consensus regarding the items' level of difficulty and the accuracy of the scoring guide. The results showed an acceptable internal consistency level (Cronbach's α= .83). The item analysis revealed that all the items had a moderate difficulty level ranging from .39 to .50. The discrimination index for most items ranged from .22 to .44. The findings suggest that the PRT is a valid and reliable instrument that can be used to measure the domain-specific skills on proportional reasoning. Both the two hypotheses (H10 and H11) for this study were therefore confirmed.

5.1.5. Study 5

Study five examined the impact of formative assessment on secondary school students' achievement and metacognition in mathematics. A quasi-experimental pretest-posttest non-equivalent group design was employed and formative assessment was conceptualized as an instructional approach comprising of five strategies. This study answered four research questions and six hypotheses. Research question one (RQ1) was on whether there was a significant difference in the students' performance on mathematics posttest between the control and intervention groups. It was hypothesized that students who were exposed to formative assessment in the intervention group will have a higher mean score on the mathematics posttest as compared to the students in the control group (H12). After controlling for pretest scores, the results revealed that students who were taught using formative assessment in the intervention group outperformed the students taught using the traditional approach in the control group hence confirming H1.

Research question two (RQ2) assessed gender influence on students' mathematics posttest and the interaction effects between gender and the type of teaching approach, and mathematics posttest. It was expected that gender will have a non-significant effect on students' mathematics posttest (H13) and that there would be no interaction effects between gender and the type of teaching approach, and mathematics posttest (H14). Analysis based on gender failed to show any remarkable differences in mathematics posttest between the experimental and the control groups. There was also no significant interaction effect between gender and the type of teaching approach, and students' mathematics posttest. Both H13 and H14 were hence confirmed.

Research question three (RQ3) assessed whether there was a significant difference in the students' ratings on their levels of metacognition between the intervention and control groups after the treatment. It was hypothesized that there will be a significant difference in the students' posttest ratings on metacognition in favor of the intervention group (H15). After controlling for pretest, the results revealed that students who were taught using formative assessment in the intervention group had higher ratings of their metacognition than students in the control groups hence confirming H15.

Research question four (RQ4) assessed how gender influenced students' metacognitive awareness after the treatment and whether there was a significant interaction between gender and the type of teaching approach, and metacognitive awareness posttest. It was hypothesized that gender will have a non-significant influence on students' metacognition posttest (H16) and that there will be no significant interaction effect between gender and the teaching approach, and metacognition posttest (H17). Analysis based on gender revealed non-significant influence and there were also no interaction effects hence confirming H16 and H17.

5.2. Recommendations

Since the studies comprising this project mainly involved the in-service mathematics teachers, the general recommendations based on the individual studies are first given and then followed by practical recommendations for the in-service and pre-service teachers in Kenya.

5.2.1. General Recommendations

The following are the recommendations based on the results of the individual studies:

1. Based on the results of the first study, it is recommended that the six factors TAFL-Q that emerged on the Kenyan sample be further refined and validated in different cultural contexts. Due to the limited time for this project, the tool was used in its current state before refining it further.
2. The second study was among the very few studies that have empirically assessed the relationship between formative assessment and metacognition. It is recommended that more studies be carried out to come up with conclusive findings. In the current case, the relationship was modeled using a sample of mathematics teachers in Kenya. It will be necessary to carry out a similar study using a sample from a different cultural context.
3. In study three, the sample for the interview was small. It is recommended that future studies utilize a larger sample for more valid results. Furthermore, classroom observations can also be used to increase the validity of the results.

4. Study four validated the PRT on a small sample of 45 students. It is recommended that further studies validate the instrument with a larger sample of students to enable generalizing the results to a larger population.
5. Study five assessed the impact of formative assessment based only on one topic in mathematics. Further research can explore other topics in mathematics using the same formative assessment approaches.

5.2.2. Practical Recommendations

These practical recommendations are given in light of the current educational reforms in Kenya. As mentioned before, the Kenyan education system is currently transitioning from the current 8-4-4 system which mainly focuses on summative evaluation to a competency-based curriculum (CBC) whose main focus is on formative assessment. This calls for a paradigm shift in the teaching approaches. This study is therefore timely and the following recommendations will be of value to both the curriculum developers and teachers in Kenya.

Recommendations for the Pre-Service Teacher Training

This study recommends that the training of pre-service teachers on formative assessment should be more emphasized in the teacher training colleges. Currently, formative assessment is generally taught as an aspect of educational assessment but more emphasis should be placed on the specific strategies involved in formative assessment. This will enable the pre-service teachers to be more prepared and better placed when they start teaching.

From the interview analysis in study three, although teachers had some understanding of the concept of metacognition, they rarely applied metacognitive strategies in the classrooms. In this line, the introduction of metacognition at the pre-service teacher training colleges is highly recommended so that the prospective teachers will be able to model the same to their students when they start teaching.

Recommendations for the In-Service Teachers

The in-service teachers should be involved in designing the programmes related to formative assessment so that they can actively take part in the implementation process. Most of the professional development programmes that have taken place in Kenya have been centralized and only involve a few stakeholders. Involving teachers in their professional development will be a great motivating factor. The adoption of school-based formative assessment training programmes which can be done at a school level is therefore recommended. This will encourage teachers' collaboration and hence get involved in action research.

Based on interview results in study three, metacognition training can be incorporated into the already ongoing in-service teacher-training program on the Strengthening of Mathematics and Science in Secondary Education (SMASSE). The SMASSE project mainly focuses on 'hands-on' activities. It will be more productive if metacognition, which is a 'minds on' activity, is incorporated into the project.

5.3. Summary

This research project was conducted between 2019 and 2021. The research was done in the context of Kenyan secondary schools and it involved mathematics teachers and grade 11 students as the participants. The focus was on formative assessment conceptualized as an instructional approach in the teaching and learning process. This dissertation, which comprises five chapters, represents the original work stemming from dedication and hard work. In chapter one, the introduction to the study is given. The focus is on the study context, problem statement, significance, and the structure of the dissertation. The results of the literature review on the themes related to the study are presented in chapter two whereas chapter three presents the aims, research questions, hypotheses, and methodology of the empirical studies.

Chapter four presents a series of five empirical studies in which 15 questions and 17 hypotheses are addressed. Study one assessed the suitability of the adapted teacher assessment for learning questionnaire in the Kenyan context. Study two examined the relationship between formative assessment and mathematics teachers' metacognitive regulation. Study three assessed teachers' perceptions and conception of metacognition. In study four, a test to measure students' proportional reasoning skills in mathematics was constructed and validated. Study five examined the impact of formative assessment on students' achievement in mathematics and their metacognitive awareness.

These studies though related are presented as separate journal articles. The results, discussions, conclusions, limitations, and implications are thus given based on individual studies. Chapter five presents a general discussion of the results by addressing each research question and the corresponding hypothesis. General and practical recommendations are also given. Since no such research has been conducted in Kenya before, the present research forms a basis for further studies on formative assessment conceptualized as an instructional approach in the teaching and learning process. The current research adds knowledge to the limited empirical evidence regarding the impact of formative assessment conceptualized as an instructional approach on students' academic achievement and metacognition in mathematics.

References

Abdellah, R. (2015). Metacognitive awareness and its relation to academic achievement and teaching performance of pre-service female teachers in Ajman University in UAE. *Procedia-Social and Behavioral Sciences, 174*(1), 560-567. https://doi.org/10.1016/j.sbspro.2015.01.707

Ab Hamid, M. R., Sami, W., & Sidek, M. H. (2017). Discriminant validity assessment : Use of Fornell & Larcker criterion versus HTMT Criterion. *Journal of Physics: Conference Series*, *890*(1), 12163. https://doi.org/10.1088/1742-6596%2F890%2F1%2F012163

Alci, B., & Karatas, H. (2011). Teacher candidates' metacognitive awareness according to their domains and sex. *International Journal of Multidisciplinary Thought*, *1*(6), 255-263. https://www.researchgate.net/profile/Hakan-Karatas/publication/309359075

Alibali, M. W., Knuth, E. J., Hattikudur, S., McNeil, N. M., & Stephens, A. C. (2007). A longitudinal examination of middle school students' understanding of the equal sign and equivalent equations. *Mathematical thinking and learning*, *9*(3), 221-247. https://doi.org/10.1080/10986060701360902

Allain, A. (2000). *Development of an instrument to measure proportional reasoning among fast-track middle school students* [Master's thesis, North Carolina State University] North Carolina State University Libraries. http://www.lib.ncsu.edu/resolver/1840.16/805

Ally, J. (2014, February, 6). *Scholars to study why graduates are shunned*. The Standard Newspaper, Kenya. https://www.standardmedia.co.ke/education/article/2000104067/scholars-to-study-why-graduates-are-shunned

Alotaibi, K.A. (2019). Teachers' perceptions of factors influence the adoption of formative assessment. *Journal of Education and Learning*, *8* (1), 74-86. https://eric.ed.gov/?id=EJ1202092

Al Shabibi, A. & Alkharusi, H. (2018). Mathematical problem-solving and metacognitive skills of 5th grade students as a function of gender and level of academic achievement. *Cypriot Journal of Educational Science*, *13*(2), 385−395. https://www.researchgate.net/publication/326000384_

Al-Wattban, M. S. (2001). *Proportional reasoning and working memory capacity among Saudi adolescents: A neo-Piagetian investigation.* (Publication No. 3036614) [Doctoral dissertation, University of Northern Colorado]. ProQuest

American Educational Research Association, American Psychological Association & National Council on Measurement and Education. (1999). *Standards for educational and psychological testing.* https://www.apa.org/science/programs/testing/standards

Amoako, I., Brembong, D. A., & Bortey, J. (2019). Knowledge of formative assessment practices among senior high school mathematics teachers in Ghana. *Open Journal of Social Science Research, 3*(3), 8-13. https://www.ajhssr.com/wp-content/uploads/2019/03/B19330813.pdf

Ananiadou, K. & Claro, M. (2009). "21st Century skills and competences for new millennium learners in OECD Countries", *OECD education working papers*, No. 41, OECD Publishing. http://dx.doi.org/10.1787/218525261154

Andersson, 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)*, 92-102. https://doi.org/10.1016/j.learninstruc.2016.12.006

Andrade, H. (2010). Students as the definitive source of formative assessment: academic self-assessment and the self-regulation of learning. In H. J. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 90–105). Routledge. https://opencommons.uconn.edu/nera_2010/25

Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & evaluation in higher education*, *32*(2), 159-181. https://doi.org/10.1080/02602930600801928.

Association for the Development of Education in Africa. (2017). Revitalizing education towards the 2030, global agenda and Africa's agenda 2063. *General synthesis report for its 2017 Triennale meeting* http://www.adeanet.org/sites/default/files/2018.02.06_general_synthesis_report_on_the_2017_triennale_final_eng.pdf

Ayan, R., & Isiksal-Bostan, M. (2019). Middle school students' proportional reasoning in real-life contexts in the domain of geometry and measurement. *International Journal of Mathematical Education in Science and Technology*, *50* (1), 65-81. https://doi.org/10.1080/0020739X.2018.1468042

Baas, D., Castelijns, J., Vermeulen, M., Martens, R., & Segers, M. (2014). The relation between assessment for learning and elementary students' cognitive and metacognitive strategy use, *British Journal of Educational Psychology*, *85*(1), 36–46. https://doi.org/10.1111/bjep.12058

Balcikanli, C. (2011). Metacognitive awareness inventory for teachers (MAIT). *Electronic Journal of Research in Educational Psychology*, 9 (3), 1309–1332. https://psycnet.apa.org/record/2011-30369-016

Beesley, A. D., Clark, T. F., Dempsey, K., & Tweed, A. (2018). Enhancing formative assessment practice and encouraging middle school mathematics engagement and persistence. *School science and mathematics*, *118*(1-2), 4-16. https://doi.org/10.1111/ssm.12255.

