Formative assessment practices in mathematics education among secondary schools in Tanzania

Florence Kyaruzi

Dissertation zum Erwerb des Doctor of Philosophy (Ph.D.)

am Munich Center of the Learning Sciences

der Ludwig-Maximilians-Universität München

Vorgelegt von

Florence Kyaruzi
aus Bukoba, Tanzania

München, den 15. Mai 2017

"Printed and/or published with the support of the German Academic Exchange Service".

1st Supervisor / Erstgutachter: Prof. Dr. Stefan Ufer

2nd Supervisor / Zweitgutachter: Prof. Dr. Jan-Willem Strijbos

3rd Supervisor / Drittgutachter: Prof. Dr. Gavin Brown

Tag der mündlichen Prüfung: 12.07.2017

Acknowledgements

This dissertation marks the end of a three-year academic journey full of vigorous scientific and personal development. I am grateful to God who accorded me with good health and countless blessings during my Ph.D. studies. Nonetheless, this dissertation would have been impossible without the support and encouragement from many individuals and funding institutions.

I would like to express my sincerest appreciation to my supervisors: Prof. Dr. Stefan Ufer, Prof. Dr. Jan-Willem Strijbos and Prof. Dr. Gavin Brown for the scholarly support, the mentoring and the formative feedback throughout my study. Their expertise, willingness to listen and constructive supervision have made this journey possible. I also thank very much Prof. Dr. Neuhaus Birgit for being a member of my Ph.D. examination committee.

Firstly, Prof. Ufer with his expertise in mathematics education and statistical analyses always showed me the right way and encouraged me to seek positive interpretations in my results. His trust in my research ideas enabled me to continually improve my work.

Secondly, I express my deepest gratitude to Prof. Strijbos, who shaped my research abilities in a positive way. His unique kind feedback from the infancy stage of our collaboration made this dissertation successful. Whenever I had a chance to get five minutes of his time in formal or informal discussions, I was privileged to learn from his expertise in feedback and content analyses techniques.

Thirdly, Prof. Brown joined the project as my international supervisor at the right time. I appreciate the effort of him supervising me so closely, disregarding the physical distance between us. His contribution shaped my knowledge about assessment and structural equation modeling techniques. I am very thankful for his critical eye on the project but also for always doing this with a smile on his face.

I would also like to thank Prof. Dr. Frank Fischer for offering me a great working space at his chair and for giving me the opportunity to get valuable feedback on my presentations in the chair meetings. I benefited greatly from my membership at the chair and the Munich Center of the Learning Sciences (MCLS).

I feel indebted to my sponsors; Tanzania Ministry of Education and Vocational Training (MoEVT) and the Deutscher Akademischer Austausch Dienst (DAAD), and to my employer, Dar es Salaam University College of Education (DUCE) for granting me a study leave and research funds.

I would like to thank all student assistants who acted as second coders of qualitative data and supported me with data collection. I was lucky to have supportive office colleagues at various times: Filitsa, Irina and Ji-hye. Finally, I am thankful to all teachers and students who participated in this study. Without their willing participation in this study, it would have been impossible for me to say anything about ‘Formative assessment practices in mathematics education in Tanzania’.

I thank my brothers and sisters and my in-laws, the family of Mr. Kalumuna Nkarage for being supportive and lovely to my family. Uniquely, I express my sincerest appreciation to my wife Diana for her love, encouragement and especially for happily taking care our son Victor and our daughter Evelyne on my absence. Victor, once you read this dissertation, you will get the answer to your question “Baba (father) what are you doing in Germany?” which you started asking me at the age of two and half years.

Florence Kyaruzi
May, 2017

Dedication

This work is dedicated to the loving memories of my late parents: Evangelina Kokutangilira and Bartazary Kyaruzi for inculcating into me the educational seed during their livelihood.

Table of Contents

Summary

Formative assessment (FA) and Assessment for Learning (AfL) as well as feedback are important practices for student learning (Black & Wiliam, 1998; Hattie & Timperley, 2007). Assessment is supposed to be a significant part of teaching and learning practices in secondary schools in Tanzania. In 1976, Tanzania introduced Continuous Assessment (CA) program in secondary schools to serve the functions of monitoring and scaffolding student learning, and partly contributing to the students' final summative assessment. Consequently, students are continuously assessed and teachers are expected to provide student with feedback about their learning progress. However, feedback is considered to be formative when it is perceived as supportive by the learner (Poulos & Mahony, 2008; Sadler, 1989) and used to improve their learning (Black & Wiliam, 1998). It is not clear how students and teachers perceive school-based assessment practices in mathematics Tanzanian secondary school, notwithstanding that there is a high failure rate in both school-based assessment and external summative assessments.

This appear to contradict Wiliam, Lee, Harrison and Black (2004) who showed that improving formative assessment produces tangible benefits in terms of student performance in externally mandated assessments. This dissertation investigates mathematics teachers' assessment practices and students' perceptions of these practices in mathematics education among secondary schools in Tanzania and how they can be improved. More specifically, student and teacher perceptions of FA and AfL were studied in light of assessment theory (Black & Wiliam, 2009), feedback theories (Hattie & Timperley, 2007; Kluger & DeNisi, 1996) and the cognitive strategy for learning from errors (Heemsoth & Heinze, 2016; Rach, Ufer, & Heinze, 2013). Two empirical studies were conducted as part of this dissertation. The dissertation addressed two general research aims: (1) To investigate Tanzanian secondary school mathematics teachers' perceptions and practices of formative assessment, as well as

students' perceptions of their teachers' practices (Study 1 reported in Chapter 2 and 3). (2) To investigate the impact of an intervention on feedback processes during mathematics education in secondary schools in Tanzania, i.e., feedback provided by the teacher, how this is perceived by students, and whether it is applied by them (Study 2 reported in Chapter 4 and 5).

In Chapter 2, the impact of secondary school students' perceptions of mathematics teachers' FA and AfL practices and feedback delivery on their feedback use, and mathematics performance was investigated. The study sampled of 2767 Form three (Grade 11) students from 48 secondary schools in the Dar es Salaam and Kilimanjaro regions in Tanzania. Surveys were combined with student focus group discussions to measure students' perceptions of FA and AfL (monitoring and scaffolding), perception of the quality of feedback delivery and their feedback use. The data were analyzed using structural equation modeling and content analysis. A four factor structural equation model with a good fit to the data indicated that student perceptions of teacher FA and AfL practices and the quality of feedback delivery strongly predicted students' perception of feedback use. More specifically, students' perceptions of the quality of feedback delivery and their perceptions of teacher scaffolding practices positively predicted their feedback use. However, students' perceptions of teacher monitoring had a negative effect on their self-reported feedback use. Qualitative results illustrate that most students valued their mathematics teacher's assessment practices. These results are consistent with the literature showing that students are likely to utilise teacher feedback when it is considered to be fair (King, Schrodt, & Weisel, 2009; Lizzio & Wilson, 2008), friendly and professionally provided (Brown & Hirschfield, 2008) and demonstrates how to correct mistakes (Shute, 2008).

In Chapter 3, the effect of mathematics teachers' FA and Assessment for Learning (AfL) perceptions and conceptions of assessment on the quality of their feedback practices was investigated. The study was conducted among 48 secondary schools in Tanzania involving fifty-four experienced mathematics teachers teaching Form three (Grade 11). Previously validated surveys were combined with interviews to investigate mathematics teachers' perceptions of FA and AfL (monitoring and scaffolding), conceptions of assessment, and feedback practices. Data were analyzed by structural equation modeling and content analysis techniques. Results from the structural equation model indicate that mathematics teachers' perceptions of FA and AfL (monitoring and scaffolding) and their conceptions of assessment purposes positively predicted the quality of their feedback practices. Interview results illustrate that mathematics teachers reported to use their students' assessment information for both formative and summative purposes. These findings support previous studies that teacher perceptions of assessment influence their actual assessment practices (Fives & Buehl, 2012). It was recommended that interventions for improving the quality of mathematics teacher's feedback practices should focus on conceptions of assessment and perceptions of their own FA and AfL practices.

In Chapter 4, the intervention study aimed at improving feedback provision by mathematics teachers via a feedback training (FBT) and to improve students' perception and use of teachers' feedback on their mathematics tests was conducted. Data was collected from Form three (Grade 11) students ($N = 251$) and their respective eight mathematics teachers. A quasi-experimental pretest, intervention (training) and posttest repeated measure design with control and experimental groups was applied. The study adopted validated scales to measure students' AfL perceptions, perceptions of teacher feedback delivery and perceptions of feedback on their mathematics test. Latent mean analyses in Mplus were used for data analysis. Results showed no significant differences in student perceptions of feedback in terms of affect and willingness to improve between the experimental and control group. Furthermore, within the experimental group analyses showed significant increases in students' perceptions of their teacher's FA and AfL practices (monitoring and scaffolding) and quality of feedback delivery after the intervention. These results replicate previous studies that it is possible to improve teacher assessment practices via interventions (Rach et al, 2013; Van de Pol, Oort, Volman, & Beishuizen, 2014).

In Chapter 5, the dissertation investigates whether a teacher-level intervention on feedback provision in mathematics classrooms supported teacher assessment practices, student perceptions of their teacher practices and student learning in whole class feedback discussions. Data were collected from the same participants as in Chapter 4. The study employed a quasi-experimental pretest, intervention (training) and posttest repeated measures design. To investigate teacher's use of student errors two lessons (pretest, posttest) were videotaped. One teacher of each school was included and half of them received a one-day training on how to apply the process model to learning from errors, and how to provide feedback in line with the Hattie and Timperley (2007) model. Student perceptions of errors and perceived teacher supports in error situations were measured. Latent means results showed that students' perceptions of teacher support in error situations significantly increased for teachers in the experimental group but did not differ for teachers in the control group. Students' perceptions of anxiety in error situations and learning orientation were not affected by the intervention. Exploratory analysis of video case studies illustrated that mathematics teachers in the experimental group appeared to be more error friendly at the posttest, and implemented more aspects of error handling strategies (e.g., citing specific errors made by students in the test, describing why errors occurred) than teachers in the control group. The findings support previous studies that it is possible to improve teacher assessment practices (Van de Pol et al, 2014) via training. However, like in previous studies (Heinze & Reiss, 2007; Rach et al., 2013) improving student use of errors was hard to change by the intervention.

1. General Introduction

Assessment and evaluation involves collecting information about variables of interest for decision making (Brookhart, 2004; Pellegrino, 2014; Popham, 2014). According to Scriven (1967), summative evaluation provided information to judge the overall value of an educational programme and formative evaluation refers to information to improve the targeted program. Extending on Scriven's definition of formative evaluation, Bloom (1969) shifted the initial focus of formative evaluation from 'program evaluation' to 'student evaluation' (Bennett, 2011; Black & Wiliam, 2003). Later on 'formative evaluation' evolved into what is now referred to as 'formative assessment' (FA). Formative Assessment (FA) and Assessment for Learning (AfL) received increased attention over the past three decades (Black & Wiliam, 1998). Sadler (1989) and Black and Wiliam (1998) define formative assessment as assessment for the purpose of learning (AfL). Ginsburg (2009) added to this definition that assessment should not be reserved for an examination of achievement after the teacher has completed instruction, but rather that assessment should be used to gain information that can help the teacher plan effective instruction.

Formative assessment represents a two-way learning process in which the students adapt their learning in response to the information provided by assessments, while the teachers adapt their teaching as well (Pat-El, Tillema, Segers, & Vedder, 2015). FA and AfL serve two functions: 'monitoring' to track student progress and 'scaffolding' to help students improve their learning (Bransford, Brown, & Cocking, 2000; Stiggins, 2005; Pat-El, Tillema, Segers, & Vedder, 2013). Monitoring entails analysing students' learning progress to foster students' self-monitoring and to optimise teaching and learning. Scaffolding involves teachers helping students to improve their learning by controlling elements of the task that are essentially beyond the student's capacity, thus permitting learners to concentrate upon and complete only those elements that are within the student's range of competence (Pat-El et al., 2013; Shepard, Hammerness, Darling-Hammond, Rust, Baratz-Snowden, Gordon, Gutierrez, & Pacheco, 2005; Wood & Ross, 1976).

In their meta-analysis on scaffolding in teacher-student interactions, Van de Pol, Volman and Beishuizen (2010) concluded that scaffolding is effective in promoting students' metacognitive and cognitive activities and providing support for student affect. FA literature provides extensive evidence that, if well implemented and well perceived by students, FA and AfL have the potential to improve student learning (Black & William, 1998, 2009; James & Pedder, 2006; Wiliam, 2011; Wiliam, Lee, Harrison, & Black, 2004). However, the success of FA and AfL also hinges on whether students perceive and utilize the guidance provided by their teacher to improve their learning strategies (Pat-El et al., 2015). Scaffolding support can also prove to be challenging for teachers because it requires diagnosing student needs and sequential fading of the support in accordance to individual development (Puntambekar & Kolodner, 2005).

1.1. Formative feedback process

Students need feedback in order to know 'how close' they are to the learning goal (Shute, 2008). The effectiveness of FA depends on the quality and usefulness of feedback provided to learners (Gronlund & Linn, 1990; Kitta & Tilya, 2010; Popham 2014; Wiggins, 1993). Entwistle (1987) and Briggs, Woodfield, Martin and Swatton (2006) argue that a useful feedback provides precise information on what is wrong and how it can be corrected. Kollar and Fischer (2010) argue that the feedback process involves various activities such as feedback provision by a teacher (or peer), feedback reception by a student and acceptance by students to apply such feedback to improve the quality of their work. Effective feedback should promote self-regulated learning and allow the learner to interact with the feedback to confirm, add to it, overwrite, tune or restructure their previous knowledge (Butler & Winne

1995; Hattie & Timperley, 2007; Jonsson, 2013; Nicol & Macfarlane-Dick, 2006). Moreover, feedback reviews show that not all feedback provided to learners is used by them (Bangert-Drowns, Kulik, Kulik & Morgan, 1991; Butler & Winne, 1995; Jonsson, 2013). Furthermore, Shute (2008) argues that the effectiveness of formative feedback depends on the student's motivation, the student's opportunity to receive timely feedback and the student's means (willingness) to use such feedback. The uptake of feedback is likely positively related to learning when the feedback is relevant and of high quality; and when the students recognize the feedback as such (Sadler, 1989; Black & Wiliam, 2009). Hence, it is important that feedback is perceived as supportive by students.

1.2. Effects of feedback perceptions on student learning

The effectiveness of feedback highly depends on student interpretations or perceptions of the feedback information. For example, Poulos and Mahony (2008) found that perceptions of feedback were related to the meaning students assigned to it, depended on how feedback was delivered, and the degree to which feedback was related to criteria, marks and grades. Students' perceptions of feedback provided by teachers (or peers) play an important role in their learning process (Poulos & Mahony, 2008; Strijbos, Narciss, & Dünnebier, 2010). For example, Strijbos, Pat-El and Narciss (2010) found that students' perception of peer feedback adequacy (fairness, usefulness and acceptance) predicted their willingness to improve and affect. Feedback needs to be perceived well by students because students' positive perceptions of feedback were related to their learning outcome (De Kleijn, Mainhard, Meijer, Brekelmans, & Pilot, 2013). In sum, students' perception of feedback appears to be a key determinant for feedback usefulness or application (Gibbs & Simpson, 2004; King, Schrodt, & Weisel, 2009).

1.3. Challenges for feedback application

In past decades several review studies have reported that not all feedback is used by students to improve their learning (Butler & Winne, 1995; Jonsson, 2013; Kluger & DeNisi, 1996). Gibbs and Simpson (2004) pointed out, for example, that some students threw away the feedback if they disliked the grade, while others were concerned only with the grade and not interested on comments on their work. Several barriers inhibit students' formative use of feedback. Firstly, feedback is likely not to be used by students when it is complex, unclear and contains technical terms (Lipnevich, Berg, & Smith, 2016; Jonsson, 2013; Narciss, 2008). Secondly, feedback is likely not useful when students do not perceive it to be useful for their future tasks (Duncan, Prowse, Wakeman, & Harrison, 2007; Jonsson, 2013; Sadler, 1989), or when there is a mismatch between students' internal feedback and external feedback from teachers (Butler & Winne, 1995). Thirdly, feedback is likely to be ineffective when students perceive that it was delivered unprofessionally (Brown & Hirschfeld, 2008). In fact, Higgins, Hartley and Skelton (2001) suggested that to promote productive use of feedback, teachers should deliver and discuss the feedback with their students in a friendly manner. Hence, student perceptions of feedback and perceptions of teacher feedback practices (including feedback delivery) appear essential for the effective use of feedback. The next section will discuss feedback models with a specific emphasis on characteristics of feedback that is likely to increase student learning.

1.4. Theoretical considerations on feedback

Peterson and Irving (2008) point out that, in educational settings, research has traditionally focused on three external types of feedback: outcome feedback or knowledge of results, corrective feedback that aims at providing the answers and process feedback which provides explicit information for adapting study strategies. However, several feedback models propose a more detailed conceptualization of effective feedback of which Kluger and DeNisi's (1996) and Hattie and Timperley's (2007) models are given detailed attention in the context of this dissertation. The Hattie and Timperley (2007) feedback model provides guidelines for delivering feedback at four levels (task, process, self-regulation, and self), while the 'Feedback Intervention Theory' (FIT) (Kluger & DeNisi, 1996) provides guidelines on how feedback influences performance based on the feedback characteristics (e.g., praise), the learning task (e.g., difficulty), and the student (e.g., level of self-efficacy). The next section discusses Kluger and DeNisi's FIT and the Hattie and Timperley model in more detail.

1.4.1. The Feedback Intervention Theory (FIT)

It is generally claimed that effective feedback should aim at reducing the gap between students' current and desired performance (Hattie & Timperley, 2007; Sadler, 1989). In their Feedback Intervention Theory (FIT), Kluger and DeNisi (1996) propose that the impact of corrective feedback on students' performance is not always positive. Kluger and DeNisi (1996) postulate four ways by which student can respond to feedback, some of which are maladaptive strategies. First, when the feedback intervention is negative (i.e., signals that the performance is below the goal/standard) students tend to increase their efforts, but only if goal is clear, high commitment is secured, and when belief in eventual success is high (Bandura, 1989; Bandura & Cervone, 1983). Second, students may abandon the standard, when they perceive a low likelihood of their actions eliminating the discrepancy between their current state and the goal/standard. This is what Eccles and Wigfield (2002) conceptualise as *feedback cost*; i.e. the negative aspects of engaging in the task of closing the gap such as performance anxiety, and fear of failure and success. Thirdly, students may change the standard in a detrimental way such as lowering the standard. Fourthly, students may reject the feedback. In summary, the FIT theory stipulates that behaviour is regulated by a comparison between the feedback and goal/standard, and that the discrepancy between them may lead to various coping strategies.

Therefore, the effectiveness of feedback depends on the kind of coping strategy a learner will adopt and that learner's perception of the feedback. The FIT indicates that the impact of feedback on student learning is probabilistic as feedback provision does not necessarily result into effective learning. Due to the uncertainty as to whether the feedback will have a positive effect on performance, Kluger and DeNisi (1996) further propose three levels at which feedback interventions operate: task learning (details of the focal task), task motivation (focal task processes), and meta-task level (involving self or personal processes). Kluger and DeNisi (1996) concluded that feedback is most effective when specific goals are set and when it focuses on the details of the focal task towards the desired learning outcome instead of focusing on the self or personal level.

1.4.2. The Hattie and Timperley feedback model

Although feedback should aim at reducing discrepancies between student's current performance and a desired goal (Hattie & Timperley, 2007; Sadler, 1989), reduction of discrepancies is not a straightforward task. The Hattie and Timperley (2007) feedback model proposes that reduction of discrepancies should involve both the teacher and students in appropriate actions. Teachers are expected to support students by providing appropriate, challenging and specific goals, and assisting students to reach goals through effective learning strategies and feedback. Students are expected to increase effort and employ more effective learning strategies, but may abandon or lower their goal(s) – especially when belief in eventual success is low. The Hattie and Timperley (2007) model proposes three key questions by which teachers and students can monitor the reduction of the discrepancy between current state of learning and the desired goal (standard). These questions prompt teachers and students to reflect on their learning process: (1) *Where am I going?* (2) *How am I going?* and (3) *Where to next?* Feedback in FA and AfL needs to go beyond the right or wrong notion of feedback, but instead inform students on their current state and how further progress can be achieved (Shute, 2008).

Based on the Hattie and Timperley (2007) model, feedback provided to students can focus on four levels: task, process, self-regulation, and self. Feedback at the task level includes descriptions that inform the student how well a task was accomplished or performed and distinguishes mistakes from correct answers (Geister, Guido, & Konradt, 2006). Feedback at the process level focuses on the learning processes and strategies that can be used to solve the task. Feedback at the self-regulation level addresses the way students monitor, direct, and regulate actions towards the learning goal. Feedback at the self-level entails remarks that are directed to the self or person, mainly for encouragement. Among the four feedback levels, feedback at the self-level is considered the least effective. Feedback at the process and self-regulation level can be especially beneficial for task that require deep processing and/or mastery, whereas feedback at the task level feedback is considered to beneficial when the task information can be used to improve strategy processing (Hattie & Timperley, 2007).

The FIT and the Hattie and Timperley models both stress that the impact of feedback on student learning depends on the feedback level or processes triggered by the feedback content, as well as students' perceptions of the feedback. Various characteristics of feedback content might have differential impact on student learning. For example, Lipnevich and Smith (2009) showed that detailed feedback had a stronger effect on student writing scores than a grade. Furthermore, Shute (2008) argues that effective feedback should be elaborated, consisting of explanations underpinning the feedback in terms of the what, how and why a students' response can be improved. In this dissertation, the Hattie and Timperley feedback model was used for teacher feedback training in Chapter 4 and 5, whereas the feedback intervention theory is used in Chapter 6 as a framework to interpret the results of the studies reported in this dissertation.

1.5. Student and teacher perceptions of assessment

There is no agreement on the distinction between perception, conception, belief and attitude and for that reason we distinguish terms perceptions of assessment and conceptions of assessment in this dissertation at the content level. For example, teacher perceptions of FA and AfL are concerned with the way teachers evaluate their own assessment practices to perform the core functions of monitoring and scaffolding student learning (e.g., "I adjust my instruction whenever I notice that my students do not understand a topic"). Meanwhile, conceptions of assessment refer to what teachers consider to be the purposes of assessment (e.g., "Assessment improves learning"). In this dissertation, student and teacher perceptions of FA and AfL were investigated in Chapter 2 and 3, whereas teacher conceptions of assessment were only studied in Chapter 3. Teachers' and students' perceptions of FA and AfL, as well as their perceptions of the current assessment practice (Fun, 2005; Gibbs & Simpson, 2003; MacLellan, 2001) play a crucial role in teaching and learning processes. Pat-El et al. (2013) point out that, due to their different roles and expertise, students and teachers may differ in their perceptions of assessment and such misalignment may lead to the misinterpretation of assessment information. Teachers and students' perceptions of FA and AfL can influence their assessment practices (Fun, 2005). For example, a study on students' perceptions about evaluation and assessment reported a reciprocal influence between student perception about assessment and their study approaches (Struyven, Dochy, & Janssens, 2005).

However, MacLellan (2001) discussed that in order to facilitate student learning, teachers need to ensure that teachers' and students' perception of assessment goals are in alignment. Pat-El et al. (2015) empirically showed such misalignment between teacher and junior vocational high school students' perceptions of FA and AfL in terms of monitoring and scaffolding practices, and that in particular high teacher efficacy and low student language proficiency contributed to such misalignment. However, in the reviewed literature, no study has systematically investigated the chain of effect from student perceptions of FA and AfL in terms of teacher monitoring and scaffolding practices to student feedback use and mathematics performance. Similarly, no study has systematically investigated the influence of teacher perceptions of monitoring and scaffolding of student learning on their assessment practices such as feedback delivery.

Conceptions are defined by Thompson (1992) as a more general mental structure, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like. Furthermore, conceptions represent different categories of ideas behind teachers' descriptions of how educational practices are experienced (Pratt, 1992). Pajares (1992) argues that teacher's conceptions of teaching, learning, and curricula strongly influence how they teach and what students can learn or achieve. This is supported by the theory of planned behaviour which provides a framework for research into human conceptions (Ajzen, 1991). The theory of planned behaviour postulates that the more positive people's beliefs are about a specific behaviour, the more they believe they can do a specific task, the more they believe it is socially acceptable to do so, and the more likely they will be able to act in accordance with their intention (Ajzen, 1991; Brown, 2008). The conceptions of assessment express strength and directions of agreement towards various purposes of assessment (Brown, 2008).

Teachers’ conceptions of assessment purposes have been extensively researched (Brown, 2004; Brown, Chaudhry & Dhamija, 2015; Brown & Hirschfeld, 2008; Gebril & Brown, 2014; Hirschfeld & Brown, 2008; Peterson & Irving, 2008; Remesal, 2011). More specifically, research in the area of teacher and student conceptions of assessment resulted into a structured understanding of four main conceptions of assessment purposes: assessment improves teaching and learning, assessment is for student accountability, assessment is for school and teacher accountability, and assessment is irrelevant to education. Brown and Hirschfeld (2008) showed for example that the conceptions that assessment is for accountability purposes predicted student reading achievement. Moreover, Brown and Hirschfield (2008) highlighted that assessment is perceived by students to be irrelevant when it is unfair, subjective and when teachers lacks professionalism in scoring student’s assignments. It is important to study teacher assessment conceptions of assessment across cultures because teacher beliefs about assessment are influenced by the demands of the context in which a teacher works (Fives & Buehl, 2012). The impact of Tanzanian mathematics teachers’ conceptions of assessment purposes was investigated in Chapter 3. In addition to student and teacher perceptions of monitoring, scaffolding and feedback practices, FA and AfL also include feedback discussions about student errors in tests and assignments and how these errors can be used to improve performance and student learning.

1.6. Learning from errors in mathematics

Mathematics learning involves students making errors (Wagner, 1981) which can have a formative function if students are supported by appropriate feedback and follow-up instruction (Ingram, Pitt, & Baldry, 2015). Errors in mathematics are defined as a fact or process that does not match a given norm (Oser & Spychiger, 2005). Errors may arise due to incorrect knowledge, application of incorrect procedures, and/or misconceptions. Errors in mathematics act as boundary markers, distinguishing between acceptable and unacceptable practices of doing mathematics (Sfard, 2007). However, errors are negatively perceived by both students and teachers as an unwelcome event, meaning that their effective use in promoting learning is rarely recognized or achieved (Borasi, 1994; Rach et al., 2013) and errors are rarely encouraged in mathematics classes (Heinze, 2005). Oser and Spychiger (2005) developed the error perceptions questionnaire to measure a student's learning orientation (student use of own errors), anxiety or fear in error situations, and perceived teacher support in error situations. Several studies that applied this questionnaire have reported the positive impact of error handling training on teacher's affective and cognitive behaviours (Heinze & Reiss, 2007; Rach et al., 2013) and student affective behaviours (Rach et al., 2013) but not on students' use of their own errors. Moreover, cross-cultural studies showed both differences and similarities in mathematics instruction across countries (Leung, 2005; Santagata, 2004), but none have focused on teaching practices in the context of whole class feedback plenary discussion of student errors in a marked test (e.g., kinds of classroom questioning by the teacher) and none have examined mathematics teaching in Africa, where teaching is heavily didactic and examination-focused. Chapter 5 in this dissertation investigate whether a teacher intervention on error handling strategies can improve teacher's error handling practices and their student's perceptions and use of errors to improve performance and learning. The intervention was guided by the cognitive strategy for learning from errors (Heemsoth & Heinze, 2016; Rach, Ufer, & Heinze, 2013).

1.7. Context of the study

The education system in Tanzania is characterized by high stake examinations which hold long-term implications to students' lives. At the end of each instructional cycle of primary and secondary education levels, students participate in an external summative examination which is centrally administered by the National Examinations Council of Tanzania (NECTA). However, to overcome the overreliance on summative examinations Tanzania introduced in 1976 a Continuous Assessment (CA) program in secondary schools. It was intended to serve as a formative practice in secondary schools and to partly contribute to student's final national examinations (NECTA, 2004; Njabili, 1999). The CA program emphasized that students should be continuously assessed and that the combined result should constitute a student's success or failure (United Republic of Tanzania, 1974). Ottevanger, Akker and Feiter (2007) pointed out that although most Sub-Saharan African countries – including Tanzania – have integrated school-based continuous assessment, teachers lack insight and proficiency in assessment skills (also referred to as 'assessment literacy'). Ottevanger et al. (2007) concluded that testing at the school level was mainly summative and hardly used for instructional purposes or to provide feedback to students. Furthermore, Kitta and Tilya (2010) noted that although several projects have been implemented by the Tanzanian Ministry of Education to support the teaching and learning process, little attention has been paid to supporting teacher assessment practices.

1.8. Rationale for the dissertation

Despite having school-based assessment which contributes to the student's final grade, students consistently underperform in their national mathematics examinations. For example, Basic Education Statistics in Tanzania (BEST) shows that for ten consecutive years (2004-2013) an average of 79% of students failed their mathematics national examinations. This seems to contradict Wiliam, Lee, Harrison and Black (2004) who showed that improving formative assessment produces tangible benefits in terms of student performance in externally mandated assessments. Hence, this raises the question whether the current continuous assessment is as formative and effective as intended. The rationale for this dissertation is twofold: (1) studying the reported advantages of FA and AfL in the Tanzanian context, and (2) to propose potential interventions for improving mathematics education in secondary schools in Tanzania. First of all, the reported advantages of FA and AfL on student learning and available empirical evidence are based on research in developed countries. However, it is important to study assessment practices across various education systems and cultures because governments enforce different policies about assessment that may influence teachers' assessment practices (Kennedy, Chan, & Fok, 2011). In fact, research showing how mathematics teachers enact and students perceive FA and AfL practices in African educational systems – including Tanzania – is scarce. Therefore, this dissertation investigates formative assessment and feedback practices in mathematics education among secondary schools in Tanzania, as well as how these could be improved.

1.9. Structure of the dissertation

This dissertation builds on the Hattie and Timperley (2007) feedback model, the cognitive process model to learning from errors (Heemsoth & Heinze, 2016; Rach, Ufer, & Heinze, 2013) and empirical studies on perceptions of FA and AfL (Pat-El et al., 2013, 2015) and conceptions of assessment (Brown, 2004; 2008) to investigate the relationship between: (1) Teacher assessment perceptions and conceptions → teacher practices, and (2) Teacher practices of FA and AfL → Student perceptions → Feedback use → Learning outcomes. Investigating these relations in the Tanzanian educational system extends the existing evidence on FA and AfL in multiple ways. Firstly, evidence from African countries is scarce in educational research in general and in the field of formative assessment, specifically. Secondly, the Tanzanian educational system with its CA program in place has a certain infrastructure for assessment interventions; however, there are doubts in the literature about the formative nature and effectiveness of this system. Thus, the analysis of the Tanzanian system may allow insights into possible problems of such systems and identify potential interventions. To investigate formative assessment practices in mathematics education among secondary schools in Tanzania, two empirical studies were conducted as part of this dissertation. These studies address two general research aims:

1) To investigate Tanzanian secondary school mathematics teachers' perceptions and practices of formative assessment, as well as students' perceptions of their teachers' practices (Study 1 reported in Chapter 2 and 3).
2) To investigate the impact of an intervention on feedback processes during mathematics education in secondary schools in Tanzania, i.e. feedback provided by the teacher, how this is perceived by students, and whether it is applied by them (Study 2 reported in Chapter 4 and 5).

The first study was conducted among 48 secondary schools, 25 schools in the Dar es Salaam region and in 23 schools in the Kilimanjaro region between September and December, 2014. More specifically the study involved a survey among fifty-four mathematics teachers and their respective 2767 Form three (Grade 11) students. A detailed account of the samples and sampling procedure is provided in the method sections of Chapter 2 and 3. Based on preliminary results of the first study (Kyaruzi, Strijbos, & Ufer, 2015) a second study was conducted, which consisted of an intervention to improve feedback provision practices by mathematics teachers. The second study involved eight secondary schools, eight mathematics teachers and 251 students in the Dar es Salaam region between February, 2016 and May, 2016.

Chapter 2 covers the student data from the first empirical study. The participants are 2767 Form three (Grade 11) students from 48 secondary schools in Dar es Salaam and Kilimanjaro regions. Previously validated surveys were combined with focus group discussions to measure students' perceptions of FA and AfL, perceptions of the quality of teacher feedback delivery, and their own perception of feedback use. Quantitative data were analyzed using structural equation modeling in Mplus, and content analysis was used to analyse focus group discussions. The following three research questions were investigated:

RQ 1. To what extent do students perceive their mathematics teachers' assessment practice as formative in terms of monitoring and scaffolding?

RQ 2. To what extent do students' perceptions of their mathematics teachers' monitoring and scaffolding practices and their perceptions of teachers' feedback delivery predict their feedback use?

RQ 3. To what extent do students' perceptions of their own feedback use predict their mathematics performance?

Chapter 3 covers the teacher data from the first empirical study. The participants are fifty-four experienced mathematics teachers teaching Form three (Grade 11) from 48 secondary schools in the Dar es Salaam and Kilimanjaro regions. Surveys were combined with interviews to measure mathematics teachers' perceptions of FA and AfL, conceptions of assessment purposes, and feedback delivery practices. Survey data were analyzed using structural equation modeling, and content analysis was used to analyse interviews. The following three research questions were investigated:

RQ 1. To what extent do secondary school mathematics teachers in Tanzania perceive their own assessment practice as formative in terms of the monitoring and scaffolding functions?

RQ 2. To what extent do secondary school mathematics teachers' perceptions of their FA and AfL practices (monitoring and scaffolding) and their conceptions about the purposes of assessment (assessment improves learning, school accountability) predict the quality of their feedback practices?

RQ 3. For what purposes do secondary school mathematics teachers typically use students' assessment information (such as student's scores in terminal and mid-term tests)?

Chapter 4 covers student and teacher data from the second study. The participants were 251 Form three (Grade 11) students and their respective eight mathematics teachers. A quasi-experimental pretest, intervention (training) and posttest repeated measures design was used. The study adopted validated scales to measure students' FA and AfL perceptions (monitoring and scaffolding), and perceptions of teacher feedback practices. Latent mean analyses techniques were used in Mplus to analyse the data. The following four research questions were investigated:

RQ 1. How do students perceive teacher feedback on their mathematics test in terms of fairness, usefulness, acceptance, willingness to improve and affect after the intervention?

RQ 2. Are student perceptions of teacher feedback on their mathematics test in terms of fairness, usefulness, acceptance, willingness to improve and affect more positive in students whose teacher received the feedback training than those whose teacher did not?

RQ 3. Do students of mathematics teachers who received the feedback training differ in the degree to which they perceive their teacher's FA and AfL practices in terms of monitoring, scaffolding and feedback delivery from students of teachers who did not receive the feedback training?

RQ 4. Are student perceptions of their feedback use more positive in students whose teacher received the feedback training than those whose teacher did not?

Chapter 5 covers student and teacher data from the second study. Data were collected from the same participants as in Chapter 4. A quasi-experimental pretest, intervention (training) and posttest repeated measures design was used. The study adopted validated scales to measure students' perceptions of errors (anxiety in error, learning orientation and perceived teacher support in errors situations), perceptions of teacher feedback delivery and perceptions of feedback. In addition, mathematics teachers' test feedback discussion was videotaped during the pretest and posttest to investigate the impact of the intervention on use of errors for learning. Latent mean analyses techniques were used in Mplus to analyse the data, and exploratory video case studies were used to illustrate how teachers typically handled student errors. The following three research questions were investigated:

RQ 1. What is the effect of the intervention on students' perception of their teacher's support in error situations?

RQ 2. What is the effect of the intervention on students' perception of their individual use of errors in learning and their anxiety in error situations?

RQ 3. Which practices of dealing with errors can be identified in lessons with a formative whole-class plenary discussion of student performance on a mathematics test?

In the remainder of this dissertation the four empirical studies will be presented in more detail, followed by a general discussion presenting overall conclusions, methodological limitations, implications for theory and practices, and an outlook for future research.

2. Students' AfL perceptions, feedback use and mathematics performance in secondary schools in Tanzania

2.1. Introduction

Formative Assessment (FA) and Assessment for Learning (AfL) received increased attention over the past three decades (Black & Wiliam, 1998). FA and AfL highlight active involvement by students in the assessment process (Black & Wiliam, 2009). The impact of FA and AfL practices depends on whether students perceive and utilize the guidance provided by their teacher to improve their learning strategies. The success of FA and AfL hinges on the student's willingness to engage in appropriate actions to close the gap between their actual performance and the target performance (Sadler, 1989) and whether students perceive and utilize the guidance provided by their teacher to improve their learning strategies (Pat-El, Tillema, Segers, & Vedder, 2015). More precisely, FA and AfL serve two functions: 'monitoring' to track student progress and 'scaffolding' to help students improve their learning (Pat-El, Tillema, Segers, & Vedder, 2013; Pat-El et al., 2015; Stiggins, 2005). FA and AfL literature provides extensive evidence that, if well implemented and well perceived by students, FA and AfL have the potential to improve student learning (Black & William, 1998, 2009; Köller, 2005; Njabili, 1999; Wiliam, 2011; Wiliam & Thompson, 2007), and especially for struggling learners (Black & Wiliam, 1998). For example, Wiliam, Lee, Harrison and Black (2004) showed that improving FA produces tangible benefits in terms of student performance in externally mandated assessments.

Feedback is a key component of FA and AfL (Shute, 2008; Wiliam, 2011) and the added value of FA and AfL – compared to summative assessment – is centred on the quality and usefulness of feedback provided to learners (Gronlund & Linn, 1990; Kitta & Tilya, 2010; Kyaruzi, 2012; Popham, 2014; Wiggins, 1993). Entwistle (1987) and Briggs, Woodfield, Martin and Swatton (2006) maintain that a useful feedback provides precise information on what is wrong and how it can be corrected by using encouragement and fair criticism. To date, it is widely acknowledged that effective feedback should promote self-regulated learning and allow the learner to interact with the feedback to confirm, add to it, overwrite, tune or restructure their previous knowledge (Butler & Winne 1995; Hattie & Timperley, 2007; Jonsson, 2013). Kollar and Fischer (2010) argue that the feedback process involves various activities such as feedback provision by a teacher (or peer), feedback reception by a student and acceptance to apply such feedback to improve their work. Notably, feedback is formative only if the information that is fed back to the learner can be used by the learner to improve his/her performance (Black & Wiliam, 2009).

2.1.1. Perceptions of formative assessment and feedback practices

Unfortunately, studies show that not all feedback provided to learners is used by them to improve their learning (Brown, Peterson, & Yao, 2016; Butler & Winne, 1995; Harris, Brown, & Harnett, 2014; Jonsson, 2013). Barriers for productive feedback use include students’ perceptions that the feedback is not useful to their future tasks, a lack of congruence between students’ preferences for feedback and the feedback provided that was provided to them, and students’ inability to understand the feedback due to technical language (Jonsson, 2013; Narciss, 2008). Therefore, it is important with respect to any assessment process that feedback is carefully provided to learners and perceived as supportive. There is no agreement on the distinction between perception, conception, belief and attitude. For example, Hattie (2015) highlights that while Australian scholars use the term ‘beliefs’, those in the United States commonly use the term ‘epistemology’ and those in Europe the term ‘conception’. In the present study we use “perceptions” as the umbrella term.

The effectiveness of formative feedback depends on student’s motivation, student’s opportunity to receive timely feedback and student’s means to use such feedback (Lipnevich, Berg, & Smith, 2016; Shute, 2008). Furthermore, the uptake of feedback is very likely only positively related to learning when the feedback is relevant and of high quality; and when the student recognizes the feedback as such. Notably, students' perceptions of feedback provided by teachers (or peers) play an important role in students' learning (Poulos & Mahony, 2008; Strijbos, Narciss, & Dünnebier, 2010). For example, King, Schrodt and Weisel (2009) showed that students' perceptions that their teacher feedback is useful was related to feedback retention (use), self-efficacy and academic performance.