Behrmann, L., & Souvignier, E. (2013). Pedagogical content beliefs about reading instruction and their relation to gains in student achievement. *European Journal of Psychology of Education*, *28*(3), 1023–1044. https://doi.org/10.1007/s10212-012-0152-3

Behrmann, L., & Souvignier, E. (2015). Effects of fit between teachers' instructional beliefs and didactical principles of reading programs. *European Journal of Psychology of Education*, *30*(3), 295-312. https://doi.org/10.1007/s10212-014-0241-6

Ben-David, A., & Orion, N. (2013). Teachers' voices on integrating metacognition into science education. *International Journal of Science Education*, *35*(18), 3161-3193. https://doi.org/10.1080/09500693.2012.697208

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in education: principles, policy & practice*, *18*(1), 5-25. https://doi.org/10.1080/0969594X.2010.513678

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, *5*(1), 7-74.https://doi.org/10.1080/0969595980050102.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability (formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5-31. https://doi.org/10.1007/s11092-008-9068-5

Bonwell, C. C., & Eison, J. A. (1991). Active learning: creating excitement in the classroom. *1991 ASHE-ERIC Higher Education Reports*: The George Washington University (ED336049). ERIC. https://eric.ed.gov/?id=ED336049

Boud, D., & Falchikov, N. (Eds.). (2007). *Rethinking assessment in higher education: Learning for the longer term* (1st ed.). Routledge. https://doi.org/10.4324/9780203964309

Braund, H., & DeLuca, C. (2018). Elementary students as active agents in their learning: an empirical study of the connections between assessment practices and student metacognition. *The Australian Educational Researcher*, *45*(1), 65-85. https://doi.org/10.1007/s13384-018-0265-z

Briggs, D. C., Ruiz-Primo, M. A., Furtak, E., Shepard, L., & Yin, Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice*, *31*(4), 13-17. https://doi.org/10.1111/j.1745-3992.2012.00251.x

Brink, M., & Bartz, D. E. (2017). Effective use of formative assessment by high school teachers. *Practical Assessment, Research, and Evaluation*, *22*(1), 8. https://doi.org/10.7275/p86s-zc41

Broadfoot, P., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). Assessment for learning: 10 principles: research-based principles to guide classroom practice. Assessment Reform Group. https://www.researchgate.net/publication/345173745_Formative_Assessment_Using_Feedback_to_Improve_Student_Learning

Brown, A. L. (1978). Knowing when, where, and how to remember: a problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology*, 1: 77–165. Erlbaum.

Brown, G. T., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.). *SAGE handbook of research on classroom assessment* (pp. 367-394). SAGE publications. http://dx.doi.org/10.4135/9781452218649.n21.

Brown, G. T, Harris, L. R, O'Quin, C., & Lane, K. E (2017). Using multi-group confirmatory factor analysis to evaluate cross-cultural research: identifying and understanding non-invariance. *International Journal of Research & Methodology in Education, 40* (1), 66-90. https://doi.org/10.1080/1743727X.2015.1070823

Bulunuz, N., Tavsanli, O., Bulunuz, M., & Karagoz, F. (2016). Achievement levels of middle school students in the standardized science and technology exam and formative assessment probes: a comparative study. *Journal of Education in Science Environment and Health*, *2*(1), 33-50. https://dergipark.org.tr/en/pub/jeseh/issue/18650/196800

Burner, T. (2016). Formative assessment of writing in English as a foreign language. *Scandinavian Journal of Educational Research*, *60*(6), 626-648. https://doi.org/10.1080/00313831.2015.1066430

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The clearinghouse: A journal of educational strategies, issues, and ideas*, *83*(1), 1-6. https://doi.org/10.1080/00098650903267784

Chan C. (2010). *Assessment: self and peer assessment*. Assessment Resources@HKU, University of Hong Kong. http://ar.cetl.hku.hk

Charalambous, C.Y., & Pitta-Pantazi, D. (2007). Drawing on a theoretical model to study students' understandings of fractions. *Educational studies in mathematics*, *64* (3), 293. https://doi.org/10.1007/s10649-006-9036-2

Choi, J., Kim, H., & Pak, S. (2018). Evaluation of automatic item generation utilities in formative assessment application for Korean high school students. *Journal of Educational Issues*, *4*(1), 68-89. https://doi.org/10.5296/jei.v4i1.12630

Cisterna, D., & Gotwals, A. W. (2018). Enactment of ongoing formative assessment: challenges and opportunities for professional development and practice. *Journal of Science Teacher Education*, *29*(3), 200-222. https://doi.org/10.1080/1046560X.2018.1432227

Clark I (2012). Formative assessment: assessment is for self-regulated learning. *Educ Psycho Rev 24*, 205–249. https://doi.org/10.1007/s10648-011-9191-6

Cohen, L., Manion, L. & Morrison, K. (2007). Research methods in education (6th ed.). Routledge. https://psycnet.apa.org/record/2007-05446-000

Creswell, J. W. (1994). *Research design and qualitative approaches*. Sage.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed approaches*. Sage. https://psycnet.apa.org/record/2008-13604-000

Crichton, H., & McDaid, A. (2016). Learning intentions and success criteria: learners' and teachers' views. *The Curriculum Journal*, *27*(2), 190-203. https://doi.org/10.1080/09585176.2015.1103278

Cross, D. I. J. (2009). Alignment, cohesion, and change: Examining mathematics teachers' belief structures and their influence on instructional practices. *Journal of Mathematics Teacher Education*, *12*(5), 325-346. https://doi.org/10.1007/s10857-009-9120-5

Csíkos, C. (2016). Strategies and performance in elementary students' three-digit mental addition. *Educational Studies in Mathematics*, *91* (1), 123-139. https://doi.org/10.1007/s10649-015-9658-3

Csíkos C., & Steklács J. (2010) Metacognition-based reading intervention programs among fourth-grade Hungarian students. In, A. Efklides, & Misailidi P (Eds.) *Trends and Prospects in Metacognition Research* (pp. 345-366). Springer. https://doi.org/10.1007/978-1-4419-6546-2_16

Csíkos, C., Szitányi, J., & Kelemen, R. (2012). The effects of using drawings in developing young children's mathematical word problem solving: a design experiment with third-grade Hungarian students. *Educational studies in mathematics*, *81*(1), 47-65. https://doi.org/10.1007/s10649-011-9360-z

Cutumisu, M., & Schwartz, D. L. (2018). The impact of critical feedback choice on students' revision, performance, learning, and memory. *Computers in Human Behavior*, *78*, 351-367. https://doi.org/10.1016/j.chb.2017.06.029

Dafik, S. B., Irvan, M., & Rohim, M. A. (2019). The analysis of student metacognition skill in solving rainbow connection problem under the implementation of research-based learning. *International Journal of Instruction*, *12*, 593-610. https://eric.ed.gov/?id=EJ1230084

Dann, R. (2014). Assessment as learning: blurring the boundaries of assessment and learning for theory, policy, and practice. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 149-166. https://doi.org/10.1080/0969594X.2014.898128

Darling-Hammond L. (2012). Policy frameworks for new assessments. In, P. Griffin, B. McGaw & E. Care (Eds.) *Assessment and Teaching of 21st Century Skills* (pp 301-339). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-2324-5_6

David, M. N., & Bwisa, H. M. (2013). Factors influencing teachers' active involvement in continuous professional development: A survey in Trans Nzoia west district, Kenya. *International Journal of Academic Research in Business and Social Sciences*, *3*(5), 224. https://hrmars.com/papers_submitted/9571/factors-influencing-teachers-active-involvement-in-continuous-professional-development-a-survey-in-trans-nzoia-west-district-kenya.pdf

Dekker, T., & Feijs, E. (2005). Scaling up strategies for change: change in formative assessment practices. *Assessment in Education: Principles, Policy & Practice*, *12* (3), 237-254. https://doi.org/10.1080/09695940500337215

Demetriou, A., Spanoudis, G., & Mouyi, A. (2011). Educating the developing mind: towards an overarching paradigm. *Educational psychology review*, *23*(4), 601-663. https://doi.org/10.1007/s10648-011-9178-3

Devine, D., Fahie, D., & McGillicuddy, D. (2013). What is 'good'teaching? Teacher beliefs and practices about their teaching. *Irish Educational Studies*, *32*(1), 83-108. https://doi.org/10.1080/03323315.2013.773228

de Winter, J. C., Dodou*, D. & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate behavioral research*, *44*(2), 147-181. https://doi.org/10.1080/00273170902794206

Dignath, C., & Büttner, G. (2018). Teachers' direct and indirect promotion of self-regulated learning in primary and secondary school mathematics classes-insights from video-based classroom observations and teacher interviews. *Metacognition and learning*, *13*(2), 127-157. https://doi.org/10.1007/s11409-018-9181-x.

Dignath-van Ewijk, C., Dickhäuser, O., & Büttner, G. (2013). Assessing how teachers enhance self-regulated learning: a multiperspective approach. *Journal of Cognitive Education and Psychology*, *12*(3), 338-358. https://doi.org/10.1891/1945-8959.12.3.338.

Dobish, M., Griffiths, J., & Meyer, R. (2017). Improving teaching and learning using the keeping learning on track professional development program and strategies. *Journal for Leadership and Instruction*, *16*(1), 22-26. https://eric.ed.gov/?id=EJ1159870

Doig, B., & Groves, S. (2011). Japanese lesson study: teacher professional development through communities of inquiry. *Mathematics teacher education and development*, *13*(1), 77-93. http://hdl.handle.net/10536/DRO/DU:30043312

Donker, A. S., De Boer, H., Kostons, D., van Ewijk, C .D., & Van der Werf, M.P. (2014). Effectiveness of learning strategy instruction on academic performance: a meta-analysis. *Educational Research Review 11* (1), 1–26. https://doi.org/10.1016/j.edurev.2013.11.002

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. https://doi.org/10.1111/j.1529-1006.2004.00018.x

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessments: the limited scientific evidence of the impact of formative assessments in education. *Practical Assessment, Research, and Evaluation*, *14*(1), 7. https://scholarworks.umass.edu/pare/vol14/iss1/7

Ebel, R. L. (1979). *Essentials of educational measurement* (3rd ed.). Englewood Cliffs. https://journals.sagepub.com/doi/abs/10.1177/001316448104100337?journalCode=epma

Edwards, A.R., Esmonde, I., & Wagner, J. F. (2011). Learning mathematics. In R.E. Mayer & P. A. Alexander (Eds.), *Handbook of Research on Learning and Instruction* (pp.197–218). Routledge. https://silo.pub/handbook-of-research-on-learning-and-instruction.html

Ekici, F., Ulutaş, B., & Atasoy, B. (2019). An investigation of preservice teachers' levels of metacognitive awareness in terms of certain variables. *Bartın Üniversitesi Eğitim Fakültesi Dergisi*, *8*(3), 1035-1054. https://dergipark.org.tr/en/pub/buefad/issue/49482/566640

Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive development inquiry. *American psychologist*, *34*(10), 906. https://psycnet.apa.org/doi/10.1037/0003-066X.34.10.906

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39. https://doi.org/10.2307/3151312.

Fu, Q. K., Lin, C. J., & Hwang, G. J. (2019). Research trends and applications of technology-supported peer assessment: a review of selected journal publications from 2007 to 2016. *Journal of Computers in Education*, *6*(2), 191-213. https://doi.org/10.1007/s40692-019-00131-x

Fujii, T. (2019). Designing and adapting tasks in lesson planning: a critical process of lesson study. In *Theory and practice of lesson study in mathematics* (pp. 681-704). Springer, Cham. https://doi.org/10.1007/978-3-030-04031-4_33

Furtak, E. M., Kiemer, K., Circi, R. K., Swanson, R., de León, V., Morrison, D., & Heredia, S. C. (2016). Teachers' formative assessment abilities and their relationship to student learning: findings from a four-year intervention study. *Instructional Science*, *44*(3), 267-291. https://doi.org/10.1007/s11251-016-9371-3

Gipps, C., Hargreaves, E., & McCallum, B. (2015). *What makes a good primary school teacher?: expert classroom strategies* (2nd ed.). Routledge. https://doi.org/10.4324/9781315648736

Gläser, K., & Riegler, P. (2015). Beginning students may be less capable of proportional reasoning than they appear to be. *Teaching Mathematics and its Applications: An International Journal of the IMA*, *34* (1), 26-34. https://doi.org/10.1093/teamat/hru025

Gliner, J.A., Morgan, G.A., & Leech, N.L. (2016*). Research methods in applied settings: An integrated approach to design and analysis* (3rd ed.). Routledge. https://doi.org/10.4324/9781315723082

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: validation with science majors and nonscience majors. *Journal of research in science teaching*, *48*(10), 1159-1176. https://doi.org/10.1002/tea.20442

Grantz, R. E., & Gruber, R. (2014). " How well did I learn what I learned?" The art of self assessment. *Journal of the Academy of Business Education*, *15*, 23. https://www.proquest.com/openview/5e500ebade1485cec07063d4afd8d608/1?pq-origsite=gscholar&cbl=2040119

Griffin, P., McGaw, B., & Care, E. (2013). Assessment and teaching of 21st century skills. Springer. http://dx.doi.org/10.1007/978-94-007-2324-5_2

Haladyna, T.M., & Rodriguez, M.C. (2013). *Developing and validating test items* (1st ed.). Routledge. https://doi.org/10.4324/9780203850381

Handal, B. & Herrington, A. (2003). Mathematics teachers' beliefs and curriculum reform. *Mathematics Education Research Journal*, *15* (1), 59-69. https://doi.org/10.1007/BF03217369

Harris, K. R., Santangelo, T., & Graham, S. (2010). Metacognition and strategies instruction in writing. In H. S. E. Waters & W. E. Schneider (Eds.), *Metacognition, strategy use, and instruction* (pp. 226–256). Guilford Press. https://psycnet.apa.org/record/2009-18875-010

Hashmi, A., Khalid, M., & Shoaib, A. (2019). A cross-sectional study of assessing metacognitive knowledge and metacognitive regulatory skills among prospective teachers and its relation to their academic achievement. *Bulletin of Education and Research, 41*(2), 215-234. https://eric.ed.gov/?id=EJ1229425

Hattie, J. (2008). Visible learning: *a synthesis of over 800 meta-analyses relating to achievement* (1st ed.). Routledge. https://doi.org/10.4324/9780203887332

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-11. https://doi.org/10.3102/003465430298487

Hattie, J., & Yates, G. C. R. (2013). *Visible learning and the science of how we learn* (1st ed.). Routledge. https://doi.org/10.4324/9781315885025

Henseler, J., Ringle, CM, & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43* (1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Heritage, M. (2007). Formative assessment: what do teachers need to know and do? *Phi Delta Kappan*, *89*(2), 140-145. https://journals.sagepub.com/doi/pdf/10.1177/003172170708900210

Heritage, M., & Heritage, J. (2013). Teacher questioning: the epicenter of instruction and assessment. *Applied Measurement in Education*, *26*(3), 176-190. https://doi.org/10.1080/08957347.2013.793190

Hermida, R. (2015). The problem of allowing correlated errors in structural equation modeling: concerns and considerations. *Computational Methods in Social Sciences*, *3*(1), 5-17. https://www.ceeol.com/search/article-detail?id=417198

Ho, R. (2006). *Handbook of univariate and multivariate data analysis and interpretation with SPSS.* Taylor and Francis, Boca. https://books.google.hu/books?hl=en&lr=&id=QMwqBgAAQBAJ&oi=fnd&pg=PP1&ots=Aa5MSNr4nQ&sig=MtupSMBdd9nGD49K-garCR74Hpc&redir_esc=y#v=onepage&q&f=false

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling : guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, *6*(1), 53–60. https://arrow.tudublin.ie/buschmanart/2/

Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (8th ed) Allyn & Bacon. https://eric.ed.gov/?id=ED423248

Howe, C., Nunes, T., & Bryant, P. (2011). The rational number and proportional reasoning: using intensive quantities to promote achievement in mathematics and science. *International Journal of Science and Mathematics Education*, *9* (2), 391-417. https://doi.org/10.1007/s10763-010-9249-9

Hsia, L. H., Huang, I., & Hwang, G. J. (2016). A web-based peer-assessment approach to improving junior high school students' performance, self-efficacy, and motivation in performing arts courses. *British Journal of Educational Technology*, *47*(4), 618-632. https://doi.org/10.1111/bjet.12248

Hungi, N., Makuwa, D., Ross, K., Saito, M., Dolata, S., Van Cappelle, F., Paviot, L, & Vellien, J. (2010). SACMEQ III project results: pupil achievement levels in reading and mathematics. *Working document*, *1*. https://doi.org/10.13140/RG.2.2.27312.35841

Ibrahim, M. & Watts, A. (2016). Assessing preservice teachers' metacognitive skills in a flipped classroom environment. In *Proceedings of EdMedia 2016--World Conference on Educational Media and Technology* (pp. 1189-1203). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/173098/.

Ige, A. M. (2014). Strategies for motivation of teachers for effectiveness amid universal basic education (UBE) programme implementation in Nigeria. *European Journal of Academic Essays*, *1*(6), 28-35. http://oaji.net/articles/2016/3667-1481649670.pdf

Inoue, N., Asada, T., Maeda, N., & Nakamura, S. (2019). Deconstructing teacher expertise for inquiry-based teaching: looking into consensus building pedagogy in Japanese classrooms. *Teaching and teacher education*, *77*, 366-377. https://doi.org/10.1016/j.tate.2018.10.016

Jacoby, J., Heugh, S., Bax, C., & Branford- White, C. (2014). Enhancing learning through formative assessment. *Innovations in education and teaching international*. *51*(1). 72-83. https://doi.org/10.1080/14703297.2013.771970

Jiang, Y., Ma, L., & Gao, L. (2016). Assessing teachers' metacognition in teaching: the teacher metacognition inventory. *Teaching and Teacher Education*, *59*, 403-413. https://doi.org/10.1016/j.tate.2016.07.014.

Johnson, C. C., Sondergeld, T. A., & Walton, J. B. (2019). A study of the implementation of formative assessment in three large urban districts. *American Educational Research Journal*, *56*(6), 2408-2438. https://doi.org/10.3102/0002831219842347

Jónsson, Í. R., Smith, K., & Geirsdóttir, G. (2018). Shared language of feedback and assessment. perception of teachers and students in three Icelandic secondary schools. *Studies in Educational Evaluation*, *56*, 52-58. https://doi.org/10.1016/j.stueduc.2017.11.003

Kafyulilo, A. C. (2013). Professional development through teacher collaboration: an approach to enhance teaching and learning in science and mathematics in Tanzania. *Africa Education Review*, *10*(4), 671-688 https://doi.org/10.1080/18146627.2013.853560

Kallio, H. Kallio, M., Virta, K., Iiskala, T., & Hotumainen, R. (2020). Teachers' support for learners' metacognitive awareness. *Scandinavian Journal of Educational Research*, 1-17. https://doi.org/10.1080/00313831.2020.1755358

Kennedy, A. (2005). Models of continuing professional development: a framework for analysis. Journal of InService Education, 31, 235–250. http://dx.doi.org/10.1080/13674580500200358

Kenney, R., Bloome, L., & Maeda, Y. (2016). Unpacking Teachers' perspectives on the purpose of assessment: beyond summative and formative. *North American Chapter of the International Group on Psychology of Mathematics Education*. *https://eric.ed.gov/?id=ED583794*

Kenya National Bureau of Statistics (2019). Population by county and sub county: Kenya *Population and Housing Census Volume 1.* Retrieved 6 December 2019 from https://www.knbs.or.ke/?wpdmpro=2019-kenya-population-and-housing-census-volume-i-population-by-county-and-sub-county

KICD. (2017). *Basic education curriculum framework*. The Republic of Kenya, KICD. https://kicd.ac.ke/wp-content/uploads/2017/10/CURRICULUMFRAMEWORK.pdf

Kiige, J. M. (2019). Challenges facing the implementation of SMASSE'S ASEI/PDSI approaches in teaching. *International Journal of Education and Multidisciplinary Studies 15*(2), 72-77. http://dx.doi.org/10.21013/jems.v15.n2.p3.

Kingston, N., & Nash, B. (2011). Formative assessment: a meta-analysis and a call for research. *Educational Measurement: Issues and practice, 30*(4), 28-37. https://doi.org/10.1111/j.1745-3992.2011.00220.x.

Kippers, W. B., Wolterinck, C. H., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education*, *75*, 199-213. https://doi.org/10.1016/j.tate.2018.06.015

Klenowski, V. (2009). Assessment for learning revisited: an Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice, (16)*3, 263-268, https://doi.org/10.1080/09695940903319646

Kline, R. B. (2015). *Principles and practice of structural equation modeling*(4th ed). Guilford publications. https://books.google.hu/books?hl=en&lr=&id=Q61ECgAAQBAJ&oi=fnd&pg=PP1&dq=Kline,+R.+B.+(2015).

KNEC. (2014). *Kenya national examination council, newsletter*, Government printer. https://qmis.knec.ac.ke/

KNEC. (2017). *Kenya national examination council, newslette*r, Government printer. https://qmis.knec.ac.ke/

Koc, I., & Kuvac, M. (2016). Preservice science teachers' metacognitive awareness levels. *European Journal of Education Studies*. https://www.oapub.org/edu/index.php/ejes/article/view/212

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: the effects of cooperative learning and metacognitive training. *American Educational Research Journal*, *40*(1), 281–310. https://doi.org/10.3102/00028312040001281

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory into practice*, *41*(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2

Krijgsman, C., Mainhard, T., van Tartwijk, J., Borghouts, L., Vansteenkiste, M., Aelterman, N., & Haerens, L. (2019). Where to go and how to get there: goal clarification, process feedback and students' need satisfaction and frustration from lesson to lesson. *Learning and Instruction*, *61*, 1-11. https://doi.org/10.1016/j.learninstruc.2018.12.005

Kyaruzi, F., Strijbos, J. W., Ufer, S., & Brown, G. T. (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. (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

Lai, E. R., & Viering, M. (2012). Assessing 21st-century skills: integrating research findings. *National Council on Measurement in Education*. https://eric.ed.gov/?id=ED577778

Lamon, S. J. (1993). Ratio and proportion: Connecting content and children's thinking. *Journal for research in mathematics education*, *24*(1), 41-61. https://doi.org/10.2307/749385

Leach, L. (2012). Optional self-assessment: some tensions and dilemmas. *Assessment & Evaluation in Higher Education*, *37*(2), 137-147. https://doi.org/10.1080/02602938.2010.515013