From a mathematics education perspective, it is argued that formative assessment occurs naturally in context of good classroom instruction (Ginsburg, 2009). Nevertheless, this is not easy to achieve. For example, in a study on learning from errors, Rach, Ufer and Heinze (2013) showed that even though students valued how their teachers' dealt with errors in the classroom and reported low fear of making errors, many of them did not use errors as a learning opportunity. The same study also showed that it is far from trivial to support teachers in delivering classroom instruction that provides students with cognitive strategies to deal with errors, so as that feedback can be used for learning. Likewise, meta-analyses indicate that substantial positive effects of FA in terms of students' mathematics achievement are not easily achieved (Bennett, 2011; Veldhuis & Van den Heuvel-Panhuizen, 2014).

Although several studies on FA and AfL have investigated the role of teacher and students' assessment perceptions in teaching and learning process (Brown, 2004; Brown, 2013; Brown, Chaudhry, & Dhamija, 2015; Gibbs & Simpson, 2003; MacLellan, 2001; Pat-El et al., 2013; Peterson & Irving, 2008), few studies have reported accounts from African (Gebril & Brown, 2014; Ndalichako, 2015) education systems. Likewise, few studies have investigated the role of perceptions of assessment in mathematics education (Adams & Hsu, 1998; Al Duwairi, 2013). In particular research on students' perception of mathematics teacher's FA and AfL practices from African educational systems and cultural contexts is missing. Therefore, the present study focuses on these practices in the context of mathematics education among secondary schools in Tanzania.

2.1.2. Education system and assessment practices in Tanzania

The education system in Tanzania is based on a “2-7-4-2-3+” schooling structure: 2 years of pre-primary school, 7 years of primary school, 4 years of Ordinary level secondary school (O-level), 2 years of Advanced level secondary school (A-level) and at least 3 years of higher education (Ministry of Education and Vocational Training, MoEVT, 2014). The education system in Tanzania is mainly characterized by high-stake examinations which hold long-term implications to students’ lives. At the end of each instructional cycle of primary and secondary education levels there is an external summative national examination, centrally administered by the National Examinations Council of Tanzania (NECTA) for certification and placement purposes. To overcome the overreliance on summative examinations, Tanzania introduced a Continuous Assessment (CA) program in 1976 in secondary schools, which was envisioned to serve as a formative practice in schools and to partly contribute to student’s final national examinations. In fact, part of the purpose statement for introducing CA in schools emphasized that: “students should be continuously assessed and the combined result is what should constitute a student’s success or failure” (United Republic of Tanzania, 1974, p. 21).

2.1.3. The present study

Despite having CA in Tanzanian secondary schools, which was envisioned to serve as a formative practice, students consistently underperform in mathematics national examinations. The Basic Education Statistics in Tanzania (BEST, 2004-2013) indicate that for ten consecutive years (2004-2013), 79% of secondary schools’ students failed their mathematics national examinations. Students' poor performance raises doubts about the formative effects of mathematics assessment practices in secondary schools. Hence, it is worth studying student perceptions of their mathematics teachers' formative assessment practices within the context of CA, as these might explain inter-individual differences in their mathematics performance (Ginsburg, 2009). Therefore, this study among Tanzanian secondary schools investigates the impact of students' perceptions of mathematics teachers' FA and AfL practices and feedback delivery on their feedback use and mathematics performance. In particular, we seek to answer three research questions:

1) To what extent do students perceive their mathematics teachers' assessment practice as formative in terms of monitoring and scaffolding?
2) To what extent do students' perceptions of their mathematics teachers' monitoring and scaffolding practices and their perceptions of teachers' feedback delivery predict their feedback use?
3) To what extent do students' perceptions of their own feedback use predict their mathematics performance?

2.1.4. Conceptual framework

Based on the theoretical conceptualizations of studies on FA and AfL (Pat-El et al., 2013; Pat-El et al., 2015; Stiggins, 2005) as well as feedback practices (Jonsson, 2013; Narciss, 2008; Shute, 2008) and feedback perceptions (King et al., 2009; Lizzio & Wilson, 2008; Poulos & Mahony, 2008; Strijbos et al., 2010) we hypothesized that: (1) student perceptions of FA and AfL, and feedback delivery, predict feedback use, and (2) student perceptions of feedback use subsequently predict mathematics performance. Figure 1 summarizes the conceptual model. The overall assumption of the model was that perceptions would influence behavior and outcomes (Ajzen, 1991; 2005). We validated the conceptual model using structural equation modeling in Mplus.

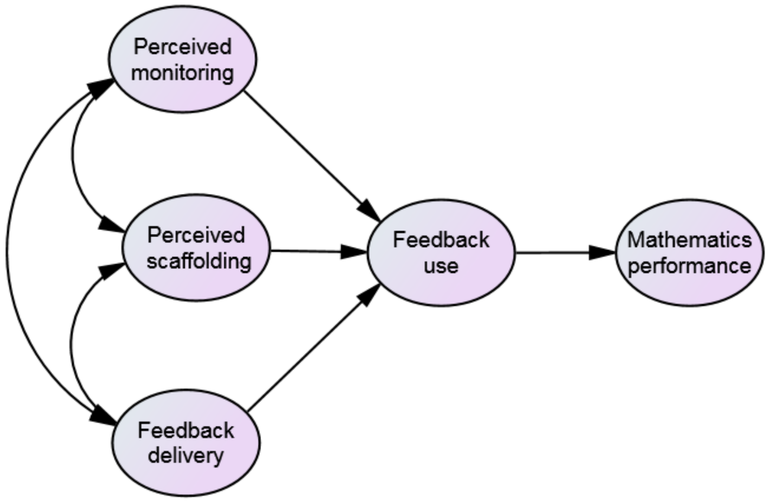

Figure 1. Conceptual model

2.2. Method

2.2.1. Participants

Data were collected in 48 secondary schools in Tanzania: 25 in the mostly urban Dar es Salaam region and 23 in the mostly rural Kilimanjaro region. Based on national educational statistics (MoEVT, 2013) the mean GPA for schools in the sampled regions (M = 4.63, SD = 0.69) did not deviate statistically from the country schools' mean GPA (M = 4.85, SD = 0.70). Three criteria were used to achieve a representative sample: school mathematics performance (high, medium, low) according to (MoEVT, 2013) school ranking, class-size (< 40, ≥ 40), and school-type (private, government). Within the 48 randomly sampled schools there were 2767 Form three (Grade 11) students (53.3% female, 46.2% male, 0.5% missing) from schools varying in mathematics performance (N_{high} = 421, N_{middle} = 997, and N_{low} = 1349). Students had an overall mean age of 16.50 (SD = 1.12) and girls were slightly younger (M = 16.31, SD = 1.04) than boys (M = 16.73, SD = 1.16), $t(2553) = -9.76$, $p < .001$, $d = .38$. The Form three class was selected for this study because it contains more teacher-based assessment practices compared to Form 1 and 2. The sample comprised of 1413 students from 30 privately run secondary schools and 1354 students from 18 government run secondary schools. This sampling process ensured that there were at least 30 groups with at least 30 participants for effective analysis of nested data (Hox, 2010).

2.2.2. Design

A mixed-method research approach with quantitative (survey) and qualitative (student focus group discussions) methods was applied (Dingyloudi & Strijbos, in press; Creswell, 2009). More specifically, we applied a correlational survey design using a two-step process (Anderson & Gerbing, 1988) of first establishing robust measurement models for each construct, followed by a structural equation model linking the constructs as outlined in the conceptual model. We complemented the quantitative analyses of survey data with content analysis of qualitative data from focus group discussions.

2.2.3. Instruments

We adopted previously validated questionnaire scales for the survey, which were adapted to the mathematics context by inserting the word ‘mathematics’ to ensure that students would focus on their mathematics teacher and his/her classroom practices. Questions for the focus group discussions were specifically developed for this study to gain some in-depth understanding of the topics covered by the questionnaire scales. Students self-reported their mathematics performance in Grade 11 terminal examinations, which is a teacher made examination. We changed the response format of the various scales which differed in response options (i.e., 4, 5, 6 or 7) to a common balanced 4-point scale: fully disagree (1), somewhat disagree (2), somewhat agree (3) and fully agree (4). See Chang (1993) for a detailed account on the advantages of a 4-point scale over a 6-point scale. We also refrained from a middle category due to its ambiguous meaning (Dunham & Davison, 1991; Kulas & Stachowski, 2009).

2.2.3.1. Questionnaires

First, to measure student perceptions of their teachers' FA and AfL practice, we used the Student Assessment for Learning Questionnaire (Pat-El et al., 2013) measuring two dimensions: 'perceived monitoring' (16 items) and 'perceived scaffolding' (12 items). Second, we adapted nine items from the 10-item feedback utility subscale of the 'Instructional Feedback Orientation Scale' (King et al., 2009); one item was excluded given a low factor loading (.32). We adapted all six items of student use of feedback subscale of the 'Assessment Experience Questionnaire' (Gibbs & Simpson, 2003) to measure students' perceptions and use of their mathematics teachers' feedback. Third, to measure students' perceptions of the way their mathematics teacher delivered feedback to them; we adapted a five items of the feedback delivery subscale of the 'Feedback Environment Scale' (Steelman, Levy, & Snell, 2004). It is noteworthy that the 'feedback delivery' subscale was below the .70 threshold for Cronbach's alpha, suggesting that robust analyses methods were required to draw valid conclusions. See Appendix A for the detailed questionnaire items. Table 1 summarises the scales that were included in this study, sample items, and the Cronbach's α from the original studies and Cronbach's α for this study.

Table 1. *Scales' sample item texts and Cronbach's α.*

Scale	k	Sample item	Cronbach α	
			Original Study	*Present Study*
Perceived monitoring	16	My mathematics teacher inquires what went well and what went badly in my work.	.89	.93
Perceived scaffolding	12	My mathematics teacher asks questions that help me gain understanding of the subject matter.	.83	.87
Feedback delivery	5	My mathematics teacher generally provides feedback in a thoughtful manner.	.62	.71[a]
Feedback use	15	I use feedback on my mathematics assignments for revising.	.85	.85

Note. k indicates number of items per scale, a - reliability after removing one negatively phrased item

The questionnaire was made available in Swahili and English. English is the language of instruction in secondary schools in Tanzania, however, many students come from Swahili homes and are often more comfortable in that language. Participants had the option of answering the English or Swahili version. In developing the Swahili questionnaire, the English version was translated by the primary researcher into Swahili and back-translated to English by two independent reviewers. Generally, more than three quarters (75.6%, N = 2102) of the students opted for the Swahili version compared to 677 students (24.4%) who answered in English.

2.2.3.2. Focus groups discussions

Six student focus group discussions (FGDs) were randomly sampled from six secondary schools. Each focus group consisted of six students each, resulting in a sub-sample of 36 Form three (Grade 11) students (*Female* = 20, *Male* = 16). The questions explored students' perceptions on type of feedback provided by their teachers, perceptions on how such feedback was provided and opportunities and/or barriers for using the provided feedback. Sample questions include: (1) Are you satisfied with the way feedback is provided to you by your mathematics teacher? Could you please explain your answer?, and (2) To what extent do you think that your mathematics teacher supports you/helps you learn from making errors in class? The average duration of the focus group discussion was 62 minutes. See Appendix B for the detailed focus group discussion questions.

2.2.4. Procedure

The study was conducted with research clearance from the University of Dar es Salaam, and regional and district offices. All participating students signed a consent form. Questionnaires were administered during the mathematics lesson by the researcher and/or with the support of two research assistants. The researcher or assistant demonstrated how to use the rating scales prior to students filling-in the questionnaire. The students needed approximately 15-25 minutes to complete the questionnaire.

Prior to data analysis, data screening was carried out to account for outliers (univariate and bivariate), as well as missing value analysis and recoding all negatively phrased items. Only 27 respondents (approx. 1%) who had more than 10% missing values were eliminated from further analysis, resulting in the sample of 2767 students (see 'participants' section). The remaining missing data were considered to be missing completely at random (MCAR) because Little's MCAR test was not statistically significant (χ^2= 48876.79, *df* = 56611, *p* = 1.00) (Peugh & Enders, 2004). We imputed missing values using the Expectation Maximization method which is considered an effective imputation method when data are MCAR (Musil, Warner, Yobas, & Jones, 2002). Comparison of the estimated statistics with the original variable statistics showed trivial differences, mostly at the third decimal point.

2.2.4.1. Multigroup invariances

Multigroup invariance testing determines whether responses to questionnaires are statistically equivalent between groups. If groups have statistically similar characteristics, then it can be concluded they have been drawn from the same population and thus comparison of scale mean scores can proceed (Wu, Li, & Zumbo, 2007). To demonstrate equivalence, a sequence of tests is applied (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000). First, the pattern of paths from latent traits to and among items has to be identical (i.e., configural equivalence). Once that is established, the equivalence of the regression weights from the latent trait to each item is determined (i.e., metric equivalence). Finally, the equivalence of the starting point or intercept of each regression at the latent scale is determined (i.e., scalar equivalence).

When large samples are involved, estimation of standard errors is very precise leading to detection of statistically significant but trivial differences in parameters between groups. To overcome this, the practical significance of the metric and scalar differences in a means and covariance structure (MACS) was determined using the dMACS procedure which determines the size of differences for each item and then allows calculation of the average factor dMACS (Nye & Drasgow, 2011). dMACS effect size values can be interpreted as trivial when |*d*MACS| < .20 and small when |*d*MACS| < .40 (Cohen, 1987; Hattie, 2009; Rosenthal, Rosnow, & Rubin, 2000).

The measurement invariance tests (see Appendix C) showed that the two language versions were configural and metric invariant, but lacked scalar equivalence. However, the average dMACS for each factor was trivial to small; that is, feedback delivery (*d*MACS = .172), perceived monitoring (*d*MACS = .095), perceived scaffolding (*d*MACS = .074) and feedback use (*d*MACS = .098) (see Appendix D or each items' dMACS value). Thus, given the average *d*MACS effect sizes for all scales across language versions, the differences were trivial. Hence, it was decided to treat the students as members of the same population and combine the data from the two language versions.

2.2.5. Analyses

2.2.5.1. Questionnaire analyses

To account for the students being in classes with a shared teacher, the hypothesized structural equation model (SEM) was estimated in Mplus version 7.31 using complex, clustered estimation options to correct standard errors for the nested nature of the data (Muthen, 1994). The interpretation of model fit was based on the following indicators: root mean squared error of approximation (RMSEA) and standardized root mean residual (SRMR) below .05 and comparative fit index (CFI) and gamma hat values above .95 indicate good fit (Byrne, 2010), while RMSEA and SRMR below .08 and CFI scores above .90 indicate acceptable fit (Hu & Bentler, 1999). Because the Chi-square statistic is overly-sensitive in large sample sizes above 250 (Byrne, 2010), we report multiple fit indices. The CFI and RMSEA are not stable estimators because CFI rewards simple models while RMSEA rewards complex models (Fan & Sivo, 2007). The Gamma hat statistic and SRMR have been shown to be stable estimators (Fan & Sivo, 2007). Importantly, as Steiger (1990) recommends, we report the 90% confidence interval for the RMSEA.

2.2.5.2. Measurement models

We estimated the measurement model for each construct in the conceptual model (Figure 1). The first measurement model constituted two factors measuring student perception of their teacher's FA and AfL practices (i.e., perceived monitoring and perceived scaffolding). This model had good fit (CFI = .95, Gamma hat = .95, RMSEA = .044 [.042, .046], SRMR = .035). Inspection of modification indices showed that two pairs of items had strong inter-correlations and overlapping content. One item from each pair was removed and the re-estimated model had improved fit (CFI = .96, Gamma hat = .96, SRMR = .031, RMSEA = .040 [.038, .042]).

Perception of feedback delivery, with five items, had good fit (CFI = .97, Gamma hat = .99, SRMR = .034, RMSEA = .076 [.062, .091]). Elimination of one item, which was negatively phrased and had a low factor loading (β = .14), improved the model fit (CFI = .999, Gamma hat = 1.00, SRMR = .010, RMSEA = .026 [.000, .052]). The measurement model for student feedback use consisted of two subscales: feedback utility (9 items) and use of feedback (6 items). However, these two scales were highly correlated (r = .978); hence, the items were combined into one new scale entitled 'feedback use'. The combined scale had acceptable fit (CFI = .90, SRMR = .041, RMSEA = .051 [.048, .055]). Eliminating three items with lowest loadings from the latent factor ($\beta < .30$) improved the model fit (CFI = .95, Gamma hat = .99, SRMR = .032, RMSEA = .041 [.037, .046]).

2.2.5.3. Focus group discussion analyses

Content analysis was used to analyze focus group discussions (FGDs). A data-derived coding scheme was developed using about ten percent of the data. Some basic coding rules were formulated to assist the segmentation and coding procedures. The threshold for segmentation agreement was 80% (Strijbos, Martens, Prins, & Jochems, 2006) and a Krippendorff's alpha value of .80 for coding reliability (Krippendorff, 2013). Two independent coders were involved in all data analysis after a 60-80 minutes training on the study rationales and the coding scheme. Two independent coding trials were performed in analyzing focus group discussions. The first coding trial analyzed all questions from two randomly sampled FGDs (one-third of all data). The segmentation agreement was 89-97%. Afterwards the segments from the coder with more segments were independently coded by each coder, resulting in a Krippendorff's alpha of .87 with a lower and upper limit of .80 and .94, respectively. Even though the segmentation agreement and reliability were above the acceptable threshold, a second coding trial was conducted to determine whether those standards were met by chance. The second coding trial analyzed 40% of randomly sampled questions from the remaining four FGDs (one-fourth of all data). The segmentation agreement for the second coding trial was 83-92%. Again segments from the coder with more segments were independently coded by each coder, leading to a Krippendorff's alpha of .88 with a lower and upper limit of .82 and .94, respectively. See Appendix E for a detailed focus group discussion coding scheme.

2.3. Results

The primary purpose of this study was to investigate the impact of students' perceptions of mathematics teachers' FA and AfL practices on their feedback use and mathematics performance among Tanzanian secondary schools. Specific results are reported with respect to research questions.

2.3.1. Student perceptions of their mathematics teacher assessment practices

Responses to four scales (i.e., perceived monitoring, perceived scaffolding, feedback delivery, and feedback use) provided insights as to how students perceived their mathematics teacher's assessment practices (Table 2).

Table 2. *Descriptive statistics per scale and their inter-correlations*

			Scale inter-correlations					
Variables	*M*	*SD*	*I*	*II*	*III*	*IV*	*V*	*ICC*
I. Perceived scaffolding	3.37	0.58	---					.16
II. Perceived monitoring	3.18	0.72	.87**	---				.20
III. Feedback delivery	3.17	0.71	.85**	.84**	---			.16
IV. Feedback use	3.50	0.48	.64**	.55**	.74**	---		.06
V. Mathematics performance	43.13%	18.58%	.16**	.13*	.15**	.16**	---	.25

Note. *N* = 2767; ICC = Intraclass correlation, ***p* < .001, * < .05

Mean scores were above somewhat agree (3.00) suggesting that students perceived their mathematics teachers' assessment practices as formative. The differences in means were generally small to medium with the largest difference being between feedback use over perceived feedback delivery (*d* = .55). However, students' performance in their mathematics terminal examination (*M* = 43.13) was – unsurprisingly given previous reports of student performance in mathematics – below the expected mean score of 50%. The large standard deviation (18.58) is most likely attributable to the systematic collection of data from students in a wide range of schools. Intraclass correlations (ICC) for each scale show how much variation in the score is attributable to membership of the same classroom (Cress, 2008; Field, 2013; Raudenbush & Bryk, 1986). ICC values above .15 suggest that significant variation is due to membership in a class that shares a mathematics teacher (Hox, 2010). ICCs were large for ‘mathematics performance’ (ICC = .25), ‘perceived monitoring’ (ICC = .20), ‘feedback delivery’ (ICC = .16), and ‘perceived scaffolding’ (ICC = .16), and small for ‘feedback use’ (ICC = .06). This provides evidence that students’ judgements depend on the practices of the individual teacher they share. This is taken into account in further analysis by including the grouping effect (see Appendix F).

2.3.2. FA and AfL perceptions, feedback use, and mathematics performance

The initial structural model consisted of 43 manifest variables and had acceptable to good fit (CFI = .92, Gamma hat = .95, SRMR = .038, RMSEA = .037 [.036, .038]). However, fit was improved by excluding items that had poor factor loadings ($\beta < .40$) and items with strong cross-loadings to related items within the same factor. This reduced the number of manifest variables from 43 to 20 and improved the fit (CFI = .98, Gamma hat = .99, SRMR = .025, RMSEA = .030 [.027, .033]). The trimmed model did not change the main conclusions of the full initial model. Appendix G gives items retained in the model and Appendix H shows excluded items.

Factors in the model had statistically significant regression weights on students’ feedback use which subsequently predicted students’ mathematics performance (see Figure 2 and Table 3). All predictor scales of feedback use were highly inter-correlated (i.e., $r > .80$). Although perceived monitoring and scaffolding were highly positively correlated ($r = .86$), as in the original study (Pat-El et al., 2013), each factor had an inverse relation to feedback use (i.e., monitoring negatively predicted use, while scaffolding positively predicted use). Further investigation of these two factors revealed a small suppressor effect by retaining the monitoring factor in the model (ΔR^2 = .012) (Paulhus, Robins, Trzesniewski, & Tracy, 2004) suggesting that the two factors had reciprocal or cooperative suppression (Conger, 1974; Langford, Schwertman, 2001; Lewis & Escobar, 1986; Watson, Clark, Chmielewski, & Kotov, 2013). This means that the structural model enhances the effect of the two scales on feedback use. This type of non-transitive structure is mathematically feasible when the correlation is not close to 1.00 and is best handled in structural equation modeling (Maruyama, 1998). Figure 2 represents a structural equation model for predictors of students' feedback use and mathematics performance.

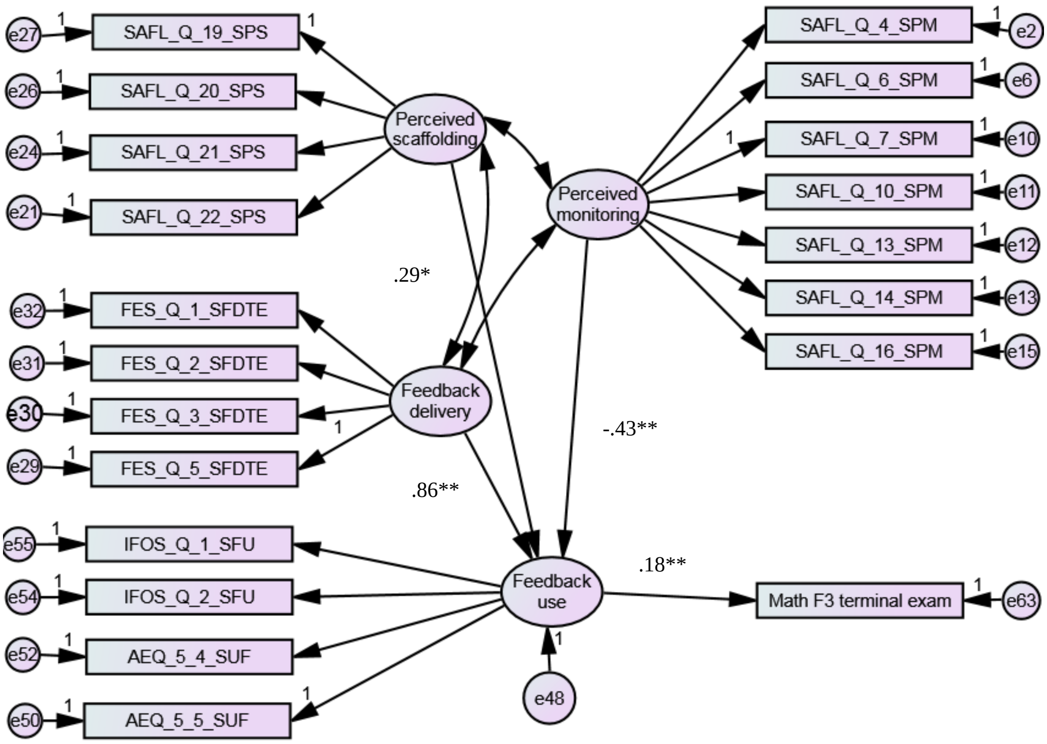

Figure 2. Predictors of students' feedback use and mathematics performance (see Appendix G for the content of the items) $**p < .001$, $*p < .05$

Combined, the four predictor scales explained almost 60% of the variance in students' self-reported feedback use and 3.2% of the variance in students' mathematics performance.

These variances represent large effects ($f^2 > .35$) for feedback use and a small effect ($f^2 < .14$) for mathematics performance (Cohen, 1992). Further mediation analyses indicate that all relations between perceived monitoring, scaffolding, feedback delivery, and mathematics performance were fully mediated via feedback use. Table 3 summarizes the estimates and standard errors for the model in Figure 2.

Table 3. *Estimates and standard errors for the prediction of students' feedback* use and *mathematics performance*

Predictors	**Feedback use**		**Mathematics performance**	
	Estimate	*S.E*	*Estimate*	*S.E*
Perceived monitoring	-.43**	.11	-	-
Perceived scaffolding	.30*	.12	-	-
Feedback delivery	.86**	.10	-	-
Feedback use	-	-	.18**	.03
R^2	.60	.05	.03	.01
f^2	1.48		.03	

Note. $**p < .001$, $*p < .05$

2.3.3. Focus group discussions results

Analysis of focus group discussions showed that majority of students valued their mathematics teachers' assessment and feedback practices, with 19 of 36 (53%) students reporting satisfaction with their mathematics teacher feedback practices because "teacher corrects our mistakes" (10 out of 19) or was "friendly delivered" (9 out of 19). However, one third (12 of 36) of the students were explicitly dissatisfied with their mathematics teachers' feedback practices. Table 4 provides examples of the reasons students gave for either being satisfied or dissatisfied with their mathematics teachers' assessment and feedback practices.

Table 4. *Examples for students' reasons for being satisfied or dissatisfied with their mathematics teachers' assessment and feedback practices*

Perception	**Summary**	**Focus group discussion excerpts**
Satisfaction		
Correction of mistakes	Ten students (28%): because their teacher tells them in the feedback what do to improve their work.	For me, I am satisfied because my teacher when giving feedback he shows me my problem and corrects me (FGD 3). I am satisfied because when we make mistakes she corrects us and advises what to do (FGD 5).
Friendly & convincing	Nine students (25%): because their teacher delivered the feedback in a friendly way.	I am satisfied because she gives us feedback very friendly and convinces us to continue studying and do more exercises (FGD 2).
Dissatisfaction		
No privacy	Two students (6%): because the teacher did not maintain privacy in the feedback they provided.	I am not satisfied. He usually does not provide feedback, when the test is marked he gives that to the class monitor (a student leader) to bring to us (FGD 6).
No correction	Six students (17%): because their teachers did not correct student errors.	I am not satisfied because when he gives the test papers he tells us to do corrections of our errors ourselves (FGD 4).
Reprimanding	Three students (8%): because their teacher reprimands low achievers.	I am not satisfied because if you get low marks you become reprimanded (FGD 6).

Note. FGD = Focus group discussion

Two major reasons were identified for students' satisfaction with teacher feedback practices, i.e. being told how to fix mistakes and receiving feedback in a supportive and friendly manner. This is consistent with recent reviews on students' feedback preferences and feedback use (Jonsson, 2013; Lipnevich et al., 2016). These approaches to providing feedback seem consistent with the idea that feedback should scaffold learning.

In contrast, dissatisfaction arose because teachers did not give the test results confidentially and personally to students and sometimes did not even provide corrections. These results replicate previous studies on the importance of confidentiality (King et al., 2009; Tierney & Koch, 2016) and the desire students have for hints on how to improve their work (Can, 2011; Lizzio & Wilson, 2008; Weaver, 2006). Dissatisfaction also came from students who reported being reprimanded by their teachers for making errors or getting low scores. It would seem that these practices are perceived as monitoring of student competence, rather than helpful scaffolding.

It was further noted that the mathematics teachers of the students included in the FGDs used various methods to inform students how to reduce the discrepancy between the desired goal and their current status. In response to question ‘*Could you please give examples of what does your mathematics teacher do when you make errors in mathematics assignments or tests*?’, the students identified four important teacher practices in situations where they had made an error. First, most students 17 of 36 (47%) perceived their mathematics teacher to provide them with cognitive support showing and/or correcting their errors. Cognitive support involved teacher practices that explicitly showed students how to fix their errors. This is illustrated by the following excerpt: *Our teacher is very friendly to us and when you do some mistakes she is ready to do corrections and teach you even personally. So I can say she is very friendly* (FGD 2).

Secondly, one-third of students 12 of 36 (33%) perceived their teacher to give them affective support in error situations. Affective support involved practices such as encouraging students and reducing fear resulting from making errors. This is illustrated by the following excerpt: *when I do errors in a test my mathematics teacher advises me and tells me to pull up my socks* [work hard] (FGD, 6). Lastly, while three students reported that their teacher ‘does nothing’ when they make an error, two students reported that their mathematics teacher reprimanded them: *When I make errors in a mathematics test as the way my fellow student said, my teacher reprimands me* (FGD 6). In general, the focus group discussions support the relationships revealed by structural equation modeling of the survey data, i.e. that perceived monitoring, scaffolding and teacher feedback delivery predict feedback use. The mechanisms teachers used to provide feedback clearly matter to students' sense of (dis)satisfaction with teacher's feedback practices.

2.4. Discussion

2.4.1. Student perceptions of FA and AfL practices

First of all, both the survey and focus group discussions showed that most secondary school students in our sample valued and considered their teacher's assessment and feedback practices to be formative. Thus, given the sample size, we expect that all Tanzanian secondary school students value their teachers' assessment practices. This replicates results from previous studies (e.g., Pat-El et al., 2015; Weaver, 2006) that students value their teacher feedback. However, according to Rach et al. (2013) student positive perception of teacher assessment practices does not necessarily result into effective performance. Although students perceived their teacher assessment practices as formative, the large ICCs indicated that students' perceptions of their mathematics teacher assessment practices varied considerably between classrooms (Kyaruzi, Strijbos, & Ufer, 2016). This means that the average scores hides considerable variation in experience, and that the overall positive ratings should not be treated as universally applicable. Students' perceptions vary; most likely in accordance with the quality of teacher practices.

2.4.2. Student use of feedback

Student perceptions of teacher FA and AfL practices and feedback delivery were strongly related to students' use of feedback. These findings are in line with previous studies (Jonsson, 2013; Lipnevich et al., 2016; Narciss, 2008; Shute, 2008; Steelman et al., 2004) which showed that feedback is effective in triggering appropriate formative responses in students when it is perceived as supportive. However, it matters to feedback use that the feedback is confidential, helpful, and friendly. The sense that feedback provided scaffolding, rather than monitoring and evaluation, was positively related to feedback use. Practically, these findings suggest that monitoring and scaffolding practices are entangled from the student perspective. Thus, an effective FA and AfL practice would be to increase scaffolding while decreasing the sense that the teacher is monitoring. This signifies that there is a practical challenge in how to increase scaffolding while teachers are also expected to evaluate and monitor student learning.

2.4.3. The negative effect of monitoring

Contrary to expectations, student perceptions of their teachers' monitoring practices had a negative impact on student feedback use. This supports Stiggins (2007) who argued that FA and AfL ought to enhance student learning rather than merely monitor student learning. Hence, this finding call for a closer examination of typical monitoring practices by mathematics teachers. The focus group discussions signalled that not all teacher monitoring practices were productive. For example, practices such as providing general feedback that does not show students how to improve their individual work reduces the perception that feedback is scaffolding. Similarly, practices such reprimanding low achievement increased students' anxiety in error situations and reduced the likelihood of students' using the feedback. These findings support previous studies (Rach et al., 2013; Van de Watering, Gijbels, Dochy, & van der Rijt, 2008) which showed that in mathematics students did not always benefit from teacher assessment practices. In light of these results, it seems that mathematics teachers ought to consider student perceptions of their feedback practices more extensively when implementing FA and AfL in mathematics classes.

2.4.4. The impact of FA and AfL perceptions on learning outcomes

The structural equation model showed that students' perceptions of feedback use had a small, statistically significant relationship to mathematics performance. More specifically, when feedback is perceived as scaffolding and well delivered by teachers it can enhance feedback use which in turn can produce the intended learning outcomes. Thus, in order for students to use feedback they need to perceive that it is helpful. This study suggests that efforts to promote students' feedback use should focus on how teachers deliver feedback and how they focus on scaffolding rather than monitoring.

A crucial finding in the examination-centred Tanzanian education system was that all relations from student perceptions of teacher FA and AfL and feedback delivery practices to mathematics performance were fully mediated via the self-reported feedback use. Increases in performance (albeit slight in this naturalistic experiment) depend on students using feedback and this depends on the quality and nature of teacher feedback practices. Therefore, teachers' efforts towards promoting good mathematics learning and/or performance should promote (a) positive feedback practices and (b) engender positive perceptions of those feedback practices. This conclusion is consistent with the literature which shows that students are likely to utilise teacher feedback when it is considered to be fair (King et al., 2009; Lizzio & Wilson, 2008), friendly and professionally provided (Brown & Hirschfield, 2008) and demonstrates how to correct mistakes (Shute, 2008).

2.4.5. Future research

Even though we triangulated findings from a self-report survey and focus group discussions, the results still need to be taken cautiously. First, the cross-sectional survey design makes it impossible to draw strong causal conclusions. Nevertheless, this study informs the design of longitudinal and intervention studies. Second, while the model depends on student ratings of teacher practices, there is no observational evidence that teachers' practices are consistent with these perceptions. Nevertheless, the current results signal the kinds of practices that need to be systematically observed in future studies. However, our results indicate that student perceived monitoring does not necessarily improve student learning which seems to be an important area for future studies investigating typical mathematics teachers' monitoring practices.

2.4.6. Theoretical contribution

Student perceptions of their teacher FA and AfL practices have a large impact on the effectiveness of instructional processes. In particular, in this contribution we noted that student perceptions of their teachers' FA and AfL practices regulate students' feedback use. These results support previous work that related student perceptions of teacher assessment practices to their performance (Brown & Hirschfield, 2008, 2007), but it further extended these results by showing that feedback use predicted students' mathematics performance. These results support the planned behaviour theory that perceptions (beliefs) influence behaviour which subsequently predicts outcomes (Ajzen, 1991). Thus, students are less likely to apply feedback when they perceive it to be irrelevant and when it lacks information on how to improve. Students' perceptions of their teacher's assessment practices – more specifically perceived scaffolding, and perception of feedback delivery – are thus very important in FA and AfL since feedback is considered formative when it is perceived as such and can be used by learners to improve their learning (Lizzio & Wilson, 2008; Poulos & Mahony, 2008; Black & Wiliam, 2009).

2.4.7. Practical contribution

Findings in this study highlight the need for increasing teacher awareness of the impact of their assessment practices on student learning. Teacher's efforts towards teaching, assessing and providing feedback to their students might be less productive if those practices are not positively perceived by their students and used to improve their learning. However, improving teacher assessment practices is far from trivial as teachers might lack knowledge in the form of concepts of assessment and of effective practices to achieve this. Moreover, large class sizes and a constant pressure to test and examine in preparation for high-stake examinations are barriers for effective formative assessment practices (Kennedy, Chan, & Fok, 2011). We argue for a professional development aimed at improving mathematics teacher's assessment practices to capitalize on the impact of assessment and feedback perceptions on student learning. Such professional development could for example encourage mathematics teachers to consider students' feelings and emotions when providing them with feedback about their mathematics performance. Strategies such as providing students with affective support in error situations (encouragement) paired with cognitive support (correcting mistakes or proving hints) could be a plausible strategy (Rach et al., 2013).

Admittedly, our results might be specific to the Tanzanian context to a certain extent. The fact that there is a large Continuous Assessment (CA) program in place might explain why a negative relation was observed between monitoring practices and student feedback use. Although this is contrary to existing formative assessment literature (Wiliam et al., 2004), it is consistent with the criticism of summative evaluations (Black & Wiliam, 1998). Thus, our results might be transferable to educational systems that apply similar assessment programmes such as India, Egypt, China and Hong Kong (see Brown et al., 2015; Gebril & Brown, 2014). This signifies the necessity for further research to take the specific educational and assessment context into account when conceptualizing teacher professional development on FA and AfL. We recommend that future studies investigate typical monitoring practices by mathematics teachers, beyond the tests from the Continuous Assessment program. Likewise, we recommend professional development aimed at improving teacher feedback and FA and AfL practices, because our results show that promoting positive student perceptions enhances feedback use and subsequently student learning gains.

3. Teacher AfL perceptions and feedback practices in mathematics education among secondary schools in Tanzania

3.1. Introduction

Formative assessment (FA) and Assessment for Learning (AfL) have become widely acknowledged as powerful tools for effective instruction (Black & Wiliam, 1998; Ecclestone, 2012). Assessment as a formal or purposeful attempt to determine students' performance during and/or after a learning phase can be used for improving the teaching and learning process, certifying students, placement of students in tracks, or for curriculum improvement (Pellegrino, 2014; Popham, 2014). Based on the ten principles of AfL first drafted by the Assessment Reform Group (ARG, 2002), the most important practices that guide teachers' implementation of AfL are: rich (classroom) questioning, feedback, peer assessment, self-assessment, and sharing learning goals and criteria of quality (Black & Wiliam, 1998, 2009; James & Pedder, 2006; Popham, 2014). However, recent research has shown that peer and self-assessment can be biased due to students' intra- and interpersonal factors (Brown & Harris, 2013, 2014), feedback is often superficial when delivered and teachers do not always ask good questions (Airasian, 1997; Barnette, Orletsky, & Sattes, 1994) or actively promote feedback seeking (Winstone, Nash, Parker, & Rowntree, 2017). FA and AfL practices serve two core functions namely: monitoring to track student progress and scaffolding to help students improve their learning (Pat-El, Tillema, Segers, & Vedder, 2013; Pat-El, Tillema, Segers, & Vedder, 2015; Stiggins, 2005). However, the nature of FA and AfL and how it leads to improved outcomes have been debated in AfL and FA literature (Bennett, 2011; Black & Wiliam, 2003, 2009; Bloom, 1969; Ginsburg, 2009; Hattie, 2009; Popham, 2014; Scriven, 1967).

3.1.1. What makes assessment formative?

The term ‘formative evaluation’ originates from Scriven’s (1967) distinction of formative evaluation to summative evaluation (Bennett, 2011; Black & Wiliam, 2003). According to Scriven (1967) summative evaluation provides information to judge the overall value of an educational program and formative evaluation refers to information to improve the targeted program. Extending on Scriven’s definition of formative evaluation, Bloom (1969) shifted the initial focus of formative evaluation from ‘program evaluation’ to ‘student evaluation’ (Bennett, 2011; Black & Wiliam, 2003). The purpose of formative evaluation is to provide feedback and corrections at each stage in the teaching and learning process (Bloom, 1969). Later on ‘formative evaluation’ evolved into what is now referred to as ‘formative assessment’ (FA).