Lee, I., & Mak, P. (2014). Assessment as learning in the language classroom. *Assessment as learning. Hong Kong: Education Bureau*. https://wlts.edb.hkedcity.net/filemanager/file/A&L3(09)_Lee%20&Mak.pdf

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, *136*(6), 1123–1135. https://doi.org/10.1037/a0021276

Louis, R. A., & Mistele, J. M. (2012). The differences in scores and self-efficacy by student gender in mathematics and science. *International Journal of Science and Mathematics Education*, *10*(5), 1163-1190. https://doi.org/10.1007/s10763-011-9325-9

Lyon, C. J., Nabors Oláh, L., & Caroline Wylie, E. (2019). Working toward integrated practice: understanding the interaction among formative assessment strategies. *The Journal of Educational Research*, *112*(3), 301-314.https://doi.org/10.1080/00220671.2018.1514359

Mahdavi, M. (2014). An overview: Metacognition in education. *International Journal of Multidisciplinary and Current Research*, *2*(6), 529-535. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1085.650&rep=rep1&type=pdf

Martinez, M. E. (2006). What is Metacognition? *Phi Delta Kappan*, *87*(9), 696–699. https://doi.org/10.1177/003172170608700916

McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: the effects of self-assessment training on performance in external examinations. *Assessment in education: principles, policy & practice*, *10*(2), 209-220. https://doi.org/10.1080/0969594032000121289

MacDonald, B. L., & Wilkins, J. L. (2016). Seven types of subitizing activity characterizing young children's mental activity. In *Qualitative research in STEM* (pp. 264-294). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315676449-19/

Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology* (2nd ed.). Houghton Mifflin Company. https://www.scirp.org/(S(vtj3fa45qm1ean45vvffcz55))/reference/ReferencesPapers.aspx?ReferenceID=537796

Merriam-Webster. (n.d.). Metacognition. In *Merriam-Webster.com dictionary*. Retrieved May 5, 2021, from https://www.merriam-webster.com/dictionary/metacognition

Mevarech Z., & Fridkin, S. (2006). The effects of IMPROVE on mathematical knowledge, mathematical reasoning, and meta-cognition. *Metacognition Learning* *1*(1), 85-97 https://doi.org/10.1007/s11409-006-6584-x

Miholic, V. (1994). An inventory to pique students' metacognitive awareness of reading strategies. *Journal of Reading*, *38*(2), 84-86. https://www.jstor.org/stable/40032273

Milligan, L. O. (2017). Education quality and the Kenyan 8-4-4 curriculum: secondary school learners' experiences. *Research in Comparative and International Education*, *12*(2), 198-212. https://doi.org/10.1177%2F1745499917711550

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: an expanded sourcebook*. Sage.https://books.google.com/books?hl=en&lr=&id=U4lU_-wJ5QEC&oi=fnd&pg=PA10&ots=kFVC5KOY_P&sig=jzM0o7WXbUmIgx6ezHEyW_pPOfs

Ministry of Education (2012). Towards a globally competitive quality education for sustainable development. *Report of the Task Force*. https://vision2030.go.ke/publication/towards-a-globally-competitive-quality-education-for-sustainable-development/

Mokhele, M. L., & Jita, L. C. (2010). South African teachers' perspectives on continuing professional development: a case study of the Mpumalanga secondary science initiative. *Procedia-Social and Behavioral Sciences*, *9*, 1762-1766. https://doi.org/10.1016/j.sbspro.2010.12.396

Morrissette, J. (2011). Formative assessment: revisiting the territory from the point of view of teachers. *McGill Journal of Education, 46(2), 247–265*. https://doi.org/10.7202/1006438ar

Moss, C. M., Brookhart, S. M., & Long, B. A. (2011). What is a shared learning target? *Educational Leadership*, *68*(6), 66-69. https://www.montanalanguageteachers.org/uploads/6/6/4/8/66489615/3_excerpts_brookhart_knowing_your_learning_target_march_2011.pdf

Muijs, D., Kyriakides, L., Van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement 25* (2): 231–256. https://doi.org/10.1080/09243453.2014.885451.

Mullis, I.V.S., & Martin, M.O. (Eds) (2017). *TIMMS 2019 Assessment frameworks*. Retrieved from Boston College, TIMSS & PIRLs International Study Center Website http://timssandpirls.bc.edu/timss2019/frameworks/

Musfirah, Y. (2019). "The use of peer assessment in speaking skill." *English Education Journal 10*(1): 67-79. http://www.e-repository.unsyiah.ac.id/EEJ/article/view/13257

Mwangi, N. I., & Mugambi, M. (2013). Evaluation of strengthening of mathematics and science in secondary education (SMASSE) program. a case study of Murang'a South District, Kenya. *International Journal of Education Learning and Development*, *1*, 46–60. http://kerd.ku.ac.ke/123456789/1175

National Council of Teachers of Mathematics, NCTM. (2015). *Characteristics of mathematical tasks.* Retrieved from, https://www.nctm.org/uploadedFiles/Conferences_and_Professional_Development/Institutes/Pre-K_Grade_12_Common_Core_Series/Beth_handouts.pdf

Naufal, M. A., Abdullah, A. H., Osman, S., Abu, M. S., & Ihsan, H. (2021). The effectiveness of infusion of metacognition in van Hiele model on secondary school students' geometry thinking level. *International Journal of Instruction*, *14*(3), 535-546. https://doi.org/10.29333/iji.2021.14331a

Naseri, M., Shoae, K. M., Effati, M, M. (2017). The effectiveness of metacognitive skills training on increasing academic achievement. *Iranian journal of educational sociology*, *1*(3), 83-88. https://www.sid.ir/en/journal/ViewPaper.aspx?ID=595701

Nasr, M., Bagheri, M. S., Sadighi, F., & Rassaei, E. (2018). Iranian EFL teachers' perceptions of assessment for learning regarding monitoring and scaffolding practices as a function of their demographics. *Cogent Education*, *5*(1), 1558916 https://doi.org/10.1080/2331186X.2018.1558916

Ngugi & Nyakweba. (2005). *Understanding SMASSE PROJECT*; a paper presented in Uasin-Gishu district SMASSE INSET cycle 2. Unpublished

Ngware, M. W., Oketch, M., & Mutisya, M. (2014). Does teaching style explain differences in learner achievement in low and high-performing schools in Kenya?. *International Journal of Educational Development*, *36*, 3-12. https://doi.org/10.1016/j.ijedudev.2014.01.004

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. https://doi.org/10.1080/03075070600572090

Nikou, S. A., & Economides, A. A. (2016). The impact of paper-based, computer-based and mobile-based self-assessment on students' science motivation and achievement. *Computers in Human Behavior, 55*, 1241-1248. https://doi.org/10.1016/j.chb.2015.09.025

Nunes, T., Bryant, P., Evans, D., Bell, D., & Barros, R. (2012). Teaching children how to include the inversion principle in their reasoning about quantitative relations. *Educational Studies in Mathematics*, *79* (3), 371-388. https://doi.org/10.1007/s10649-011-9314-5

Nunes, T., & Bryant, P. (2009). Paper 4: Understanding relations and their graphical representation. *Key understandings in mathematics learning. Nuffield Foundation*. https://www.scribd.com/document/458609573/Nunes-Bryant-Watson-2009-Mathematics-learning-1-8-pdf

Nunes, T., & Bryant, P. (2015). The development of quantitative reasoning. In L. S. Liben & U. Müller (Eds.), *Handbook of child psychology and developmental science* (7th ed., Vol. 2. Cognitive Process, pp. 715-764). https://doi.org/10.1002/9781118963418.childpsy217

OECD. (2005). *Formative assessment: improving learning in secondary classrooms. policy brief*. OECD. https://www.oecd.org/education/ceri/35661078.pdf

OECD. (2009). Teaching practices. Teachers' beliefs and attitudes, in creating effective teaching and learning environments: *First results from TALIS, OECD publishing* https://dx.doi.org/10.1787/9789264068780-6-en

OECD. (2013). *Teaching and learning international survey*, *TALIS 2013*: conceptual Framework. OECD http://www.oecd.org/education/school/TALIS%20Conceptual%20Framework_FINAL.pdf

OECD. (2014). *The high-level group on the modernization of higher education: Report to the European Commission on new modes of learning and teaching in higher education*. OECD. https://op.europa.eu/en/publication-detail/-/publication/fbd4c2aa-aeb7-41ac-ab4c-a94feea9eb1f

Ono, Y., & Ferreira, J. (2010). A case study of continuing teacher professional development through lesson study in South Africa. *South African Journal of education*, *30*(1). https://doi.org/10.15700/saje.v30n1a320

Owen, L. (2016). The impact of feedback as a formative assessment on student performance. *International Journal of Teaching and Learning in Higher Education 28*(2): 168-175. https://files.eric.ed.gov/fulltext/EJ1111131.pdf

Ozan, C., & Kıncal, R. Y. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Educational Sciences: Theory & Practice*, *18*(1), 85–118. http://dx.doi.org/10.12738/estp.2018.1.0216

Öz, H. (2014). Turkish teachers. practices of assessment for learning in the english as a foreign language classroom. *Journal of Language Teaching and Research*, *5*(4), 775. http://dx.doi.org/10.4304/jltr.5.4.775-785

Özsoy, G. (2011). An investigation of the relationship between metacognition and mathematics achievement. *Asia Pacific Education Review*, *12*(2), 227-235. http://dx.doi.org/10.1007/s12564-010-9129-6

Özsoy, G., & Günindi, Y. (2011). Prospective preschool teachers' metacognitive awareness. *İlköğretim Online*, *10*(2), 430-440. http://ilkogretim-online.org.tr/

Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: theoretical and practical connotations, when it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology, 11*(2), 551-576. http://dx.doi.org/10.14204/ejrep.30.12200.

Panadero, E., & Alonso-Tapia, J. (2014). How do students self-regulate? Review of Zimmerman's cyclical model of self-regulated learning. *Anales De Psicología / Annals of Psychology*, *30*(2), 450-462. https://doi.org/10.6018/analesps.30.2.167221

Panadero, E., & Broadbent, J. (2018). Developing evaluative judgment: a self-regulated learning perspective. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing evaluative judgment: Assessment for knowing and producing quality work*. Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315109251-9/

Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: a review of known unknowns and potential directions. *Educational Psychology Review*, *28*(4), 803-830. https://doi.org/10.1007/s10648-015-9350-2.

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: four meta-analyses. *Educational Research Review*, *22*, 74-98. https://doi.org/10.1016/j.edurev.2017.08.004

Paris, S. G., & Jacobs, J. E. (1984). The benefits of informed instruction for children's reading awareness and comprehension skills. *Child Development*, *55*(6), 2083–2093. https://doi.org/10.2307/1129781

Partnership for 21st century skills (2009). *P21 framework definitions*. Retrieved from http://www.p21.org/storage/documents/P21_Framework_Definitions.pdf

Pat-El, R. J., Tillema, H., & Segers, M. (2013). Validation of assessment for learning questionnaires for teachers and students. *British Journal of Educational Psychology*, *83* (1), 98–113. https://doi.org/10.1111/j.2044-8279.2011.02057.x

Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2015). Multilevel predictors of differing perceptions of assessment for learning practices between teachers and students. *Assessment in Education: Principles, Policy & Practice*, *22*(2), 282-298. https://doi.org/10.1080/0969594X.2014.975675

Pepper, D. (2011) Assessing key competences across the curriculum and Europe. *European Journal of Education*, *46*(3), 335–353. https://doi.org/10.1111/j.1465-3435.2011.01484.x

Philipp, R. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 257–315). Information Age Publishing. https://books.google.hu/books?hl=en&lr=&id=B_onDwAAQBAJ&oi=fnd&pg=PR1&dq

Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2018a). Implementation of formative assessment–effects of quality of programme delivery on students' mathematics achievement and interest. *Assessment in Education: Principles, Policy & Practice*, *25*(2), 160-182. https://doi.org/10.1080/0969594X.2016.1170665.

Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2018b). Interplay of formative assessment and instructional quality—interactive effects on students' mathematics achievement. *Learning environments research*, *21*(1), 61-79. https://doi.org/10.1007/s10984-017-9240-2.

Pontefract, C., & Hardman, F. (2005). The discourse of classroom interaction in Kenyan primary schools. *Comparative education*, *41*(1), 87-106. https://doi.org/10.1080/03050060500073264

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K. & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *A product from the ESRC methods programme Version*, *1*, p.b92. https://doi.org/10.13140/2.1.1018.4643.

Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2019). Formative assessment in mathematics: mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction*, *60*, 154-165. https://doi.org/10.1016/j.learninstruc.2018.01.004

Ramatlapana, K. A. (2009). Provision of in-service training of mathematics and science teachers in Botswana: teachers' perspectives. *Journal of Mathematics Teacher Education*, *12*(2), 153-159. https://link.springer.com/content/pdf/10.1007/s10857-009-9101-8.pdf

Randel, B., Apthorp, H., Beesley, A., Clark, T., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes. *The Journal of Educational Research*, 109(5), 491-502. http://dx.doi.org/10.1080/00220671.2014.992581

Republic of Kenya. (2010). *The constitution of Kenya.* Government Printer.

Republic of Kenya. (2012). *Sessional paper No.14 of 2012: reforming education and training sectors in Kenya.* Kenya Ministry of Higher Education, Science and Technology

Republic of Kenya. (2014). *National education sector plan (NESP).* Ministry of education science and technology. Government Printer

Ringle, C., Da Silva, D., & Bido, D. (2015). Structural equation modeling with the SmartPLS. *Bido, D., da Silva, D., & Ringle, C. (2014). Structural equation modeling with Smart pls. Brazilian Journal of Marketing*, *13* (2). https://ssrn.com/abstract=2676422

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and instruction*, *21*(2), 267-280. https://doi.org/10.1016/j.learninstruc.2010.07.004

Rotsaert, T., Panadero, E., Estrada, E., & Schellens, T. (2017). How do students perceive the educational value of peer assessment in relation to its social nature? A survey study in Flanders. *Studies in Educational Evaluation*, *53*, 29-40. https://doi.org/10.1016/j.stueduc.2017.02.003

Rotsaert, T., Panadero, E., & Schellens, T. (2018). Peer assessment use, its social nature challenges, and perceived educational value: a teachers' survey study. *Studies in Educational Evaluation*, *59*, 124-132. https://doi.org/10.1016/j.stueduc.2018.07.001

Rotsaert, T., Panadero, E., Schellens, T., & Raes, A. (2018). "Now you know what you're doing right and wrong!" Peer feedback quality in synchronous peer assessment in secondary education. *European Journal of Psychology of Education*, *33*(2), 255-275. https://doi.org/10.1007/s10212-017-0329-x.

Sach, E. (2012). Teachers and testing: An investigation into teachers' perceptions of formative assessment. *Educational Studies*, *38*(3), 261-276. https://doi.org 10.1080/03055698.2011.598684

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, *34*(2), 159-179. https://doi.org/10.1080/02602930801956059

Saito, H., & Inoi, S. I. (2017). Junior and senior high school EFL teachers' use of formative assessment: A mixed-methods study. *Language Assessment Quarterly*, *14*(3), 213-233. https://doi.org/10.1080/15434303.2017.1351975

Sayed, Y., & Kanjee, A. (2013). Assessment in Sub-Saharan Africa: challenges and prospects, *Assessment in Education: Principles, Policy & Practice*, *20*(4), 373-384 https://doi.org/10.1080/0969594X.2013.849056

Scheiber, C., Reynolds, M. R., Hajovsky, D. B., & Kaufman, A. S. (2015). Gender differences in achievement in a large, nationally representative sample of children and adolescents. *Psychology in the Schools*, *52*(4), 335-348. https://doi.org/10.1002/pits.21827

Schoenfeld, A. H. (2010). Research methods in (mathematics) education. In *Handbook of international research in mathematics education* (pp. 481-533). Routledge. https://doi.org/10.4324/9780203930236

Schraw, G. (1998). Promoting general metacognitive awareness: *Instructional Science 26*: 113–125. https://doi.org/10.1023/A:1003044231033

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary educational psychology*, *19*(4), 460-475. https://doi.org/10.1006/ceps.1994.1033

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational psychology review*, *7*(4), 351-371. https://doi.org/10.1007/BF02212307

Schumaker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Lawrence Erlbaum. https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9781410610904&type=googlepdf

Schunk, D. H. (2012). *Learning theories: an educational perspective* (6th ed). Pearson. http://thuvienso.bvu.edu.vn/handle/TVDHBRVT/15440

Sharma, R., Jain, A., Gupta, N., Garg, S., Batta, M., & Dhir, S. K. (2016). Impact of self-assessment by students on their learning. *International Journal of Applied and Basic Medical Research*, *6*(3), 226. https://doi.org/ 10.4103/2229-516X.186961

Shepard, L.A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 624–646). Greenwood Publishing Group.

Shute, V. J., & Becker, B. J. (2010). *Innovative assessment for the 21st century*. Springer. https://doi.org/10.1007/978-1-4419-6530-1_1

Singh, P. (2000). Understanding the concepts of proportion and ratio constructed by two grade six students. *Educational Studies in Mathematics, 43* (3), 271-292. https://doi.org/10.1023/A:1011976904850

Siswati, B. H., & Corebima, A. D. (2017). The effect of education level and gender on students' metacognitive skills in Malang, Indonesia. *Advances in Social Sciences Research Journal*, *4*(4). https://doi.org/10.14738/assrj.44.2813

SMASSE project (2008). *Statistical analysis on SMASSE project impact assessment survey*, JICA: Unpublished.

Sperling, R. A, Howard, B. C, Miller, L. A, & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. *Contemporary Educational Psychology*, *27* (1), 51-79. https://doi.org/10.1006/ceps.2001.1091

Spiller, D. (2012). Assessment matters: self-assessment and peer assessment. *The University of Waikato*, *13*. Retrieved from, www.sciepub.com/reference/202268.

Spruce, R., & Bol, L. (2015). Teacher beliefs, knowledge, and practice of self-regulated learning. *Metacognition and Learning*, *10*(2), 245-277. https://doi.org/10.1007/s11409-014-9124-0.

Staub, F. C., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology, 94*(2), 344-355 https://psycnet.apa.org/doi/10.1037/0022-0663.94.2.344

Stel, van der Stel, M., & Veenman, M. V. J. (2010). Development of metacognitive skillfulness: a longitudinal study. *Learning and Individual Differences 20* (3): 220–224 https://doi.org/10.1016/j.lindif.2009.11.005.

Stewart, P. W., Cooper, S. S., & Moulding, L. R. (2007). Metacognitive development in professional educators. *The Researcher*, *21*(1), 32-40. http://www.nrmera.org/wp-content/uploads/2016/02/Researcherv21n1Stewart.pdf

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and teacher education*, *17*(2), 213-226. https://doi.org/10.1016/S0742-051X(00)00052-4

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.

Sulaiman, T., Rahim, A., Syrene, S., & Yan, K. (2021). Primary Science Teachers' Perspectives about Metacognition in Science Teaching. *European Journal of Educational Research*, *10*(1), 75-84. https://eric.ed.gov/?id=EJ1284376

Tanner, K. D. (2012). Promoting student metacognition. *CBE—Life Sciences Education*, *11*(2), 113-120.https://doi.org/10.1187/cbe.12-03-0033

Taras, M. (2010). Student self-assessment: processes and consequences. *Teaching in higher education*, *15*(2), 199-209. https://doi.org/10.1080/13562511003620027

Tenório, T., Bittencourt, I. I., Isotani, S., & Silva, A. P. (2016). Does peer assessment in on-line learning environments work? A systematic review of the literature. *Computers in Human Behavior*, *64*, 94-107. https://doi.org/10.1016/j.chb.2016.06.020

Teo, T. (2013). Examining the psychometric properties of the epistemic belief inventory (EBI). *Journal of Psychoeducational Assessment*, *31*(1), 72–79. https://doi.org/10.1177%2F0734282912449436

Thienngam, S., Promlek, A., & Thongsaard, K. (2020). Influence of teachers' metacognitive skills on development of early-childhood students. *Australian Journal of Teacher Education (Online)*, *45*(1), 19-30. https://search.informit.org/doi/abs/10.3316/informit.190615314597838

Tjoe, H., & de la Torre, J. (2014). The identification and validation process of proportional reasoning attributes an application of a cognitive diagnosis modeling framework. *Mathematics Education Research Journal*, *26* (2), 237-255. https://doi.org/10.1007/s13394-013-0090-7

Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education*, *38*(3), 323-342. https://doi.org/10.1080/03054985.2012.689693.

Tsivitanidou, O. E., Constantinou, C. P., Labudde, P., Rönnebeck, S., & Ropohl, M. (2018). Reciprocal peer assessment as a learning tool for secondary school students in modeling-based learning. *European Journal of Psychology of Education*, *33*(1), 51-73. https://doi.org/10.1007/s10212-017-0341-1

Turner, J. C., Christensen, A., & Meyer, D. K. (2009). Teachers' beliefs about student learning and motivation. In L.J. Saha & A.G. Dworkin. (eds.), *International handbook of research on teachers and teaching* (pp. 361-371). Springer. https://doi.org/10.1007/978-0-387-73317-3_23

UNESCO. (2014). *Global Citizenship Education: Preparing Learners for the Challenges of the 21st Century*. UNESCO. https://books.google.com/books?id=VvxnBAAAQBAJ&printsec=frontcover&dq=UNESCO.+(2013).

UNESCO. (2015). Education for all 2000–2015: Achievements and challenges. UNESCO. https://en.unesco.org/gem-report/report/2015/education-all-2000-2015-achievements-and-challenges

Usher, B. L. (2019). *Analyzing teachers' experiences with metacognition communication with students about classroom assessment and pedagogy* (Publication No. 13879111) [Doctoral dissertation, Grand Canyon University]. ProQuest

Van Dooren, W., De Bock, D., Hessels, A., Janssens, D., & Verschaffel, L. (2005). Not everything is proportional: effects of age and problem type on propensities for overgeneralization. *Cognition and instruction*, *23* (1), 57-86. https://doi.org/10.1207/s1532690xci2301_3

Van Dooren, W., De Bock, D., & Verschaffel, L. (2010). From addition to multiplication… and back. The development of students' additive and multiplicative reasoning skills. *Cognition and Instruction*, *28*(3), 360–381. https://doi.org/10.1080/07370008.2010.488306

Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education*, *81*, 123-132. https://doi.org/10.1016/j.compedu.2014.10.001

Van der Kleij, F. M. (2019). Comparison of teacher and student perceptions of formative assessment feedback practices and association with individual student characteristics. *Teaching and Teacher Education*, *85*, 175-189. https://doi.org/10.1016/j.tate.2019.06.010.

Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: a meta-analysis. *Review of educational research*, *85*(4), 475-511. https://doi.org/10.3102%2F0034654314564881

Van der Nest, A., Long, C., & Engelbrecht, J. (2018). The impact of formative assessment activities on the development of teacher agency in mathematics teachers. *South African Journal of Education*, *38*(1). http://dx.doi.org/10.15700/saje.v38n1a1382

Van de Walle, J. A., & Lovin, L. H. (2006). Teaching student-centered mathematics: Grades 5-8. Montreal Pearson

Van Gennip, N. A., Segers, M. S., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, *20*(4), 280-290. https://doi.org/10.1016/j.learninstruc.2009.08.010.

Vattøy, K. D., & Smith, K. (2019). Students 'perceptions of teachers' feedback practice in teaching English as a foreign language. *Teaching and Teacher Education*, *85*, 260-268. https://doi.org/10.1016/j.tate.2019.06.024

Veenman, M. V. J. (2011). Learning to self-monitor and self-regulate. In R.E. Mayer & P A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 197-218). Routledge.

Veenman, M. V. J. (2013). Assessing metacognitive skills in computerized learning environments. In: R. Azevedo., & V. Aleven. (eds), *International handbook of metacognition and learning technologies.* Springer International handbooks Education, vol28. Springer. https://doi.org/10.1007/978-1-4419-5546-3_11

Veenman, M. V. J., & Beishuizen, J. J. (2004). Intellectual and metacognitive skills of novices while studying texts under conditions of text difficulty and time constraint. *Learning and Instruction 14* (6), 621–640. https://doi.org/10.1016/j.learninstruc.2004.09.004

Veenman, M. V. J., & Spaans, M. A. (2005). The relation between intellectual and metacognitive skills: age and task differences. *Learning and Individual Differences* 15 (2), 159–176. https://doi.org/10.1016/j.lindif.2004.12.001

Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and instruction*, *14*(1), 89-109. https://doi.org/10.1016/j.learninstruc.2003.10.004

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach. P. (2006). metacognition and learning: conceptual and methodological issues. *Metacognition Learning 1*(1)3–14. https://doi.org/10.1007/s11409-006-6893-0

Vogelzang, J., & Admiraal, W. F. (2017). Classroom action research on formative assessment in a context-based chemistry course. *Educational Action Research*, *25*(1), 155-166. https://doi.org/10.1080/09650792.2016.1177564

Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: implications for national curriculum policies. *Journal of Curriculum Studies*, *44* (3), 299-321. https://doi.org/10.1080/00220272.2012.668938

Vrugt, A., & Oort, F. J. (2008). Metacognition, achievement goals, study strategies, and academic achievement: pathways to achievement. *Metacognition and learning*, *3*(2), 123-146. https://doi.org/10.1007/s11409-008-9022-4

Wafubwa, R. N. (2020). Role of formative assessment in improving students' motivation, engagement, and achievement: a Systematic review of literature. *International Journal of Assessment and Evaluation 28* (1), 17-31. https://doi.org/10.18848/2327-7920/CGP/v28i01/17-31

Wafubwa, R. N., & Csíkos, C. (2021). Formative assessment as a predictor of mathematics teachers' levels of metacognitive regulation. *International Journal of Instruction*, *14*(1), 983-998. https://doi.org/10.29333/iji.2021.14158a

Wafubwa, R. N., & Obuba, E. (2015). Influence of strengthening mathematics and science in secondary education (SMASSE) in-service education and training (INSET) on the attitude of students towards mathematics performance in public secondary schools of Rangwe Division, Homa-Bay Sub-County-Kenya. *Journal of Education and Practice 6* (26), 57-62. https://eric.ed.gov/?id=EJ1077453

Wafubwa, R. N., Opoku-Sarkodie, R., & Csíkos, C. (2020). Construction and validation of a test for measuring students' proportional reasoning on rates, ratios, & proportions. *International Journal of Pedagogy, Policy and ICT in Education*, (*8*), 53-76. https://www.ajol.info/index.php/ijp/article/view/198696

Widiastuti, I. A. M. S., Mukminatien, N., Prayogo, J. A., & Irawati, E. (2020). Dissonances between teachers' beliefs and practices of formative assessment in EFL classes. *International Journal of Instruction*, *13*(1), 71-84. 71-84. https://doi.org/10.29333/iji.2020.1315a

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: a five-step guide for novices. *Australasian Journal of Paramedicine*, *8* (3), 1-11. http://dx.doi.org/10.33151/ajp.8.3.93

Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press. https://www.magonline.net/files/132972297.pdf

Wiliam, D., & Thompson, M. (2008). Integrating assessment with learning: what will it take to make it work? In C. A. Dwyer (Ed.), *The future of Assessment: Shaping teaching and learning* (pp. 53-82). Routledge.

Williams, J. P., & Atkins, J. G. (2009). The role of metacognition in teaching reading comprehension to primary students. In *Handbook of metacognition in education* (pp. 38-56). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203876428-10/

Wilson, N. S., and Bai, H. (2010). The relationships and impact of teachers' metacognitive knowledge and pedagogical understandings of metacognition. *Metacognition Learning*, 5, 269-288. https://doi.org/10.1007/s11409-010-9062-4

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.2014.990416.

Xiao, Y. (2017). Formative assessment in a test-dominated context: how test practice can become more productive. *Language Assessment Quarterly*, *14*(4), 295-311. https://doi.org/10.1080/15434303.2017.1347789.

Xiao, Y., & Yang, M. (2019). Formative assessment and self-regulated learning: how formative assessment supports students' self-regulation in English language learning. *System*, *81*, 39-49. https://doi.org/10.1016/j.system.2019.01.004.

Yin, X., & Buck, G. A. (2019). Using a collaborative action research approach to negotiate an understanding of formative assessment in an era of accountability testing. *Teaching and Teacher Education*, *80*, 27-38. https://doi.org/10.1016/j.tate.2018.12.018.

Young, A., & Fry, J. D. (2008). Metacognitive awareness and academic achievement in college students. *Journal of the Scholarship of Teaching and Learning*, *8*(2), 1-10. https://eric.ed.gov/?id=EJ854832

Young, J. E., & Jackman, M. G. A. (2014). Formative assessment in the Grenadian lower secondary school: teachers' perceptions, attitudes and practices. *Assessment in Education: Principles, Policy & Practice*, *21*(4), 398-411. https://doi.org/10.1080/0969594X.2014.919248

Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance: an introduction and an overview. In B. J. Zimmerman & D. H. Schunk (Eds.), *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (pp. 1–12). Routledge.

Appendices

Appendix A1: Formative Assessment Mathematics Lesson Plan Guide for Teachers

Topic

Subtopic

Duration

Number of students

Date

RATIONALE……………………………………………………………………………………….
…………………………………………………………………………………………………
……………………………………………………………………………………………....

Stage	Strategy	Teaching action	Purpose
Step 1 Where the learner is going (10 minutes)	Making learning targets and standards for success clear	Present the problem to the students. Use a whole class discussion to clarify learning intentions and possible strategies for solving the problem.	To elicit ideas for possible ways of solving the problem
Step 2 Where the learner is right now (15 minutes)	Discussions peer assessment self-assessment	Use questions to enable students to determine where they are. Provide hints	Diagnose the student's strengths and limitations inspire students to reflect on their work to make sure it makes sense
Step 3 How to get there (15minutes)	Self-assessment peer-assessment teacher's feedback	Show and discuss solutions to the problems. Discuss the special features	Show and illustrate different strategies Demonstrate how the problem strategies can be applied in authentic situations

Appendix A2: Formative Assessment Training Schedule

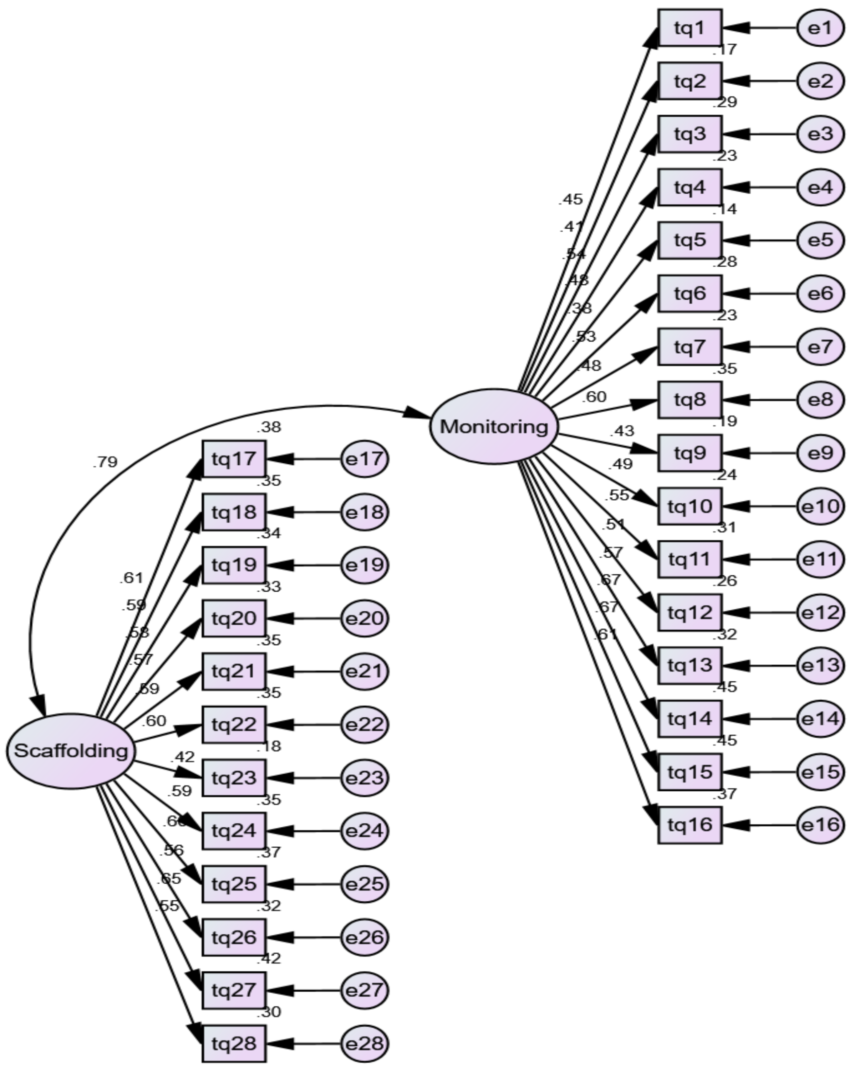
Day	Session	Content
Day 1	Session 1	• Introduction • Different concepts of formative assessment • Teachers' conception and experiences of formative assessment
	Session 2	• The conceptual framework for the study (Black & Wiliam, 2009)- five strategies, three agents, and three processes
Day 2	Session 1	• A detailed discussion of each strategy and how they can promote learning (based on research reports) • Discussion on how the strategies can enhance metacognition (research evidence) • Teachers views on different strategies
	Session 2	• Implementation of the strategies • Implementation challenges and how they can be addressed • Lesson plan guide

Appendix B: Teacher Assessment for Learning Questionnaire (TAFL-Q)

1. I encourage my students to reflect upon how they can improve their assignments.
2. After a test, I discuss the answers given with each student.
3. While working on their assignments, I ask my students how they think they are doing.
4. I involve my students in thinking about how they want to learn at school.
5. I give my students the opportunity to decide on their learning objectives.
6. I ask my students to indicate what went well and what went badly concerning their assignments.
7. I encourage students to reflect upon their learning processes and how to improve
8. I inform my students on their strong points concerning learning.
9. I inform my students on their weak points concerning learning.
10. I encourage my students to improve on their learning processes.
11. I give students guidance and assistance in their learning.
12. I discuss assignments with my students to help them understand the content better.
13. I discuss with my students the progress they have made.
14. After an assessment, I inform my students on how to improve their weak points.
15. I discuss with my students how to utilize their strengths to improve on their
16. Together with my students, I consider ways on how to improve on their weak points.
17. I adjust my instruction whenever I notice that my students do not understand a topic.
18. I provide my students with guidance to help them gain an understanding of the content taught.
19. During my class, students are given the opportunity to show what they have learned.
20. I ask questions in a way my students understand.
21. By asking questions during class, I help my students gain an understanding of the content taught.
22. I am open to student contributions in my class.
23. I allow my students to ask each other questions during class.
24. I ensure that my students know what areas they need to work on to improve their results.
25. I give my students opportunities to ask questions.
26. My students know what the evaluation criteria for their work are.
27. I ensure that my students know what they can learn from their assignments.
28. I can recognize when my students reach their learning goals.