An ongoing discussion in the FA literature is whether FA is a process, instrument, interpretation and/or use of assessment information (e.g., Bennett, 2011). In our view formative assessment is the thoughtful application of a purposefully selected methodology or instrument that fosters the interpretation of student performance to inform teachers and students about the learning progress (Bennett, 2011; Popham, 2014). For example, Popham (2014) argues that tests are used to collect information that can lead teachers or students to adjust their actions accordingly, but tests themselves are not formative or summative by default; it depends on the purpose for which the assessment is used. We consider FA and AfL as an instructional strategy concerned with how teachers and students use assessment-elicited evidence – such as student’s test scores or responses in classroom discussions – to regulate their teaching processes and learning tactics. We argue that it is the decision made by the teacher and their students’ regarding the use of assessment-elicited evidence that makes assessment formative or not. Studies show that teachers possess predetermined beliefs about assessment which are rooted in their experience of assessment as assessors or past assessee which can influence the degree to which their assessment practice is more or less formative (Fives & Buehl, 2012; Fulmer, Lee, & Tan, 2015; Pajares, 1992).

3.1.2. Teacher perceptions of FA and AfL practices and conceptions of assessment

There is no agreement on the distinction between perception, conception, belief and attitude. For example, Hattie (2015) highlights that while Australian scholars use the term ‘beliefs’, those in the United States commonly use the term ‘epistemology’ and those in Europe the term ‘conception’. In the present study we do distinguish between the terms perceptions and conceptions, but we do so at the content level.

This means that in our view teacher perceptions of FA and AfL are concerned with the way teachers evaluate their own practices to perform the core functions of monitoring and scaffolding student learning (e.g., “I adjust my instruction whenever I notice that my students do not understand a topic”). Monitoring practices entail analysing student learning progress to foster students’ self-monitoring by finding challenges and opportunities to optimise teaching and learning. Meanwhile, scaffolding involves teachers helping students to improve their learning by controlling elements of the task that are essentially beyond the student’s capacity (Pat-El et al., 2013). Wood, Bruner and Ross (1976) support that scaffolding can be achieved when the teacher control elements of the task that are essentially beyond the learner’s capacity, thus permitting him to concentrate upon and complete only those elements that are within his range of competence. Kyaruzi, Strijbos, & Ufer (2016) showed that students’ perceptions of their mathematics teacher’s monitoring and scaffolding practices were significantly related to their mathematics achievement.

Meanwhile, in our view conceptions of assessments refer to what teachers consider to be the general purposes of assessment (e.g., “Assessment improves learning”). It has been shown that teachers' conceptions about the nature and purposes of assessment strongly influence how they teach and what students can actually learn or achieve (Barnes, Fives, & Dacey, 2015, 2017; Pajares, 1992). It is also proposed by the theory of planned behavior that conceptions (beliefs) and perceptions influence behavior and outcomes (Ajzen, 1991, 2005). Studies in the area of teacher conceptions of assessment have consistently identified four main conceptions about the purposes of assessment: (1) assessment improves teaching and learning, (2) assessment makes students accountable for learning, (3) assessment makes schools and teachers accountable, and (4) assessment is irrelevant to education (Brown, 2002, 2004, 2011).

First, the conception that assessment improves teaching and learning is the central argument for AfL (Black & Wiliam, 1998; Popham, 2014) and requires teachers to use evidence about student learning to support them further (Brown, 2004). Second, the conception that assessment makes schools and teachers accountable, presumes that assessment is used to account for the use of society's resources (Brown, 2004). More specifically, it utilizes assessment results to demonstrate publicly that teachers or schools are doing a good job and may impose consequences for schools or teachers for not reaching required standards (Brookhart, 1994; Firestone, Mayrowetz, & Fairman, 1998). Third, student accountability conception is evidenced by assignment of grades, checking off student performance against criteria, placing students into classes based on performance, as well as various qualification examinations for graduation or placement to higher levels of educational opportunity (Brown, 2004). The conception of assessment as irrelevant**,** regards assessment as a bad practice. In light of this conception, assessment is usually understood as a formal, organized process of evaluating student performance and has no legitimate place in the teaching and learning process (Brown, 2004). For example, Brown, Chaudhry and Dhamija (2015) argues that assessment is perceived as irrelevant when assessment processes are considered inadequate or irrelevant to the teachers' ability to improve student learning (e.g., assessment forces teachers to teach in away against their conception).

Studies have investigated the role of teacher conceptions of assessment and assessment perceptions on teaching and learning process (e.g., Brown, 2004; Brown et al., 2015; Gibbs & Simpson, 2003; MacLellan, 2001; Pat-El et al., 2013). However, few studies have reported accounts from African educational systems (e.g., Gebril & Brown, 2014; Kitta, 2014; Ndalichako, 2015). Furthermore, comparatively few studies provide accounts of teachers' assessment perceptions and their determinants in mathematics education (e.g., Adams & Hsu, 1998; Al Duwairi, 2013; Ginsburg, 2009; Rach, Ufer, & Heinze, 2013). Teacher conceptions are activated by the contextual demands in which the teacher is operating (Fives & Buehl, 2012). Although Tanzania is similar to India and Hong Kong as they all have a strong examinations system, it has unique contextual factors such as large class size (Semali & Mehta, 2012; Sumra & Katabaro, 2014) and challenges related to the language of instruction in secondary schools in Tanzania; i.e., English is used instead of the native Swahili language that students speak in their home environment (Brock-Utne, 2007; Qorro, 2007, 2013).

3.1.3. Teacher feedback practices

The quality of feedback practices (i.e., 'feedback delivery' and 'promote feedback seeking') is essential in regulating student learning, because the more considerate the feedback source is when providing feedback, the more likely an individual is to accept and respond to the feedback provided (Duijnhouwer, Prins, & Stokking, 2012; Fedor, Eder, & Buckley, 1989; Gregory & Levy, 2015; King, Schrodt, & Weisel, 2009; Strijbos, Pat-El, & Narciss, 2010; Steelman, Levy, & Snell, 2004). It is important that the feedback is delivered in an effective way, such as maximising clarity of information (Winstone et al., 2017).

Feedback seeking foster students to identify areas they need help and seek feedback that align with their learning needs (Carless, Salter, Yang, & Lam, 2010). However, students are likely to seek feedback if the social dynamics of the classroom or the teacher promotes feedback seeking behaviours (Neitzel & Davis, 2014). Similarly, there is limited information on how to promote students' behaviour from being passive feedback receivers to active feedback seekers (Winstone et al., 2017). For example, Kyaruzi (2012) noted that feedback practices differ among secondary school mathematics teachers in Tanzania. Teacher feedback delivery and promoting feedback seeking are important practices of a quality feedback.

3.1.4. Education system in Tanzania

The education system in Tanzania is centralized and utilizes one curriculum across the country. The formal education system in Tanzania is hierarchical in nature with a "2-7-4-2-3+" schooling structure: 2-years of pre-primary school, 7 years of primary education, 4 years of ordinary level secondary education (O-level), 2 years of advanced level secondary education (A-level) and at least 3 years for higher education (Ministry of Education and Vocational Training, MoEVT, 2014). The education system in Tanzania is mainly characterized by high-stake examinations which hold long-term implications to students' lives. At the end of each instructional cycle of primary and secondary education levels there is an external summative national examination, centrally administered by the National Examinations Council of Tanzania (NECTA) for certification and placement purposes. Tanzania has nearly achieved 100% enrollment of all primary school aged children (Sumra & Rajani, 2006), which automatically increased student enrollment in secondary education. Recent data indicate that the enrolment in Form 1 (Grade 9) has increased from 147,490 students in 2006 to 514,592 students in 2013 (BEST, 2014, p. 4). This dramatic increase raised the teaching load for science and mathematics teachers who are fewer compared to other instructional domains (Semali & Mehta, 2012).

Despite school improvement programmes such as the Secondary Education Development Programme, (Ministry of Education, 2008), mathematics education in secondary schools in Tanzania has suffered from low passing rate for a long time (BEST, 2014; Kitta, 2004). Several studies have examined general educational challenges in Tanzania that might explain the low passing rate: (a) the transition from Swahili as the language of instruction in primary schools to English in secondary schools (Qorro, 2007; Vuzo, 2007), (b) the large class sizes due to increased student enrollment (BEST, 2014), (c) the curriculum content overload (Kitta & Tilya, 2010), and (d) the lack of in-service teacher professional development (Komba, 2007). Further challenges includes the lack of assessment skills to implement effective school based assessment (Osaki, Hosea, & Ottevanger, 2004).

3.1.5. Assessment in secondary schools in Tanzania

In 1976 Tanzania introduced a *Continuous Assessment* (CA) program in secondary schools to overcome the overreliance on summative examinations and was envisioned to serve as a formative practice in schools and to partly contribute to student's final national examinations. CA emphasizes that students should be continuously assessed and the combined result is what should constitute a student's success or failure (United Republic of Tanzania, 1974). More specifically, CA intended to serve two functions: (a) monitor and scaffold student learning by providing feedback on classroom assignments and tests (formative), and (b) partly contribute to a student's final summative examinations. Teachers as key implementers of CA in schools are supposed to provide feedback to their students and help them bridge the gap between current performance and the desired standard.

3.1.6. The present study

FA and AfL literature provides extensive evidence that, if well implemented by teachers and well perceived by students, FA and AfL have the potential to improve student learning (Black & Wiliam, 1998; Njabili, 1999; Wiliam, Lee, Harrison, & Black, 2004; Wiliam, 2011), and especially for struggling learners (Black & Wiliam, 1998). Although Continuous Assessment (CA) in Tanzania was supposed to be formative, analysis of Basic Education Statistics in Tanzania (BEST, 2004-2013) indicate that for ten consecutive years 79% of secondary schools' students failed their mathematics national examinations. Students' poor performance and the challenges mathematics teachers face in the Tanzanian education system combined raise doubts about the formative effects of secondary school mathematics teachers' assessment practices, such as the quality of feedback practices. The present study investigates the effect of mathematics teachers' FA and AfL perceptions, and their conceptions of assessment purposes on the quality of feedback practices. More specifically, the study examines three research questions:

1) To what extent do secondary school mathematics teachers in Tanzania perceive their own assessment practice as formative in terms of the monitoring and scaffolding functions?
2) To what extent do secondary school mathematics teachers' perceptions of their FA and AfL practices (monitoring and scaffolding) and their conceptions about the purposes of assessment (assessment improves learning, school accountability) predict the quality of their feedback practices?
3) For what purposes do secondary school mathematics teachers typically use students' assessment information (such as student's scores in terminal and mid-term tests)?

3.2. Method

3.2.1. Participants

The study was conducted among 48 secondary schools in Tanzania: 25 in the mostly urban Dar es Salaam region and 23 in the mostly rural Kilimanjaro region. Based on national educational statistics (MoEVT, 2013) the mean GPA for schools performance in the sampled regions (*M* = 4.63, *SD* = 0.69) did not deviate statistically from the country schools' mean GPA (*M* = 4.85, *SD* = 0.70). Three criteria were used to achieve a representative sample: school mathematics performance (high, medium, low) according to school ranking (MoEVT, 2013), class-size (< 40, ≥ 40), and school-type (private, government). Within the 48 randomly sampled school there were 54 mathematics teachers (years of teaching experience: *M* = 10.87, *SD* = 10.39) from schools varying in mathematics performance (N_{high} = 8, N_{middle} = 19, N_{low} = 27) performing schools. The sample constituted 16.7% female and 83.3% male teachers with an overall mean age of 37.26 (*SD* = 10.96) with a range of 23 to 66 years. The Form three class (Grade 11) was selected because it contains more teacher-based assessment practices compared to Form 1 and 2. Teachers taught classes with typical class sizes (*M* = 49 students, *SD* = 20.49) and had a typical teaching load in term of teaching periods per week (*M* = 22.05, *SD* = 7.33, with range a range of 6-38 periods; one period is equivalent to 40 minutes).

3.2.2. Design

A mixed-method research approach was applied, combining quantitative (survey) and qualitative (interviews) methods (Leech & Onwuegbuzie, 2009). Specifically, we employed a concurrent embedded design where qualitative and quantitative data were simultaneously collected and analyzed to complement each other (Creswell, 2009; Dingyloudi & Strijbos, in press). We complemented the quantitative analyses of survey data with content analysis of qualitative data from teacher interviews.

3.2.3. Instruments

3.2.3.1. Questionnaires

We adopted previously validated questionnaire scales for the survey, which were adapted to the mathematics context by inserting the word ‘mathematics’ to ensure that teachers would reflect on their mathematics students. Firstly, we used the Teacher Assessment for Learning Questionnaire (TAFLQ) to measure teacher’s perceptions of their AfL practice (Pat-El et al., 2013) in terms of ‘perceived monitoring’ and ‘perceived scaffolding’. Secondly, we adopted the ‘feedback delivery’ and ‘promoting feedback seeking’ subscales from the Feedback Environment Scale (FES) (Steelman et al., 2004) to measure the quality of teachers’ feedback practices. Third, mathematics teachers’ conceptions of assessment were measured using the Teacher Conceptions of Assessment survey (TCoA-III) (Brown, 2004) consisting of main four sub-scales. Only two sub-scales were sufficiently reliable: ‘assessment improves student learning’ and ‘assessment improves school quality’, whereas two main sub-scales (‘assessment is irrelevant (ignored/bad)’, and ‘assessment is for student accountability’) were excluded from analysis. We adapted the response format of the various scales which differed in response options (i.e., 5, 6 or 7) to a common balanced 4-point scale: fully disagree (1), somewhat disagree (2), somewhat agree (3) and fully agree (4). We also refrained from a middle category due to its ambiguous meaning (Dunham & Davison, 1991; Kulas & Stachowski, 2009). It is noteworthy that the ‘promoting feedback seeking’, ‘feedback delivery’, and ‘school accountability’ subscales were below the .70 threshold for Cronbach’s alpha, suggesting that robust analyses methods were required to draw valid conclusions. See Appendix I for detailed questionnaire items. Table 5 summarises the adopted scales, number of items per scale, a sample item per scale, and the Cronbach's α from the original studies and Cronbach's α for this study.

Table 5. *Sample items and scales' Cronbach's α*

Scale	**k**	**Sample item**	**Cronbach's α**	
			Original Study	*Present Study*
Perceived monitoring	16	I ask my students to indicate what went well and what went badly concerning their assignments.	.87	.82
Perceived scaffolding	12	I adjust my instruction whenever I notice that my students do not understand a topic.	.77	.77
Feedback delivery	5	I am supportive when giving my students feedback about their mathematics performance.	.86	.58
Promote feedback seeking	4	I encourage my students to ask for feedback whenever they are uncertain about their mathematics performance.	.84	.45
School accountability	6	Assessment is a good way to evaluate a school.	.77	.58
Improve-learning	10	Assessment helps students improve their learning.	.83	.71

Note. k = number of items per scale

3.2.3.2. Interviews

The interview questions were specifically developed for the present study to again some in-depth understanding of the topics covered in questionnaire scales. The interview focused on two main goals: (a) teacher teaching practices and testing practices such: teacher reactions to student errors, teacher perceptions of FA and AfL practices, and (b) Teacher perception of student experiences with teaching and testing practices such as: perceptions of student reaction on teacher feedback. For this study teachers' responses to the question "For what purposes mathematics teachers typically use students' assessment information (such as student's scores in terminal and mid-term tests)" were analysed. The average duration of the interviews was 27 minutes. See Appendix J for detailed interview questions.

3.2.4. Procedure

The research was conducted with research clearance from the University of Dar es Salaam. A participating teacher signed a consent form. Questionnaires were administered to mathematics teachers by the researcher or by one of two research assistants. The researcher or assistant demonstrated how to use the rating scales prior to the mathematics teachers filling-in the questionnaire. The teachers needed approximately 20 minutes to complete the questionnaire.

Prior to data analyses, data screening was carried out to account for outliers (univariate and bivariate), as well as missing value analysis and recoding of all negatively phrased items. Data were considered to be missing completely at random (MCAR) because Little’s MCAR test was not statistically significant ($\chi^2 = 101.67$, $df = 1732$, $p = 1.00$) (Peugh & Enders, 2004). We imputed for missing values using Expectation Maximization (EM) method, which is considered an effective imputation method when data are MCAR (Musil, Warner, Yobas, & Jones, 2002). Investigation of the EM estimated statistics such as items means showed minimal differences to the un-estimated data (i.e., differences noticeable at the .001 level).

3.2.5. Analyses

3.2.5.1. Questionnaire analyses

Structural equation modeling (SEM) using scale means was used to estimate the impact of mathematics teachers’ perceptions of their FA and AfL practices and conceptions of assessment on the quality of their feedback practices. The SEM approach was preferred over normal regressions because it provides a stronger framework to account for response bias and takes into account non-random measurement errors (Comşa, 2010). To account for the small sample size, the structural model utilized the scale means instead of variables

(Little, Cunningham, Shahar, & Widaman, 2002; Marsh, Hau, Balla, & Grayson, 1998). The interpretation of model fit was based on the following indicators: root mean squared error of approximation (RMSEA) and standardized root mean residual (SRMR) below .05 and comparative fit index (CFI) and gamma hat values above. 95 indicate good fit (Byrne, 2010), while RMSEA and SRMR below .08 and CFI scores above .90 indicate acceptable fit (Hu & Bentler, 1999). Because the Chi-square statistic is overly-sensitive in large sample sizes above 250 (Byrne, 2010), we report multiple fit indices. The CFI and RMSEA are not stable estimators because CFI rewards simple models while RMSEA rewards complex models (Fan & Sivo, 2007). The Gamma hat statistic and SRMR have been shown to be stable estimators (Fan & Sivo, 2007). As recommended by Steiger (1990) we report the 90% confidence interval for the RMSEA.

3.2.5.2. Interview analyses

Content analysis was used to analyse the interviews. A data-derived coding scheme was developed using about ten percent of all interviews. The threshold for segmentation agreement was 80% (Strijbos, Martens, Prins, & Jochems, 2006) and a Krippendorff's alpha value of 0.80 for coding reliability (Krippendorff, 2013). Two independent coders were involved in all data analysis after a 50 minutes training on the study rationales and the coding scheme. Four iterations of independent coding trials were performed in analyzing interviews. The first coding trial analyzed seven randomly sampled interviews (equivalent to 13% of all interviews), but had a low segmentation agreement of 67-70%. Three more coding trials were performed; each used six randomly sampled interviews (equivalent to 11% of all interviews). The second coding trial had a segmentation agreement of 83-86%, but independent coding of the segmentation from the coder with more segments resulted in a Krippendorff's alpha of .63 with a lower and upper limit of .51 and .75, respectively. The third coding trial had the segmentation agreement of 71-74%. The fourth coding trial had a segmentation agreement of

87-89%. Afterwards, all segments from the coder with more segments were independently coded by each coder, leading to a Krippendorff's alpha of .88 with a lower and upper limit of .78 and .96, respectively. See Appendix K for the detailed interview coding scheme.

3.4. Results

The primary purpose of this study was to investigate the effect of mathematics teachers' FA and AfL perceptions (monitoring and scaffolding) and conceptions of assessment on the quality of their feedback practices. Specific results are reported with respect to research questions. Survey results are first reported followed by interview results.

3.4.1. Mathematics teachers' perceptions of their FA and AfL practices

The first research question sought to investigate mathematics teachers' FA and AfL perceptions. Responses to the FA and AfL perceptions scales (i.e., monitoring, scaffolding), conception scales (i.e., improve learning, school accountability) and the quality of feedback practices (i.e., 'feedback delivery' and 'promote feedback seeking') provided insights as to how mathematics teachers perceived their assessment practices. Table 6 summarises descriptive results on teacher perceptions and conceptions of their FA and AfL practices.

Table 6. *Descriptive statistics per scale and their inter-correlations*

Variables	**Descriptive**					**Scale inter-correlations**				
	N	*Min*	*Max*	*M*	*SD*	*I*	*II*	*III*	*IV*	*V*
I. Perceived monitoring	54	2.44	4.00	3.61	0.31	-				
II. Perceived scaffolding	54	2.75	4.00	3.64	0.32	.72**	-	-		
III. School accountability	54	2.17	4.00	3.29	0.42	.30*	.19	-		
IV. Improve Learning	54	2.40	4.00	3.41	0.39	.45**	.35**	.57**	-	
V. Feedback delivery	54	2.50	4.00	3.49	0.43	.49**	.56**	.29*	.17	
VI. Promote feedback seeking	54	2.50	4.00	3.54	0.44	.30*	.30*	.27*	.15	.37**

Note. I & II = AfL perceptions, III & IV = conceptions of assessment, ** $p < .01$, * $p < .05$

Table 6 reveals that mathematics teachers' assessment perceptions were above somewhat agree (3.00) for all scales, suggesting that the mathematics teachers evaluated positively their own FA and AfL practices, conceptions of assessment and the quality of feedback practices. Furthermore, the inter-correlations indicate that: (a) mathematics teachers' FA and AfL perceptions had medium to high positive correlation with the quality of feedback delivery, and (b) the conception that assessment promoted school accountability was positively correlated with the quality of feedback delivery.

3.4.2. Teachers' FA and AfL perceptions, conceptions of assessment and quality of feedback practices

The effect of mathematics teachers' FA and AfL perceptions (scaffolding and scaffolding) and conceptions of assessment (school accountability and improve learning) on the quality of teacher feedback practices ('feedback delivery' and 'promote feedback seeking') was estimated in a structural equation model (see Figure 3).

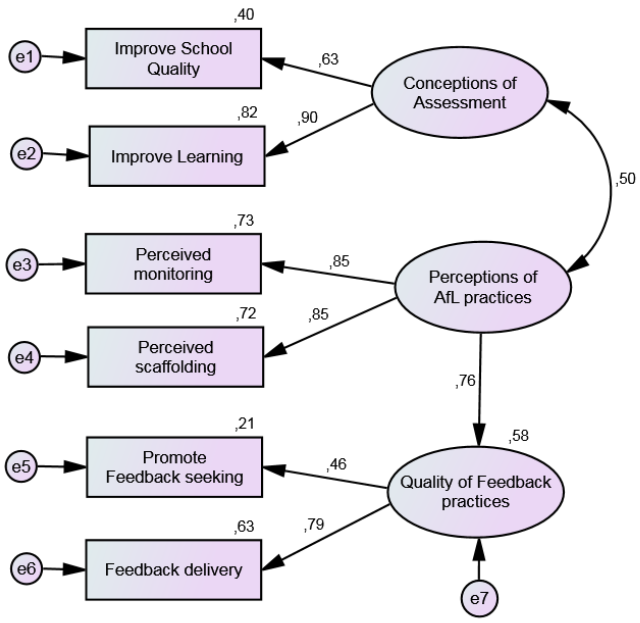

Figure 3. Teacher's FA and AfL perception and assessment conceptions as predictors of the quality of their feedback practices

The model had good fit: CFI = .981, Gamma hat = .989, SRMR = .053 and RMSEA = .069 [.000, .192]. Mathematics teachers' conceptions of assessment purposes were positively and highly correlated with the perceptions of their own AfL practices (r = .50). Combined, teacher's conceptions of assessment and perceptions of AfL explained 58% of the variance in the quality of their feedback practices. This represents a large effect (f^2 = 1.38; Cohen, 1992).

3.4.3. Mathematics teachers' use of student's assessment information

With the help of the interview we aimed to investigate how the mathematics teachers used students' assessment information such as tests scores and classroom discussions. Essentially, this was motivated by the claim that it is the purpose for which the assessment information is used (by the teacher and students) that makes the assessment to be formative. Table 7 summarises six key themes and sample interview excerpts on teachers' use of students' assessment information.

Table 7. *Mathematics teachers' (N = 54) use of their students' assessment information*

Key themes	Interview excerpts
1. Show students how to improve (44%)	I use assessment information to: Do corrections to all students in the class and I normally involve students who are doing better to do corrections on the board so that other students can be encouraged (Teacher 38).
2. Devise teaching approaches (30%)	I use assessment information to: Evaluate myself if what I taught was understood by my students or not. If students perform poor, I prepare remedial classes so that I can re-teach students who scored below the average (Teacher 25).
3. Ability grouping (20%)	In our school we normally use student's scores first of all in ranking students. Secondly, student's scores are bases for student promotion or retention in the same class (Teacher 53).In our school we identify and separate slow learners so that they can get a special attention; they are special classes (Teacher 8).
4. Accountability reports (17%)	Normally assessment analysis goes far to inform parents (Teacher 5). I use assessment to: Collect marks for Continuous Assessment (CA) in order to meet our school development (Teacher 17).
5. Motivate high achievers (17%)	Sometimes we award best students and try to assist slow learners. In awarding the best students it helps the slow learners to work hard (Teacher 52).
6. Reprimand low achievers (4%)	I always use assessment results to reprimand students who drop in their performances (Teacher 46).

Table 7 shows that mathematics teachers reported using their students' assessment information to make various decisions. Analyses of interviews resulted into six main themes on how mathematics teachers use their students' assessment information: (1) to show students how to improve, (2) to devise their teaching approaches, (3) to categorise students into ability groups, (4) to compose accountability reports (to parents, school authority, etc.), (5) to motivate high achievers, and (6) to reprimand low achievers. Firstly, teachers reported a formative use of student assessment information such as reflections on their teaching practices, improving their teaching approaches, correcting student errors and conducting remedial classes to support weaker students. Secondly, summative practices were also reported such as ability grouping (if no specific support was provided to each ability group), accountability reports, and using assessment to reprimand low achievers.

Using assessment information as a motivation for high achievers may have positive impact on learning, if and only if, it leads to positive changes in students' effort, engagement or self-efficacy (Kluger & DeNisi, 1996). Furthermore, it was noted that ability grouping can be a formative practice when intended to provide students with extra support as evidenced in excerpt ("In our school we identify and separate slow learners so that they can get a special attention; they are special classes"; Teacher 8). Ability grouping was a non-formative practice if it was used solely for ranking students "We normally use student's scores first of all in ranking students"; Teacher 53). In general, the mathematics teachers reported more formative than summative uses of their students' assessment information.

3.5. Discussion

This study investigated the effect of mathematics teachers' FA and AfL perceptions (monitoring and scaffolding) and conceptions of assessment (school accountability, improve learning) on the quality of their feedback practices. The first research question sought to investigate the extent to which mathematics teachers perceived their assessment practice as formative in terms of the monitoring and scaffolding of student learning. The mathematics teachers had a positive perception, indicating that they perceived their own assessment practices were formative. Moreover, this indicates that Tanzanian mathematics teachers value their own FA and AfL practices, which replicates findings from previous studies on teacher perceptions of their own assessment practices (e.g., Pat-El et al., 2015; Rach et al., 2013; Veldhuis & Van den Heuvel-Panhuizen, 2014). Furthermore, our results also indicate that mathematics teachers had positive conceptions of assessment, i.e. that the purpose of assessment was to improve student learning and promote school accountability. This is consistent with previous studies indicating that teachers consider the purpose of assessment to be that of improving student learning and promoting school accountability (Brown, 2004, 2006: Barnes et al., 2017).

The second research question investigated the extent to which mathematics teachers' perceptions of their FA and FA practices (monitoring and scaffolding) and conceptions of assessment (school accountability, improve learning) predicted the quality of their feedback practices. The structural equation model indicates that mathematics teachers' perceptions of FA and AfL and their conceptions of assessment were highly correlated, and combined they strongly predicted their quality of feedback practices. These findings support previous studies that perceptions of assessment are related to teacher assessment practices (Fives & Buehl, 2012). These findings are consistent with Van de Pol, Oort, Volman and Beishuizen (2014) who found that scaffolding is an important practice for improving teacher assessment practices and student learning.

The third research question sought to identify typical uses of student's assessment information by mathematics teachers. In line with Al Duwairi (2013), our interview results also showed that mathematics teachers used students' assessment information for both formative and summative purposes. However, how mathematics teachers balance summative and formative uses of student assessment information is an important issue for future research. Furthermore, the interview results indicate that mathematics teacher's assessment practices were rooted in their conceptions of assessment purposes. For example, the majority of mathematics teachers reported to use their student assessment information to reflect on their teaching approaches and to provide feedback on their students' learning; both activities are considered core elements of a formative assessment practice (Black & Wiliam, 2003; 2009; Ginsburg, 2009; Hattie, 2009). The observed role of teacher conceptions of assessment adds to previous studies that assessment improves teaching and learning (Brown, 2004; Brown et al., 2015). The results are also consistent with Ndalichako (2015) who reported that 60% of 2047 Tanzanian secondary school teachers agreed that the purpose of classroom assessment was to improve teaching and learning processes. Additionally, other reported uses such as accountability reports to parents and students' ability grouping, provide further support that conceptions of assessment promoted student and school accountability (Brown, 2006; Firestone et al., 1998). Surprisingly, some mathematics teachers reported to use their student's assessment information to reprimand low achievers. Although such practices were reported by only a few mathematics teachers, and reported by a few students in Chapter 2, these practices are highly discouraged. Teachers should be encouraged to use errors in mathematics tests or assignment to inform students on how to improve (Rach et al., 2013), or provide educational counselling instead of reprimanding low achieving students (Yaghambe & Tshabangu, 2013).

3.5.1. Methodological limitations

Our results should be interpreted in light of some limitations. Firstly, we mainly used self-report data from surveys and teacher interviews. Future research could further substantiate our findings with other measures such as observational data. Secondly, the reliability of the ‘feedback delivery’ and ‘promote feedback seeking’ scales was below the typical threshold (i.e., $\alpha > .70$), which indicate that our results should be interpreted cautiously. However, we applied structural equation modeling which is a robust technique and takes into account random or non-random measurement errors. Thirdly, based on the relatively small sample of mathematics teachers we cannot generalise our findings beyond this sample. Additionally, we suggest that these results may be substantiated by observational and longitudinal studies to examine other potential factors that might influence the quality of teacher feedback practices.

3.5.2. Theoretical and practical implications

Mathematics teachers are aware that effective formative assessment demands both teachers and students to reflect on the assessment information. However, if mathematics teachers only reflect on this information but students do not utilize the feedback provided by their teachers, FA and AfL practices are apt to fail (Pat-El et al., 2013). Our results indicate that mathematics teachers had positive perceptions of their own FA and AfL practices and conceptions of assessment, and that combined they predicted the quality of their feedback practices. Thus, these results support the planned behaviour theory that conceptions (beliefs) influence behaviour (Ajzen, 1991). Furthermore, qualitative results from the interviews support that mathematics teachers reported various uses of student assessment information (Gronlund & Linn, 1990). The self-reported uses of assessment information aligned with established teacher conceptions of assessment that assessment improves teaching and learning (Black & Wiliam, 1998; Brown, 2002; Popham, 2014), and that assessment promotes school and student accountability (Brown, 2004, 2006, 2011). In sum, mathematics teachers’ perceptions of their FA and AfL practices, and their conceptions of assessment positively and strongly predicted the quality of feedback practices. We recommend interventions for improving the quality of teacher feedback practices to capitalize on teacher conceptions of assessment and perceptions of their FA and AfL practices.

4. Impact of feedback training on students' perceptions of mathematics teachers' feedback practices among secondary schools in Tanzania

4.1. Introduction

Formative Assessment (FA) and Assessment for Learning (AfL) received increased attention over the past three decades (Black & Wiliam, 1998; Wiliam, 2011) as both assessment practices stress the diagnostic purpose of assessment (i.e., improve student learning) and the active involvement of the student. Formative assessment involves certain kinds of teacher-student interactions such as 'monitoring' to track student progress and 'scaffolding' to help students improve their learning (Pat-El, Tillema, Segers, & Vedder, 2013; Pat-El, Tillema, Segers, & Vedder, 2015; Pat-El et al., 2015; Stiggins, 2005) and requires that students perceive these interactions as formative and make use of them to improve their learning (Heinze & Reiss, 2007). Several studies (Can, 2011; De Kleijn, Mainhard, Meijer, Brekelmans, & Pilot, 2013; Fun, 2005; Montgomery & Baker, 2007; Nesbit & Burton, 2006; Pat-El et al., 2013; Pat-El et al., 2015; Weaver, 2006) support the idea that students' perceptions of their teachers' assessment and feedback practices are essential for their learning gains. Assessment information should lead to 'instructional adjustment' by both the teacher and students to improve learning. Therefore, it is necessary to support student's positive perception and use of feedback information.

4.1.1. Role of student perception of feedback

The shift from teacher centred to learner centred education places more emphasis on the role of students' experiences during the instructional process (Maclellan, 2001; Smyth, 2004). Kollar and Fischer (2010) argue that the feedback process involves various activities such as feedback provision by a teacher (or peer), feedback reception by a student and acceptance by students to apply such feedback to improve the quality of their work. Students' perception of feedback (Lizzio & Wilson, 2008: Poulos & Mahony, 2008; Strijbos, Narciss, & Dünnebier, 2010) is assumed to influence their feedback application (De Kleijn et al., 2013; King, Schrodt & Weisel, 2009). Brown and Hirschfield (2008) discussed for example that student perceptions of their teacher's assessment and feedback practices influences the way they learn and what they can actually achieve. Lizzio and Wilson (2008) noted that students are readily able to describe the qualities of assessment feedback that they do and do not value. Strijbos, Narciss and Dünnebier (2010) developed a survey to measure student perceptions of feedback and Strijbos, Pat-El and Narciss (2010) showed that secondary school students' perception of feedback adequacy (fairness, usefulness and acceptance) predicted their willingness to improve and affect. Although it is widely accepted that feedback helps students acknowledge and learn from their mistakes (Hattie & Timperley, 2007), students face several obstacles in understanding their teacher written (and/or oral) feedback, which can be detrimental to the actual use of feedback (Jonsson, 2013). Students' perceptions of feedback appear to be an important aspect for feedback use.

Although the issue of what constitutes effective feedback is highly debated (Shute, 2008) some feedback models seem to characterize what constitutes a formative feedback. Our theoretical consideration on what constitutes effective feedback was guided by the Hattie and Timperley (2007) feedback model. The authors identify four levels of feedback: task (i.e., whether work was correct or incorrect), process (i.e., comments about the processes or strategies underpinning the task), self-regulation (i.e., reminders to students about strategies they can use to improve their own work) and self (i.e., non-specific praise and comments about effort). From their review, they demonstrated that task, process and self-regulation feedback all contribute to learning outcome gains, while feedback at the self-level generally does not. Additionally, Narciss (2008) and Shute (2008) both advocate for elaborated feedback with explanations underpinning 'what', 'how' and 'why' when providing feedback on students' work. In regard to this study, we investigated if teacher feedback practices can be improved by the training on improving feedback practices. The central question focuses on how can we change student perceptions and their use of teacher feedback?

4.1.2. Interventions on teacher assessment and students' use of feedback

Changing teachers' assessment and feedback practices and students' use of feedback information seems to be possible by interventions on the student level (Borasi, 1994; Duijnhouwer, Prins, & Stokking, 2012; Rach, Ufer & Heinze, 2013; Winstone, Nash, Parker, & Rowntree, 2017), but there is scarce evidence for interventions on the teacher level in mathematics education (Heemsoth & Heinze, 2016). In most cases students may need advice on understanding and using feedback before engaging with it (Jonsson, 2013; Weaver, 2006). Winstone et al. (2017) highlight the need for specific scaffolding to support student use of feedback. Supporting teacher feedback practices can promote student feedback perceptions, because the delivery of feedback and how feedback related to criteria, marks and grades was found to be related to student perceptions of feedback (Poulos & Mahony, 2008). However, it is still an open issue whether interventions that focus on specific AfL situations have an effect on students' perceptions and use of feedback information. For example, Pat-El et al (2013) developed a survey with two sub-scales to measure monitoring and scaffolding perceptions of teacher FA and AfL practices. Also, since feedback recipients are more likely to use a well delivered feedback (Fedor, Eder, & Buckley, 1989), Steelman, Levy and Snell (2004) proposed a scale for measuring perceptions of feedback delivery. In general, although several studies have investigated student perceptions of feedback in secondary education (Peterson & Irving, 2008) and higher education (De Kleijn et al., 2013; Duijnhouwer et al., 2012; Poulos & Mahony, 2008; Weaver, 2006), to our knowledge research from African educational systems and cultural contexts is missing.

4.1.3. The present study

Student perceptions of their teacher's assessment and feedback practices appear to play a central role for students' use of feedback and their learning. This study examines a situation that involves scaffolding students learning in their perception and use of feedback on a written test. Tanzania introduced in 1976 *Continuous Assessment* (CA) in secondary schools, which was envisioned to serve as a formative practice. However, Basic Education Statistics in Tanzania (BEST, 2004-2013) indicate that for ten consecutive years (2004-2013) an average of 79% of secondary schools students failed their mathematics national examinations. Such persistent failure still happens even though mathematics teachers reported to practice various assessment practices which are supposed to be formative. For example, Kyaruzi, Strijbos and Ufer (2015) noted that 93% of the 54 mathematics teachers in their sample self-reported that they provide feedback on students' mathematics tests. Moreover, in Chapter 2 it was found that student perception of teacher feedback delivery practices strongly predicted students' feedback use (ß = .86). Yet, in focus group discussions students voiced some dissatisfaction with their mathematics teacher feedback practices (see Chapter 2), which indicates that despite the assumed formative nature of current mathematics assessment practices in Tanzanian secondary schools, teachers' feedback practices and/or students' utilization of feedback might be suboptimal. Hence, an intervention in the shape of a training to improve mathematics teachers' feedback practices is a sensible first step to improve students' perception and use of their teacher's feedback on their mathematics test. The aim of the present study was to determine whether a feedback training intervention could improve formative feedback practices of Tanzanian secondary school mathematics teachers, and the degree to which students perceived their teacher's practice as formative. Specifically, we seek to answer the following questions:

1) How do students perceive teacher feedback on their mathematics test in terms of fairness, usefulness, acceptance, willingness to improve and affect after the intervention?
2) Are student perceptions of teacher feedback on their mathematics test in terms of fairness, usefulness, acceptance, willingness to improve and affect more positive in students whose teacher received the feedback training than those whose teacher did not?
3) Do students of mathematics teachers who received the feedback training differ in the degree to which they perceive their teacher's FA and AfL practices in terms of monitoring, scaffolding and feedback delivery from students of teachers who did not receive the feedback training?
4) Are student perceptions of their feedback use more positive in students whose teacher received the feedback training than those whose teacher did not?

4.2. Method

This study was conducted in eight secondary schools including four government and four private run schools in the Dar es Salaam region in Tanzania. This region was sampled because according to the National Examinations Council of Tanzania statistics (NECTA, 2014), school performance in the Dar es Salaam region (M = 1.64, SD = 0.63) did not deviate from the overall national mathematics performance (M = 1.55, SD = 0.65). Based on the NECTA (2014) statistics, the Dar es Salaam region had 191 secondary schools with more than 40 students, classified according to performance as: 10 (5%) high performing, 173 (91%) middle performing, and 8 (4%) low performing schools. Because error analysis has been reported as being especially beneficial for lower performing students, this study was conducted among four high and four low performing schools. In the sampling of schools, we prioritized schools with mixed gender (boys and girls) to maximize gender representativeness; fifteen schools (7 high and 8 low performing) met this criterion. Next, four schools were randomly sampled from each stratum and in each school category two schools were randomly assigned to the experimental and control group. One mathematics class (Grade 11) was randomly sampled from each school and all students in the sampled class were included in the study.

4.2.1. Participants

The sample consisted of Form three (Grade 11) students (*N* = 251) and their respective mathematics teachers (*N* = 8). Table 8 provides a description of students' and teachers personal and school characteristics.