Note. Adopted from, Validation of Assessment for Learning Questionnaires for teachers and students by Pat-El et al., 2013, *British Journal of Educational Psychology*, *83* (1), 98–113, adopted with permission.

Appendix B1: Confirmatory Factor Analysis for a two-Factor TAFL-Q (N=180)



Appendix B2: Range of the Item-Total Correlations for the 19 Items six-Factor TAFL-Q (N=213)

Item-Total Statistics

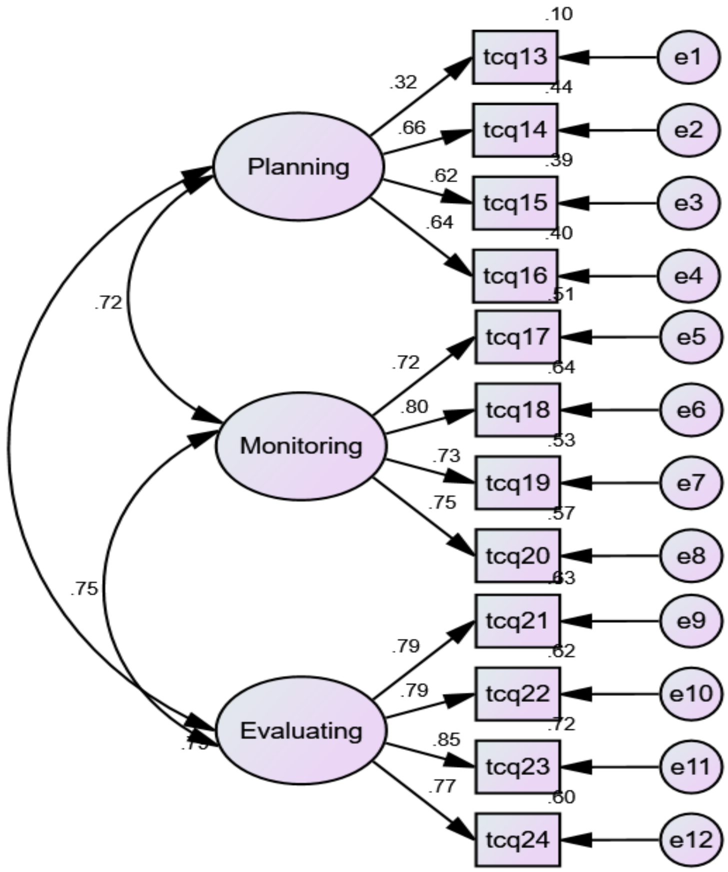
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
After a test, i discuss the answer given with each student	53.3568	93.391	.399	.868
I involve my students in thinking about how they want to learn at school	53.3897	91.013	.485	.864
I give my students the opportunity to decide on their learning objectives	54.1502	92.062	.380	.870
I inform my students on their strong points concerning learning	53.1080	91.748	.567	.861
I inform my students on their weak points concerning learning	53.0798	94.545	.433	.866
I encourage my students to improve on their learning processes	52.8967	96.197	.385	.867
I discuss assignments with my students to help them understand the content better	52.8873	95.893	.412	.866
I discuss with my students the progress they have made	52.9765	95.240	.454	.865
After an assessment, i inform my students on how to improve their weak points	53.1127	91.846	.580	.860
Together with my students, i consider ways on how to improve on their weak points	53.1174	93.019	.594	.860
I adjust my instructions whenever i notice that my students do not understand a topic	53.1221	91.004	.563	.861
I provide my students with guidance to help them gain an understanding of the content taught	53.0610	91.907	.601	.860
I ask questions in a way my students understand	52.8779	95.994	.428	.866
By asking questions during class, I help my students gain an understanding of the content taught	52.7418	96.277	.450	.865
I am open to student contribution in my class	52.7230	95.730	.440	.865
I allow my students to ask each other questions during class	53.3756	93.254	.402	.868
I ensure that my students know what areas they need to work on in order to improve their results	53.1408	93.037	.532	.862
My students know what the evaluation criteria for their work are	53.4789	91.109	.499	.863
I ensure that my students know what they can learn from their assignments	53.2629	91.459	.574	.860

Appendix C: Metacognitive Awareness Inventory for Teachers (MAIT)

1. I am aware of the strengths and weaknesses in my teaching
2. I know what skills are most important to be a good teacher
3. I know what I am expected to teach
4. I have control over how well I teach
5. I try to use teaching techniques that worked in the past
6. I have a specific reason for choosing each teaching technique I use in the class
7. I am aware of what teaching techniques I use while I am teaching
8. I use helpful teaching techniques automatically
9. I use my strengths to compensate for my weaknesses in my teaching
10. I can motivate myself to teach when I need to teach
11. I use different teaching techniques depending on the situation
12. I know when each teaching technique I use will be most effective
13. I pace myself while I am teaching to have enough time
14. I teach my specific goals before I start teaching
15. I ask myself questions about the teaching materials I am going to use
16. I organize my time to best accomplish my teaching goals
17. I ask myself periodically if I meet my teaching goals while I am teaching
18. I find myself assessing how useful my teaching techniques are while I am teaching
19. I check regularly to what extent my students comprehend the topic while I am teaching
20. I ask myself questions about how well I am doing while I am teaching
21. I ask myself how well I have accomplished my teaching goals once I have finished
22. I ask myself if I could have used different techniques after each teaching experience
23. After teaching a point, I ask myself if I would use it more effectively next time
24. I ask myself if I have considered all possible techniques after teaching a point

Note. Adopted from, Metacognitive awareness inventory for teachers (MAIT), by C. Balcikanli, 2011, *Electronic Journal of Research in Educational Psychology, 9* (3), 1309–1332, adopted with permission.

Appendix C1 Confirmatory Factor Analysis for MAIT (N=180)



Appendix C2: Range of the Item-Total Correlations for the MAIT scale (N=213)

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1. I am aware of the strengths and weaknesses in my teaching	71.2676	146.310	.266	.900
2. I know what skills are most important in order to be a good teacher	71.2347	146.539	.278	.900
3. I know what I am expected to teach	71.1362	147.269	.263	.900
4. I have control over how well I teach	71.0610	144.558	.384	.898
5. I try to use teaching techniques that worked in the past	71.3944	145.315	.302	.900
6. I have a specific reason for choosing each teaching technique I use in the class	71.2207	142.475	.483	.896
7. I am aware of what teaching techniques I use while I am teaching	71.1549	144.386	.399	.898
8. I use helpful teaching techniques automatically	71.3474	142.615	.428	.897
9. I use my strengths to compensate for my weaknesses in my teaching	71.3239	142.494	.482	.896
10. I can motivate myself to teach when I really need to teach	71.3803	141.595	.459	.897
11. I use different teaching techniques depending on the situation	71.3944	139.636	.531	.895
12. I know when each teaching technique I use will be most effective	71.4789	140.911	.527	.895
13. I pace myself while I am teaching in order to have enough time	71.5869	143.253	.376	.899
14. I teach my specific goals before I start teaching	72.0000	136.311	.498	.896
15. I ask myself questions about the teaching materials I am going to use	71.5117	139.647	.511	.896
16. I organize my time to best accomplish my teaching goals	71.4648	140.373	.542	.895
17. I ask myself periodically if I meet my teaching goals while I am teaching	71.5962	135.931	.660	.892
18. I find myself assessing how useful my teaching techniques are while I am teaching	71.5634	136.398	.655	.892
19. I check regularly to what extent my students comprehend the topic while I am teaching	71.4789	138.194	.605	.894
20. I ask myself questions about how well I am doing while I am teaching	71.5258	136.581	.601	.893
21. I ask myself how well I have accomplished my teaching goals once I have finished	71.6901	133.922	.657	.892
22. I ask myself if I could have used different techniques after each teaching experience	71.9577	134.682	.599	.893
23. After teaching a point, I ask myself if I would use it more effectively next time	71.9484	132.125	.643	.892
24. I ask myself if I have considered all possible techniques after teaching a point	71.8873	134.176	.587	.894

Appendix D: Interview Schedule

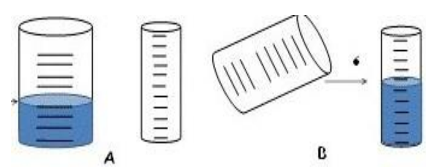
1. What would you consider as your strengths and weaknesses in teaching mathematics?
2. What do you consider before deciding on a particular method to use in your teaching?
3. How do you make use of your strengths and weaknesses?
4. How often do you prepare your lesson plans?
5. How often do you evaluate your teaching goals at the end of each lesson?
6. How often do you question whether you are meeting your teaching objectives while teaching?
7. To what extent did your teaching approach predict students' achievement at the end of term one exams?
8. Do you consider yourself a metacognitive teacher? Briefly explain?

Appendix E: Proportional Reasoning Test-PRT (30 Marks)
Time: 1 hour

Instructions

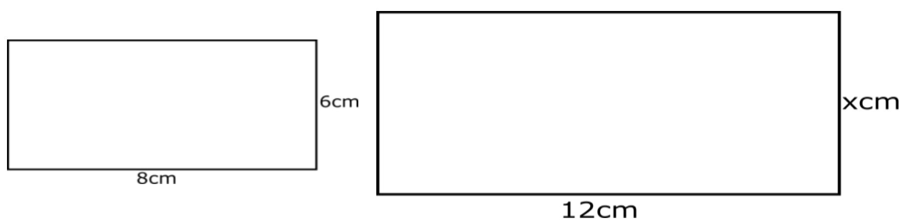
- You are required to attempt all the questions in this test
- Read the questions carefully before attempting them
- The marks are equally distributed (3 marks for each question)
- Show all the work in the space provided after each question.
- Correct procedure/ working will be awarded even if the answer is wrong

1. Last week, Mary answered 24 out of 30 questions correctly in an exam. This week, she answered 20 out of 24 questions correctly in another exam. For which exam did Mary have better results? Explain your answer.
2. Nafula bought 3 lollypops at 12 shillings and Atieno bought five lollypops at 20 shillings. Who used less amount of money? Explain your answer.
3. How many glasses of orange juice can you make with 12 cups of water if eight 8 cups of water can produce 14 glasses of orange juice? Show your calculations.
4. The diagrams below show two tins of different sizes but marked with the same scale on each of them. Oil is poured into the broad tin until it reaches the fourth mark. When the same oil is poured into the small tin as demonstrated in diagram B, it rises to the sixth mark. If both tins are emptied and oil is poured into the broad tin until it reaches the sixth mark, to what level can this oil rise if it is poured into the small cylinder?



5. A group of 7 girls shares 3 chapatis equally and another group of 9 boys shares 4 chapatis equally. Who gets a bigger size of chapati, a girl or a boy? Explain your answer.
6. Mary has the option of working in Mombasa or Nairobi. She discovered that the workers in Mombasa work 8 hours per day and receive Ksh 24 000 every 15 days while those in Nairobi work 6 hours per day and receive Ksh 20 000 every 12 days. If Mary decides to work for 20 days, which job option will be best for her? Explain your answer.