Table 8. *Demographics of participating students and teachers*

Demographic	**Total**	**Control**	**Experimental**
Students	251	121	130
Gender			
Male	135	68	67
Female	116	53	63
Age	16.29 (0.95)	16.07 (0.84)	16.49 (1.00)
Male	16.42 (0.93)	15.81 (0.68)	16.41 (1.07)
Female	16.14 (0.96)	16.28 (0.91)	16.57 (0.94)
School Performance			
High	135	67	68
Low	116	54	62
Teachers	8	4	4
Gender			
Male	7	4	3
Female	1	0	1
Age	42.50 (9.38) range: 32-57	41.25 (3.95) range: 38-47	43.75 (13.62) range: 32-57
School Performance			
High	4	2	2
Low	4	2	2
Highest Qualification			
Bachelor degree	6	2	4
Diploma in education	2	2	0
Teaching load/ hours per week	22.75 (9.38) range: 12-38	26 (9.78) range: 18-38	19.50 (9.00) range: 12-30
Class size	64.37 (28.15) range: 37-115	52.00 (15.41) range: 37-70	76.00 (34.91) range: 40-115

Note. Mean (standard deviation)

4.2.2. Design

A quasi-experimental pretest, intervention (training) and posttest repeated measures design with experimental and control groups was adopted. Students filled in questionnaires after each lesson. Using intact classes, two groups were formed through random assignment: Control (*N* = 121) and Experimental (*N* = 130). The four mathematics teachers in the experimental group participated in a one-day intensive feedback training (FBT) for improving their feedback practices. To ensure equity, after the posttest data collection, the four mathematics teachers in the control group received the same one-day feedback training. The time interval between the intervention (training) and posttest measures was approximately one month. Figure 4 summarizes the overall research design.

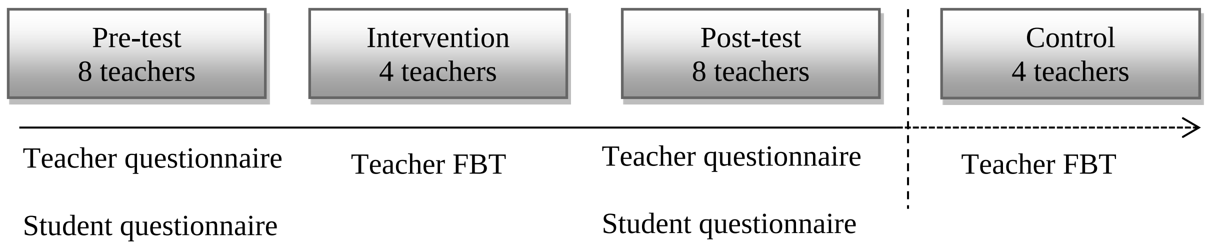

Figure 4. General research design (FBT = Feedback training)

4.2.3. Intervention

The feedback training (FBT) contained concepts and practical sessions focusing on how mathematics teachers can improve feedback provision using the Hattie and Timperley (2007) model, and how to apply cognitive strategies for learning from errors (Heemsoth & Heinze, 2016; Rach, Ufer & Heinze, 2013). More specifically, mathematics teachers practiced how to provide feedback at the task level, process level and self-regulation levels of the Hattie and Timperley model. In regard to cognitive strategies for learning from errors, mathematics teachers brainstormed and practiced how to identify student errors, describe and explain student errors (why errors), how to correct student errors, and how to develop strategies for avoiding similar errors. The training used typical samples of written mathematics feedback on students' tests collected during pretest lessons. The mathematics teachers engaged in identifying good/bad features in the feedback they provided on the sampled student tests based on theoretical insights regarding the four feedback levels (Hattie & Timperley, 2007), use of feedback prompts (Gan & Hattie, 2014) and features of formative feedback (Shute, 2008). Appendix L provides an overview of the training contents and activities.

4.2.4. Instruments

Questionnaires were administered in two classroom contexts: (a) during a normal mathematics lesson (pretest), and (b) during a classroom lesson when the mathematics teacher discussed students' results on a test (posttest) which involved a sequence of tasks (student receive test results, students respond to a feedback perceptions questionnaire on the test, teacher conducts a feedback plenary discussion, and students respond to a short questionnaire about the perception of the plenary discussion). First, during the pretest a general questionnaire measured (a) students' perceptions of their teacher's AfL practices in terms of monitoring and scaffolding (Pat-El et al., 2013), (b) student perceptions of teacher feedback delivery (Steelman et al., 2004), and (c) student's own perceptions of their feedback use (Gibbs & Simpson, 2003; King et al., 2009). Second, students responded to a feedback perceptions questionnaire (Strijbos, Narciss et al., 2010; Strijbos, Pat-El et al., 2010) immediately after their teacher issued them with feedback on their mathematics test. This feedback perceptions questionnaire asked students to indicate how they perceived the teacher feedback in terms of fairness, usefulness, acceptance, willingness to improve, and affect. Third, after the feedback plenary discussion on the mathematics test the students completed a short questionnaire to measure the fidelity of the teacher's feedback discussion (Jacobs, Garnier, Gallimore, Hollingsworth, Givvin, Rust, Kawanaka, Smith, Wearne, Manaster.., & Stigler, 2003) and items to measure students' perceptions of the feedback plenary discussion. All adopted scales were previously validated, adapted to a balanced and symmetrical 6-point scale ranging from: *completely disagree* (1) to *completely agree* (6). See Appendices M, N and O for the detailed questionnaire items. Table 9 summarises scales, number of items per scale, a sample item, and the Cronbach's α from the original studies and Cronbach's α for this study.

Table 9. *Sample items and scales' Cronbach's α*

Scale	**k**	**Sample item**	**Cronbach's α**		
			Original Study	*Present study*	
				Pretest	*Posttest*
Perceived monitoring	16	My mathematics teacher inquires what went well and what went badly in my work.	.89	.93	.93
Perceived scaffolding	12	My mathematics teacher provides me with hints to help understand the subject matter.	.83	.87	.89
Feedback utility	9	I usually reflect on my mathematics teacher's feedback.	.85	.79	.79
Feedback delivery	4	My mathematics teacher considers my feelings when giving me feedback about my mathematics performance.	.86	.77	.77
Feedback adequacy	9	I am satisfied with this feedback.	.89	.81	.82
Affect	6	I felt confident after this feedback on my mathematics test.	.67	.78	.78
Willingness to improve	3	I am willing to improve my mathematics performance.	.71	.66	.64
Authenticity of feedback plenary discussion	4	Was the videotaped plenary feedback discussion typical/ representative for the lessons your teacher normally teaches?	-	.62[a]	.59[a]
Perception of plenary feedback discussion	5	After this plenary feedback discussion, I now know how I can correct most of my mistakes.	-	.81	.82

Note. k = number of items per scale, a = alpha after removing one negatively phrased item

It is noteworthy that authenticity of feedback delivery and willingness to improve scales were below the .70 threshold for Cronbach's alpha, suggesting that revisions to the scales were needed to accurately reflect the responses of the participants.

4.2.5. Procedure

The research was conducted with research clearance from the University of Dar es Salaam. All participants signed a consent form. Questionnaires were administered at two phases: before (pretest) and after (posttest) the intervention. During the pretest, 326 respondents answered the general questionnaire, however, 61 students did not participate in all data collection points and were eliminated; thus, leaving the sample with 265 respondents. Furthermore, fourteen respondents with more than 10% missing data per session were also removed from analysis leaving the final sample with 251 respondents. The remaining data were considered to be missing completely at random (MCAR) because Little's MCAR test was not statistically significant ($\chi 2 = 10190.60$, $df = 32739$, $p = 1.00$) (Peugh & Enders, 2004). Missing values were imputed with the Expectation Maximisation (EM) procedure which is considered to be an effective imputation method when data are completely missing at random (Musil, Warner, Yobas, & Jones, 2002). Comparison of the estimated statistics with the original variable statistics showed trivial differences, mostly at the third decimal point.

4.2.5.1. Measurement models

As each scale was adopted from a previously published inventory, measurement models were tested using confirmatory factor analysis. The interpretation of model fit was based on the following indicators: root mean squared error of approximation (RMSEA) and standardized root mean residual (SRMR) below .05 and comparative fit index (CFI) and gamma hat values above .95 indicate good fit (Byrne, 2010), while RMSEA and SRMR below .08 and CFI scores above .90 indicate acceptable fit (Hu & Bentler, 1999). Because the Chi-square statistic is overly-sensitive in large sample sizes above 250 (Byrne, 2010), we report multiple fit indices. The CFI and RMSEA are not stable estimators because CFI rewards simple models while RMSEA rewards complex models (Fan & Sivo, 2007). We therefore report Gamma hat statistic (Fan & Sivo, 2007). Importantly, as Steiger (1990) recommends, we report the 90% confidence interval for the RMSEA. Table 10 summarizes the measurement models after trimming the model by removing items with poor factor loading (< .40), lowest contribution to the factor (R^2 < .30), and sometimes a negatively phrased item which loaded poorly on a factor (see Brown, 2004a; Carlson, Wilcox, Chou, Yang, Chang, Blanchard, Marterella, Kuo, & Clark, 2011).

Table 10. *Measurement models during pretest and posttest*

Measurement models	**Time**	**k**	**SRMR**	**RMSEA [90%CI]**	**CFI**	**Gamma hat**
Perception of AfL (monitoring & scaffolding)	Pretest	16	.035	.056 [.043,.069]	.965	.961
	Posttest	16	.051	.088 [.077, .100]	.916	.909
Feedback delivery	Pretest	4	.024	.069 [.000,.154]	.991	.995
	Posttest	4	.047	.168 [.098, .248]	.951	.973
Perception of feedback (willingness and affect)	Pretest	8	.065	.077 [.049, .105]	.941	.973
	Posttest	8	.044	.049 [.005, .080]	.974	.989
Perception of feedback adequacy (fairness, usefulness and acceptance)	Pretest	9	.090	.116 [.094, .139]	.893	.933
	Posttest	9	.081	.144 [.122, .167]	.843	.900
Authenticity and perception of the feedback plenary	Pretest	7	.082	.149 [.120, .180]	.904	.924
	Posttest	7	.073	.128 [.098,.159]	.929	.943
Feedback use	Pretest	9	.032	.014 [.000,.076]	.991	1.00
	Posttest	9	.020	.128 [.000,.076]	1.00	1.00

Note. k = number of items, SRMR = Standardized Root Mean Residual; RMSEA = Root Mean Squared Error of Approximation; CFI = Comparative Fit Index

With reference to Table 10, it should be noted that although fit quality was mixed the scales were used as specified; however, the validity of inferences might be affected by this level of fit to the data in some scales. Based on Gamma hat all models had acceptable to good fit.

4.2.5.2. Measurement invariance

Measurement invariance is a prerequisite of comparison between groups and measurement occasions (Reise, Widman, & Pugh, 1993; Vandenberg & Lance, 2000; Wu, Li & Zumbo, 2007; McArdle, 2007). Thus, invariance tests were conducted to generate evidence for scale comparability. The invariance of a scale across measurement occasions is evaluated by establishing whether fixing model parameters (e.g., factor regression weights, covariances, factor intercepts, or residuals) as equivalent results in a statistically significant difference in model fit (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000). For measurement invariance to be demonstrated the difference in χ^2, taking into account the difference in *df*, should not be statistically significant (i.e., $p > 0.05$) and the difference in CFI should be Δ CFI ≤ 0.01 (Brown & Chai, 2012).

Table 11 summarizes the measurement invariance results for the four interaction groups; i.e. the control and experimental group per measurement occasion: pretest control (C1), posttest control (C2), pretest experimental (E1), posttest-experimental (E2). First, two scales had strong (i.e., configural, metric and scalar) measurement invariance: (a) student feedback perceptions in terms of willingness to improve and affect, and (b) authenticity of feedback plenary discussion. Second, two scales had weak (i.e., configural and metric) measurement invariance: perception of feedback delivery and perceptions of FA and AfL-monitoring and scaffolding. Thirdly, perception of feedback use and perception of feedback adequacy (fairness, usefulness, and acceptance) were non-invariant. Metric invariance is sufficient for using a factor in repeated measures analyses because intercepts are likely to change due to intervention effects (McArdle, 2007; Wu, Li, & Zumbo, 2007), therefore we estimated between groups analyses for strong invariant scales, and only within group comparisons for weak invariant scales.

4.2.6. Analyses

Having established a plausible set of scales, latent mean comparisons were used to evaluate differences between groups and measurement occasions. Latent mean analyses (LMA) were used to assess the latent mean differences in scale means relative to the pretest-control group as a reference group (set to 0). This approach provides a stronger framework to account for response bias and takes into account random or non-random measurement errors (Comşa, 2010; Marsh, Guo, Parker, Nagengast, Asparouhov, Muthén & Dicke, 2017; Sass, 2011). In LMA the mean of a latent factor is computed and differences of other groups with a similar latent factor are estimated as z-score differences to the reference group (Hussein, 2010). The Wald χ^2 test of parameter constraints was used to assess whether the differences in latent means among the groups were significant or not.

Table 11. *Multigroup measurement invariances for the four interaction groups*

Models	**Comparison**	**SRMR**	**RMSEA**	X^2	**df**	**Δ** X^2	**Δ df**	**p-value**	**CFI**	**Δ CFI**
1. Perception of AfL: monitoring and scaffolding										
Unconstrained (configural)-A	B vs. A	0.060	0.071	879.385	536	58.429	48	**0.144**	0.907	
Measurement weights (metric)-B	C vs. A	0.088	0.069	937.654	584	157.142	102	0.000	0.904	**0.003**
Measurement intercepts (scalar)-C	B vs. C	0.104	0.071	1037.646	638	104.039	54	0.000	0.892	0.015
2a. Perceptions of feedback: Fairness, usefulness and acceptance										
Unconstrained (configural)-A	B vs. A	0.090	0.136	319.313	96	43.407	18	0.001	0.861	
Measurement weights (metric)-B	C vs. A	0.108	0.132	362.719	114	73.851	45	0.004	0.845	0.016
Measurement intercepts (scalar)-C	B vs. A	0.115	0.119	393.164	141	30.445	27	**0.295**	0.843	0.018
2b. Student perception of feedback: Willingness and affect										
Unconstrained (configural)-A	B vs. A	0.066	0.099	168.997	76	21.341	18	**0.263**	0.877	
Measurement weights (metric)-B	C vs. A	0.097	0.087	182.673	94	37.859	42	**0.653**	0.883	0.006
Measurement intercepts (scalar)-C	B vs. C	0.101	0.069	195.46	118	15.015	24	**0.920**	0.893	0.010
3. Authenticity and perception of feedback plenary										
Unconstrained (configural)-A	B vs. A	0.075	0.143	185.799	52	26.784	15	**0.031**	0.912	
Measurement weights (metric)-B	C vs. A	0.107	0.132	212.583	67	46.898	36	**0.106**	0.904	**0.008**
Measurement intercepts (scalar)-C	B vs. C	0.112	0.114	232.697	88	20.114	21	**0.514**	0.905	**0.007**
4. Perception of feedback delivery										
Unconstrained (configural)-A	B vs. A	0.043	0.164	34.902	8	10.767	9	**0.292**	0.952	
Measurement weights (metric)-B	C vs. A	0.076	0.116	45.669	17	30.005	21	**0.092**	0.949	**0.003**
Measurement intercepts (scalar)-C	B vs. C	0.106	0.099	64.907	29	19.238	12	**0.083**	0.936	0.016
5. Perception of feedback use										
Unconstrained (configural)-A	B vs. A	0.030	0.000	18.547	20	26.815	12	0.008	1.000	
Measurement weights (metric)-B	C vs. A	0.112	0.058	45.361	32	46.968	27	**0.010**	0.973	0.027
Measurement intercepts (scalar)-C	B vs. C	0.142	0.056	65.514	47	20.153	15	**0.166**	0.963	0.037

Note. SRMR = Standardized Root Mean Residual; RMSEA = Root Mean Squared Error of Approximation; CFI = Comparative Fit Index; values marked in bold indicate that the statistical equivalence assumption holds (difference in χ^2, taking into account the difference in *df*, should not be statistically significant (i.e., $p > 0.05$) and/or the difference in CFI should be Δ CFI ≤ 0.01).

4.3. Results

The current study aimed at improving feedback provision by mathematics teachers via a feedback training (FBT) and improving students' perception and use of their teachers' feedback on their mathematics tests. For quality control, we assessed if teachers implemented the feedback plenary discussion as much as they usually implement their routine lessons and how it was perceived by their students. At pretest students in the experimental (*M* = 4.88, *SD* = 1.16) and control (*M* = 5.01, *SD* = 1.05) groups perceived their mathematics teachers to implement the feedback plenary discussion in an authentic manner as they normally do in their lessons. Likewise, students in the experimental (*M* = 5.45., *SD* = 0.89) and control group (*M* = 5.60, *SD* = 0.58) groups perceived the feedback plenary discussion to be useful. At posttest, students in the experimental (*M* = 4.98, *SD* = 1.05) and control (*M* = 4.93, *SD* = 1.04) groups were also positive about the authenticity of the videotaped lessons. Also, at posttest, students in the experimental (*M* = 5.43., *SD* = 0.79) and control group (*M* = 5.40, *SD* = 0.92) groups perceived the feedback plenary discussion to be useful. Specific results are reported in order of research questions.

4.3.1 Student perceptions of feedback on mathematics test

The first research question investigated how students perceived teacher feedback on their mathematics test in terms of fairness, usefulness, acceptance, willingness to improve, and affect before and after the feedback training intervention. Based on measurement invariance results in Table 11, latent mean comparisons were estimated for the student perceptions of willingness to improve and affect scales. First, there was neither a significant difference in students' willingness to improve (Wald $\chi^2_{(1,\ N\ =\ 251)} = 0.25$, $p = .874$) nor in students' affect (Wald $\chi^2_{(1,\ N\ =\ 251)} = 1.755$, $p = .185$) between the pretest and posttest in the whole sample.

Table 12 summarizes the latent means for student perceptions of feedback in terms of willingness to improve and affect by group and measurement occasion.

Table 12. *Latent means for student perceptions of willingness to improve and affect by group and measurement occasion*

Scales **Groups**	**C1**	**C2**		**E1**		**E2**		
		Estimate	*SE*	*Estimate*	*SE*	*Estimate*	*SE*	*Wald p*
Willingness to improve	—	.048	.125	.017	.127	-.004	.119	.673
Affect	—	.002	.139	.105	.128	-.145	.132	.169

Note. C1 was set to 0, SE = standard error; C1 = Pretest control; C2 = Posttest control; E1 = Pretest experimental; E2 = Posttest-experimental

By chance, the experimental groups were somewhat more positive in terms of willingness to improve and affect than the control group at the pretest. By the end of the intervention, however, the experimental group had a lower score for both willingness to improve and affect than both the pretest and posttest scores for the control group.

Further analyses for the main effect of the group were estimated by means of within group comparisons to answer the second research question. Analyses for the main effect of the group showed that there were neither significant differences in willingness to improve (Wald $\chi^2_{(1,\ N\ =\ 130)}$ = .039, p =. 844) nor in affect (Wald $\chi^2_{(1,\ N\ =\ 130)}$ = .052, p = .819) among students whose teacher received the feedback training and whose teacher did not. Furthermore, interaction analyses did not indicate significant interaction effects in willingness to improve (Wald $\chi^2_{(1,\ N = 251)}$ = .178, p = .673) or affect (Wald $\chi^2_{(1,\ N = 251)}$ = 1.889, p = .169) between groups and measurement occasions. Therefore, our results show no significant differences in the development of students' perceptions of their teacher feedback in terms of their willingness to improve or affect.

4.3.2. Students' perception of AfL and feedback delivery

The third research question investigated the extent to which students of mathematics teachers who received the feedback training differed in the degree to which they perceived their teacher's FA and AfL practices in terms of monitoring, scaffolding and feedback delivery from students of teachers who did not receive the feedback training. Because the perception of FA and AfL and perception of feedback delivery scales were weakly invariant (see Table 11), between group comparisons were not estimated. As a result, we estimated within group (E1 vs. E2 and C1 vs. C2) longitudinal comparisons. Table 13 summarizes the within group results for student perception of FA and AfL (monitoring, scaffolding) and feedback delivery in the control and experimental groups.

Table 13. *Latent means for within group students' perceptions of AfL and feedback delivery*

Scales	**E1 vs. E2**				**C1 vs. C2**			
	E1	*E2 Estimate*	*SE*	*Wald p*	*C1*	*C2 Estimate*	*SE*	*Wald p*
Perceived monitoring	.000	.434 *	.104	.000	.000	.093	.128	.468
Perceived scaffolding	.000	.323 **	.118	.006	.000	.123	.126	.329
Feedback delivery	.000	.388 *	.115	.001	.000	.140	.135	.297

Note. E1 and C1 was set to 0; * p < .005, ** p < .05 for Wald χ^2 test; SE = standard error; C1 = Pretest control; C2 = Posttest control; E1 = Pretest experimental; E2 = Posttest-experimental

With reference to Table 13, the mean comparisons within the experimental group showed a significant pre-post increases in student perception of monitoring (Wald $\chi^2_{(1, N = 130)}$ =17.56 , p = .006), perception of scaffolding (Wald $\chi^2_{(1, N = 130)}$ =7.44 , p = .006), and perception of feedback delivery (Wald $\chi^2_{(1, N = 130)}$ = 11.42, p = .001). Contrary, the Wald χ^2 test within the control group (C1 vs. C2) showed that there were non-significant pre-post differences in student perceptions of monitoring (Wald $\chi^2_{(1, N = 121)}$ = .527, p = .468), scaffolding (Wald $\chi^2_{(1, N = 121)}$ = .953, p = .329) and feedback delivery (Wald $\chi^2_{(1, N = 121)}$ = 1.088, p = .297) practices. Thus, the intervention had significant impact on student perceptions of their teacher's FA and AfL practices (i.e., monitoring and scaffolding) and feedback delivery in the experimental group.

4.3.3. Student perception of own feedback use

The fourth research question investigated whether the intervention improved student formative use of teacher feedback. Based on measurement invariance tests, the student perception of feedback use scale was only configurally invariant; as a result, we could not estimate between groups' latent means and answer research question four. However, additional analyses showed that within group (E1 vs. E2 and C1 vs. C2) comparisons were invariant. The within experimental group model for student perception of feedback use had SRMR = .129, CFI = .999, and RMSEA = .008 [.000, .009] and Gamma hat = 1.000, while the within control group model had SRMR = .069, CFI = .969, and RMSEA = .040 [.000, .094], and Gamma hat =. 989. Table 14 summarizes the latent means of student perceptions of feedback use at both measurement occasions in the experimental and control group.

Table 14. *Latent means for within group students' perceptions of feedback use*

Scales	**E1 vs. E2**				**C1 vs. C2**			
	E1	*E2 (Estimate)*	*SE*	*Wald p*	*C1*	*C2 Estimate*	*SE*	*Wald p*
Feedback use	.000	.296 *	.112	.009	.000	.065	.114	.646

Note. E1 and C1 was set to 0; * $p < .005$, ** $p < .05$ for Wald χ^2 test; SE = standard error;
C1 = Pretest control; C2 = Posttest control; E1 = Pretest experimental; E2 = Posttest-experimental

The mean comparisons within the experimental group showed a significant increase in student perception of feedback use (Wald $\chi^2_{(1,\, N\, =\, 130)} = 6.83$, $p = .009$) but not within the control group (Wald $\chi^2_{(1,\, N\, =\, 121)} = 0.210$, $p = .646$). Hence, students whose teacher received the feedback training significantly perceived more feedback use than students whose teacher did not receive the feedback training.

4.3.4. Follow up intercept analyses

We conducted follow up measurement invariance analyses for the non-invariant scales (perception of feedback adequacy and perception of feedback use) and weak-invariant scales (perception of monitoring, perception of scaffolding and perception of feedback delivery). More specifically, we estimated invariances within four comparisons: starting point (E1 vs. C1), changes within the control group (C2 vs. C1), changes at the posttest (E2 vs. C2), and changes within the experimental group (E1 vs. E2). All comparisons were invariant between groups at the pretest (starting point) and for the pretest and posttest within control group comparisons. Additional analyses were conducted for non-invariant comparisons: (a) scales with divergent end points (perception of feedback fairness, usefulness, acceptance & perception of feedback use), and (b) scales with incomparable end points (perception of feedback fairness, usefulness, acceptance; perception of monitoring and scaffolding). The intercept analysis for non-invariant scales explored whether the non-invariance pattern can be attributed to intervention effects or bias.

4.3.4.1. Between groups (divergent) comparisons at the posttest (C2 vs. E2)

The intercepts were further assessed for scales with divergent end points between the experimental and control group at the posttest. Table 15 summarizes the intercepts for each subscale using the metric model from the measurement invariance results (C2 vs. E2).

Table 15. *Intercept analysis for scales with divergent end points (C2 vs. E2)*

Subscales	**Intercepts at posttest**	
	Control (C2)	*Experimental (E2)*
Perception of feedback use	5.150	5.229
Fairness	4.152	4.267
Usefulness	4.854	4.929
Acceptance	4.395	4.627

Note. C2 = Posttest- control; E2 = Posttest-experimental

Table 15 reveal that the experimental group had a consistently higher intercept for each subscale (perception of feedback use, fairness, usefulness, and acceptance) than the control group at the posttest. Although further analyses are warranted, it can be argued that the intervention might have changed the intercept at the posttest for respondents in the experimental group compared to the control group.

4.3.4.2. Incomparability within the experimental group (E1 vs. E2)

It was investigated whether the scales' incomparability within experimental group suggests that the intervention might have changed the scales' intercepts. Table 16 summarizes the intercepts for each subscale using the metric model from the measurement invariance results (E1 vs. E2).

Table 16. *Intercept analysis for incomparable comparisons in the experimental group*

Subscales	**Intercept for experimental group**	
	Pretest	*Posttest*
Monitoring	4.400	4.887
Scaffolding	4.420	4.920
Fairness	4.492	4.267
Usefulness	4.985	4.929
Acceptance	4.658	4.627

Table 16 reveals that the incomparable end point between the pretest and posttest in the experimental group for student in perception of monitoring and scaffolding can be attributed to the effect of the intervention because these subscales had consistently higher intercepts at the posttest compared to the pretest. However, the intercepts decreased for the subscales measuring student perceived feedback fairness and were almost unchanged for feedback usefulness and acceptance.

4.4. Discussion

This aim of the present study was to determine whether a feedback training intervention could improve formative feedback practices of Tanzanian secondary school mathematics teachers, and the degree to which students perceived their teacher's practice as formative. The first and second research question investigated how students perceived teacher feedback on their mathematics test in terms of fairness, usefulness, acceptance, willingness to improve and affect – before and after the feedback training intervention – and if the feedback perceptions were more positive among students whose teacher received the feedback training than those whose teacher did not. Our findings show that students' perceptions of the feedback on their mathematics tests neither changed between the pretest and posttest nor that it changed within the control or experimental group. These results suggest that the intervention either did not improve the teachers' written feedback on students' tests or students did not notice such improvement. Future video analyses and detailed analyses of teacher comments on student tests could further elucidate these results.

Apart from closer examination of classroom practice through video analyses and teacher comments on students' tests, contextual factors such as large class size might have limited teachers' capacity to providing detailed comments in student tests. In fact, during the training the teachers voiced several potential barriers that might hinder them from fully implementing the training materials; most notably the large amount of teaching hours per week and large class sizes. These results support that the typical large class sizes in schools in African educational systems impose strong challenges for effective teacher assessment and feedback practices (Akyeampong, Lussier, Pryor, & Westbrook, 2013; Kitta & Tilya, 2010; Ottevanger, Akker, & Feiter, 2007). In particular, given the large classroom size the mathematics teachers could not provide detailed feedback on individual students' test (of which its perceptions were assessed) and they instead conducted a whole class feedback plenary discussion.

The third research question investigated if students of mathematics teachers who received the feedback training differed in the degree to which they perceived their teacher's monitoring, scaffolding and feedback delivery practices from students of teachers who did not receive the feedback training. Within experimental group analyses showed that students significantly differed in their perception of their teacher's monitoring, scaffolding, and feedback delivery practices between the pretest and posttest measures, while similar analyses showed no such differences for the control group. These results imply that the feedback training intervention improved those teachers' assessment practices and that students are capable of identifying changes in their teachers' assessment practices (Lizzio & Wilson, 2008). More specifically, this results replicate the findings by Van de Pol, Volman, Oort and Beishuizen (2014) who found that exposing teachers to a training of how to scaffold student learning improved their diagnostic strategies and the quality of their support to students.

The fourth research question investigated if students of mathematics teachers who received the feedback training differed in the degree to which their perception of feedback use from students of teachers who did not receive the feedback training. Results from the within group analyses support that students of mathematics teachers who received the feedback training significantly perceived more use of their teacher feedback than students of teachers who did not. Thus, student perceptions of feedback use can be improved via teacher training in effective feedback practices. It is essential that students positively perceive the feedback, because evidence from previous studies indicates that when teacher feedback is positively perceived the likelihood increases that students are willing to use the feedback formatively to close the gap between current and desired level (Pat-El et al., 2013; Neitzel & Davis, 2014; Sadler, 1989; Rakoczy, Harks, Klieme, Blum, & Hockweber, 2013; Winstone et al., 2017).

4.4.1. Theoretical contribution

The findings show that it is possible – to some extent – to improve mathematics teachers' assessment and feedback practices through interventions. Our findings support that students are capable of identifying changes in their teacher assessment practices such as effective feedback delivery, and perceived monitoring and scaffolding. Furthermore, our results replicate that it is possible to improve teacher assessment practices (Van de Pol, Oort, Volman, & Beishuizen, 2014). However, this seems to be harder for central constructs such as feedback quality (or in earlier studies, cognitive support) and easier for affective or atmospheric variables (Heinze & Reiss, 2007; Rach et al., 2013). Nevertheless, we replicate results found in other studies in a different context (i.e., formative discussion of a mathematics test) after a rather short intervention and in a very short time frame given the long chain of effects from teacher training to student perceptions. Yet, we contend that the quite immediate short-term effects of the intervention warrant further research on long-term effects in the future.

4.4.2. Limitations and implications

Even though we systematically drew our sample and applied an intervention design, our results should be interpreted bearing into mind two main limitations. First, as in other repeated measures designs, some participants were dropped from further analyses due to student absenteeism at all data-collection occasions. Nevertheless, the missing data were missing completely at random; hence, missing data did not affect the validity of our main conclusions. Second, some analyses for between group comparisons were not feasible due to non-measurement invariance and as a result within group analyses were estimated. Despite these limitations we are confident that the sophisticated analysis methods such latent mean analyses in Mplus are sufficiently robust, also taking care of invariances and missing data strengthens our main findings.

4.4.3. Conclusion and recommendations for practice

In general, the intervention was effective in two ways. First, despite the very short intervention (six hours in one day) on teacher level it was still able to generate positive effects that were noticed by the students; a result similar to that of the study by Rach et al. (2013) which applied a much more intense intervention. Second, it then follows that if such interventions could be paired with follow-up training to teachers they can substantially improve their assessment and feedback practices. As our findings clearly show that teacher assessment practices such as monitoring and scaffolding students' learning can be improved via interventions, we therefore recommend pre-service and in-service training programmes on assessment literacy to enhance teachers' formative assessment and feedback practices.

5. Impact of training on students' perceptions of errors in mathematics learning among secondary schools in Tanzania

5.1 Introduction

Formative Assessment (FA) and Assessment for Learning (AfL) received increased attention over the past three decades (Black & Wiliam, 1998) as both assessment practices stress the diagnostic purpose of assessment (i.e., improve student learning) and the active involvement of the student. Formative assessment involves certain kinds of teacher-student interactions such as 'monitoring' to track student progress and 'scaffolding' to help students improve their learning (Pat-El, Tillema, Segers, & Vedder, 2013, 2015; Stiggins, 2005) and requires that students perceive these interactions as formative and can make use of them to improve their learning (Heinze & Reiss, 2007). Classroom questioning is one of the most important assessment-like interactions (Popham, 2014) that requires teachers to possess a wide array of evidence-eliciting techniques, such as questions that tap into declarative, procedural, schematic, and strategic knowledge (Li, Ruiz-Primo, & Shavelson, 2006). In formative assessment, questions can originate from either the teacher or students. More importantly, students can ask questions that provide incidental opportunities for a teacher to conduct an 'assessment conversation' so as to gather evidence about student's level of understanding (Ruiz-Primo, 2011).

However, research into the quality of teachers' questioning suggests that a large percentage of teacher questions only require simple recall of factual knowledge (Airasian, 1997; Barnette, Orletsky, & Sattes, 1994; Gall, 1984; Kloss, 1988). For example, Gall (1970) found that 60% of teachers' questions required factual recall, 20% were procedural, and only 20% required deeper understanding by the students. Furthermore, students know that the teacher already knows the answer to the question and does not expect any original, divergent, or integrative thinking on the part of the student; thus, they do not think deeply about the answers (Gipps, 1994; Torrance & Pryor, 1998; Wade & Moje, 2000). As a consequence, much of daily classroom questioning is 'knowledge telling' rather than 'knowledge making', even though there is good evidence that when teachers use questions that require rich, divergent, higher-order thinking, enhanced learning takes place (Black & Wiliam, 2009; Black, Harrison, Lee, Marshall, & Wiliam, 2003; Gall, 1984; Wood, 1988).

From a mathematics perspective, formative assessment occurs naturally in the context of good classroom instruction (Ginsburg, 2009; Veldhuis & Van den Heuvel-Panhuizen, 2014). Mathematics learning involves students' making errors (Wagner, 1981) which can be formative if students are supported by appropriate feedback and follow-up instruction (Ingram, Pitt, & Baldry, 2015). Nevertheless, meta-analyses indicate that positive effects of FA in terms of student achievement are not easily achieved (Bennett, 2011; Veldhuis & Van den Heuvel-Panhuizen, 2014). For example, Rach, Ufer, and Heinze (2013) showed in a study on learning from errors that even though students valued the way in which their teachers' dealt with errors in their classroom, and even though students reported a low fear of making errors, many students did not use errors as a learning opportunity. The same study showed that it is far from trivial to support teachers in terms of classroom instruction that provides students with cognitive strategies to deal with errors.

5.1.1. Theoretical framework for learning from errors

The theory of negative knowledge postulates that individuals possess two complementary types of knowledge: (a) positive knowledge about correct facts and procedures, and (b) negative knowledge about incorrect facts and procedures (Minsky, 1994). Errors belong to negative knowledge. Oser and Spychiger (2005) define errors in mathematics as a fact or process that does not match a given norm. Errors may arise due to incorrect knowledge, application of incorrect procedures, and/or misconceptions. Errors in mathematics act as boundary markers, distinguishing between acceptable and unacceptable practices of doing mathematics (Sfard, 2007). However, errors are negatively perceived by both students and teachers as an unwelcome event, meaning that their effective use in promoting learning is rarely recognized or achieved (Borasi, 1994; Rach et al., 2013) and errors are rarely encouraged in mathematics classes (Heinze, 2005). Some theoretical perspectives on learning do not invite students and teachers to treat errors in a positive light. In particular, classical behaviorism states that errors should not be discussed with learners to protect them from internalizing erroneous situations (Skinner, 1961), while cognitivism insist that effective teaching should explore learners' errors and misconceptions and capitalize on them during instruction (Borasi, 1994; VanLehn, 1999).

Based on empirical studies Rach et al. (2013) propose a model describing two different ways of dealing with errors. The model postulates that four practices are essential for effective learning from errors: (1) identify or describe the error (sensibility), (2) understand the error or explain it (analysis), (3) correct the error (correction), and (4) develop strategies for avoiding similar errors in the future (prevention). These practices can be subsumed under a pragmatic outcome-oriented approach or an analytic process-oriented approach to learning from errors (see Figure 5). In practice many mathematics teachers and learners are satisfied with the pragmatic outcome-oriented approach with just error correction once an error is identified. Unfortunately, the pragmatic outcome-oriented approach is not as effective as the analytic process-oriented approach; the latter involves learning from the error through error analysis and error prevention strategies before correcting the error (Heemsoth & Heinze, 2016). The decision between the two approaches depends on students' appraisals of error situations and aspects of teacher behaviour in error situations (Rach et al, 2013; Santagata, 2004).

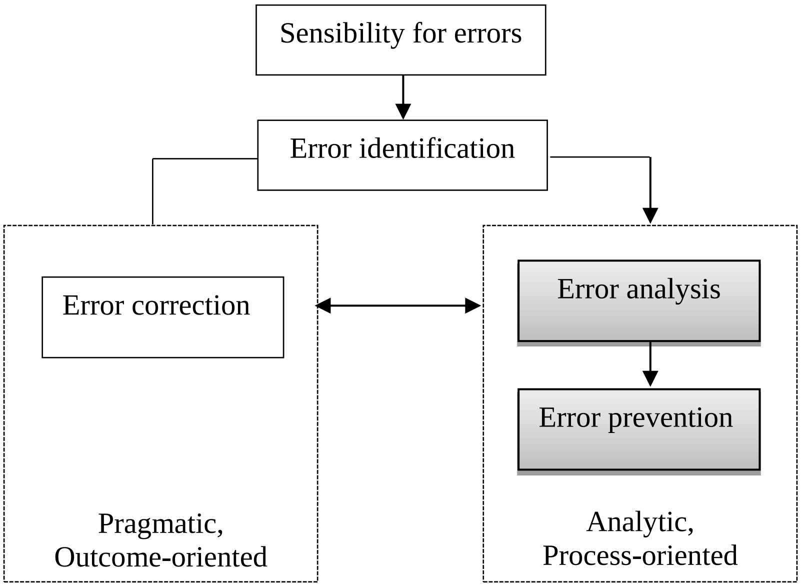

Figure 5. The model for learning from errors (adopted from Rach et al., 2013, p. 23)

Theoretical consideration on what constitutes effective feedback was guided by the Hattie and Timperley feedback model. Hattie and Timperley (2007) identify four levels of feedback, i.e. task (i.e., whether work was correct or incorrect, descriptive comments about the substance of the work), process (i.e., comments about the processes or strategies underpinning the task), self-regulation (i.e., reminders to students about strategies they can use to improve their own work) and self (i.e., non-specific praise and comments about effort). From their review, they demonstrated that task, process and self-regulation feedback all contribute to learning gains, while self-feedback generally does not. Likewise, Shute (2008) concludes in her feedback review that a formative feedback should focus on the task and be elaborated to enhance learning. Additionally, Gan and Hattie (2014) showed that feedback prompts can enhance students' knowledge of error detection. Hence, combined the process model for learning from errors (Rach et al, 2013; Heemsoth & Heinze, 2016) and the Hattie and Timperley (2007) feedback model seems a powerful approach to uncover the degree to which teachers can effectively relate to student errors as part of teacher feedback practices.

5.1.2. Learning from errors and teaching practices in mathematics

Several studies have investigated how teachers and students can use errors for effective learning (Santagata, 2004; Schwartz & Hartman, 2007; Stigler, Gonzales, Kawanaka, Knoll, & Serrano, 1999). Oser and Spychiger (2005) developed the error perceptions questionnaire to measure student learning orientation (student use of their own errors), anxiety or fear of errors, and perceived teacher support in error situations. Several studies that applies this questionnaire have reported the positive impact of error handling training on teacher's affective and cognitive behaviours (Heinze & Reiss, 2007; Rach et al., 2013) and student affective behaviours (Rach et al., 2013) but not on students' use of their own errors. Additionally, Heemsoth and Heinze (2016) conducted a student focused intervention which showed that student reflection on their own errors had the potential to improve their procedural and conceptual mathematics knowledge but not students' use of their own errors. In fact, Siegler and Chen (2008) have argued that although reflection on errors is important, reflection on errors is more demanding for students than reflection on a correct solution; which might explain the lack of effects on students' use of their own errors.

Moreover, cross-cultural studies showed both differences and similarities in mathematics instruction across countries (Leung, 2005; Santagata, 2004). For example, Stigler and Hiebert (1999) noted that individual student problem solving was more used by mathematics teachers in Japan, teacher-moderated classroom discussion was more used by mathematics teachers in Germany, and that US teachers preferred to present problems to students, explain the solution procedures and ask students to solve similar problems independently. Thus, mathematics teachers' teaching patterns differ more across countries than within a country. Despite several studies examining teaching practices in mathematics classrooms (Ingram, Pitt, & Baldry, 2015; Prediger & Erath, 2014; Roller, 2016; Santagata, 2004; Santagata & Yeh, 2014), none have focused on teaching practices in the context of whole class feedback plenary discussion of student errors in a marked test (e.g., kinds of classroom questioning by the teacher) and none have examined mathematics teaching in Africa, where teaching is heavily didactic and examination-focused (Akyeampong, Lussier, Pryor, & Westbrook, 2013; Ottevanger, Akker, & Feiter, 2007; Tilya & Mafumiko, 2010).

5.1.3. Mathematics education in Tanzania

In 1976 Tanzania introduced *Continuous Assessment* (CA) in secondary schools, which was intended to be formative. However, Basic Education Statistics in Tanzania (BEST, 2004-2013) indicates that for ten consecutive years (2004-2013) an average of 79% of secondary schools students failed their mathematics national examinations. Such persistent failure still happens even though mathematics teachers reported to practice various formative assessment practices. For example, Kyaruzi, Strijbos and Ufer (2015) noted that 93% of the 54 mathematics teachers in their sample self-reported that they provide feedback on students' mathematics tests. Yet, students in focus group discussions in Chapter 2 voiced some dissatisfaction with their mathematics teacher's feedback practices. This suggests that despite the assumed formative nature of current mathematics assessment practices in Tanzanian secondary schools, teachers' feedback practices and/or students' utilization of feedback is suboptimal.

5.1.4. The present study

Although evidence from Kyaruzi, Strijbos and Ufer (2015) indicated that the majority of mathematics teachers corrected students' errors after the test in the form of whole class feedback plenary discussions, little is known about how formative those discussions and interactions were. To overcome this, an intervention for improving mathematics teachers' error handling practices was developed. This study seeks to establish whether a teacher-level intervention to improve their scaffolding of student learning from errors, affected students' perceptions of their teacher's error handling practices and their use of student errors for learning in whole class discussions. More specifically, the following research questions were examined:

1) What is the effect of the intervention on students' perception of their teacher's support in error situations?
2) What is the effect of the intervention on students' perception of their individual use of errors in learning and their anxiety in error situations?
3) Which practices of dealing with errors can be identified in lessons with a formative whole-class plenary discussion of student performance on a mathematics test?

5.1.5. Hypotheses

Based on the reviewed literature, the following hypotheses were derived. First, we hypothesized that students in the experimental group will perceive more teacher support in error situations after the intervention than students in the control group (Hypothesis 1). Since research evidence show that it is difficult to foster student use of their own errors, we expect small effects in student use of their own errors for learning (Hypothesis 2a), but expect less anxiety in error situations for student in the experimental group than students in the control groups (Hypothesis 2b). Since research evidence shows that mathematics teacher's teaching patterns are more similar within a country than across countries (Santagata, 2004; Stigler, Gonzales, Kawanaka, Knoll, & Serrano, 1999), and also Slavin (2008) shows that it is difficult to change teacher teaching practices through an intervention of less than 12 weeks. Therefore, we expect only small changes – if at all – in mathematics teachers' general pedagogical approach to feedback discussion (Hypothesis 3).

5.2. Method

5.2.1. Participants

The study was conducted in the Dar es Salaam region of Tanzania. The region was sampled because according to statistics from the National Examinations Council of Tanzania (NECTA, 2014), secondary schools in the Dar es Salaam region (M = 1.64, SD = 0.63) did not deviate from the overall national mathematics performance (M = 1.55, SD = 0.65). Based on NECTA (2014), the Dar es Salaam region had 191 secondary schools with (≥ 40) students, classified according to performance as: 10 (5%) high performing, 173 (91%) middle performing, and 8 (4%) low performing schools.

Because error analysis has been reported as having a greater benefit for lower performing students, this study was conducted among four high and four low performing schools. In sampling schools, we prioritized schools with mixed gender (boys and girls) to maximize gender representativeness and fifteen schools (7 high and 8 low) performing schools met this criterion. In total eight schools were sampled: four high performing and four low performing schools. One Form three mathematics class (Grade 11) was randomly sampled from each school and all students in the sampled class were invited to freely participate in the study. The sample consisted of eight classrooms resulting in eight mathematics teachers and their respective Form three (Grade 11) students (N = 251). Table 17 provides a description of students' and teachers' personal and school characteristics.

Table 17. *Demographics of participating students and teachers*

Demographic	Total	Control	Experimental
Students	251	121	130
Gender			
Male	135	68	67
Female	116	53	63
Age	16.29 (0.95)	16.07 (0.84)	16.49 (1.00)
Male	16.42 (0.93)	15.81 (0.68)	16.41 (1.07)
Female	16.14 (0.96)	16.28 (0.91)	16.57 (0.94)
School performance			
High	135	67	68
Low	116	54	62
Teachers	8	4	4
Gender			
Male	7	4	3
Female	1	0	1
Age	42.50 (9.38) range: 32-57	41.25 (3.95) range: 38-47	43.75 (13.62) range: 32-57
School performance			
High	4	2	2
Low	4	2	2
Highest qualification			
Bachelor degree	6	2	4
Diploma in education	2	2	0
Teaching load/ hours per week	22.75 (9.38)	26 (9.78)	19.50 (9.00)
Class size	64.37 (28.15)	52.00 (15.41)	76.00 (34.91)

Note. Mean (standard deviation)

5.2.2. Design

A quasi-experimental repeated measures within-subject pretest, intervention (training) and posttest design was conducted. At the pretest and posttest the teachers' plenary feedback discussion of a mathematics test was videotaped with two cameras. Two teachers from each school-performance category (high, low) were randomly assigned to the experimental and control group. Since we used intact classes, this also formed two student groups: experimental (N = 130) and control (N = 121). The four mathematics teachers in the experimental group participated in a one-day intensive feedback training (FBT) for improving their feedback practices and formative use of students' errors. To ensure equity, the four mathematics teachers in the control group were provided with the same one-day feedback training after the posttest data collection. The time interval between the intervention (training) and posttests measures was approximately one month. Figure 6 summarizes the overall research design.

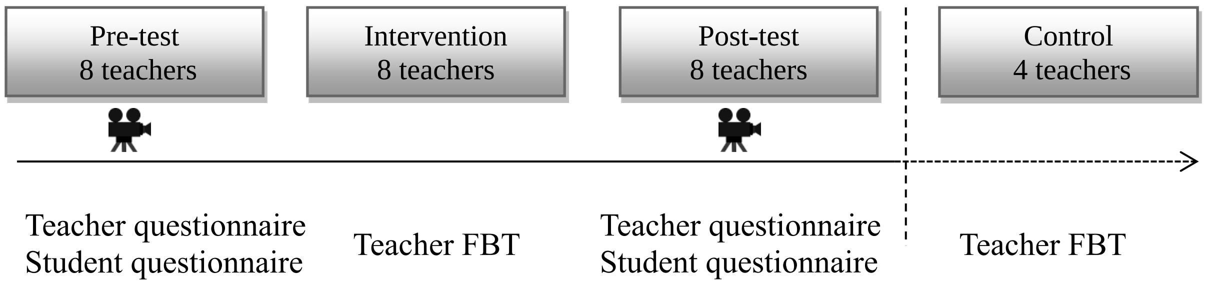

Figure 6. General research design (Note. = videotaped lesson, FBT = Feedback training)

5.2.2.1. Intervention

The feedback training (FBT) contained theory and practical sessions focusing on how mathematics teachers can improve and provide feedback using the Hattie and Timperley (2007) model, and how to apply the cognitive strategy for effective learning from errors (Heemsoth & Heinze, 2016; Rach et al., 2013). More specifically, mathematics teachers practiced how to provide feedback at the task level, process level and self-regulation level (Hattie & Timperley, 2007). With respect to the cognitive strategies for learning from errors, teachers brainstormed and practiced how to identify student errors, understand student errors (why errors), correct student errors, and how to develop strategies for avoiding similar errors. The training used typical samples of written mathematics feedback on students' tests collected from mathematics classes during the pretest (see Appendix L for a detailed description of the training content, activities, and materials).

5.2.3. Instruments

5.2.3.1. Questionnaires

At the pretest and the posttest, a questionnaire was administered to measure student perceptions of errors and perceived teacher support in error situations (Spychiger, Küster, &

Oser, 2006) comprising three subscales: anxiety in errors, student learning orientation, and teacher support in error situation. At the posttest and pretest students also completed a short questionnaire to measure the fidelity of their teacher's feedback plenary discussion (Jacobs et al., 2003). All scales were previously validated, adapted to the mathematics context and measured on a balanced and symmetrical 6-point scale ranging from *completely disagree* (1) to *completely agree* (6). It is noteworthy that anxiety, teacher support in error situations, and authenticity of feedback delivery were below the .70 threshold for Cronbach's alpha, suggesting that revisions to the scales were needed to accurately reflect the responses of the participants. See Appendices M and O for the detailed questionnaire items. Table 18 summarises the adopted scales, number of items per scale, a sample item per scale, and the Cronbach's α from the original studies and Cronbach's α for this study.

Table 18. *Sample items and scales' Cronbach's α*

Scale	**k**	**Sample item**	**Cronbach's α**		
			Original Study	*Present Study*	
				Pretest	*Posttest*
Learning orientation	8	If I do something wrong in mathematics class I perceive this as an opportunity to learn.	.71	.75	.76
Anxiety	5	I feel ashamed when I make a mistake in front of the class in mathematics.	.78	.49	.54
Teacher support in errors	7	If I make a mistake in mathematics class, my teacher discusses it with me in a way that I really learn from it.	.79	.65	.56
Authenticity of feedback plenary	4	Was the videotaped plenary feedback discussion typical/ representative for the lessons your teacher normally teaches?	-	.62[a]	.59[a]
Perception of feedback discussion	5	After this plenary feedback discussion, I now know how I can correct most of my mistakes.	-	.81	.82

Note. k = number of items per scale, a- after removing one negatively phrased item

5.2.3.2. Video recording

Two video cameras; teacher and student focused camera were used to collect data using the guidelines recommended by TIMSS 1999 (Jacobs et al., 2003) and the IPN Study (Seidel, Prenzel, Dalehefte, Meyer, Lehrke, & Duit, 2005).

5.2.4. Procedure

The study was conducted with research clearance from the University of Dar es Salaam. All participants signed a consent form. Students filled in questionnaires after each lesson. Questionnaires were administered before (pretest) and after (posttest) the intervention. Although 326 respondents answered the questionnaire during the pretest, 61 students did not participate in all data collection points and were eliminated from the study. Furthermore, 14 respondents with more than 10% missing data per session were also removed from analysis, leaving the final sample with 251 respondents. The remaining data were considered to be missing completely at random (MCAR) because Little's MCAR test was not statistically significant ($\chi 2$ = 10190.60, df = 32739, p = 1.00) (Peugh & Enders, 2004). Missing values were imputed with the expectation maximisation (EM) procedure which is considered to be an effective imputation method when data are completely missing at random (Musil, Warner, Yobas, & Jones, 2002).

5.2.4.1. Measurement models

As each scale was adopted from a previously validated inventory, measurement models were tested using confirmatory factor analysis. The interpretation of model fit was based on the following indicators: root mean squared error of approximation (RMSEA) and standardized root mean residual (SRMR) below .05 and comparative fit index (CFI) and gamma hat values above .95 indicate good fit (Byrne, 2010), while RMSEA and SRMR below .08 and CFI scores above .90 indicate acceptable fit (Hu & Bentler, 1999). Because the Chi-square statistic is overly-sensitive in large sample sizes above 250 (Byrne, 2010), we report multiple fit indices. The CFI and RMSEA are not stable estimators because CFI rewards simple models while RMSEA rewards complex models (Fan & Sivo, 2007). The Gamma hat statistic and SRMR have been shown to be stable estimators (Fan & Sivo, 2007). As Steiger (1990) recommends, we report the 90% confidence interval for the RMSEA.

The measurement model for the three inter-correlated scales measuring students' 'Learning orientation', 'Anxiety in error situations', and 'Teacher support in error situations' had relatively poor fit from the pretest data (i.e., SRMR = .054, CFI = .825, RMSEA = .067 [.057, .077]). By removing eight items with a poor factor loading (< .40) the fit improved substantially (i.e., SRMR = .047, CFI = .95, Gamma hat = .98 and RMSEA = .054 [.034, .072]). The same model had also a good fit at the posttest (i.e., SRMR = .053, CFI = .93, Gamma hat = .96 and RMSEA = .067 [.050, .085]). The measurement model for students' perceptions and authenticity of the feedback plenary discussion (Jacobs et al., 2003) had poor fit at the pretest (i.e., SRMR = .076, CFI = .88, RMSEA = .122 [.101, .144]. Removing two items with low contribution improved fit (i.e., SRMR = .082, CFI = .90, Gamma hat = .92, RMSEA = .149 [.120, .180]). The same model at the posttest had similar acceptable fit (i.e., SRMR = .073, CFI =. 93, Gamma hat = .94, RMSEA = .128 [098, .154]). Although the quality of model fit was mixed, all scales were used as specified while remaining aware that the validity of inferences is affected by this level of fit to the data.

5.2.4.2. Measurement invariances

Measurement invariance is a prerequisite of comparison between groups and measurement occasions (McArdle, 2007; Reise, Widman, & Pugh 1993; Vandenberg & Lance, 2000; Wu, Li, & Zumbo, 2007). Hence, invariance tests were conducted to generate evidence for scales' comparability. The invariance of a model across the measurement occasions is evaluated by establishing whether fixing model parameters (e.g., factor regression weights, covariances, factor intercepts, or residuals) as equivalent results in a statistically significant difference in model fit (Cheung & Rensvold, 2002; Vandenberg &

Lance, 2000). To demonstrate measurement invariance, the difference in χ^2, taking into account the difference in *df*, should not be statistically significant (i.e. $p > 0.05$) and/or the difference in CFI should be $\Delta CFI \leq 0.01$ (Brown & Chai, 2012). All measurement models were strongly invariant (see Appendix P), hence latent mean analyses were feasible.

5.2.5. Analyses

5.2.5.1. Survey analyses

Having established a plausible set of scales, latent mean comparisons were used to evaluate differences between groups and measurement occasions. Latent mean analyses (LMA) were used to assess the difference in scale means relative to the pretest-control group as reference group (set to 0). The approach provides a stronger framework to account for response bias and takes into account random or non-random measurement errors (Comşa, 2010; Marsh, Guo, Parker, Nagengast, Asparouhov, Muthén & Dicke, 2017; Sass, 2011). In LMA the mean of a latent factor is computed and differences of other groups with a similar latent factor are estimated as z-score differences to the reference group (Hussein, 2010). The Wald test of parameter constraints was used to assess whether the differences in latent means were significant or not.

5.2.5.2. Video analyses

An initial inductive analysis of 50% of the videotaped lessons was performed to extract common patterns in mathematics teachers' pedagogical approaches to feedback plenary discussion. Three main approaches were identified: (1) student-centred approach (inviting students to do corrections on the blackboard), (2) teacher-centred approach (teacher doing corrections on the blackboard), and (3) individual student marking scheme approach (teacher provides a marking scheme to each student). These three approaches were subsequently used to code teacher's practices. To determine the accuracy of coding, four teacher videos from two schools were randomly sampled from the experimental and control group and analyzed by a second coder (a doctoral student). This resulted in 87.5% agreement leading a Krippendorff's alpha of .72 with a lower and upper limit of 0.00 and 1.00, respectively. The remaining video data were analyzed by the lead author.

Apart from the observed common patterns on how mathematics teachers conducted the feedback plenary discussions, excerpts from the videotaped lessons at the pretest and posttest were selected to illustrate teacher-student error handling strategies (i.e., describing the error, analyzing the error, correcting the error, and creating situations for preventing more errors) were implemented in lessons.

5.3. Results

5.3.1. Authenticity of classroom practice

Since videotaping can disrupt normal teaching patters, we assessed the authenticity of the videotaped lesson and students' perceived usefulness of the feedback plenary discussion. In general, at pretest students in the experimental (M = 4.88, SD = 1.16) and control (M = 5.01, SD = 1.05) groups were positive about the authenticity of the videotaped feedback plenary discussions indicating that students perceived the videotaped lessons to reflect the normal mathematics lessons. Likewise, students in the experimental (M = 5.45, SD = 0.89) and control group (M = 5.60, SD = 0.58) groups perceived the feedback plenary discussion to be useful. At posttest, students in the experimental (M = 4.98, SD = 1.05) and control (M = 4.93, SD = 1.04) groups were also positive about the authenticity of the videotaped lessons. Similarly, at posttest, students in the experimental (M = 5.43., SD = 0.79) and control group (M = 5.40, SD = 0.92) groups perceived the feedback plenary discussion to be useful.

5.3.2. Student perceptions of errors and perceived teacher support in error situations

Table 19 summarizes the scales' manifest and latent means. The descriptive statistics show that with the exception of 'anxiety in error situations', raw means were close to or above mostly agree (5.00) suggesting that the students had a positive learning orientation and perceived their mathematics teachers' as supportive in error situations. By chance, the experimental groups were somewhat more positive at the pretest for learning orientation and less positive anxiety and teacher support in error situations than the control group. By the end of the intervention, however, the experimental group had a higher score for both learning orientation and teacher support in error situations and a lower score for anxiety than both the pretest and posttest scores in the control group. However, the gains for the experimental group over the control group were quite small, with only the difference in student perception of teacher support in error situations being statistically significant over the control pretest group.

Table 19. *Manifest and latent means for scales*

	Descriptive					**Control**		**Experimental**	
Scales	*Pretest*		*Posttest*			*Pre*	*Post*	*Pre*	*Post*
	M	*SD*	*M*	*SD*	*d*				
1. Learning orientation	4.89	0.89	5.03	0.86	.16	—	.172	.081	.236
2. Anxiety in error situations	2.14	1.13	2.13	1.18	-.01	—	-.079	-.111	.084
3. Teacher support in error situations	4.73	1.28	5.01	1.02	.24	—	.144	-.161	.264**
4. Mathematics performance	48.93	23.69	47.10	25.31	-.12	—	—	—	—

Note. ** $p < .01$ for Wald χ^2 test

Next, within group comparisons were conducted for student perception of teacher support in errors situations. The change in student perceptions of teacher support in error situations within the experimental group was moderate (z = .356, Wald $\chi^2_{(1, N = 130)} = 10.86$, $p < .005$), while the change within the control group was not statistically significant (z = .139, Wald $\chi^2_{(1,\ N\ =\ 121)} = 1.097$, $p = .30$). Thus, students in the experimental group changed moderately and became more positive in their perception that their mathematics teacher was error friendly and used errors formatively.

With reference to Table 19, it is evident that students' performance in their mathematics tests at pretest (M = 48.93, SD = 23.69) and posttest (M = 47.10, SD = 25.31) were slightly below the expected mean score of 50% but with higher standard deviations. However, there was no gain in performance. Furthermore, overall the correlations of the three student perceptions scales to mathematics performance were either statistically not significant or very weak (i.e., r < .20) (Table 20). The overall effect of anxiety in error situations was negatively related to student mathematics performance at the posttest in the control group but not in the experimental group before or after the intervention (Table 20).

Table 20. *Correlations between the scales at pretest and posttest within the control and experimental group*

Scales	Control				Experimental			
	1	2	3	4	1	2	3	4
1. Learning orientation	—	-.61**	.57**	.15	—	-.59**	.52**	.06
2. Anxiety in error situations	-.49**	—	-.39**	-.22*	-.41**	—	-.24**	.03
3. Teacher support in error situations	.49**	-.37**	—	.11	.41**	-.27**	—	.11
4. Mathematics performance	.01	.05	.07	—	.13	.06	.06	—

Note. Below diagonal = pretest correlations, Above diagonal = posttest correlations, ** p < .01, * p <.05

5.3.3. Teacher practices of handling student errors

Qualitative analysis of videotaped lessons showed that mathematics teachers employed three main pedagogical approaches to feedback plenary discussions. The student-centred approach was observed in 8 of 16 (50%) lessons; four lessons in the experimental group and four in the control group. In the student-centred approach the teacher invited and encouraged students to solve mathematical questions on the blackboard. The teacher provided students with scaffolding support only if most students failed to solve the question. The teacher-centred approach was identified in 6 of 16 (38%) lessons; four lessons in the experimental group and two in control group. In the teacher-centred approach the teacher personally solved questions on the blackboard, mainly with little student involvement. The individual student marking scheme approach was observed in 2 of 16 (12%) lessons; both from the same teacher. In this approach the teacher provided each student with a marking scheme for the purpose of self-corrections. In general, despite the intervention, teachers in the experimental group maintained their general pedagogical approach during the posttest. Nevertheless, exploratory case studies of teacher practices showed improvement at posttest.

5.3.4. Exploratory case studies of feedback plenary discussions

To gain more insight in the actual feedback plenary discussions, we performed two case studies to illustrate some differences in the application of error handling strategies by teachers in the experimental and control group. The two case studies were randomly chosen one from each group. The same teacher was observed at pretest and posttest. In the observed case studies, one mathematics task from a *functions and relations* topic that was common in lessons of both teachers was analysed. Tables 21 and 22 contain excerpts of the error handling strategies employed by the same teacher (Teacher 1) in the experimental group.

Table 21. *Error handling practices (Teacher 1) experimental pretest*

Time	Activity	Error Strategy
01:52	Question1: Given the relation, R= {(a, m), (b, m), (c, m), (d, n), (c, n)}. Find: (a) the domain and range of R, T: We know that a domain is represented by the first entry, and range is denoted by the second entry. T: Domain = {a, b, c, d}. T: Range is given by the second entry. What are the second entries? T: Range = {m, n} Question 1(b): Draw the pictorial representation of R. T: (The teacher draws the pictorial representation of R)	-
05:37	If R= {(x, y): y = 2x-3, x ∈ *R*}. (a) Find the domain and range (b) Draw the graph of R. T: Thus is a linear relation with no boundaries. T. For a linear function with no boundaries, domain will be what?	-
06:21	S: Domain will be 0, S2: Domain ={x: x ∈ *R*}. T: Very good: T: When writing the solution use mathematical notations, *some of you were using words*. Of course it is fine, but always use mathematical notations.	1
06:50	T: What about the range? (Looking at students) *if you got it correctly tell what you wrote* S: Range ={y: y ∈ *R* } Question (2b) Draw the graph of R= {(x, y): y = 2x-3, x ∈ *R*}.	-
08:35	T: This is a straight line, you need only two points and then you join them by using a ruler. *I will not use the intercepts because we will get fractions which are difficult to plot.* (The teacher guides students to draw the graph)	4

Table 22. *Error handling practices (Teacher 1) experimental posttest*

Time	Activity	Error Strategy
04:20	Question 1: Given the relation, R= {(x, y): y≤x+1, 0≤ y≤ 2, x ≤ 4}. Find (a) R^{-1} (inverse of R), (b) Draw the graph of R. T: Do you remember the principle? If you want to find the inverse of R, first, interchange x and y, then make y the subject. Don't alter the inequalities the inequalities remain as they are. T: The teacher writes: R^{-1}= {(x, y): x-1≤y, 0≤ x≤ 2, y ≤ 4}. T: *Some of you treated each part of the R independently* as: R^{-1} = {(x, y): x-1≤y}, R^{-1}= {(x, y): 0≤ x≤ 2}, R^{-1}= {(x, y): y ≤ 4}. That is wrong. I asked you where is R^{-1}?	-
06:30	T: *You can work each part independently but at the end you are supposed to write R^{-1} as one set.* (b) Draw the graph of R, R= {(x, y): y≤x+1, 0≤ y≤ 2, x ≤ 4}.	2
07:26	T: When we are dealing with inequalities it means we are going to compare the lesser side and the greater side. So you must have the boundaries. T: Is the boundary included or excluded?	3
11:36	S: Included T: Why it is included? S: Included because of the equal sign in y ≤ x+1	-
22:07	T: The teacher draws the graph involving students.	-
22:40	T: This is what you were supposed to do. Many of you were having problem. T: *Drawing the graph of R^{-1}, use similar procedures as we used for R.*	4
24:04	T*: You were supposed to find the domain of R but many you solved for the domain of R^{-1}*	1

Note. S = Student, T = Teacher; Error strategies: 1 = Describe error, 2 = Explain error, 3 = Correct error, 4 = Prevent error (generalize)

At pretest the teacher explicitly described the error made by students (*"some of you were using words instead of mathematical terms"*) and highlighted a situation where students were more likely to make more errors *("I will not use intercepts because we will get fractions which are difficult to plot"*). During the posttest, the same teacher integrated more error handling strategies. First, the teacher described a student error by citing specific errors made by students in the test (*"some of you treated each part of the R independently, many of you solved for the domain of* R^{-1} *instead of domain of R"*). Secondly, the teacher corrected student errors (*"You can work each part independently but at the end you were supposed to write* R^{-1} *as one set"*) and generalized the solution strategy to other questions (*"to draw the graph of* R^{-1}*, use similar procedures as we used for R'*). Generally, at posttest, although the teacher's pedagogical approach did not change completely, the teacher appears more error friendly and used more strategies for learning from errors than at pretest.

Tables 23 and 24 contain excerpts of the error handling strategies employed by the same teacher (Teacher 7) in the control group during the pretest.

Table 23. *Error handling excerpt from the control group at the pretest (Teacher 7)*

Time	**Activity**	**Strategy**
14:30	Qn2: Draw the graph of the inverse of the relation R= {(x, y): x+y≤0, y≥x-1} and state the domain and range.	-
16:11	S: To find $R^{-1,}$ the first step you interchange x and y variables.	-
14:20	T: Yes, correct	-
16:58	S: Then you make y the subject, no, make x the subject	-
16:59	T: Make subject x or y?	-
17:00	S: Other students-says, make y the subject	-
17:18	T: Are you making the subject x or y? You make y the subject	-
17:01	S: R^{-1}= {(x, y): y≤-x, y ≥x+1}.	-
17:15	T: Correct.	-
19:36	S: The next step is to draw the table of values	-
20:15	T: *Most of you were confusing on drawing the graph and shading the required area.*	1
30:52	T: Will it be a smooth or dotted line (inclusion or exclusion of boundary points)?	-
31:00	S: Smoothen line	-
30:00	T: Why smooth line?	-
30:30	S: Because there is = in ≤ (< or =)	-
31:40	T: Yes. We draw a smooth line because of ≤	-
31:56	S: So we have to test for the required region.	-
32:00	T: Who can give us a point to test the required region?	-
32:28	S: Use (0,0)	-
32:40	T: *We cannot use (0, 0) because it is a point on the line, choose another point below or above the line please.*	2
32:40-	S: (A student correctly draws the graph and shades the required region	-
44:00	S: Student says domain represent all real numbers of x	
44:36	T: Are you convinced that domain is all real numbers?	-
42:02	*T: You were supposed to shade the area that satisfies both graphs*	1
47:00	T: Based on the graph, Domain = {x: x ≤ 0.5}	-
47:32	T: Why should we use ≤ and not ≥?	-
49:20	S: Because from the graph all values of x are less than 0.5	-
49:28	S: Range is all real numbers of y.	-
49:47	T: Thank you for your presentation, clap hands for him.	-
50:05	T: Was there any reason for those who scored 0, 5 or 20%	-
50:42	T: Student X what was a problem for you?	-
50:53	SX: I didn't understand question number 2	-
51:00	T: Did you attend the class when I taught, function? If you don't understand a lesson ask me or ask your fellow.	-
51:30	T: Student Y, you were supposed to get 100% but you got 70%, what was the problem?	-
51:43	SY: I did not understand question number one.	-
52:22	T: *Some of you drew the graph of R instead of* R^{-1}	1

Note. S = Student, T = Teacher; Error strategies: 1 = Describe error, 2 = Explain error, 3 = Correct error, 4 = Prevent error (generalize)

Table 24. Error handling excerpt from the control group at the posttest (Teacher 7)

Time	Activity	Strategy
	Draw the graph of a function $f(x) = \begin{cases} 2x+1, & x \leq 2 \\ 5, & 0 < x \leq 4 \end{cases}$ and state the domain and range.	
26:50	S: We use table of values to find points for plotting a graph.	-
33:00	S: In drawing the second part of the graph, 0 is exclusive.	-
33:29	T: Yes, how do you indicate that?	-
33:50	S: Indicated by the open circle above the closed circle.	-
33:55	T: Very good, that's how it should appear.	-
34:15	S: From the graph domain and range were indicated.	-
35:35	T: Yes, that is how it was supposed to be.	-
	T: Clap for all who presented on the board.	
36:00	T: What was wrong, all of you were supposed to get 100%, what was the problem?	-
	S: Time was limited	
36:20	T: No, that is not true. *You don't revise your notice*.	-

Note. S = Student, T = Teacher; Error strategies: 1 = Describe error, 2 = Explain error, 3 = Correct error, 4 = Prevent error (generalize)

At pretest the teacher employed some error handling strategies such as describing student errors (*"some of you drew the graph of R instead of* R^{-1}*"*), explaining why it is an error (*"we cannot use the origin (0,0) to test our inequality because it is a point in the line"*), however, these practices did not appear during posttest. Secondly, at the posttest the teacher seemed to be error unfriendly and blamed students for their poor performance (*"you don't revise your notice"*). Generally, the teacher in the control group (teacher 7) was somehow aware of error handling strategies such as describing the student error before correcting them, although he could not sustain them at posttest as was the case for the teacher in the experimental group (teacher 1).

5.4. Discussion

This study examined students' perceptions of their teacher's feedback and support in error situations. It was also investigated whether a one-day teacher professional development intervention brought about visible changes in how teachers helped students deal with errors. The first research question investigated the effect of the intervention on students' perception of their teacher's support in error situations. Based on the mean score, students perceived their teacher's support in error situations as above or close to mostly agree, implying that students in both the control and experimental group had positive perceptions of their teacher's support in error situations. At posttest, the experimental group had a statistically significant, but moderate positive change in perceived teacher support in error situations, whereas there was no other visible difference associated with the intervention. These results confirm Hypothesis 1 and replicate the results from previous studies that teacher support in error situations can be improved by interventions (Heinze & Reiss, 2007; Rach et al, 2013).

The second research question investigated the effect of the intervention on students' perception of their individual use of errors in learning and their anxiety in error situations. The results show that students in our sample did not fear errors and considered to use errors for learning. However, the gains for the experimental group over the control group were quite small, with neither significant difference in student anxiety in error situations nor in their learning orientation over the control pretest group. Thus, the intervention did not improve students' use of errors in learning and their anxiety in error situations; Hypothesis 2a is rejected. The results with respect to students' use of errors supports that the positive impact on students' use of errors for learning is not easily achieved (Heinze & Reiss, 2007; Rach et al., 2013). The challenge in stimulating student learning from own errors is in line with the evidence that student learning from own errors is more challenging than learning from correct situations (Siegler & Chen, 2008). Furthermore, although Rach et al (2013) showed that it is possible to improve student anxiety in error situations via error handling interventions, similar to the study by Heinze and Reiss (2007) our results do not provide evidence for this. Thus, students in both the control and experimental groups did not significantly change their anxiety in error situations over the pretest groups; Hypothesis 2b is rejected. Finally, unlike Rach et al. (2013), Borasi (1994) and Heemsoth and Heinze (2016) who showed that promoting an error friendly environment may lead to productive learning outcomes, our data does not provide evidence for this. This might be accounted for by the small sample of schools and the non-uniformity of the teacher-made mathematics tests.

The third research question aimed to uncover which practices of dealing with errors can be identified in lessons with a formative whole-class plenary discussion of student performance on a mathematics test. Exploratory case studies illustrated some differences between teachers in the experimental and control group. The teachers in the experimental group demonstrated more error handling strategies – such as citing more student errors in the test, describing why the error occurred and generating situations where similar error might occur – than teachers in the control group. The exploratory case studies of excerpts from videotaped feedback plenary discussion at the pretest and posttest illustrate that the potential of interventions to affect teacher practices (Van de Pol, Oort, Volman, & Beishuizen, 2014).

Generally, as cross-cultural studies support that mathematics instructions have similarities within countries (Santagata, 2004, 2005; Leung, 2005), mathematics teachers in our sample consistently employed teacher-centred and learner-centred approaches to feedback plenary discussions. Predominantly, the teachers' approach to feedback plenary discussion was teacher-centred, which supports previous studies from Tanzania (Kitta & Tilya, 2010; Njabili, 1999; Tilya & Mafumiko, 2010) that even though the curriculum emphasizes learner-centred approaches, teaching remains to be teacher-centred. These practices are culturally rooted and can be hardly changed with a short intervention. Future research could investigate whether a longer and more intensive training as well as continued support during the intervention improve their practices.

5.4.1. Theoretical contribution

The study showed that it is possible – to some extent – to improve mathematics teachers' practices of students support in error situations via a short-term intervention. Also, the findings support that students are readily capable of identifying the qualities of assessment practices they do and do not value (Lizzio & Wilson, 2008). Finally, although it is difficult to change the main teacher pedagogical approach, the exploratory case studies illustrate that a short-term intervention can provide visible qualitative improvements in teacher error handling strategies.

5.4.2. Limitations and implications

Although we systematically drew our sample and applied an intervention design, the results should be interpreted bearing into mind several limitations. First, the intervention was conducted among few schools and only involved eight teachers. Second, even though schools were randomly assigned to experimental and control groups, the number of sampled schools and teachers limits generalisations beyond our sample. Third, the duration of the intervention was below the 12 weeks recommendation by Slavin (2008) for effective changes in teacher practices.

5.4.3. Conclusion and recommendations for practice

Based on our findings we encourage teachers and students to use student errors formatively to improve the instructional process. Moreover, teachers are encouraged to utilize the analytic model for learning from errors; in particular linking their feedback discussions to typical examples of student errors that were observed when marking tests. Teachers could be encouraged to compute statistics to students on the success/failure rate for each test-question to help them identify areas where students need more help. As indicated by our findings, the intervention mainly improved teachers' error handling practices to the extent that their improvement was noted by students. Future studies may investigate effective interventions for improving individual student use of errors for learning and reducing student anxiety in error situation.

6. General discussion and conclusions

Formative assessment (FA) and Assessment for Learning (AfL) as well as feedback are essential practices for student learning (Black & Wiliam, 1998; Hattie & Timperley, 2007). The change from a testing culture to an assessment culture emphasizes that assessment should be an integral of the teaching and learning process (MacLellan, 2001; Smyth, 2004). The Assessment Reform Group (ARG, 2002) first proposed ten AfL principles to assist teachers' implementation of assessment for learning. Among other practices, the principles stressed that assessment should be adaptive and constructive. Drawing from the ten AfL principles, five AfL practices have been emphasized in the FA and AfL literature: sharing goals and quality criteria, feedback, classroom questions, peer and self-assessment, and formative use of summative tests (Black & Wiliam, 2009; Popham, 2014). These five AfL practices perform two main functions: scaffolding and monitoring student learning (Stiggins, 2005; Pat-El, Tillema, Segers, & Vedder, 2013, 2015). Subsequent research has shown, however, that each of these practices may be subject to partial or incomplete assumptions. For example, feedback is often superficial, self-oriented and not effective, and teachers do not always ask good questions or lack assessment literacy (Airasian, 1997; Barnette, Barnette, Orletsky, & Sattes, 1994).

Moreover, several studies show that FA and AfL as well as feedback practices are not always well perceived by students and teachers (Butler & Winne, 1995; Gibbs & Simpson, 2003; Jonsson, 2013; Pat-El et al., 2013, 2015; Sadler, 1989). While the constructive alignment of assessment perceptions between student and teachers is highly emphasized for improved learning outcomes (Biggs, 1996), this is not easily achieved (Pat-El et al., 2013, 2015). In fact, formative assessment is a two-way process demanding that both teacher and students use assessment information with the aim of how to improve *current* assessment practices (Pat-El et al., 2015). Improving *current* assessment practices is the main difference between formative assessment and summative assessment, because formative assessment stresses the ‘right now’ and improvement of what is being taught or learned (Popham, 2014).

Assessment is supposed to be a significant part of teaching and learning practices in secondary schools in Tanzania. In secondary schools there exists school-based Continuous Assessment (CA) which is supposed to serve the two previously mentioned functions of monitoring and scaffolding student learning and also partly contribute to the students’ final summative assessment. It is not clear how students and teachers perceive school-based assessment practices in mathematics in Tanzanian secondary school, notwithstanding that there is a high failure rate in both school-based assessment and external summative assessments. This dissertation investigates mathematics teachers’ assessment practices and students’ perceptions of these practices in mathematics education among secondary schools in Tanzania and how they can be improved. More specifically, both student and teacher perceptions of FA and AfL were studied in light of assessment theory (Black & Wiliam, 2009), feedback theories (Hattie &Timperley, 2007; Kluger & DeNisi, 1996) and the cognitive strategy for learning from errors (Heemsoth & Heinze, 2016; Rach, Ufer, & Heinze, 2013). Two empirical studies were conducted as part of this dissertation. These studies addressed two general research aims:

(1) To investigate Tanzanian secondary school mathematics teachers’ perceptions and practices of formative assessment, as well as students’ perceptions of their teachers’ practices (Study 1 reported in Chapter 2 and 3).

(2) To investigate the impact of an intervention on feedback processes during mathematics education in secondary schools in Tanzania, i.e., feedback provided by the teacher, how this is perceived by students, and whether it is applied by them (Study 2 reported in Chapter 4 and 5).

The following sections provide a summary of both two studies across four thesis chapters and their respective findings, followed by an integrated discussion, and conclusions across the studies. The chapter concludes with methodological limitations, implications for theory and practice, and directions for future research.

6.1. Summary of studies

6.1.1. Chapter 2: Student perceptions and practices of AfL and mathematics performance

The effect of students' perceptions of their mathematics teachers' AfL practices on their feedback use and mathematics performance was investigated by means of structural equation modeling. A total of 2767 Form three (Grade 11) students were sampled from 48 secondary schools in the Dar es Salaam and Kilimanjaro regions. To achieve a representative sample, sampling of schools used multi-stage random sampling involving multiple criteria: school performance (high, middle and low); school ownership (government vs. private); school location (rural vs. urban), gender of respondents (boys vs. girls), and school class size (below or above 40 students per class). Students responded to surveys measuring: (a) student perceptions of FA and AfL in terms of perceived monitoring and scaffolding (Pat-El et al., 2013), (b) feedback utility (King et al., 2009) and student use of feedback (Gibbs & Simpson, 2003), and (c) perception of feedback delivery (Steelman et al., 2004) subscales. All subscales were adapted to the mathematics context and partially refined. Questionnaires were combined with six focus group discussions involving thirty-six students from six different secondary schools. The data were analyzed using structural equation modeling and content analysis.

Descriptive results show that students positively evaluated their mathematics teacher's assessment practices. A four factor structural equation model (SEM) indicated a good fit to the data with CFI = .98, Gamma hat = .99, SRMR = .025, and RMSEA = .030 [.027, .033]. The SEM results indicated that student perceptions of teacher FA and AfL practices and the quality of feedback delivery strongly predicted students' feedback use. More specifically, students' perceptions of the quality of feedback delivery and their perceptions of teacher scaffolding practices positively predicted their self-reported feedback use. However, students' perceptions of teacher monitoring had a negative effect on their self-reported feedback use. Further analyses showed that the effect of students' perceptions of teachers' FA and AfL practices (monitoring, scaffolding, feedback delivery) on mathematics performance was fully mediated by students' perception of feedback use. The results from student focus group discussions illustrated that most students valued their mathematics teachers' assessment and feedback practices, especially when feedback was delivered in a constructive manner (i.e., considering student feelings) or showed students how to improve the quality of their work. However, some students reported to be dissatisfied with their mathematics teacher's assessment and feedback practices, especially when poor performance was coupled with reprimanding. It was concluded that students' perceptions of teacher scaffolding practices and the quality of feedback delivery predicted student feedback use, which predicted students' mathematics performance.