7. Your father decides to give a piece of land as an inheritance to your three brothers Joe, Alex, and Peter in the ratio 4:5:3. Peter being the firstborn feels he has already accumulated enough wealth and therefore decides to share his portion equally with Joe and Alex. What will be the ratio of Joe's share to Alex's share? Show your work.
8. In a mixture of 60 litres, the ratio of orange concentrate to water is 7:5. If the principal of a school wants to make orange juice for the students in the ratio of 3:2, how many liters of water should he add to the mixture? Show your work.
9. The figures below show two similar rectangles. The height of the first rectangle is 6cm and the width is 8cm. The width of the second rectangle is 12cm. Explain how you would find the height of the larger rectangle.



10. The heights of two trees taken three years ago were eight feet for the tree (I) and ten feet for the tree (II). When the heights were taken today, tree (I) was 14 feet and tree (II) was 16 feet. Which of the trees increased most over the past three years?

Appendix F: Junior Metacognitive Awareness Inventory (Jr. MAI)

	Item	NA	R	S	O	A
1	I know when I understand something					
2	I can make myself learn when I need to					
3	I try to use ways of studying that has worked for me before					
4	I know what the teacher expects me to learn					
5	I learn best when I already know something about the topic					
6	I draw pictures or diagrams to help me understand while learning					
7	When I am done with my schoolwork, I ask myself if I learned what I wanted to learn					
8	I think of several ways to solve a problem and then choose the best one.					
9	I think about what I need to learn before I start working.					
10	I ask myself how well I am doing while I am learning something new.					
11	I pay attention to important information					
12	I learn more when I am interested in the topic					
13	I use my learning strengths to make up for my weaknesses					
14	I use different learning strategies depending on the task					
15	I occasionally check to make sure I'll get my work done on time					
16	I sometimes use learning strategies without thinking					
17	I ask myself if there was an easier way to do things after I finish a task					
18	I decide what I need to get done before I start a task					

Adapted from (Sperling et al., 2002 p.26)

Appendix G: Letter from the County Director of Education

REPUBLIC OF KENYA

MINISTRY OF EDUCATION, SCIENCE AND TECHNOLOGY
State Department of Education – Bungoma County

When Replying please quote
e-mail: bungomacde@gmail.com

County Director of Education
P.O. Box 1620-50200
BUNGOMA

Ref No: BCE/DE/19/VOL.II1/56

Date: 8TH JULY, 2019

TO WHOM IT MAY CONCERN

RE: AUTHORITY TO CARRY OUT RESEARCH – RUTH NANJEKHO WAFUBWA – REF NO. 18/2019

The bearer of this letter Ruth Nanjekho Wafubwa of University of Szeged has been authorized to carry out research on ***"Impact of formative assessment on mathematics student's achievement in Kenyan secondary schools"*** a period ending ***August 2019.***

Kindly accord her the necessary assistance

CALLEB OMONDI
For: COUNTY DIRECTOR OF EDUCATION
BUNGOMA COUNTY

Appendix H: Certificate of Ethical Approval

NACOSTI ACCREDITED

ERC/PhD/005/2020

Pwani UNIVERSITY

ETHICS REVIEW COMMITTEE

ACCREDITTED BY THE NATIONAL COMMISSION FOR SCIENCE, TECHNOLOGY AND INNOVATION (NACOSTI, KENYA)

CERTIFICATE OF ETHICAL APPROVAL

THIS IS TO CERTIFY THAT THE PROPOSAL SUBMITTED BY:

RUTH N. WAFUBWA

REFERENCE NO:
ERC/PhD/005/2020

ENTITLED:

Impact of formative assessment on student's achievement in mathematics in Bungoma County, Kenya

TO BE UNDERTAKEN AT:
BUNGOMA COUNTY, KENYA

FOR THE PERIOD
FROM: 03/20/2020 TO: 03/19/2021

HAS BEEN **APPROVED** BY THE ETHICS REVIEW COMMITTEE

AT ITS SITTING HELD AT PWANI UNIVERSITY, KENYA

ON THE 03/19/2020

CHAIRMAN	SECRETARY	LAY MEMEBER

Pwani UNIVERSITY

Ethics Review Committee,
Pwani University, www.pu.ac.ke, email: t.rewe@pu.ac.ke, j.ndiso@pu.ac.ke tell: 0719 182218, 0720785791
The ERC, Giving Integrity to Research for Sustainable Development

Appendix I: Research License

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **984696**

Date of Issue: **01/April/2020**

RESEARCH LICENSE

This is to Certify that Miss.. Ruth Nanjekho Wafubwa of University of Szeged, has been licensed to conduct research in Bungoma on the topic: Impact of formative assessment on mathematics students' achievement in Kenyan secondary schools for the period ending : 01/April/2021.

License No: **NACOSTI/P/20/4723**

984696

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

Appendix J: Letter of Consent to Teachers

Study Title: Formative Assessment and Metacognition in the Teaching and Learning of Mathematics In Kenyan Secondary Schools: Teachers' Perceptions, Students' Achievement, and Students' Metacognition

Dear Teacher,

I am a Ph.D. student at the Institute of Educational Sciences, University of Szeged, Hungary. I plan to carry out the above study on formative assessment. This study will involve secondary school students and teachers of Bungoma County. The study aims to evaluate the preparedness of mathematics teachers in implementing competency-based assessments and to suggest the best formative assessment practices that can be used in mathematics classrooms. It's hoped that the findings of this study will contribute knowledge towards the development of an assessment framework in mathematics instruction for secondary students in Kenya. This study will be carried out in the context of the competence-based curriculum (2.6.3.3) in investigating formative assessment methods that teachers of mathematics can employ to promote the key competencies in mathematics.

All data collected in this study will be kept under lock and key. Responses will be kept confidential, and the use of pseudonyms enforced. After the study, students, teachers, and other stakeholders are entitled to free access to the published reports.

CONSENT

I………………………………………. agree to participate in the study.

Signature: ……………………………..   Date: ………………………………………

Thank you very much for your support and consideration.

Ruth N Wafubwa

Email: ruthnanje@gmail.com **Cell: +254724737497/+36 204087958**

Relevant Publications

Journal Articles

Wafubwa, R. N., Csíkos, C., & Opoku-Sarkodie, R. (in press). In-service mathematics teachers' conception and perceptions of metacognition in their teaching experience. *SN Social Sciences*.

Wafubwa, R. N., & Csíkos, C. (2022). Impact of formative assessment instructional approach on students' mathematics achievement and their metacognitive awareness. *International Journal of Instruction, 15*(2), 119-138.

Wafubwa, R. N. (2021). Challenges of teaching and assessing the 21st-century competencies in Africa: a focus on the Kenyan new curriculum of basic education. *East African Journal of Education Studies*, *3*(1), 96-105. https://doi.org/10.37284/eajes.3.1.332

Wafubwa, R. N. (2020). Role of formative assessment in improving students' motivation, engagement, and achievement: a Systematic review of literature. *International Journal of Assessment and Evaluation 28* (1), 17-31. https://doi.org/10.18848/2327-7920/CGP/v28i01/17-31

Wafubwa, R. N., & Csíkos, C. (2021). Formative assessment as a predictor of mathematics teachers' levels of metacognitive regulation. *International Journal of Instruction*, *14*(1), 983-998. https://doi.org/10.29333/iji.2021.14158a

Wafubwa, R. N., & Csíkos, C. (2021). Assessing the suitability of the adapted teacher assessment for learning questionnaire in the Kenyan context. *African Journal of Research in Mathematics, Science and Technology Education*, 1-13. *https://doi.org/10.1080/18117295.2021.1899490*

Wafubwa, R. N., & Ochieng, P. O. (2021). Students' perception of teachers' use of formative assessment strategies in mathematics classrooms. *Elementary Education Online*, *20*(2), 123-132. http://ilkogretim-online.org/

Wafubwa, R. N., & Obuba, E. (2015). Influence of strengthening mathematics and science in secondary education (SMASSE) in-service education and training (INSET) on the attitude of students towards mathematics performance in public secondary schools of Rangwe Division, Homa-Bay Sub-County-Kenya. *Journal of Education and Practice 6* (26), 57-62. https://eric.ed.gov/?id=EJ1077453

Wafubwa, R. N., Opoku-Sarkodie, R., & Csíkos, C. (2020). Construction and validation of a test for measuring students' proportional reasoning on rates, ratios, & proportions. *International Journal of Pedagogy, Policy and ICT in Education*, (*8*), 53-76. https://www.ajol.info/index.php/ijp/article/view/198696

Conference Papers

Wafubwa, R. N. (2021). Formative assessment as an instructional approach in mathematics classrooms: impact on students' achievement and metacognition. In Molnár, Gyöngyvér; Tóth, Edit (ed.) *The answers of education to the challenges of the future: XXI. National Educational Science Conference Szeged, November 18-20, 2021. Program, lecture summaries. Szeged, Hungary*: Scientific Committee for Pedagogy of the Hungarian Academy of Sciences , Institute of Education , University of Szeged (2021) 690 p. pp. 288-288. , 1 p.

Wafubwa, R. N. & Purevjav, D. (2021). Kenyan secondary school science and mathematics teachers' professional development needs. In Molnár, Gyöngyvér; Tóth, Edit (ed.) *The answers of education to the challenges of the future: XXI. National Educational Science Conference Szeged, November 18-20, 2021. Program, lecture summaries. Szeged, Hungary:* Scientific Committee for Pedagogy of the Hungarian Academy of Sciences , Institute of Education, University of Szeged (2021) 690 p. pp. 687-687. , 1 p.

Wafubwa, R. N. (2021). Mathematics teachers' perceived levels of metacognition and students' achievement in mathematics. In JURE 2021: *Education and Citizenship: Learning and Instruction and Shaping of Futures*; pp.1-2

Wafubwa, R. N. (2021). Construction and validation of a test to measure students' proportional reasoning in Mathematics. In *EARLI 2021: Education and Citizenship: Learning and Instruction and Shaping of Futures*; pp. 261

Wafubwa, R. N. (2020). Perceptions of formative assessment in mathematics classrooms: a case of Kenyan secondary school students. In J. Krisztián (ed) *Educational Science - Answers to the Challenges of the New Millennium: 13th International Conference on Education and Practice* [13th Training and Practice International Conference on Educational Science]: Program and Abstracts: Program and Abstracts Gödöllő, Hungary: SZIE (2020) pp. 88

Wafubwa, R. N. Formative assessment as a tool for improving students' motivation and engagement in secondary schools. In V. Aranka, A. Helga, M-K, & Zsófia (eds.) *Educational Science - Horizons and Dialogues. Abstract volume: XIX. National Conference on Education Pécs, Hungary: Pedagogical Scientific Committee of the Hungarian Academy of Sciences*, University of Pécs, Faculty of Arts, Institute of Education (2019) pp. 586

Wafubwa, R. N. (2019). Metacognition, self-assessment and students' achievement: Key issues and considerations. In M. E. Katalin & D. Katinka (eds.) *PÉK 2019 [CEA 2019] XVII. 17th Conference on Educational Assessment: Program and Abstracts Szeged, Hungary:* University of Szeged, Doctoral School of Education

Declaration

I hereby declare that I am the sole author of this dissertation. To the best of my knowledge, this dissertation contains no material previously published by any other person except where due acknowledgment has been made. This dissertation contains no material which has been accepted as part of the requirements of any other academic degree program.