6.1.2. Chapter 3: Teacher perceptions of assessment and quality of feedback delivery

The aim of this chapter was to investigate among fifty-four experienced mathematics teachers' the effect of perceptions of the functions of FA and AfL (monitoring and scaffolding) and conceptions of assessment purposes on the quality of their feedback practices. These mathematics teachers taught students reported in Chapter 2 and thus were sampled from 48 secondary schools in Dar es Salaam and Kilimanjaro regions. Surveys and interviews were combined to measure teacher: (a) perceptions of FA and AfL (monitoring and scaffolding) practices, (b) conceptions of assessment purposes, and (c) feedback practices (feedback delivery and promoting feedback seeking). The data were analyzed by means of structural equation modeling and content analysis. Results from the structural equation model indicated that mathematics teachers' perceptions of FA and AfL (monitoring and scaffolding) and their conceptions of assessment purposes (assessment improves learning, school accountability) predicted the quality of feedback practices. The interview results illustrated that (a) mathematics teachers reported to use their students' assessment information for both formative and summative purposes, and (b) teacher's use of students' assessment information (e.g., improve teaching, accountability reports to parents) were rooted in their conceptions of assessment purposes. It was concluded that mathematics teachers' perceptions of their FA and AfL practices (monitoring and scaffolding) and their conceptions of assessment purposes (assessment improve learning, school accountability) predicted the quality of their feedback practices.

6.1.3. Chapter 4: Impact of intervention on students' perceptions of feedback

The intervention aimed at improving feedback provision by mathematics teachers via a one-day feedback training (FBT) centered around the Hattie and Timperley (2007) feedback model, and to improve students' perception and/or use of their teacher's feedback on their mathematics test. Data were collected from 251 Form three (Grade 11) and their respective eight mathematics teachers from eight secondary schools in the Dar es Salaam region. A quasi-experimental pretest, intervention (training) and posttest repeated measures design with an experimental and a control group was conducted. Schools were randomly sampled based on their performance (high, low). Previously validated scales were adopted measuring: (a) student FA and AfL perceptions (monitoring and scaffolding), (b) student perceptions of the quality of feedback delivery, and (c) student perceptions of feedback on their mathematics test (in terms of fairness, usefulness, acceptance, willingness to improve, and affect). The data were analysed with latent mean analyses. Measurement invariances between groups and measurement occasions were estimated to ensure valid mean comparisons.

The results showed no significant differences between pre-post measures in student perceptions of feedback, in terms of affect and willingness to improve, between the experimental and control group in the entire sample. Furthermore, within the experimental group the analyses showed significant increases in students' perceptions of their teacher's FA and AfL practices (monitoring and scaffolding) after the intervention. Furthermore, analyses within the experimental group showed significant increases in student perceptions of the quality of feedback delivery after the intervention. Moreover, students' perceptions of their own feedback use significantly increased in the experimental group after the intervention. However, similar comparisons within the control group showed no significant differences in student perceptions of teacher FA and AfL practices (monitoring and scaffolding), and student perceptions of the quality of teacher feedback delivery after the intervention. Similarly, students in the control group did not show significant differences in the perceptions of their own feedback use after the intervention. It was concluded that the intervention improved mathematics teacher's FA and AfL practices, and the quality of feedback delivery, as well as students' perceptions of their self-reported feedback use.

6.1.4. Chapter 5: Impact of intervention on students' perceptions of support in error situations

The intervention sought to establish whether the feedback training (FBT) intervention improved students' perceptions of their mathematics teacher's support in error situations, and student use of their own errors for learning. Data were collected from the same participants as in Chapter 4. A quasi-experimental pretest, intervention (training) and posttest repeated measures design was conducted. To investigate teacher's use of student errors two lessons (pretest, posttest) were videotaped. Mathematics teachers in the experimental group participated in a one-day feedback training, which also covered cognitive strategies for learning from errors (the analytic model). Surveys were administered before and after the intervention to measure: (a) student anxiety in error situations, (b) student use of errors for learning (learning orientation), and (c) perceived teacher support in error situations (error friendliness). Latent means analysis showed that students' perceptions of teacher support in error situations significantly increased within the experimental group at the posttest but did not change significantly within the control group. Furthermore, students' perceptions of anxiety in error situations and their individual use of errors for learning (learning orientation) were not affected by the intervention. Finally, exploratory case studies illustrated that mathematics teachers in the experimental group appeared to be more error friendly at the posttest, and implemented more aspects of error handling strategies (e.g., citing specific errors made by students in the test, describing why errors occurred) than teachers in the control group. It was concluded that the intervention improved teachers' error handling practices to the extent that their improvement was noted by students.

6.2. A synthesis of the studies

The general aim of this dissertation was to investigate formative assessment practices in mathematics education among secondary schools in Tanzania. Specifically, all empirical studies examined mathematics teachers' and students' variables of interest, including: (a) students' perceptions of mathematics teachers' FA and AfL practices, (b) students' perceptions of quality of teacher feedback practices and their use of teacher feedback, (c) mathematics teachers' perceptions and conceptions of their own assessment practices, (d) student perceptions of feedback on mathematics tests, (e) the impact of an intervention on teachers' feedback practices and students' perceptions of those practices, and (f) the impact of an intervention on students' perceptions of teacher support in error situations. Essentially, the main psychometric difference among the studies is that Study 1 (Chapter 2 and 3) used a 4-point balanced scale suitable in large samples (N = 2767) while Study 2 (Chapter 4 and 5) used a 6-point balanced scale to increase variance due to a smaller sample (*N* = 251). Hence, caution should be applied to the comparison and integration of the findings will therefore only be made for the overarching themes. In the following subsections, general findings from Study 1 and 2 will be discussed and some common integrated conclusions drawn.

6.2.1. Student perceptions of their mathematics teacher AfL practices

Students' perception of their teacher assessment practices is important for effective student learning (Fun, 2005). This dissertation examined student perceptions of FA and AfL in terms of teacher monitoring and scaffolding. Findings in Chapter 2 and 3 indicated that students positively evaluated their teacher assessment practices to be formative and they perceived more teacher scaffolding practices than monitoring. These findings are consistent with previous research which showed that students positively evaluated their teacher's assessment practices (Pat-El et al., 2015; Rach et al., 2013). Furthermore, these findings support those by Brazeal, Brown and Couch (2016) who showed that secondary school students had positive perceptions of formative assessment (FA) and that perceived FA was an essential tool for promoting their learning process and learning outcomes. Although students perceived their teacher assessment practices to be formative, there were large variations among Tanzanian students which may indicate that teacher practices systematically vary among mathematics classes (Kyaruzi, Strijbos, & Ufer, 2016). It can be concluded that secondary school students perceived their teacher's assessment practices to be formative, particularly their scaffolding practices.

6.2.2. Student perceptions of the quality of feedback delivery and feedback use

Student perceptions of the quality of their mathematics teacher's feedback delivery were considered an important ingredient for effective learning. Steelman et al. (2004) showed that effective feedback should be delivered in a professional manner, considering the feedback recipient's emotions in order to attain positive outcomes. Results from the structural equation analysis in Chapter 2 showed that student perceptions of the quality of feedback delivery predicted their self-reported feedback use. Qualitative results from student focus group discussions illustrate that students considered an effective feedback to be one that is delivered constructively showing what and how to improve and considers their feelings. Thus, teachers should consider students' emotions when providing students with feedback on their performance; especially for low achieving students in order to promote their feedback use. This implies that teachers need to have effective strategies in communicating feedback information about student learning.

These conclusions are consistent with previous studies that the more considerate the feedback source is perceived when providing feedback, the more likely an individual is to accept, and respond to the feedback information (Gregory & Levy, 2015; Fedor et al., 1989; Steelman et al., 2004; Strijbos, Pat-El, & Narciss, 2010). It can be concluded that student perceptions of the quality of their mathematics teacher feedback delivery regulated their learning processes such as feedback use. These results advance and extend previous studies on the role of positive perceptions of feedback delivery in the work place to feedback delivery in school contexts, showing that student perceptions of the quality of feedback delivery predicts their self-reported feedback use. Moreover, evidence from the intervention study in Chapter 4 and 5 showed that the quality of teacher feedback practices can be improved by interventions.

6.2.3. Students' AfL perceptions, feedback use and mathematics performance

Assessment perceptions influence students' learning practices and actual performances (Brown & Hirschfield, 2008). Hence, it was essential to assess how student perceptions of FA and AfL predicted their perceived feedback use and their mathematics performance. Findings in Chapter 2 showed that student perceptions of teacher scaffolding and quality of feedback delivery had a positive impact on their feedback use, while perceptions of monitoring had a negative impact on their feedback use. These findings provide a mixed message on the impact of scaffolding and monitoring perceptions on student feedback use. Positive results on the effect of scaffolding support findings in the meta-analysis conducted by Van de Pol et al. (2010), which showed that scaffolding is effective at promoting metacognitive activities of students (Mercer, Dawes, Wegerif, & Sams, 2004), cognitive activities of students (Murphy & Messer, 2000), and students' affect (Turner, Meyer, Cox, Logan, DiCintio & Thomas, 1998).

However, the negative impact of student perceived monitoring on their feedback use supports some empirical studies which showed that assessment practices are not always positively related to student learning outcomes (Kluger & DeNisi, 1996). Students in our sample benefited more from their teacher scaffolding than from monitoring practices. A closer scrutiny of scaffolding and monitoring items showed that scaffolding items activated external support to the learner (e.g., "My teacher gives me hints …"), while monitoring items activated students' inner capabilities (e.g., "My teacher encourages me to reflect on my learning."). These results imply that effective student self-regulated practices such as monitoring may not necessarily promote effective learning and that effective classroom practices depend heavily on the teacher (Cowie & Harrison, 2016).

6.2.4. Teacher conceptions of assessment and perceptions of AfL practices

Effective formative assessment practices ought to be well perceived by students and teachers (MacLellan, 2001), because teacher beliefs filter, frame and guide their practices (Fives & Buehl, 2012). In this dissertation, mathematics teachers' positively evaluated their own assessment practices. These results also support Brown (2002, 2011) who showed that teachers moderately agreed that assessment improves teaching and learning. More specifically, mathematics teachers evaluated their monitoring and scaffolding practices more positive (Chapter 3) than the way their practices were perceived by their students (Chapter 2).

This was expected, as Pat-El et al. (2015) argues that teachers tend to overestimate their assessment practices compared to their students because they have a wide expertise which might contribute to such incongruence of perceptions. However, effective formative assessment requires matched perceptions between teachers and their students because the mismatch in FA and AfL perceptions might lead to ineffective practices (MacLellan, 2001; Pat-El et al., 2015). Similar to student perceptions in Chapter 2, teacher perceptions of FA and AfL (scaffolding and monitoring) practices and their conceptions of assessment (assessment improves learning, school accountability) were related to the quality of feedback practices in Chapter 3. The influence of teacher perceptions of assessment on their actual practices is supported by the literature (Pajares, 2002; Fives & Buehl, 2012).

6.2.5. Impact of interventions on student perceptions of errors

Mathematics learning involves making errors which can be effectively used to promote student learning if paired with effective cognitive and affective support (Rach et al., 2013). As reported in Chapter 5, students' perceptions of their teacher's error handling practices were improved by the intervention, but the intervention did not improve student perceptions of anxiety in error situations and student use of errors for learning. These results support previous studies that teacher error handling practices can be improved via interventions (Heemsoth & Heinze, 2016; Rach et al., 2013). Also, the finding that the intervention did not improve student use of errors for learning supports previous studies which indicate that it is challenging to foster student use of errors for learning (Heinze & Reiss, 2007; Rach et al, 2013). However, the findings on student anxiety in error situations is inconsistent with Rach et al. (2013) who showed that error handling interventions reduced student anxiety in error situations. Hence, it can be argued that teacher error handling practices can be improved by training, but changing student perceptions of errors in particular their use of errors for learning warrants further investigation. In fact, Siegler and Chen (2008) argue that the lack of student use of errors for learning might be due to such learning from own errors being more challenging than learning from correct situations.

Another potential explanation for the findings with respect to student anxiety in error situations might be the different nature of the intervention in this dissertation compared to that in previous studies. As part of the intervention in this dissertation students reported their perceptions of errors in the context of a marked test which might have influenced student anxiety to be more stable compared to previous studies. Furthermore, the intervention (one day) was markedly shorter than the duration in previous studies that report significant improvement in student anxiety in error situations (Rach et al., 2013).

6.2.6. Impact of interventions on teacher and student perceptions of AfL

The results of the intervention indicated that student perceptions of their mathematics FA and AfL practices (i.e., monitoring and scaffolding) and feedback practices improved within the experimental group but not within the control group between pretest and posttest measures. These results support previous studies (Borasi, 1994; Heemsoth & Heinze, 2016; Rach et al., 2013; Van de Pol et al., 2014) that a teacher level intervention can improve student perceptions of teacher assessment practices. It can be concluded that teacher assessment practices and student perceptions of assessment can be improved by error handling interventions. The following sections will address methodological limitations, followed by theoretical and practical implications. The dissertation will end with recommendations for future research, and general concluding remarks.

6.3. Methodological limitations

This dissertation has several strengths such as use of different data collection methods: surveys, interviews, focus group discussions and video observations which helped to obtain broader perspective on the phenomenon of interest. However, there are some limitations which might restrict the conclusions that can be drawn. More specifically, limitations in relation to measurement invariance, missing values, and research design and relationships among constructs are discussed.

6.3.1. Measurement invariance

The dissertation involved various grouping aspects that can influence measurement accuracy, such as questionnaire language (English vs. Swahili), research condition (experimental group vs. control group) and measurement occasion (pretest vs. posttest). Measurement invariance testing was essential to determine whether students' responses were statistically equivalent between groups (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000). As a result, some scales were non-invariant and could not be used for between group latent mean comparisons. Thus, although measurement invariance was carefully handled, it restricted potential analyses. Future studies may further investigate sources of measurement invariance in the scales employed in this dissertation and how they can be reduced or avoided.

6.3.2. Missing data

Some statistical analyses procedures require that no data is missing in the dataset (Schafer & Graham, 2009). In particular, in this dissertation, structural equation modeling requires no missing data. The missing data were high (23%) in the repeated measures study (Study 2) and low (1%) in the cross-sectional survey (Study 1). The Expectation Maximisation (EM) method was used to estimate and impute missing values. Although the data were found to be missing completely at random (MCAR), missing data reduced the number of participants – in particular in the repeated measures study in Chapter 4 and 5. However, the data were found to be missing completely at random, hence missing data did not systematically affect the main conclusions.

6.3.3. Challenges related to research design

The research design had two main challenges. First, Study 1 (Chapter 2 and 3) mainly used a cross-sectional survey design which makes it impossible to draw strong causal conclusions. Nevertheless, this design was enhanced by qualitative data such as teacher interviews and student focus groups discussions. Secondly, although conducting an experimental study in a real educational setting can be beneficial (Study 2; Chapter 4 and 5), individuals tend to manipulate their behaviours as a result of being aware that they are observed (Cook, 1967). To ensure fidelity, we administered items to measure the authenticity of the feedback plenary discussions (Seidel et al., 2005; Jacobs et al., 2003). Descriptive mean values indicated that videotaped lessons were authentic and similar to untaped lessons. However, results should be interpreted cautiously, because to some extent video observations might have influenced teacher's typical practices.

6.3.4. Relationship of constructs from different frameworks

The dissertation utilized different construct to measure student and teacher perceptions of FA and AfL and conceptions of assessment purposes. Some of these constructs (e.g., feedback utility vs. use of feedback) were difficult to differentiate; in particular if they were jointly estimated in the same analysis. However, the modification indices in structural equation modeling (see Chapter 2) helped to identify items with overlapping content or meaning, which was one of the main reasons for trimming the models and eliminate such items. Inspection of item correlations served the same purpose of identifying conflicting scales. Moreover, some constructs such as assessment perceptions and conceptions of assessment purposes were contextualized for this dissertation due to their inconsistent use in the literature.

6.4. Implications for theory and practice

Theoretical and practical advantages of formative assessment and feedback practices have been reported in various studies (Black & Wiliam, 1998; Hattie & Timperley, 2007). This dissertation provided insight into the role of student and teacher FA and AfL perceptions on their assessment practices and learning outcomes among secondary schools in Tanzania. Theoretical and practical implications are identified next.

6.4.1. Theoretical implications

The aims of this dissertation were twofold: (1) investigate mathematics teachers' perceptions and practices of formative assessment, as well as students' perceptions of their teachers' practices, and (2) investigate the impact of an intervention on improving mathematics teacher's assessment practices and student feedback use. The role of formative assessment and feedback for enhancing teaching and learning process is highly emphasized in the literature (Black & Wiliam, 1998; Hattie & Timperley, 2007; Popham, 2014). However, feedback reviews show that feedback and formative assessment are not always well perceived by students (Jonsson, 2013; Kluger & DeNisi, 1996).

The findings in Chapter 2 showed that student perceptions of their teachers' FA and AfL practices regulate students' feedback use. These findings support previous work on the role of student perceptions of feedback and their subsequent regulation (Duijnhouwer, Prins, & Stokking, 2012; Poulos & Mahony, 2008; Strijbos, Narciss, & Dünnebier, 2010), but it further extended these results by showing that feedback use predicted students' mathematics performance. Students' perceptions of their teacher's assessment practices – more specifically perceived scaffolding, and perception of feedback delivery – are thus very important in FA and AfL, since feedback is considered formative when it is perceived as such and used by learners to improve their learning (Lizzio & Wilson, 2008; Wiliam, 2009). Furthermore, the main findings support the theory of planned behaviour which postulates that perceptions of practices influence behaviours which subsequently influence outcomes (Ajzen, 1991, 2005).

The findings presented in Chapter 3 indicated that mathematics teachers are aware that effective formative assessment demands both teachers and students to reflect on the assessment information (Pat-El et al., 2015). Furthermore, it has been shown that mathematics teachers had positive perceptions of their own FA and AfL practices. Finally, it was noted that mathematics teachers' conceptions of assessment purposes and their perceptions of FA and AfL practices predicted the quality of their feedback practices. Thus, these results further support the theory of planned behaviour that beliefs influence behaviors (Ajzen, 1991). It has been show that, to a large extent, mathematics teachers' assessment practices were aligned with the well-established teacher conception that assessment improves teaching and learning (Black & Wiliam, 1998; Brown, 2002; Popham, 2014), and the conception that assessment promotes school and student accountability (Brown, 2004, 2006, 2011). Despite the need for improved teacher FA and AfL practices, few studies have established interventions to support mathematics teacher assessment practices (Heinze & Reiss, 2007; Rach et al., 2013; Van de Pol et al., 2014). The results in Chapter 4 indicated that it is possible, to some extent, to improve mathematics teachers' assessment and feedback practices through an intervention; in particular mathematics teacher's scaffolding, monitoring, and feedback practices. Furthermore, these results replicate recent research that it possible to improve teacher assessment practices by interventions (Van de Pol et al., 2014).

Although several studies have applied video observation to investigate how teachers and students can use errors for effective learning (Santagata, 2004; Stigler & Hiebert, 1999), few studies have done so in the African context. In Chapter 5 it was shown that the intervention on error handling strategies improved student perceptions of teacher support in error situations. This finding replicates conclusions from previous studies that teacher support in error situations can be improved by interventions (Heinze & Reiss, 2007; Rach et al, 2013). However, there was no other visible difference associated with the intervention with respect to student perceptions of anxiety in error situations and student use of errors for learning. The results with respect to the lack of student use of errors are consistent with previous studies which showed that that the positive impact on students' use of errors for learning is not easily achieved (Heinze & Reiss, 2007; Rach et al., 2013). These results are in line with results in Chapter 2 in the sense that students are more likely to benefit from teacher support than when they need to depend on their own capabilities. An open question that remains for future research is how student use of errors for learning can be improved.

6.4.2. Practical implications

This dissertation demonstrated that students' and teachers' perceptions of formative assessment influence their assessment practices. Practical implications for this dissertation are based on the major findings across studies. In Chapter 2, it was noted that perception of teacher scaffolding and feedback delivery and were strong positive predictors of student feedback use. Firstly, we suggest mathematics teachers to deliver feedback in a thoughtful manner considering students' psychological aspects such as their emotions. Secondly, mathematics teachers are urged to scaffold their students' learning using prompts containing hints in situations that are beyond student capabilities before fading their support. Also, mathematics teachers are encouraged to provide feedback showing what needs to be done to improve students' progress. Finally, teachers are urged to promote scaffolding at the expense of monitoring practices.

Findings in this dissertation have implications for policy, professional development and teacher education. Professional development for improving teacher's FA and AfL and feedback practices are recommended. The results in Chapter 4 and 5 showed that even a short-term intervention can provide some improvement in teacher feedback practices that can be noticed by their students. It is argued that professional development for improving mathematics teacher's assessment practices should capitalize on the impact of assessment and feedback perceptions on student learning. Such professional development could for example encourage mathematics teachers to consider students' emotions when providing them with feedback about their mathematics performance. Strategies such as providing students with affective support in error situations (encouragement) paired with cognitive support (correcting mistakes or proving hints) could be a plausible strategy (Rach et al., 2013). The Tanzania Ministry of Education is encouraged to equip teachers with professional development on effective assessment and feedback practices from the Continuous Assessment (CA) program. Likewise, pre-service teacher education should develop or enhance their training by incorporating feedback models promoting effective assessment and feedback practices.

6.5. Open questions and research outlook

6.5.1. Formative assessment and high stakes examinations

The need for a shift from a testing to an assessment culture is sufficiently evident for educational systems that thrives on high stakes examinations, such as that in Tanzania. Evidence in this dissertation shows that Tanzanian secondary school mathematics teachers used students' assessment information for both summative and formative functions. It is recommended that future studies investigate how teachers regulate the conflicting functions of these two extremes of assessment. In the Tanzanian context this could additionally focus on how teacher might effectively balance the formative and summative use of school based Continuous Assessment program. Neesom (2009) showed that teachers might perceive summative and formative assessment practices differently. Hence, future studies may investigate what perceptions teachers and students have on summative assessments and how these perceptions influence their assessment practices.

6.5.2. Student formative use of errors

Errors are negatively perceived by both students and teachers. As a result, the effective use of errors in promoting learning is rarely recognized or achieved (Borasi, 1994; Rach et al., 2013). Some evidence in this dissertation points out that students do not fear errors in mathematics tests, assignments and/or in classroom discussions. However, it is not clear whether student and teacher perceptions of errors are related to their perceptions of formative or summative assessment. While evidence from this dissertation indicates that it is possible to improve teacher support in students' error situations, it is not clear how to improve student's use of their own errors for learning (Heemsoth, & Heinze, 2016). Future studies may further investigate effective interventions for improving student use of errors for learning.

6.5.3. Teacher monitoring practices in mathematics education

Feedback is considered to be formative when it is perceived well (Poulos & Mahony, 2008) and used by students to improve their learning (Black & Wiliam, 1998). Although students had positive perceptions to their mathematics teacher monitoring practices, this dissertation showed that monitoring negatively predicted student feedback use. Some studies have pointed out that use of ineffective feedback originates from a mismatch between student internal feedback and external feedback from teachers (Butler & Winne 1995), and when it is not professionally delivered (Brown & Hirschfeld, 2008). Future studies may investigate typical mathematics teachers' monitoring practices and how they could be improved.

6.6. General conclusion

The studies presented and discussed in this dissertation showed the importance of students' and mathematics teachers' perceptions of FA and AfL on their learning and teaching practices. In order to improve student feedback use, research in this dissertation shows that teachers should improve their scaffolding and feedback practices. Monitoring student learning is important FA and AfL practice; however, research evidence in this dissertation showed that students did not benefit from teacher monitoring practices. In addition, studies in this dissertation showed that it is possible for interventions to improve teacher's monitoring and scaffolding of student learning, their feedback practices (feedback delivery, promoting feedback seeking), and teacher support in student error situations. Similar to student perceptions, this dissertation showed that teacher assessment practices are influenced by their perceptions of AfL and FA practices and conceptions of assessment purposes. In order to promote student feedback use, this dissertation suggests that teachers should be encouraged to consider student perceptions of their assessment practices.

References

Adams, T. L., & Hsu, J. Y. (1998). Classroom assessment: Teachers' conceptions and practices in mathematics. *School Science and Mathematics*, *98*(4), 174-80.

Airasian, P. W. (1997). *Classroom assessment* (3rd ed.). New York: McGraw-Hill.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179-211. doi:10.1016/0749-5978(91)90020-T

Ajzen, I. (2005). *Attitudes, personality and behavior* (2nd ed.). New York: Open University Press.

Akyeampong, K., Lussier, K., Pryor, J., & Westbrook, J. (2013). Improving teaching and learning of basic maths and reading in Africa: Does teacher preparation count? *International Journal of Educational Development*, *33*(3), 272-282. doi.org/10.1016/j.ijedudev.2012.09.006

Al Duwairi, A. (2013). Secondary teachers' conceptions and practices of assessment models: The case for mathematics teachers in Jordan. *Education*, *134*(1), 124-133.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended 2-step approach. *Psychological Bulletin*, *103*(3), 411-423. doi: 10.1037/0033-2909.103.3.411

Assessment Reform Group (ARG, 2002). *Assessment for Learning: 10 Principles*. Retrieved from http://www.assessment-reform-group.org.uk

Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of Personality and Social Psychology,* 45 (5), 1017-1028.

Bandura, A. (1989). Social cognitive theory. In R. Vasta, *Annals of child development. 6, 1-60.*

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*(2), 213-238. doi:10.2307/1170535

Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives & M. Gregoire Gill (Eds.), *International handbook of research on teacher beliefs* (pp. 284-300). New York: Routledge.

Barnes, N., Fives, H., & Dacey, C. M. (2017). U.S. teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education, 65,* 107-116. doi:10.1016/j.tate.2017.02.017

Barnette, J., Orletsky, S., & Sattes, B. (1994, July). *Evaluation of teacher classroom questioning behaviors.* Paper presented at the Third Annual National Evaluation Institute Field Symposium, Gatlinburg, TN.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25. doi:10.1080/0969594X.2010.513678

BEST (2004-2013). *Basic Education Statistic in Tanzania (National data)*. Retrieved from http://www.moe.go.tz/index.php?option=com_docman&task=cat_view&gid=358&Itemid=619

BEST (2014). *Basic Education Statistics in Tanzania (National data)*. Prime Minister's Office Regional Administration and Local Government: Dodoma, Tanzania.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education,* 32, 347-364.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, *5*(1), 7-74.

Black, P., & Wiliam, D. (2003). In praise of educational research: formative assessment: *British Educational Research Journal, 29 (5), 623-637.*

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational*

Assessment, Evaluation and Accountability, *21*(1), 5-31.doi:10.1007/s11092-008-9068-5

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead, UK: Open University Press.

Bloom, B.S. (1969). Some theoretical issues relating to educational evaluation. In R.W. Tyler (Ed.), *Educational evaluation: New roles, new means. The 63rd yearbook of the National Society for the Study of Education* (Volume 69(2), pp. 26-50). Chicago, IL: University of Chicago Press.

Borasi, R. (1994). Capitalizing on errors as 'springboards for inquiry': A teaching experiment. *Journal for Research in Mathematics Education*, *25*(2), 166-208. doi:10.2307/749507

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school* (Rev. ed.). Washington, DC: National Academies Press.

Brazeal, K. R., Brown, T. L., & Couch, B. A. (2016). Characterizing student perceptions of and buy-in toward common formative assessment techniques. *CBE Life Sciences Education*, *15*(4),1-14.

Briggs, M., Woodfield, A., Martin, C., & Swatton, P. (2006). *Assessment for learning and teaching in primary schools.* Glasgow: Bell & Bain.

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, *7*(4), 279-301. doi:10.1207/s15324818ame07042

Brock-Utne, B. (2007). Learning through a familiar language versus learning through a foreign language: A look into some secondary school classrooms in Tanzania. *International Journal of Educational Development*, *27*(5), 487-498. doi:10.1016/j.ijedudev.2006.10.004

Brown, G. T. L. (2002). *Teachers' conceptions of assessment.* Unpublished doctoral dissertation, University of Auckland, Auckland, NZ.

Brown, G. T. L. (2004a). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological Reports*, *94*(3), 1015-1024. doi:10.2466/PR0.94.3.1015-1024

Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy and Practice, 11*(3), 301-318. doi:10.1080/0969594042000304609

Brown, G. T. L. (2006). Integrating teachers' conceptions: Assessment, teaching, learning, curriculum and efficacy. In A. P. Prescott (Ed.), *The concept of self in education, family and sports* (pp. 1-49). Hauppauge, NY: Nova Science Publishers.

Brown, G. T. L., & Hirschfield, G. F. (2007). Students' conceptions of assessment and mathematics: Self-regulation raises achievement. *Australian Journal of Educational & Developmental Psychology*, *7*(1), 63-74.

Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice*, *15*(1), 3-17. doi:10.1080/09695940701876003

Brown, G. T. L. (2008). *Conceptions of assessment*: *Understanding what assessment means to teachers and students*. New York: Nova Science Publishers.

Brown, G. T. L. (2011). New Zealand prospective teacher conceptions of assessment and academic performance: Neither student nor practicing teacher. In R. Kahn, J. C. McDermott & A. Akimjak (Eds.) *Democratic access to education* (pp. 119-132). Los Angeles, CA: Antioch University Los Angeles, Department of Education.

Brown, G. T. L., & Chai, C. (2012). Assessing Instructional Leadership: A Longitudinal Study of New Principals. *Journal of Educational Administration, 50*(6), 753-772.

Brown, G. T. L. (2013). Student conceptions of assessment across cultural and contextual differences: University students perspectives of assessment from Brazil, China, Hong Kong, and New Zealand. In G. A. D Liem & A. B. I. Bernardo (Eds.), *Advancing* *cross-cultural perspectives on educational psychology* (pp. 143-167). Charlotte, NC: Information Age.

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367-393). Thousand Oaks, CA: Sage.

Brown, G. T., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research*, *2*(1), 22-30. doi:10.14786/flr.v2i1.24

Brown, G. T. L., Peterson, E. R., & Yao, E. (2016). Student conceptions of feedback: Impact on self-regulation, self-efficacy, and academic achievement. *British Journal of Educational Psychology, 86*(4), 606-629. doi:10.1111/bjep.12126

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245-281. doi:10.2307/1170684

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications and programming*. Routledge: London.

Can, G., & Walker, A. (2011). A model for doctoral students' perceptions and attitudes toward written feedback for academic writing. *Research in Higher Education*, *52*(5), 508-536. doi:10.1007/s11162-010-9204-1

Carless, D., Salter, D., Yang, M., & Lam, J. (2010). Developing sustainable feedback practices. *Studies in Higher Education*,1, 1-13. doi: 10.1080/03075071003642449

Carlson, M., Wilcox, R., Chou, C., Yang, F., Chang, M, Blanchard, J...Clark, F. (2011). Psychometric properties of reverse-scored items on the CES-D in a sample of ethnically diverse older adults. *Psychological Assessment*, 23(2)558-562. doi:10.1037/a0022484

Chang, L. (1993). Using confirmatory factor analysis of multi-method data to assess the psychological equivalence of 4-pont and 6-pont Likert-type scales. In proceedings of the *Annual Meeting of the National Council on Measurement in Education*. Atlanta: U.S Department of Education.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159. doi:10.1037/0033-2909.112.1.155

Cohen, J. W. (1987). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Comşa, M. (2010). How to compare means of latent variables across Countries and Waves: Testing for invariance measurement. An application using Eastern European Societies. *Sociologia*, *42*(6), 639-669.

Conger, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement*, *34*(1), 35-46.

Cook, D. L. (1967). *Impact of the Hawthorne effect in experimental design in educational research*. The Ohio State University, Ohio.

Cowie, B., & Harrison, C. (2016). Classroom processes that support effective assessment. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 335-350). New York: Routledge.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative and mixed approaches*. Los Angeles: Sage.

De Kleijn, R. A. M., Mainhard, M. T., Meijer, P. C., Brekelmans, M., & Pilot, A. (2013). Master's thesis projects: Student perceptions of supervisor feedback, *Assessment & Evaluation in Higher Education*, 38 (8), 1012-1026. doi:10.1080/02602938.2013.777690

Dingyloudi, F., & Strijbos, J. W. (2017). Mixed methods research as a pragmatic toolkit: Understanding versus fixing complexity in the learning sciences. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman & P. Reimann (Eds.), *The international handbook of the learning sciences* (pages to be assigned). New York: Routledge/Taylor & Francis.

Duijnhouwer, H., Prins, F. J., & Stokking, K. M. (2012). Feedback providing improvement strategies and reflection on feedback use: Effects on students' writing motivation, process, and performance. *Learning and Instruction*, *22*(3), 171-184. doi: 10.1016/j.learninstruc.2011.10.003

Duncan, N., Prowse, S., Wakeman, C., & Harrison, R. (2007). 'Feed-forward': Improving students' use of tutors' comments: Corrigendum. *Assessment and Evaluation in Higher Education*, *32*(4), 513. doi:10.1080/02602930701510329

Dunham, T. C., & Davison, M. L. (1991). Effects of scale anchors on student ratings of instructors. *Applied Measurement in Education*, *4*(1), 23-35.

Eccles, J.S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53,109-32

Ecclestone, K. (2012). From testing to productive student learning: implementing formative assessment in Confucian-heritage settings. *Assessment in Education: Principles, Policy & Practice*, *19*(2), 275-276. doi:10.1080/0969594X.2011.632863

Entwistle, N. (1987). *Understanding classroom learning*. London: Hodder & Stoughton.

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*(3), 509-529. doi:10.1080/00273170701382864

Fedor, D. B., Eder, R. W., & Buckley, M. R. (1989). The contributory effects of supervisor intentions on subordinate feedback responses. *Organizational Behavior and Human Decision Processes*, *44*(3), 396-414. doi:10.1016/0749-5978(89)90016-2

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage Publications.

Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 95-113.

Fives, H., & Buehl, M. M. (2012). Spring cleaning for the 'messy' construct of teachers' beliefs: What are they? Which have been examined? What can they tell us?. In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, M. Zeidner, ... M. Zeidner (Eds.), *APA Educational Psychology Handbook: Individual differences and cultural and contextual factors* (Vol. 2, pp. 471-499). Washington, DC: American Psychological Association. doi:10.1037/13274-019

Fulmer, G. W., Lee, I. C. H., & Tan, K. H. K. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy, and Practice* 10.1080/0969594X.2015.1017445

Fun, N. L. (2005). *Hong Kong academics' and students' perceptions of assessment purposes and practices* (The LOAP questionnaire survey report). Hong Kong: Institute of Education.

Gall, M. D. (1970). The use of questions in teaching. *Review of Educational Research*, *40,* 707-721.

Gall, M. D. (1984). Synthesis of research on teachers' questioning. *Educational Leadership, 42*(3), 40-47.

Gan, M. S., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*, *42*(6), 861-878. doi:10.1007/s11251-014-9319-4

Gebril, A., & Brown, G. T. L. (2014). The effect of high-stakes examination systems on teacher beliefs: Egyptian teachers' conceptions of assessment. *Assessment in Education: Principles, Policy and Practice*, *21*(1), 16-33. doi:10.1080/0969594X.2013.831030

Geister, S., Konradt, U., & Hertel, G. (2006). Effects of process feedback on motivation, satisfaction, and performance in virtual teams. *Small Group Research*, *37*(5), 459-489. doi:10.1177/1046496406292337

Gibbs, G., & Simpson, C. (2003). Measuring the response of students to assessment: The assessment experience questionnaire. *Paper presented at the 11th Improving Student Learning Symposium*, Hinckley, UK.

Ginsburg, H. P. (2009). The challenge of formative assessment in mathematics education: Children's minds, teachers' minds. *Human Development*, *52*(2), 109-128. doi:10.1159/000202729

Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.

Gregory, J. B., & Levy, P. E. (2015). Feedback delivery and the role of the feedback provider. In Gregory, J. B.,& Levy, P.E (Eds.), *Using feedback in organizational consulting* (pp. 47-62). Washington, DC: American Psychological Association. doi:10.1037/14619-005

Gronlund, N. E., & Linn, R, L. (1990). *Measurement and evaluation in teaching*. New York: Macmillan.

Harris, L. R., Brown, G. T. L., & Harnett, J. (2014). Understanding classroom feedback practices: A study of New Zealand student experiences, perceptions, and emotional responses. *Educational Assessment, Evaluation and Accountability*, *26*(2), 107-133. doi:10.1007/s11092-013-9187-5

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81-112. doi:10.3102/003465430298487

Hattie, J. (2015). Review of the international handbook of research on teachers' belief. *The Australian Educational and Developmental Psychologist*, *32*(1), 91-92. doi:10.1017/edp.2015.5

Heemsoth, T., & Heinze, A. (2016). Secondary school students learning from reflections on the rationale behind self-made errors: A field experiment. *Journal of Experimental Education*, *84*(1), 98-118. doi: 10.1080/00220973.2014.963215

Heinze, (2005). Mistake-handling activities in the mathematics classroom. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 3, pp. 105-112. Melbourne: PME.

Heinze, A., & Reiss, K. (2007). Mistake-handling activities in the mathematics classroom: Effects of an in-service teacher training on students' performance in Geometry. In Woo, J. H., Lew, H. C., Park, K. S. & Seo, d. Y. (Eds.). *Proceedings of the 31st Conference of the International Group for the Psychology of Mathematics Education*, Vol. 3, pp. 9-16. Seoul: PME.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York: Routledge.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1-55. doi:10.1080/10705519909540118

Hussein, M. H. (2010). The peer interaction in primary school questionnaire: Testing for measurement equivalence and latent mean differences in bullying between gender in Egypt, Saudi Arabia and the USA. *Social Psychology of Education: An International Journal*, *13*(1), 57-76.

Ingram, J., Pitt, A., & Baldry, F. (2015). Handling errors as they arise in whole-class interactions. *Research in Mathematics Education*, *17*(3), 183-197. doi:10.1080/14794802.2015.1098562

Jacobs, J., Garnier, H., Gallimore, R., Hollingsworth, H., Givvin, K.B., Rust, K., Kawanaka, T.,…Stigler, J. (2003). *Third International Mathematics and Science Study 1999 Video Study Technical Report: Volume 1-Mathematics* (Technical Report, NCES 2003-012). Washington, DC: National Center for Education Statistics.

James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *The Curriculum Journal*, *17*(2), 109-138.

Jenkins, J. O. (2010). A multi-faceted formative assessment approach: better recognising the learning needs of students. *Assessment and Evaluation in Higher Education*, *35*(5), 565-576. doi:10.1080/02602930903243059.

Jonsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education, 14*(1), 63-76. doi:10.1177/1469787412467125

Kennedy, K. J., Chan, J. K. S., & Fok, P. K. (2011). Holding policy-makers to account: Exploring 'soft' and 'hard' policy and the implications for curriculum reform. *London Review of Education*, *9*(1), 41-54. 10.1080/14748460.2011.550433

King, P. E., Schrodt, P., & Weisel, J. J. (2009). The Instructional Feedback Orientation Scale: Conceptualizing and validating a new measure for assessing perceptions of instructional feedback. *Communication Education*, *58*(2), 235-261. doi:10.1080/03634520802515705

Kitta, S. (2004). *Enhancing mathematics teachers' pedagogical content knowledge and skills in Tanzania.* Unpublished doctoral dissertation, University of Twente, Enschede, the Netherlands.

Kitta, S. (2014). Science teachers' perceptions of classroom assessment in Tanzania: An exploratory study. *International Journal of Humanities Social Sciences and Education, 1*(12)51-55.

Kitta, S., & Tilya, F. (2010). The status of learner-centred learning and assessment in Tanzania in the context of the competence-based curriculum. *Papers in Education and Development Journal of the School of Education, University of Dar es Salaam, 29, 77-90*.

Kitta, S., & Tilya, F. (2010). The status of learner-centred learning and assessment in Tanzania in the context of the competence-based curriculum. *Papers in Education and Development Journal*, *29,* 77-90.

Kloss, R. J. (1988). Toward asking the right questions: The beautiful, the pretty, and the messy ones. *The Clearing House, 61*, 245-248.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254-284. doi:10.1037/0033-2909.119.2.254

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction*, *20*(4), 344-348. doi:10.1016/j.learninstruc.2009.08.005

Köller, O. (2005). Formative assessment in classrooms: A review of the empirical German literature. In J. Looney (Ed.), *Formative assessment: Improving learning in secondary classrooms* (pp. 265–279). Paris, France: Organization for Economic Co-Operation and Development.

Komba, W. L. M. (2007). Teacher professional development in Tanzania: Perceptions and practices. *Papers in Education and Development*, *27*, 1-27.

Krippendorff, K. (2013). *Content analysis: An introduction to its methodology* (3rd ed.). Los Angeles: Sage.

Kulas, J. T., & Stachowski, A. A. (2009). Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality*, *43*(3), 489-493. doi:10.1016/j.jrp.2008.12.005

Kyaruzi, F. (2012). Effectiveness of mathematics continuous assessment feedback in secondary schools in Tanzania. *Journal of Education Science and Humanities*, *1*(2), 53-65.

Kyaruzi, F., Strijbos, J. W., & Ufer, S. (2015, August). *Formative assessment practices in mathematics education among secondary schools in Tanzania*. Poster presentation at the 16th biennial EARLI conference, Limassol, Cyprus.

Kyaruzi, F., Strijbos, J. W., & Ufer, S. (2016, August). Students' assessment for learning perceptions and their mathematics achievement in Tanzanian secondary schools. In C. Csíkos, A. Rausch & J. Szitányi (Eds.), *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 155-162). Szeged, Hungary: PME.

Langford, E., Schwertman, N., & Owens, M. (2001). Is the property of being positively correlated transitive? *The American Statistician, 55*, 322-325.

Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & Quantity: International Journal of Methodology*, *43*(2), 265-275. doi:10.1007/s11135-007-9105-3

Leung, F. S. (2005). Some characteristics of East Asian mathematics classrooms based on data from the TIMSS 1999 Video Study. *Educational Studies in Mathematics, 60*(2), 199-215. doi:10.1007/s10649-005-3835-8

Lewis, J. W, & Escobar, L. A. (1986). Suppression and enhancement in bivariate regression. *Journal of the Royal Statistical Society*, 35 (1), 17-26.

Li, M., Ruiz-Primo, M. A., & Shavelson, R. J. (2006). Towards a science achievement framework: The case of TIMSS 1999. In Howie, F S. & Plomp, F T. (Eds.), *Contexts of learning mathematics and science: Lessons learned from TIMSS* (pp. 291–311). London, UK: Routledge.

Lipnevich, A. A., & Smith, J. K. (2009). 'I really need feedback to learn:' Students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation and Accountability*, *21*(4), 347-367. doi:10.1007/s11092-009-9082-2

Lipnevich, A. A., Berg, D. A. G., & Smith, J. K. (2016). Toward a model of student response to feedback. In G. T. L. Brown & L. R. Harris (Eds.), *The handbook of human and social conditions in assessment* (pp. 169-185). New York: Routledge.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*(2), 151-173. doi:10.1207/S15328007SEM0902_1

Lizzio, A., & Wilson, K. (2008). Feedback on assessment: students' perceptions of quality and effectiveness. *Assessment and Evaluation in Higher Education*, *33*(3), 263-275. doi:10.1080/02602930701292548

MacLellan, E. (2001). Assessment for learning: The differing perceptions of tutors and students. *Assessment and Evaluation in Higher Education*, *26*, 307-318. doi:10.1080/02602930120063466

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2017). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, doi:10.1037/met0000113.

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*(2), 181-220. doi: 10.1207/s15327906mbr3302_1

Marshall, B.,& Drummond, J.M.(2006). How teachers engage with Assessment for Learning: Lessons from the classroom. *Research Papers in Education, 21, (2), 133-149.*

Maruyama, G. M. (1998). *Basics of structural equation modeling*. Thousand Oaks, CA: Sage.

McArdle, J. J. (2007). Five steps in the structural factor analysis of longitudinal data. In R. Cudeck, R. C. MacCallum, R. Cudeck, R. C. MacCallum (Eds.). *Factor analysis at 100: Historical developments and future directions* (pp. 99-130). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: Ways of helping children to use language to learn science. *British Educational Research Journal*, 30, 359-377.

Ministry of Education (Tanzania). (2008). *Education Sector Development Programme Document*. Dar es Salaam: Ministry of Education.

Ministry of Education and Vocational Training (MoEVT) (2013). *Certificate of secondary education examination school rank.* Dar es Salaam: National Examination Council of Tanzania. Retrieved from http://www.necta.go.tz/matokeo_html_2013/CSEE2013/olevel.htm

Ministry of Education and Vocational Training (MoEVT, 2014). Sera ya elimu na mafunzo [Education training and policy]. Dar es Salaam: MoEVT.

Minsky, M. (1994). Negative expertise. *International Journal of Expert Systems*, *7*(1), 13-19.

Montgomery, J. J., & Baker, W. (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing*, *16*(2), 82-99.

Murphy, N., & Messer, D. (2000). Differential benefits from scaffolding and children working alone. *Educational Psychology*, *20, 17-31.*

Musil, C. M., Warner, C. B., Yobas, P. K., & Jones, S. L. (2002). A comparison of imputation techniques for handling missing data. *Western Journal of Nursing Research*, *24*(7), 815-829.doi: 10.1177/019394502762477004

Muthen, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research*, 22(3), 376-398.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer & M. P. Driscoll (Eds.), *Handbook of Research on Educational Communications and Technology* (3rd ed., pp. 125-143). Mahwah, NJ: Lawrence Erlbaum.

National Examinations Council of Tanzania (2004): *Thirty years of the national examinations of Tanzania*: Dar es Salaam: NECTA.

National Examinations Council of Tanzania (2013). *Certificate of secondary education examination school rank*: Dar es Salaam: Ministry of Education and Vocational Training. From http://www.necta.go.tz/matokeo_html_2013/CSEE2013/olevel.htm

National Examinations Council of Tanzania (NECTA, 2014). *Certificate of Secondary Education Examinations (CSEE) schools ranking*: http://www.necta.go.tz/brn

Ndalichako, J. L. (2015). Secondary school teachers' perceptions of assessment: *International Journal of Information and Education Technology*, *5*(5), 326-330.

Neesom, A. (2009). Report on teacher's perceptions of formative assessment: London, QCA.

Neitzel, C., & Davis, D. (2014). Direct and indirect effects of teacher instruction and feedback on student adaptive help-seeking in upper-elementary literacy classrooms. *Journal of Research in Education*, *24*(1), 53-68.

Nesbit, P. L., & Burton, S. (2006). Student justice perceptions following assignment feedback. *Assessment and Evaluation in Higher Education*, *31*(6), 655-670.

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.

Njabili, A. F. (1999). *Public examinations: A tool for curriculum evaluation.* Dar es Salaam: Mture Educational.

Nye, C. D., & Drasgow, F. (2011). Effect size indices for analyses of measurement equivalence: Understanding the practical importance of differences between groups.

Journal of Applied Psychology, *96*(5), 966-980.

Osaki, K., Hosea, K., & Ottevanger, W. (2004). *Reforming Science and Mathematics education in Sub-Sahara Africa: Obstacles and opportunities*. Dar es Salaam: TEAMS Project, University of Dar es Salaam.

Oser, F., & Spychiger, M. (2005). *Lernen ist schmerzhaft. Zur Theorie des negativen Wissens und zur Praxis der Fehlerkultur* [Learning is painful. On the theory of negative knowledge and the practice of error culture]. Weinheim, Germany: Beltz.

Ottevanger, W., Akker, J., & Feiter, L. (2007). Developing science, mathematics, and ICT education in Sub-Saharan Africa: Patterns and promising practices (Working paper no. 101): The World Bank: Washington.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, *62*(3), 307-332. doi:10.3102/00346543062003307

Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2013). Validation of assessment for learning questionnaires for teachers and students. *British Journal of Educational Psychology, 83*, 98-113. doi: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 and Practice, 22*(2), 282-298. doi:10.1080/0969594X.2014.975675

Paulhus, D. L., Robins, R. W., Trzesniewski, K. H., & Tracy, J. L. (2004). Two replicable suppressor situations in personality research. *Multivariate Behavioral Research*, *39*(2), 303-328.

Pellegrino, J. W. (2014). Assessment as a positive influence on 21st century teaching and learning: A systems approach to progress. *Psicologia Educativa*, *20*(2), 65-77. doi:10.1016/j.pse.2014.11.002

Peterson R. E. & Irving, E. S. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction*, 18, 238-250.

Peugh, L. J., & Enders, C.K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement*: Review of Educational Research*, *74*(4), 525-556.

Popham, W. J. (2014). *Classroom assessment: What teachers need to know* (7th ed.). Harlow, US: Pearson.

Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. *Assessment and Evaluation in Higher Education*, *33*(2), 143-154. doi:10.1080/02602930601127869

Prediger, S., & Erath, K. (2014). Content, interaction, or both? Synthesizing two German traditions in a video study on learning to explain in mathematics classroom microcultures. *Eurasia Journal of Mathematics, Science and Technology Education, 10*(4), 313-327. doi:10.12973/eurasia.2014.1085a

Puntambekar, S., & Kolodner, J. (2005). Toward implementing distributed scaffolding: Helping students learn science from design: *Journal of Research in Science Teaching,* 42(2), 185-217.

Qorro, M. (2007). Language of instruction and its effects on the quality of education. *Papers in Education and Development*, *27,* 56-78.

Qorro, M. (2013). Language of instruction in Tanzania: Why are research findings not heeded? *International Review of Education, 59*(1), 29-45. doi: 10.1007/s11159-013-9329-5

Rach, S., Ufer, S., & Heinze, A. (2013). Learning from errors: Effects of teachers' training on students' attitudes towards and their individual use of errors. *Proceedings of the National Academy of Sciences, 8*(1), 21-30.

Rakoczy, K., Harks, B., Klieme, E., Blum, W., & Hochweber, J. (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning and Instruction*, *27*, 63-73. doi:10.1016/j.learninstruc.2013.03.002

Raudenbush, S., & Bryk, S.A. (1986). A hierarchical model for studying school effects: *Sociology of Education*, 59, 1-17.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*(3), 552-566.

Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, *27*(2), 472-482. doi:10.1016/j.tate.2010.09.017

Roller, S. A. (2016). What they notice in video: A study of prospective secondary mathematics teachers learning to teach. *Journal of Mathematics Teacher Education*, *19*(5), 477-498. doi:10.1007/s10857-015-9307-x

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge, UK: Cambridge University Press.

Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, *37*(1), 15-24. doi:10.1016/j.stueduc.2011.04.003

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119-144.

Santagata, R. (2004). 'Are you joking or are you sleeping?' Cultural beliefs and practices in Italian and U.S. teachers' mistake-handling strategies. *Linguistics and Education*, *15*(1-2), 141-164. doi:10.1016/j.linged.2004.12.002

Santagata, R. (2005). Practices and beliefs in mistake-handling activities: A video study of Italian and US mathematics lessons. *Teaching and Teacher Education*, *21*(5), 491-508. doi:10.1016/j.tate.2005.03.004

Santagata, R., & Yeh, C. (2014). Learning to teach mathematics and to analyze teaching effectiveness: Evidence from a video- and practice-based approach. *Journal of Mathematics Teacher Education*, *17*(6), 491-514. doi:10.1007/s10857-013-9263-2

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, *29*(4), 347-363.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147-177. doi:10.1037/1082-989X.7.2.147

Shepard, L., Hammerness, K., Darling-Hammond, L., Rust, F., Baratz-Snowden, J., Gordon, E., & ... Pacheco, A. (2005). Assessment. In L. Darling-Hammond, J. Bransford, L. Darling-Hammond, J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275-326). San Francisco, CA, US: Jossey-Bass.

Schwartz, D.L., & Hartman, K. (2007). It is not television anymore: Designing digital video for learning and assessment. In R. Goldman, S. Derry, R. Pea, & B. Barron (Eds.), *Video Research in the Learning Sciences* (pp.335-348). Mahwah, NJ: Erlbaum.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne & M. Scriven (Eds.), *Perspectives of curriculum evaluation*. Chicago: Rand McNally.

Seidel, T., Prenzel, M., Dalehefte, M. I., Meyer, H. C., Lehrke, M., Duit, R. (2005). The IPN video study: An overview. In Seidel, T., Prenzel, M., & Korbag, M. (Eds), *How to run a video study* (pp.7-19). Waxmann: Münster.

Semali, L. M., & Mehta, K. (2012). Science education in Tanzania: Challenges and policy responses. *International Journal of Educational Research*, *53*, 225-239. 10.1016/j.ijer.2012.03.012

Sfard, A. (2007). When the rules of discourse change, but nobody tells you: Making sense of mathematics learning from a commognitive standpoint. *Journal of the Learning Sciences*, *16*(4), 565-613. doi:10.1080/10508400701525253.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153-189. doi:10.3102/0034654307313795

Siegler, R. S., & Chen, Z. (2008). Differentiation and integration: Guiding principles for analyzing cognitive change. *Developmental Science*, *11*(4), 433-448. doi:10.1111/j.1467-7687.2008.00689.x

Skinner, B. F. (1961). Why we need teaching machines. Harvard Educational Review, 31, 377-398.

Slavin, R.E. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 77,454-499.

Smyth, K. (2004). The benefits of students learning about critical evaluation rather than being summatively judged. *Assessment & Evaluation in Higher Education*, *29*(3), 369-378.

Spychiger, M., Küster, R., & Oser, F. (2006). Dimensionen von Fehlerkultur in der Schule und deren Messung: Der Schülerfragebogen zur Fehlerkultur im Unterricht für Mittel und Oberstufe [Dimensions of error culture in education and its measurement: The student questionnaire on error culture in secondary education]. *Schweizerische Zeitschrift für Bildungswissenschaften*, *28*(1), 87-110.

Steelman, L. A., Levy, P. E., & Snell, A. F. (2004). The feedback environment scale: Construct definition, measurement, and validation. *Educational and Psychological Measurement*, *64*(1), 165-184.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173-180. doi:10.1207/s15327906mbr2502_4

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, *87*(4), 324-328.

Stiggins, R. (2007). Assessment through the student's eyes. *Educational Leadership*, 65(8), 22-26.

Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study: Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States*. Washington, DC: National Center for Educational Statistics.

Stigler, J.W., & Hierbert, J. (1991). *The teaching gap*. Free Press: New York.

Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers and Education*, *46*(1), 29-48.

Strijbos, J.W, Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, *20*(4), 291-303. doi:10.1016/j.learninstruc.2009.08.008

Strijbos, J. W., Pat-El, R. J., & Narciss, S. (2010, May). Validation of a (peer) feedback perceptions questionnaire. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell, & T. Ryberg (Eds.), *Proceedings of the 7th International Conference on Networked Learning* (pp. 378-386) [Electronic Proceedings, ISBN 978-1-86220-225-2]. Aalborg, Denmark: Aalborg University.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education*, *30*(4), 325-341. doi:10.1080/02602930500099102

Sumra, S., & Katabaro, J. K. (2014). Declining quality of education: Suggestions for arresting and reversing the trend. *ESRF Discussion Paper*, *63*(3), 1-53.

Sumra, S., & Rajani, R. (2006). *Secondary education in Tanzania: Key policy challenges*. Dar es Salaam, Tanzania: HakiElimu.

Thompson, A. (1992). Teachers' beliefs and conceptions. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). New York : Macmillan.

Tierney, R. D., & Koch, M. J. (2016). Privacy in classroom assessment. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 267-283). New York: Routledge.

Tilya, F., & Mafumiko, F. (2010). The compatibility between teaching methods and competence-based curriculum in Tanzania. *Papers in Education and Development*, 29, 37-54.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom.* Buckingham, UK: Open University Press.

Turner, J. C., Meyer, D. K., Cox, K. E., Logan, C., DiCintio, M., & Thomas, C. T. (1998). Creating contexts for involvement in mathematics. *Journal of Education and Psychology, 90, 730-745*.

United Republic of Tanzania. (URT, 1974). *The Musoma resolution*. Dar es Salaam: Government Printers.

Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, *22*(3), 271-296. doi:10.1007/s10648-010-9127-6

Van de Pol, J., Volman, M., Oort, F., & Beishuizen, J. (2014). Teacher scaffolding in small-group work: An intervention study. *Journal of the Learning Sciences*, *23*(4), 600-650.

doi:10.1080/10508406.2013.805300

Van de Watering, G., Gijbels, D., Dochy, F., & van der Rijt, J. (2008). Students' assessment preferences, perceptions of assessment and their relationships to study results. *Higher Education: The International Journal of Higher Education and Educational Planning*, *56*(6), 645-658.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational Research Methods, 3(4), 4-70.

VanLehn, K. (1999). Rule-learning events in the acquisition of a complex skill: An evaluation of cascade. *Journal of the Learning Sciences*, *8*(1), 71-125. doi:10.1207/s15327809jls0801_3

Veldhuis, M., & Van den Heuvel-Panhuizen, M. (2014). Exploring the feasibility and effectiveness of assessment techniques to improve student learning in primary mathematics education. In C. Nicol, S. Oesterle, P. Liljedahl & D. Allan (Eds.), *Proceedings of the 38th Conference of the International Group for the Psychology of Mathematics Education and the 36th Conference of the North American Chapter of the Psychology of Mathematics Education* (Vol. 5, pp. 329-336). Vancouver, Canada: PME.

Vuzo, M. (2007). *Revisiting the language of instruction policy in Tanzanian secondary schools: A comparative study of geography classes taught in Kiswahili and English*. Unpublished doctoral dissertation, University of Oslo, Unit for Comparative and International Education, Oslo, Norway.

Wade, S. E., & Moje, E. B. (2000). The role of text in classroom learning. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (Eds.), *Handbook of Reading Research Volume III* (pp. 609-627). Mahwah, NJ: LEA.

Wagner, R. F. (1981). Remediating common math errors. *Academic Therapy*, *16*(4), 449-453.

Watson, D., Clark, L. A., Chmielewski, M., & Kotov, R. (2013). The value of suppressor effects in explicating the construct validity of symptom measures. *Psychological Assessment, 25*(3), 929-941.

Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*, 31(3), 379-394.

Wiggins, G. (1993). *Assessing student performances*. San Francisco: Josey-Bass.

Wiliam, D. (1999). Formative assessment in mathematics: Feedback. *Equals: Mathematics and Special Educational Needs, 5*(3), 8-11.

Wiliam, D. (2007). *What does research say the benefits of formative assessment are?* Reston: The National Council of Teachers of Mathematics.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3-14.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.). *The future of assessment: Shaping teaching and learning,* (pp.53-82). Mahwah, NJ: Erlbaum.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice, 11*(1), 49-65. doi:10.1080/0969594042000208994

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, *52*(1), 17-37. doi: 10.1080/00461520.2016.1207538

Wood, D. (1988). *How Children Think and Learn*. Oxford, UK: Blackwell.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89-100.

Wu, D. A,. Li, Z., & Zumbo, B. N. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data: *Practical Assessment Research & Evaluation*, *12*(3), 2-26.

Yaghambe, R. S., & Tshabangu, I. (2013). Disciplinary networks in secondary schools: Policy dimensions and children's rights in Tanzania**.** *Journal of Studies in Education*, *3*(4), 42-56.

Appendix A. Chapter 2 (Study 1) - Student English questionnaire

Through this questionnaire, you are invited to provide information regarding how you experience mathematics assessment practices. The questionnaire covers some general information (such as your age), your beliefs about assessment, how you perceive formative assessment practices, and how you perceive feedback practices in your mathematics class. Some questions are similar to those asked to your mathematics teacher. There are no right or wrong answers. Your answers will remain anonymous.

1. Date: ______ (day) ___________ (month) ______ (year)

2. Gender: ☐ Male ☐ Female

3. Age ______ (**years**)

4. School name: ______________________

5. Nationality: _________________

6. Form/Class: _________________

7. Stream: ____________________

8. What was your **Mathematics** score or grade in Form two national examination: ________

9. What was your **English** score or grade in Form Two national examination: ____________

10. What was your **Mathematics** score or grade in last Form three terminal examination:____

11. What was your **English** score or grade in the last Form three terminal examination: _____

To be completed by the researcher

<table>
<tr><td>Schoolnumber</td><td></td><td rowspan="2">District</td><td rowspan="2"></td><td rowspan="2">Region</td><td rowspan="2"></td></tr>
<tr><td>Respondentgroup</td><td></td></tr>
<tr><td>School ownership</td><td></td><td>Lettercode ST</td><td></td><td>School category</td><td></td></tr>
</table>

Read each statement carefully and tick (√) inside the box alongside your best option among ***fully disagree, somewhat disagree, somewhat agree or fully agree.***

		Fully disagree	**Somewhat disagree**	**Somewhat agree**	**Fully agree**
SCoA-VI_1_SIS	I pay attention to my assessment results in order to focus on what I could do better next time.	☐	☐	☐	☐
SCoA-VI_2_SSC	Assessment encourages my class to work together and help each other.	☐	☐	☐	☐
SCoA-VI_3_SIB(R)	Assessment is unfair to students.	☐	☐	☐	☐
SCoA-VI_4_SEA	Assessment results show how intelligent I am.	☐	☐	☐	☐
SCoA-VI_5_SIT	Assessment helps teachers track my progress.	☐	☐	☐	☐
SCoA-VI_6_SSP	Assessment is an engaging and enjoyable experience for me.	☐	☐	☐	☐
SCoA-VI_7_SIG(R)	I ignore assessment information.	☐	☐	☐	☐
SCoA-VI_8_SIT	Assessment is a way to determine how much I have learned from teaching.	☐	☐	☐	☐
SCoA-VI_9_SIT	Assessment is checking off my progress against achievement objectives or standards.	☐	☐	☐	☐
SCoA-VI_10_SIS	I make use of the feedback I get to improve my learning.	☐	☐	☐	☐
SCoA-VI_11_SEQ	Assessment provides information on how well schools are doing.	☐	☐	☐	☐
SCoA-VI_12_SSC	Assessment motivates me and my classmates to help each other.	☐	☐	☐	☐
SCoA-VI_13_SIB(R)	Assessment interferes with my learning.	☐	☐	☐	☐
SCoA-VI_14_SIS	I look at what I got wrong or did poorly on to guide what I should learn next.	☐	☐	☐	☐
SCoA-VI_33_SEA	Assessment tells my parents how much I have learnt.	☐	☐	☐	☐
SCoA-VI_15_SIS	I use assessments to take responsibility for my next learning steps.	☐	☐	☐	☐
SCoA-VI_16_SEA	Assessment results predict my future performance.	☐	☐	☐	☐

		Fully disagree	**Somewhat disagree**	**Somewhat agree**	**Fully agree**
SCoA-VI_17_SSC	Our class becomes more supportive when we are assessed.	☐	☐	☐	☐
SCoA-VI_18_SIB(R)	Teachers are over-assessing.	☐	☐	☐	☐
SCoA-VI_19_SIS	I use assessments to identify what I need to study next.	☐	☐	☐	☐
SCoA-VI_20_SEA	Assessment is important for my future career or job.	☐	☐	☐	☐
SCoA-VI_21_SSC	When we do assessments, there is a good atmosphere in our class.	☐	☐	☐	☐
SCoA-VI_22_SIB(R)	Assessment results are not very accurate.	☐	☐	☐	☐
SCoA-VI_23_SIT	My teachers use assessment to help me improve.	☐	☐	☐	☐
SCoA-VI_24_SEQ	Assessment measures the worth or quality of schools.	☐	☐	☐	☐
SCoA-VI_25_SSC	Assessment makes our class cooperate more with each other.	☐	☐	☐	☐
SCoA-VI_26_SIB(R)	Assessment is value-less.	☐	☐	☐	☐
SCoA-VI_27_SIT	Teachers use my assessment results to see what they need to teach me next.	☐	☐	☐	☐
SCoA-VI_28_SSC	When we are assessed, our class becomes more motivated to learn.	☐	☐	☐	☐
SCoA-VI_29_SIG(R)	I ignore or throw away my assessment results.	☐	☐	☐	☐
SCoA-VI_30_SIT	Assessment shows whether I can analyse and think critically about a topic.	☐	☐	☐	☐
SCoA-VI_31_SPS	I find myself really enjoying learning when I am assessed.	☐	☐	☐	☐
SCoA-VI_32_SIG(R)	Assessment has little impact on my learning.	☐	☐	☐	☐
SAFL-Q_1_SPM	My mathematics teacher encourages me to reflect on how I can improve my assignments.	☐	☐	☐	☐
SAFL-Q_2_SPM	After examining my test results, my mathematics teacher discusses the answers I gave to the test with me.	☐	☐	☐	☐

		Fully disagree	Somewhat disagree	Somewhat agree	Fully agree
SAFL-Q_3_SPM	Whilst working on my mathematics assignments, my mathematics teacher asks me how I think I am doing.	☐	☐	☐	☐
SAFL-Q_4_SPM	My mathematics teacher allows me to think about what I want to learn in school.	☐	☐	☐	☐
SAFL-Q_5_SPM	My mathematics teacher gives me the opportunity to decide on my own learning objectives.	☐	☐	☐	☐
SAFL-Q_6_SPM	My mathematics teacher inquires what went well and what went badly in my work.	☐	☐	☐	☐
PISA-2004_1_SSE	I am sure I can understand even the most difficult topic in mathematics.	☐	☐	☐	☐
SAFL-Q_7_SPM	My mathematics teacher encourages me to reflect on my learning process and to think about how to improve next time.	☐	☐	☐	☐
SAFL-Q_8_SPM	My mathematics teacher stresses my strengths concerning learning.	☐	☐	☐	☐
SAFL-Q_9_SPM	My mathematics teacher identifies my weaknesses concerning learning.	☐	☐	☐	☐
SAFL-Q_10_SPM	I am encouraged by my mathematics teacher to improve my learning process.	☐	☐	☐	☐
SAFL-Q_11_SPM	My mathematics teacher gives me guidance to assist my learning.	☐	☐	☐	☐
SAFL-Q_12_SPM	My mathematics teacher discusses assignments with me to help me understand the subject matter better.	☐	☐	☐	☐
PISA-2004_2_SSE	I am convinced that I can understand even the most difficult topic taught by my mathematics teacher.	☐	☐	☐	☐
SAFL-Q_13_SPM	My mathematics teacher discusses with me the progress I make.	☐	☐	☐	☐
SAFL-Q_14_SPM	After each assessment my mathematics teacher informs me how to improve the next time.	☐	☐	☐	☐
SAFL-Q_15_SPM	My mathematics teacher discusses with me how to exploit my strengths to improve my assignment.	☐	☐	☐	☐
SAFL-Q_16_SPM	My mathematics teacher and I consider ways to improve my weak points.	☐	☐	☐	☐

		Fully disagree	**Somewhat disagree**	**Somewhat agree**	**Fully agree**
SAFL-Q_17_SPS	When I do not understand a topic, my mathematics teacher tries to explain it in a different way.	☐	☐	☐	☐
SAFL-Q_18_SPS	My mathematics teacher provides me with hints to help understand the subject matter.	☐	☐	☐	☐
PISA-2004_3_SSE	I am convinced that I can achieve good results in mathematics homework and exams.	☐	☐	☐	☐
SAFL-Q_19_SPS	During mathematics class I have an opportunity to show what I have learned.	☐	☐	☐	☐
SAFL-Q_20_SPS	My mathematics teacher asks questions in a way I understand.	☐	☐	☐	☐
SAFL-Q_21_SPS	My mathematics teacher asks questions that help me gain understanding of the subject matter.	☐	☐	☐	☐
SAFL-Q_22_SPS	My mathematics teacher allows for my contribution during the lesson.	☐	☐	☐	☐
SAFL-Q_23_SPS	I have the opportunity to ask my classmates questions during the mathematics lesson.	☐	☐	☐	☐
SAFL-Q_24_SPS	My mathematics teacher makes me aware of the areas I need to work on to improve my results.	☐	☐	☐	☐
PISA-2004_4_SSE	I know exactly what to do at home in order to understand the mathematics instruction.	☐	☐	☐	☐
SAFL-Q_25_SPS	There is an opportunity to ask questions during the mathematics lesson.	☐	☐	☐	☐
SAFL-Q_26_SPS	I am aware of the criteria by which my math assignment will be evaluated.	☐	☐	☐	☐
SAFL-Q_27_SPS	When I receive a mathematics assignment it is clear to me what I can learn from it.	☐	☐	☐	☐
SAFL-Q_28_SPS	My mathematics assignments allow me to show what I am capable of.	☐	☐	☐	☐
IFOS-Q_1_SFU	I think feedback from my mathematics teacher is vitally important in improving my mathematics performance.	☐	☐	☐	☐

		Fully disagree	Somewhat disagree	Somewhat agree	Fully agree
FES-Q_1_SFDTE	My mathematics teacher is supportive when giving me feedback about my mathematics performance.	☐	☐	☐	☐
FES-Q_16_SFKST	When I ask for feedback on my mathematics performance, my fellow students generally do not give me the information right away.	☐	☐	☐	☐
IFOS-Q_5_SFU	I think that feedback from my mathematics teacher provides clear direction on how to improve my performance.	☐	☐	☐	☐
FES-Q_11_SFKTE (R)	My mathematics teacher is often annoyed when I directly ask for feedback about my mathematics performance.	☐	☐	☐	☐
FES-Q_5_SFDTE	My mathematics teacher is tactful when giving me feedback about my mathematics performance.	☐	☐	☐	☐
IFOS-Q_6_SFU	Feedback from my mathematics teacher can be a valuable form of praise.	☐	☐	☐	☐
AEQ-5_4_SUF	The feedback on my mathematics assignments prompts me to revise instructional material covered earlier in the course.	☐	☐	☐	☐
FES-Q_2_SFDTE	My mathematics teacher considers my feelings when giving me feedback about my mathematics performance.	☐	☐	☐	☐
IFOS-Q_8_SFU	Feedback from my mathematics teacher motivates me to improve my mathematics performance.	☐	☐	☐	☐
FES-Q_13_SFKTE	I feel comfortable asking my mathematics teacher for feedback about my mathematics performance.	☐	☐	☐	☐
AEQ-5_6_SUF(R)	When I receive written feedback on my mathematics assignments I tend to only read the marks.	☐	☐	☐	☐
FES-Q_12_SFKTE	When I ask for feedback on my mathematics performance, my mathematics teacher generally gives me the information right away.	☐	☐	☐	☐
IFOS-Q_7_SFU	I pay careful attention to instructional feedback from my mathematics teacher.	☐	☐	☐	☐

		Fully disagree	Somewhat disagree	Somewhat agree	Fully agree
FES-Q_3_SFDTE	My mathematics teacher generally provides feedback in a thoughtful manner.	☐	☐	☐	☐
IFOS-Q_3_SFU	I listen carefully when my mathematics teacher provides oral feedback.	☐	☐	☐	☐
FES-Q_18_SFKST	My fellow students encourage me to ask for feedback whenever I am uncertain about my mathematics performance.	☐	☐	☐	☐
AEQ-5_3_SUF	The feedback on my mathematics assignments helps me with subsequent assignments.	☐	☐	☐	☐
FES-Q_4_SFD(R)	My mathematics teacher provides feedback in an intimidating manner.	☐	☐	☐	☐
AEQ-5_1_SUF	I read written feedback on my mathematics assignments carefully to understand what the feedback is saying.	☐	☐	☐	☐
IFOS-Q_2_SFU	I usually reflect on my mathematics teacher’s feedback.	☐	☐	☐	☐
FES-Q_15_SFKST (R)	My fellow students are often annoyed when I ask them for feedback on my mathematics performance.	☐	☐	☐	☐
IFOS-Q_4_SFU	I am extremely encouraged by positive feedback from my mathematics teacher.	☐	☐	☐	☐
AEQ-5_5_SUF	I use feedback on my mathematics assignments for revising.	☐	☐	☐	☐
IFOS-Q_10_SFU	I feel relieved when I receive positive feedback from my mathematics teacher.	☐	☐	☐	☐
FES-Q_14_SFKTE	My mathematics teacher encourages me to ask for feedback whenever I am uncertain about my mathematics performance.	☐	☐	☐	☐
FES-Q_SFKST	I feel comfortable asking my fellow students for feedback about my mathematics performance.	☐	☐	☐	☐
AEQ-5_2_SUF	I use the feedback on my mathematics assignments to review what I have done.	☐	☐	☐	☐

Appendix B. Chapter 2 (Study 1) - Student focus group discussions guide questions

The purpose of FGD with students is to validate student questionnaires items and scales. The FGD will be conducted in 6 secondary schools (3 schools from each region) and among those schools two schools from each school category of HP, LP and MP. The FGD will involve 6-10 students randomly chosen among Form three students. Specifically, the FGD will aim to gain deeper insight on how students experience mathematics assessment practices.

<u>A: Student's assessment and testing practices</u>

1. How do you view mathematics in your class as a subject? (simple, interesting, complicated, challenging etc.) Could you please explain why? (!)

2. Do you believe that you can do well in the final national mathematics examinations? (!)

3. If mathematics was an optional subject in Form three, would you choose to study it or not? Could you please explain your answer? / Would you like to study mathematics at an advanced level secondary school or even further?

4. Could you please give examples of what does your mathematics teacher do when you make errors in mathematics assignments or tests? (!)

5. What kind of feedback is provided, if any, by your mathematics teacher (scores only, scores with comments or comments only)? How do you use such mathematics feedback? (!)

<u>B: Perceived teacher teaching and testing practices</u>

6. To what extent do you think that your mathematics teacher supports you/help you learn from while making errors in class? (!)

7. Are you satisfied with the way feedback is provided to you by your mathematics teacher? Could you please explain your answer? (!)

8. To what extent do you understand the feedback given to you on mathematics assignments or during mathematics class by your teacher? Could you please explain your answer? (!)

9. What do you think could be done to improve formative assessment in the context of mathematics teaching and learning?

10. How do you view the way your mathematics teacher gives you feedback on how well you are doing in mathematics class?

Note. (!) = Priority questions

Appendix C. Chapter 2 (Study 1) - Language measurement invariances test.

Model	SRMR	RMSEA	Df	Chi	Δdf	ΔChi	P	CFI	ΔCFI
Unconstrained	.0444	.030	1708	5995.484				.915	
Measurement weights	.0539	.030	1747	6130.327	39	134.843	.0000	.914	-.001
Measurement intercepts	.0521	.032	1790	6893.929	43	763.602	.0000	.899	-.015
Structural weights	.0534	.032	1793	6913.624	3	19.695	.0001	.899	.000
Structural covariances	.0665	.032	1799	6945.837	6	32.213	.0000	.898	-.001
Structural residuals	.0666	.032	1800	6957.384	1	11.547	.0006	.898	.000
Measurement residuals	.0833	.034	1843	7903.521	43	946.137	.0000	.880	-.018

Appendix D. Chapter 2 (Study 1) - Item-level degree of non-measurement invariance (dMACS).

Scale/Factor	**Item Code**	**Item**	**Area**	**Sqrt(Area)**	**SD**	**dMACS**
Feedback delivery	SDFTE1	1	0.004	0.063	0.882	0.071
	SDFTE2	2	0.001	0.036	1.067	0.034
	SDFTE3	3	0.020	0.143	0.934	0.153
	SDFTE5	4	0.175	0.418	0.970	0.431
Perceived monitoring	SAF1SPM	1	0.007	0.083	0.994	0.084
	SAF2SPM	2	0.044	0.210	1.104	0.190
	SAF3SPM	3	0.017	0.130	1.080	0.120
	SAF4SPM	4	0.007	0.083	1.110	0.075
	SAF6SPM	5	0.007	0.081	1.025	0.079
	SAF7SPM	6	0.014	0.120	0.954	0.126
	SAF8SPM	7	0.054	0.233	0.940	0.248
	SAF9SPM	8	0.000	0.008	0.955	0.008
	SAF10SPM	9	0.010	0.101	0.921	0.110
	SAF11SPM	10	0.002	0.045	0.892	0.051
	SAF12SPM	11	0.003	0.055	0.949	0.058
	SAF13SPM	12	0.005	0.074	1.060	0.069
	SAF14SPM	13	0.007	0.083	1.031	0.081
	SAF15SPM	14	0.012	0.109	1.006	0.108
	SAF16SPM	15	0.001	0.026	1.043	0.025
Perceived scaffolding	SAF17SPS	1	0.007	0.084	0.936	0.090
	SAF18SPS	2	0.007	0.085	0.919	0.092
	SAF19SPS	3	0.015	0.121	0.867	0.139
	SAF20SPS	4	0.001	0.026	0.897	0.029
	SAF21SPS	5	0.003	0.050	0.836	0.060
	SAF22SPS	6	0.001	0.037	0.892	0.042
	SAF23SPS	7	0.012	0.109	0.921	0.118
	SAF24SPS	8	0.004	0.062	0.934	0.066
	SAF25SPS	9	0.001	0.029	0.761	0.038
	SAF26SPS	10	0.003	0.054	0.931	0.058
	SAF27SPS	11	0.004	0.061	0.768	0.079
Feedback use	IFO1SFU	1	0.006	0.075	0.697	0.107
	IFO2SFU	2	0.008	0.092	0.810	0.113
	IFO3SFU	3	0.000	0.007	0.756	0.009
	IFO4SFU	4	0.001	0.028	0.886	0.032
	IFO5SFU	5	0.002	0.042	0.841	0.051
	IFO7SFU	6	0.004	0.066	0.721	0.091
	IFO8SFU	7	0.005	0.071	0.781	0.091
	AE1SUF	8	0.005	0.073	0.813	0.089
	AE2SUF	9	0.001	0.034	0.717	0.048
	AE3SUF	10	0.057	0.239	0.791	0.302
	AE4SUF	11	0.006	0.081	0.849	0.095
	AE5SUF	12	0.012	0.107	0.755	0.142

Note. *dMACS* = differences in a means and covariance structure (MACS), dMACS effect size values can be interpreted as trivial when |*d*MACS| < .20 and small when |*d*MACS| < .40).

Appendix E. Chapter 2 (Study 1) - Student focus group discussions coding scheme

Main Category	**Code**	**Sub-Category**	**Description / Definition**	**Examples**
1. Students' Mathematics Self-Efficacy (SMSE)	SMSE:SIMP	Simple (1)	Students' self-perception of learning mathematics as: interesting subject, simple, usual, difficult, complicated, or a frustrating subject.	Mathematics is the subject which is very simple or easy if you study it.
	SMSE:INTERESTING	Interesting (2)		Mathematics is interesting for me because it easy to study applying formula.
	SMSE:DIFF	Difficult/ Complicated (3)		For me I see mathematics is a difficult subject because it contains a lot of calculations.
	SMSE:CHAL	Challenging (4)		My comment is that mathematics is the challenging subject.
1.1. Reasons for Students' Mathematics Self-Efficacy (R-SMSE)	RSMSE:SIMP-FORM	Simple-Use Formula (1)		Is a simple subject…it has a formula
	RSMSE:INTERESTING	Interesting-Interesting (2)		Mathematics is interesting for me because it easy to study applying formula.
	RSMSE:DIFFICULT-MC	Difficult-Many calculations (3)		Difficult subject because it contains a lot of calculations.
	RSMSE:CHAL	Challenging-More Efforts (4)		Challenging subject because if you study mathematics seriously you must pass but if you don't study it seriously you cannot pass.
	RSMSE: NOT INTERE:	Not interested in maths (5)		Because in our school the students they are not interested in mathematics subject.
	RSMSE: FEW TEACH.	Few Teachers (6)		Because of few teachers
2. Students' mathematics self-efficacy in National Examinations (SMSE-NE)	SMSE-NE:DW-Yes	Can Do Well (1)		
	SMSE-NE:NDW-No	Not Do Well (2)		
2.1. Reasons for (R-SMSE-NE)	SMSE-NE:DW-PRAC	Do Well-Practice a Lot (1)	Students' self-belief in being able to do well (pass) mathematics tasks/ national examinations. To note down all reported causes for student perceived pass/failure in mathematics.	*You know practice makes perfect*, so even if am not doing well mathematics, there is still a year before I do my final national examinations, so if I keep on practicing every day, I believe I will do well.
	SMSE-NE: DW-FPLAN	Do Well-Future Plans With It (2)		I will do well in mathematics because I have the plan with it.
	SMSE-NE:DW-AEF	Do Well-Aware of Examination Format (3)		First I understand the format of the national examinations how it comes.
	SMSE-NE:NDW-DIFF	Not Do Well-Difficult Subject (4)		I use my all powers in order to do well in mathematics but when the results come Ahaa…., So due to that I can say that I don't hope to get good marks in mathematics subject.
	SMSE-NE:NDW-FMT	Not Do Well-Few Mathematics Teachers (5)		Because of few teachers am not sure if I will do well.
3. Students' Willingness Studying Mathematics (SWSM)	SWSM:OPT-MAT	Opt-Studying Mathematics (1)		
	SWSM:NOT OPT-MAT	Not Opt-Studying Mathematics (2)		

3.1. Reasons for Students' Willingness Studying Mathematics (SSWSM)	SWSM:NOM-DIFF	Not Opt Mathematics-Difficult (1)	Identify if students are willingly studying mathematics or studying it because it is a compulsory subject. Further analysis to identify if students are 'planning' to pursue mathematics related courses. Reasons for opting (carrier aspirations) /not-opting for studying mathematics (e.g., avoiding failure).	I could not choose it because mathematics requires many efforts.
	SWSM:NOM-DSN	Not Opt Mathematics-Dislike Mathematics (2)		I wouldn't choose mathematics, let me say the truth, I am running away from calculations because one question you calculate for more than five minutes.
	SWSM:NOM-SAT	Not Opt Mathematics-too Abstract Topics (3)		I can't choose that subject because there are many topics …which will not help me in my life so topics such as logarithm when I ask myself I use a lot of time to learn this topic but in my life which place am I going to use this topic?
	SWSM:OFM-SIMP	Opt Mathematics-Simple (4)		I could choose it because it is the simple subject.
	SWSM:OFM-PADV	Opt Mathematics-Perceived Advantages (5)		I will choose mathematics because it is a very challenging subject and it is believed that when the student knows how to solve mathematics question he can do anything else.
	SWSM:OFM-CASP	Opt Mathematics-Career Aspirations (6)		Of course I will choose it because… my future dream and the course which I want to study. For me I want one day to study computer engineering so I need physics and mathematics; science subjects.
4. Perceived Teacher Reactions to Students' Errors (PTRSE)	PTRSE:COGN SUP	Cognitive Support (1)	Students' perceived teacher support to their errors in mathematics tasks/tests. Analysis of teacher response/support if includes: cognitive support (showing/correcting students errors), affective support (encouraging students and reducing fear resulting from errors), or teacher negative reactions such as reprimand.	After doing the examinations he comes in class, he does corrections and if there are topics we are laying behind we didn't understand he repeats them again.
	PTRSE:AFFECT SUP	Affective Support (2)		He does not punish me he just gives me his opinions/advises on how I can just move from lowest to higher performance.
	PTRSE:PUN-CORP	Reprimand (3)		S: For the first teacher something he has taught and you have missed the question he will punish you severely. R: May be what kind of reprimand? S: He uses normal sticks.
	PTRSE:PUN-TASKS+	Reprimand-More Tasks (4)		He gives me some sort of questions to do always that is why I like it now compared to the previous time.
	PTRSE:DO NOTHING	Does nothing (5)		He does nothing.
5. Type of Feedback Provided (TFBP)	TFBP:SO-NC	Scores Only-No Comments (1)	Analysis of type of mathematics teacher feedback (comments, scores, comments and scores) provided to students.	In the examinations papers we get scores only (tick and cross).
	TFBP:SO+CO	Scores & Comments (2)		Both advise (comments) and marks.
5.1. Student Feedback Use (SFBU)	SFBU:LPFB	Look for Peer Feedback (1)	Further analysis of what students do when they receive teacher feedback (understand it, look for more	I try to follow maybe my friend ask, I try to see the errors then I list them down I go to my friend…telling me maybe the way she did.
	SFBU:LPTH	Look for Teacher Help (2)		On myside when I get mathematics feedback the first thing I look is the grade… if it is poor I follow the teacher

			help from peer / teacher, do nothing).	to correct me.
	SFBU:IB-STRA	If Bad Score-Strategize (3)		If the feedback is bad I continue to practice in order to make it to be good.
	SFBU: Do-N	Does-Nothing (4)		When I get the paper I just wait for a teacher to do correction then if he or she doesn't come I put it in the drawer.
6. Student Perceived Teacher Support in Error Situation (SPTS-ER)	SPTS-ER:SAW/T	Show Alternative Way/Techniques (1)	Reported student perceived teacher support in error situations. Teacher support in error situation may include: identifying error, providing alternative way for error correction, advice students to reduce emotions in error situation.	Because when you do your quiz wrong your teacher corrects you that you should do it this way.
	SPTS-ER:ADVIC	Advise (2)		And advise you where to study more.
	SPTS-ER:LEXT	Large Extent (3)		For me I think it is 70%.
	SPTS-ER:SEXT	Small Extent (4)		I think 25% percent.
7. (Di) Satisfaction with Teacher Feedback-/Delivery (D/STFBD)	SSFBD: SATISFIED	SSFB: Satisfied (1)		
	SSFBD: DISSATISFIED	DSFB: Dissatisfied (2)		
7.1. (Di) Satisfaction with Teacher Feedback-/Delivery (STFB/D)	SSFBD:DE-WHAT	Delivery-Tell What To Do (1)	Student (di) satisfaction with the way teacher provides feedback about their mathematics learning. Highlight if supporting way of providing feedback such as considering student emotions, providing room for improvement in the feedback are practised.	I am satisfied with the way teacher gives us feedback because when giving us feedback tell us how to improve.
	SSFBD:DE-FR+CO	Delivery-Friendly & Convincing (2)		Yes, I am satisfied because she is very friend to us and convinces us to continue studying and do more exercises.
	DSFBD:NS-NPU	Dissatisfied-Reprimand (3)	Detect if unsupportive ways of feedback provision such as not considering student feelings, feedback not showing how to solve the problem, and providing personal focused feedback are practised.	For me am not satisfied because as my fellow student said that if you get low marks you are going to be reprimanded and if you are reprimanded it means for you it is painful so that it can make you to dislike the subject.
	DSFBDNS-NCO	Dissatisfied-No Corrections (4)		He never come and say you got this wrong, no comments, no what.

8. Student Feedback Understanding (SFBU)	SFBU:UND-COM	Understand-Comments (1)	Identify if students report to understand the feedback provided by their mathematics teachers. This could be related to whether the feedback content consists of comments for improvement or crosses and ticks without comments.	For me I understand through the comments which he writes in my mathematics examination.
	SFBU:NUND-CR+TS	Not Understand-No Comments (Crosses & Ticks) (2)		I cannot learn anything because we can't understand that we deserve that 80% or 20% we don't know.
9. Students' Views for Improving Mathematics FA Practices (SVIFA-PR)	SVIFA-PR:+TE+EXA	More Tests & Exams (1)	Students' views on how to improve FA practices. Account for possible constraints for successful implementing FA, and suggestions for overcoming such challenges from student point of view.	On myside the number (frequency) of mathematics examinations/tests should be increased because …mathematics needs a lot of practice
	SVIFA-PR:+MATTEACHR	More Mathematics Teachers (2)		To improve (increase) many teachers of mathematics.
	SVIFA-PR:PC-MDIFF	Perception Change-Mathematics Is Difficult (3)		Some of the students should be given some counselling to change the mind-sets of/ that mathematics is difficult they should erase that thing from their minds because mathematics is easy when you practice.
	SVIFA-PR:+QMATBOOKS	Increase Quality Mathematics Books (4)		In my views I think number of books in our schools should be increased … we cannot practice much because we don't have many books.
	SVIFA-PR:PRO-F-T	Professional & Friendly-Teachers (5)		To have the teacher who can teach us or can give us assignments or feedback in order to understand and to have materials which we can study can make us perfect.
	SVIFA-PR:PE-COOPR	Peer Cooperation (6)		Students should form group discussions because in group discussion everyone gives the ideas...make students to perform well.
10. Student Perception of Teacher Feedback Delivery	SPTFD:BAD PRACT	Bad Practices (1)		Our teacher gives feedback in a very poor way and this is because he has a heavy teaching load.
	SPTFD:GOOD PRACT:	Good Practices (1)		I like the way our madam (teacher) gives us feedback because she gives us kindly even though you performed badly she doesn't shout at you, she doesn't get mad at you she always encourage you.

Appendix F. Chapter 2 (Study 1) - Syntax for a four factors structural equation model

```
TITLE:  A four factors model: Perceptions of AfL, feedback use and mathematics performance;
DATA:  FILE IS 161222_SPSS_Student_no Missing.dat;
LISTWISE IS OFF;
FORMAT IS FREE;
VARIABLE:   NAMES ARE
          ID SFDTE1 SFDTE2 SFDTE3 SFDTE5 SFDTE4r Respgrnb Msf3te SAF1SPM
         SAF2SPM SAF3SPM SAF4SPM SAF5SPM SAF6SPM SAF7SPM SAF8SPM SAF9SPM
         SAF10SPM SAF11SPM SAF12SPM SAF13SPM SAF14SPM SAF15SPM SAF16SPM
         SAF17SPS SAF18SPS SAF19SPS SAF20SPS SAF21SPS SAF22SPS SAF23SPS
         SAF24SPS SAF25SPS SAF26SPS SAF27SPS SAF28SPS IFO1SFU IFO2SFU IFO3SFU
         IFO4SFU IFO5SFU IFO6SFU IFO7SFU IFO8SFU
         IFO10SFU AE1SUF AE2SUF AE3SUF AE4SUF AE5SUF AE6SUFr;
USEVARIABLES ARE
         SFDTE1 SFDTE2 SFDTE3 SFDTE5 Respgrnb Msf3te SAF4SPM SAF6SPM SAF7SPM
         SAF10SPM SAF13SPM SAF14SPM SAF16SPM SAF19SPS SAF20SPS SAF21SPS
         SAF22SPS IFO1SFU IFO2SFU AE4SUF AE5SUF;
CATEGORICAL ARE;

IDVARIABLE IS ID;
 MISSING IS ALL (9999);
CLUSTER IS Respgrnb;                     ! Respgrnb denotes teacher/class code

ANALYSIS:
TYPE IS COMPLEX;
MODEL:
! MEASUREMENT MODEL:
SF_FDTE BY SFDTE1* SFDTE2 SFDTE3 SFDTE5;
SF_FDTE@1;             ! SF_FDTE denoted student perceptions of feedback delivery scale
SF_SPM BY SAF4SPM* SAF6SPM SAF7SPM   SAF10SPM SAF13SPM SAF14SPM SAF16SPM;
         SF_SPM@1;              ! SF_SPM denotes perceived monitoring scale
SF_SPS BY SAF19SPS* SAF20SPS SAF21SPS SAF22SPS;
 SF_SPS@1;             ! SF_SPS denotes student perceived scaffolding scale
SI_SFU BY IFO1SFU* IFO2SFU   AE4SUF AE5SUF;
SI_SFU@1;              ! SI_SFU denotes student feedback use scale

! STRUCTURE MODEL:
SI_SFU ON SF_FDTE SF_SPM SF_SPS;
   Msf3te ON SI_SFU;
   SF_SPS WITH SF_SPM;
   SF_SPS WITH SF_FDTE;
   SF_SPM WITH SF_FDTE;

!  MODEL INDIRECT:          ! This syntax test for mediation effect
   Msf3te VIA SI_SFU SF_SPS;
   Msf3te VIA SI_SFU SF_SPM;
   Msf3te VIA SI_SFU SF_FDTE;

OUTPUT:
    samp res stdyx mod;
```

Appendix G. Chapter 2 (Study 1) - Items retained in scales

Scale	**Item code**	Item content
Perceived monitoring	SAFL_Q_4_SPM	My mathematics teacher allows me to think about what I want to learn in school.
	SAFL_Q_6_SPM	My mathematics teacher inquires what went well and what went badly in my work.
	SAFL_Q_7_SPM	My mathematics teacher encourages me to reflect on my learning process and to think about how to improve next time.
	SAFL_Q_10_SPM	I am encouraged by my mathematics teacher to improve my learning process.
	SAFL_Q_13_SPM	My mathematics teacher discusses with me the progress I make.
	SAFL_Q_14_SPM	After each assessment my mathematics teacher informs me how to improve the next time.
	SAFL_Q_16_SPM	My mathematics teacher and I consider ways to improve my weak points.
Perceived scaffolding	SAFL_Q_19_SPS	During mathematics class I have an opportunity to show what I have learned.
	SAFL_Q_20_SPS	My mathematics teacher asks questions in a way I understand.
	SAFL_Q_21_SPS	My mathematics teacher asks questions that help me gain understanding of the subject matter.
	SAFL_Q_22_SPS	My mathematics teacher allows for my contribution during the lesson.
Feedback delivery	FES_Q_1_SFDTE	My mathematics teacher is supportive when giving me feedback about my mathematics performance.
	FES_Q_2_SFDTE	My mathematics teacher considers my feelings when giving me feedback about my mathematics performance.
	FES_Q_3_SFDTE	My mathematics teacher generally provides feedback in a thoughtful manner.
	FES_Q_5_SFDTE	My mathematics teacher is tactful when giving me feedback about my mathematics performance.
Feedback use	AEQ_5_4_SUF	The feedback on my mathematics assignments prompts me to revise instructional material covered earlier in the course.
	AEQ_5_5_SUF	I use feedback on my mathematics assignments for revising.
	IFOS_Q_2_SFU	I usually reflect on my mathematics teacher's feedback.
	IFOS_Q_1_SFU	I think feedback from my mathematics teacher is vitally important in improving my mathematics performance.

Appendix H. Chapter 2 (Study 1) - Items removed from scales

Scale	Item code	Item content
Perceived monitoring	SAFL_Q_1_SPM	My mathematics teacher encourages me to reflect on how I can improve my assignments.
	SAFL_Q_2_SPM	After examining my test results, my mathematics teacher discusses the answers I gave to the test with me.
	SAFL_Q_3_SPM	Whilst working on my mathematics assignments, my mathematics teacher asks me how I think I am doing.
	SAFL_Q_5_SPM	My mathematics teacher gives me the opportunity to decide on my own learning objectives.
	SAFL_Q_8_SPM	My mathematics teacher stresses my strengths concerning learning.
	SAFL_Q_9_SPM	My mathematics teacher identifies my weaknesses concerning learning.
	SAFL_Q_11_SPM	My mathematics teacher gives me guidance to assist my learning.
	SAFL_Q_12_SPM	My mathematics teacher discusses assignments with me to help me understand the subject matter better.
	SAFL_Q_15_SPM	My mathematics teacher discusses with me how to exploit my strengths to improve my assignment.
Perceived scaffolding	SAFL_Q_17_SPS	When I do not understand a topic, my mathematics teacher tries to explain it in a different way.
	SAFL_Q_18_SPS	My mathematics teacher provides me with hints to help understand the subject matter.
	SAFL_Q_23_SPS	I have the opportunity to ask my classmates questions during the mathematics lesson.
	SAFL_Q_24_SPS	My mathematics teacher makes me aware of the areas I need to work on to improve my results.
	SAFL_Q_25_SPS	There is an opportunity to ask questions during the mathematics lesson.
	SAFL_Q_26_SPS	I am aware of the criteria by which my math assignment will be evaluated.
	SAFL_Q_27_SPS	When I receive a mathematics assignment it is clear to me what I can learn from it.
	SAFL_Q_28_SPS	My mathematics assignments allow me to show what I am capable of.
Feedback delivery-teacher	FES_Q_4_SFDr	My mathematics teacher provides feedback in an intimidating manner.
Feedback use	IFOS_Q_1_SFU	I think feedback from my mathematics teacher is vitally important in improving my mathematics performance.
	IFOS_Q_5_SFU	I think that feedback from my mathematics teacher provides clear direction on how to improve my performance.
	IFOS_Q_6_SFU	Feedback from my mathematics teacher can be a valuable form of praise.
	IFOS_Q_8_SFU	Feedback from my mathematics teacher motivates me to improve my mathematics performance.
	AEQ_5_6_SUFr	When I receive written feedback on my mathematics assignments I tend to only read the marks.
	IFOS_Q_7_SFU	I pay careful attention to instructional feedback from my mathematics teacher.
	IFOS_Q_3_SFU	I listen carefully when my mathematics teacher provides oral feedback.
	AEQ_5_3_SUF	The feedback on my mathematics assignments helps me with subsequent assignments.
	AEQ_5_1_SUF	I read written feedback on my mathematics assignments carefully to understand what the feedback is saying.
	IFOS_Q_10_SFU	I feel relieved when I receive positive feedback from my mathematics teacher.
	AEQ_5_2_SUF	I use the feedback on my mathematics assignments to review what I have done.
	IFOS_Q_1_SFU	I think feedback from my mathematics teacher is vitally important in improving my mathematics performance.

Appendix I. Chapter 3 (Study 1) - Teacher questionnaire

Through this questionnaire, you are invited to provide information regarding your formative assessment practices. The questionnaire covers some general information (such your teaching experience), your beliefs about assessment, your formative assessment practices and your feedback practices in your mathematics class. Some questions are similar to those asked to your students. There are no right or wrong answers. Your answers will remain anonymous.

1. Date: ______ **(day)** ___________ **(month)** ______ **(year)**

2. Answer the following questions. Together they form an anonymous code:

	The last letter of your first name
	The first letter of your surname
	The first letter of your month of birth
	The first letter of your favourite food

3. Gender**:** ☐ Male ☐ Female 4. Age ______ **(years)**

5**.** What is your highest education qualification level? 6. Nationality_______________

☐ Diploma ☐ Bachelor ☐ Master degree

☐ Doctoral ☐ Other (please specify) _____________________________________

7. What is your mathematics teaching experience (**in years**) ______________

8. In which Forms do you teach mathematics? Please **list them all** ____________________

9. What is your average teaching load (periods) **per week** _________

10. What is the average number of students in your mathematics class (es) __________

To be completed by the Researcher

Schoolnumber		District		Region	
Respondentgroup					
School ownership		Lettercode TE		School category	

Read each statement carefully and tick (√) inside the box alongside your best option among ***fully disagree, somewhat disagree, somewhat agree or fully agree.***

		Fully disagree	Somewhat disagree	Somewhat agree	Fully agree
COA-III_1_TSQ	Assessment provides information on how well schools are doing.	☐	☐	☐	☐
COA-III_2-TPG	Assessment places students into categories.	☐	☐	☐	☐
COA-III_3_TIS	Assessment is a way to determine how much students have learned from teaching.	☐	☐	☐	☐
COA-III_25_TBA(R)	Assessment interferes with teaching.	☐	☐	☐	☐
COA-III_7_TIN(R)	Assessment forces teachers to teach in a way against their beliefs.	☐	☐	☐	☐
COA-III_26_TBA(R)	Assessment has little impact on teaching.	☐	☐	☐	☐
COA-III_13_TIS	Assessment feeds back to students their learning needs.	☐	☐	☐	☐
COA-III_16_TIN(R)	Assessment is unfair to students.	☐	☐	☐	☐
COA-III_4_TIS	Assessment provides feedback to students about their performance.	☐	☐	☐	☐
COA-III_10_TSQ(R)	Assessment is an accurate indicator of a school's quality.	☐	☐	☐	☐
COA-III_5_TIS	Assessment is integrated with teaching practice.	☐	☐	☐	☐
COA-III_8-_TIN	Teachers conduct assessments but make little use of the results.	☐	☐	☐	☐
COA-III_27_TBA(R)	Assessment is an imprecise process.	☐	☐	☐	☐
COA-III_22_TIS	Assessment helps students improve their learning.	☐	☐	☐	☐
COA-III_11_TPG	Assessment is assigning a grade or level to student work.	☐	☐	☐	☐
COA-III_12_TIS	Assessment establishes what students have learned.	☐	☐	☐	☐
COA-III_17_TINr)	Assessment results are filed and ignored.	☐	☐	☐	☐

		Fully disagree	Somewhat disagree	Somewhat agree	Fully agree
COA-III_14_TIS	Assessment information modifies ongoing teaching of students.	☐	☐	☐	☐
COA-III_15_TSQ	Assessment results are consistent.	☐	☐	☐	☐
COA-III_18_TIN	Teachers should take into account the error and imprecision in all assessment.	☐	☐	☐	☐
COA-III_19_TSQ	Assessment is a good way to evaluate a school.	☐	☐	☐	☐
COA-III_20_TIS	Assessment determines if students meet qualifications standards.	☐	☐	☐	☐
COA-III_6_TSQ	Assessment results are trustworthy.	☐	☐	☐	☐
COA-III_21_TIS	Assessment measures students' higher order thinking skills.	☐	☐	☐	☐
COA-III_9_TIN(R)	Assessment results should be treated cautiously because of measurement error.	☐	☐	☐	☐
COA-III_23_TIS	Assessment allows different students to get different instruction.	☐	☐	☐	☐
COA-III_24_TSQ	Assessment results can be depended on.	☐	☐	☐	☐
TAFL-Q_1_TPM	I encourage my students to reflect upon how they can improve their assignments.	☐	☐	☐	☐
TAFL-Q_2_TPM	After a test, I discuss the answers given with each student.	☐	☐	☐	☐
TAFL-Q_3_TPM	While working on their assignments, I ask my students how they think they are doing.	☐	☐	☐	☐
TAFL-Q_4_TPM	I involve my students in thinking about how they want to learn at school.	☐	☐	☐	☐
TAFL-Q_5_TPM	I give my students the opportunity to decide on their learning objectives.	☐	☐	☐	☐
TAFL-Q_6_TPM	I ask my students to indicate what went well and what went badly concerning their assignments.	☐	☐	☐	☐
TAFL-Q_7_TPM	I encourage students to reflect upon their learning processes and how to improve their learning.	☐	☐	☐	☐
PISA-2004_1_TTE	I believe that I can inspire students to solve new mathematics problems.	☐	☐	☐	☐
TAFL-Q_8_TPM	I inform my students on their strong points concerning learning.	☐	☐	☐	☐
TAFL-Q_9_TPM	I inform my students on their weak points concerning learning.	☐	☐	☐	☐

		Fully disagree	Somewhat disagree	Somewhat agree	Fully agree
TAFL-Q_10_TPM	I encourage my students to improve on their learning processes.	☐	☐	☐	☐
TAFL-Q_11_TPM	I give students guidance and assistance in their learning.	☐	☐	☐	☐
TAFL-Q_12_TPM	I discuss assignments with my students to help them understand the content better.	☐	☐	☐	☐
TAFL-Q_13_TPM	I discuss with my students the progress they have made.	☐	☐	☐	☐
TAFL-Q_14_TPM	After an assessment, I inform my students on how to improve their weak points.	☐	☐	☐	☐
PISA-2004_2_TTE	Even when I am not feeling well, I can still teach my students well.	☐	☐	☐	☐
TAFL-Q_15_TPM	I discuss with my students how to utilize their strengths to improve on their assignment.	☐	☐	☐	☐
TAFL-Q_16_TPM	Together with my students, I consider ways on how to improve on their weak points.	☐	☐	☐	☐
TAFL-Q_17_TPS	I adjust my instruction whenever I notice that my students do not understand a topic.	☐	☐	☐	☐
TAFL-Q_18_TPS	I provide my students with guidance to help them gain understanding of the content taught.	☐	☐	☐	☐
TAFL-Q_19_TPS	During my class, students are given the opportunity to show what they have learned.	☐	☐	☐	☐
TAFL-Q_20_TPS	I ask questions in a way my students understand.	☐	☐	☐	☐
TAFL-Q_21_TPS	I ask questions during class that help my students gain understanding of the content taught.	☐	☐	☐	☐
PISA-2004_3_TTE	I am confident that I can develop creative ideas to change unfavorable teaching structures.	☐	☐	☐	☐
TAFL-Q_22_TPS	I am open to student contribution in my class.	☐	☐	☐	☐
TAFL-Q_23_TPS	I allow my students to ask each other questions during class.	☐	☐	☐	☐
TAFL-Q_24_TPS	I ensure that my students know what areas they need to work on in order to improve their results.	☐	☐	☐	☐
TAFL-Q_25_TPS	I give my students opportunities to ask questions.	☐	☐	☐	☐

		Fully disagree	Somewhat disagree	Somewhat agree	Fully agree
TAFL-Q_26_TPS	My students know what the evaluation criteria for their work are.	☐	☐	☐	☐
TAFL-Q_27_TPS	I ensure that my students know what they can learn from their assignments.	☐	☐	☐	☐
TAFL-Q_28_TPS	I can recognize when my students reach their learning goals.	☐	☐	☐	☐
PISA-2004_4_TTE	I am confident that I can maintain a good relationship with all students (including problematic students) when I invest the effort.	☐	☐	☐	☐
IFOS-Q_2_TFU	I think my students usually reflect on my feedback.	☐	☐	☐	☐
FES-Q_4_TFDTE(R)	My students feel intimidated when I provide feedback on their mathematics performance.	☐	☐	☐	☐
FES-Q_18_TFKTS	My students encourage each other to ask for feedback whenever they are uncertain about their mathematics performance.	☐	☐	☐	☐
AEQ-5_1_TUF	My students read my written feedback on their mathematics assignments carefully to understand what the feedback is saying.	☐	☐	☐	☐
FES-Q_15_TFKTS(R)	My students are often annoyed when fellow students ask them for feedback about their mathematics performance.	☐	☐	☐	☐
IFOS-Q_10_TFU	My students feel relieved when they receive positive feedback.	☐	☐	☐	☐
AEQ-5_5_TUF	My students use my feedback on their mathematics assignments for revising.	☐	☐	☐	☐
IFOS-Q_7_TFU	My students pay careful attention to instructional feedback that I provide.	☐	☐	☐	☐
AEQ-5_6_TUF(R)	When my students receive my written feedback on their mathematics assignments they tend to only read the marks.	☐	☐	☐	☐
IFOS-Q_3_TFU	My students listen carefully when I provide oral feedback.	☐	☐	☐	☐
AEQ-5_2_TUF	My students use my feedback on their mathematics assignments to review what they have done.	☐	☐	☐	☐
IFOS-Q_4_TFU	My students are extremely encouraged by positive feedback.	☐	☐	☐	☐
FES-Q_16_TFKTS	When my students ask for feedback about their mathematics performance from fellow students, they receive the information right away	☐	☐	☐	☐

		Fully disagree	Somewhat disagree	Somewhat agree	Fully agree
FES-Q_17_TFKTS	My students feel comfortable asking fellow students for feedback about their mathematics performance.	☐	☐	☐	☐
FES-Q_1_TFDTE	I am supportive when giving my students feedback about their mathematics performance.	☐	☐	☐	☐
IFOS-Q_1_TFU	I think my feedback is vitally important in improving student mathematics performance.	☐	☐	☐	☐
FES-Q_2_TFDTE	I consider my students' feelings when giving them feedback.	☐	☐	☐	☐
TFKTT_1	I often discuss a mathematics lesson with my fellow mathematics teachers.	☐	☐	☐	☐
IFOS-Q_5_TFU	I think that my feedback provides clear direction on how my students can improve their mathematics performance.	☐	☐	☐	☐
FES-Q_3_TFDTE	I generally provide feedback in a thoughtful manner.	☐	☐	☐	☐
IFOS-Q_6_TFU	My feedback can be a valuable form of praise.	☐	☐	☐	☐
FES-Q_5_TFDTE	I am tactful when giving my students feedback about their mathematics performance.	☐	☐	☐	☐
AEQ-5_4_TUF	My feedback on mathematics assignments prompts my students to revise instructional material covered earlier in the course.	☐	☐	☐	☐
FES-Q_11_TFKTE(R)	I feel annoyed when my students ask for feedback on their mathematics performance.	☐	☐	☐	☐
IFOS-Q_8_TFU	My feedback motivates my students to improve their mathematics performance.	☐	☐	☐	☐
FES-Q_12_TFKTE	When my students ask for feedback on their mathematics performance, I generally give the information right away.	☐	☐	☐	☐
TFKTT_3	I feel comfortable asking my fellow mathematics teachers for feedback on my teaching.	☐	☐	☐	☐
FES-Q_13_TFKTE	I feel comfortable when my students ask for feedback about their mathematics performance.	☐	☐	☐	☐
TFKTT_2	I often discuss the quality of mathematics tests that I construct with my fellow mathematics teachers.	☐	☐	☐	☐

		Fully disagree	Somewhat disagree	Somewhat agree	Fully agree
FES-Q_14_TFKTE	I encourage my students to ask for feedback whenever they are uncertain about their mathematics performance.	☐	☐	☐	☐
TFKTT_4	I feel comfortable asking my fellow mathematics teachers for feedback on tests that I construct.	☐	☐	☐	☐
AEQ-5_3_TUF	My feedback on mathematics assignments helps my students with subsequent assignments.	☐	☐	☐	☐

When you think about ASSESSMENT which of these kinds of PRACTICES do you have in mind?

--- Please tick all that apply ---

☐ Unplanned observation

☐ Marked student mathematics homework

☐ Student self-assessment (e.g., student marking his/her assignment using textbook answers)

☐ Peer assessment (e.g., students marking an assignment from a fellow student)

☐ Conferencing (e.g., individual student remedial consultations)

☐ Teacher made written test (e.g., mid-term, terminal and annual examinations)

☐ Standardized test (e.g., mock examination, national examinations)

☐ Feedback on student tests

☐ 1-3 hour examination (e.g., national examinations)

☐ Oral question & answer during mathematics class time

☐ Planned observation (e.g., running record, checklist, record of student scores)

☐ Student written work (e.g., mathematics homework)

☐ Portfolio/scrapbook

Appendix J. Chapter 3 (Study 1) - Interview with mathematics teachers guide questions

This interview aims to collect data that triangulate with and elaborate on questionnaire data. In particular, the interview questions seek to gain deeper insight regarding teacher assessment perceptions and practices.

<u>A: Teaching practices and testing practices</u>

1. How do you view the teaching of mathematics in your class? (simple, usual, frustrating, or complicated task, etc.) (!)
2. How do you inform your students about their errors in the mathematics written tests/ and or in-class assignments? Please can you give one example on how you respond to student errors during class?

4. What is FA to you as teacher of mathematics?
5. How do you view FA/CA? How do you think FA contributes to your student's mathematics learning, if so? (!)
6. Do you think there is a necessity to improve FA/CA in mathematics teaching or learning? If yes, how could it be done?
7. Do you construct tests or adopt available tests? If you adopt, from which sources? If you do not adopt, please explain your answer.
8. To what extent do you use Continuous Assessment (CA) guidelines if you develop mathematic tests yourself? (!)
9. To what extent do you conduct departmental standardization of mid-term, terminal and annual mathematics tests? How is this standardization achieved and maintained? (!)
10. Could you please give any example on how do you typically use student's assessment information such as their scores in terminal and mid-term test? (!)

<u>B: Perceived student experiences with teaching and testing practices</u>

11. Do you believe that your students can do well in the final national mathematics examinations? Why do you think so? (!)
12. How do you think your students treat your mathematics feedback?
13. How do you deal with your students' reactions to the feedback you provide?

Note. (!) = Priority questions

Appendix K. Chapter (Study 1) -Teacher interview coding scheme

Main Category	**Code**	**Sub-Category**	**Description / Definition**	**Examples**
1.Teaching Efficacy (TEF)	TEF:SIMP	Simple (1)	Teachers' belief in being able to use various teaching strategies, to respond to students' questions, or engage students at adequate levels of competence. Reported teacher perceptions of teaching mathematics as interesting task, simple, usual, difficult, complicated, or a frustrating task.	Teaching mathematics is a simple task to me.
	TEF:INTERESTING	Interesting (2)		Normally, I enjoy teaching mathematics.
	TEF:USUAL	Usual (3)		To me teaching mathematics is a usual task.
	TEF:DIFF	Difficult (4)		In some situations teaching mathematics is very difficult.
	TEF:CHAL	Challenging (5)		I can say it is a challenging task
	TEF:FRUST	Frustrating (6)		Teaching is simple, results are frustrating
1.1 Reasons for TEF	SIMP: EXPERTISE	Expertise/Professionalism/Experience (1)	Reasons for reported teacher perceptions of teaching mathematics as interesting task, simple, usual, difficult, complicated, or a frustrating task.	For me it is a usual task because I am a professional in mathematics so for me it is a usual one.
	SIMP: +RESOURCES	Available resources (2)		We are luck here we have some of the materials but I am sure in other grounds (schools) they don't have those materials.
	SIMP: INTEREST	Enjoy Teaching Mathematics (3)		I love mathematics myself so I can't see that it is a very tough thing.
	DIFF: LACK T-RE	Lack teaching resource /aids (4)		There is no enough teaching aids for instance topics such as of 'earth as the sphere' and 'circles' there are no enough instruments.
	DIFF: BIG CLASS SIZE	Big class size/Teaching load (5)		First of all they have to reduce the number of students in a class because am supposed to teach only 45 students in a class but now I teach almost 200 students.
	DIFF: POOR FOUN	Poor students' maths foundation/background (6)		Sometimes, I just use much of time in making a clear concept so that everyone can understand.
2.Teacher Reaction to Errors (TRNE)	TRNE:+VE COG SUPP	Cognitive Support (1)	Teachers' response to student errors in mathematics tasks/tests. Analysis of teacher response if includes: cognitive support (showing and correcting students errors), affective support (encouraging students and reducing emotions resulting from errors), or negative reactions such as reprimand and task based reprimand such as assigning more tasks to students.	After marking if there are some errors I go in class to correct those errors.
	TRNE:+VE AF SUPP	Affective Support (2)		So I try to give them hope keep practicing.
	TRNE:-VE PUN-CORP	Reprimand (3)		Also we reprimand them if they score let say below 20.
	TRNE:-VE PUN+TASKS	Reprimand-More Tasks (4)		If gave 10 questions in the assignment, I can raise the questions up to 50.

4. Teacher FA Knowledge (TFAK)	TFAK:TES+ EXM	Tests& Examinations (1)	Teacher awareness of FA as an integral in the instructional process such as feedback practices, or perceiving FA as summative examinations and tests not-integrated in the instructional process.	FA are those tests we give to students...like terminal, mid-term tests and annual.
	TFAK:CL-PR	Classroom Practices (2)		I think it is the assessment within the lesson.
	TFAK:FB-T	Feedback for Teaching (3)		I can use FA to see whether what I taught was successful or not.
5. Teacher FA Perception (TFAP)	TFAP-GFBT	Good-Feedback to Teaching/ Learning (1)	Positive perceptions appreciate FA as a tool for providing immediate feedback to the instructional process (to teachers and students). Negative perception considers FA as a boring task, too demanding in providing feedback to students. Constraints for successful FA such as heavy teaching load, student negative perception towards mathematics are categorised under negative perceptions of FA.	With FA at each stage of teaching you realize mistakes to be corrected. FA really supports my students to have good performance in examinations.
	TFAP-GVA	Good-Valid Student Ability (2)		Students' summative results may not reflect the actual ability of the student as FA does.
	TFAP-BHTL	Bad-Heavy Teaching /Marking Load (3)		Sometimes a lot of examinations may interfere normal teaching.
	TFAP-BSP	Bad-Student Perception /Cooperation (4)		Sometimes student perceptions make mathematics assessment difficult.
6:Improve FA Practices (TIFAPR)	TIFAPR:+IT_ASS	Increase Tests & Other Assessment methods (1)	Teachers' views on how to improve FA practices. Account for possible constraints for successful implementing FA, and suggestions for overcoming possible challenges.	If there could be frequent exams this will help them to study and maybe help them to improve. We use tests, but we can find other ways of assessing them.
	TIFAPR:+FB	More Feedback (2)		Seriously make check up to the students what they do/what they fail to do.
	TIFAPR:R-TL_TL	Reduce Teaching/Testing Load (3)		Reduce the number of tests so that you don't resort most time in assessing rather than in teaching.
	TIFAPR:NIN	No Improvement Needed (4)		For the time the way it is done here I think it is the best.
	TIFAPR:ACT ON	Act On Assessment Info (5)		Sometimes...we give the test but we don't consider implementing the results we are just recording it.
7.Test Practices Construct/ Adopt (TTPR)	TTPR:CO-Q	Construct Own-Questions (1)	Testing practices showing how mathematics teachers prepare tests, either (i) construct own tests or (ii) adopt tests/questions from other sources.	In fact we have to construct our own tests.
	TTPR:A-VS	Adopt _Various Sources(2)		I adopt questions from various past papers and from various schools.
8.Using Assessment Guidelines (TUAG)	TUAG:NO-G	No-Guidelines (1)	Identification of available assessment guidelines for supporting teacher test construction practices.	Mhhhh, anyway these are not so much applied...for Form three we are not following those formats.
	TUAG:U_NF_S	Use-NECTA Format/Syllabus(2)		Yes, because the distribution of marks should be according to the national exams. Yes, I use the syllabus to measure the student development.

	TUAG:U-OWNG	Use-Own Guidelines(3)		Use my own knowledge from teacher's college on how a test should look like.
8.1. Perceived Usefulness of CA Guidelines (TUAG)	TPUG:UQT	Useful Quality Test(1)	Reported mathematics teacher perceived usefulness and challenges of the assessment guidelines.	I think in most cases they are helpful in the sense that they make the teacher construct a standard exam
	TUAG:NOT	Not Useful (2)		I am the examination officer here and I am an experienced teacher, and then sometimes it is not necessary for the head of department to cross-check.
9. Departmental Standardisation of Exams (DSTE): Who Does that?	DSTE:S-DM	Some Department Members (1)	Finding out process available in schools for monitoring the quality of teacher made tests such as departmental standardisation (moderation) of mathematics tests. A description on members of the departmental standardisation committee (if exist).	That one is happening ……….between teachers.
	DSTE:HOD	Head of Department only (2)		Every teacher after constructing the test it goes through the head of department.
	DSTE:NOT-D	Not-Done (3)	No standardisation of test is done.	What I prepare is what I give students.
9.1. Perceived Usefulness of Standardisation (DSTE)	DSTE:U-IQ	Useful-Increase Quality (1)	Teacher perception towards test moderation or/construction process as (constructive or not).	The process is good...the idea from more than one teacher is constructive.
	DSTE:NU-LQ	Not Useful-Lower Quality (2)		It takes time because sometimes we have a lot of things to do.
10.Using Assessment Information (TUAI)	TUAI:SSHTI	Show Students How to Improve (1)	All reported use of assessment information such as student scores in examinations or tests. Analysis of assessment based decisions (re-teaching difficult topics, feedback to students, standardisation of results, reports to parents, reprimand, etc.) will be further categorised into potential 'formative' and 'potential 'summative' assessment based decisions.	From how they have performed...I can perceive weaknesses to different students ...help them accordingly.
	TUAI:DMTA	Devise My Teaching Approaches / Re-Teach Difficult Topics (2)		I just go through the questions and see the way they have performed...can tell me what to be done next time. I try to solve different questions which are very difficult to them.
	TUAI:CSIAG	Categorise Students into Ability Groups (3)		In our school we separate them...we are just identifying them as slow learners…. get a special attention.
	TUAI:ACR_P_NE	Accountability Reports (Parents/NECTA) (4)		Use to prepare reports for NECTA. They are well implemented...we use to fill the parents' reports.
	TUAI:MOT-HAC	Motivate-High Achievers (5)		If students register a positive deviation in comparison with the previous exams…we motivate them
	TUAI:PUN-LAC	Reprimand-Low Achievers(6)		Those making mistake because of laziness I usually take measures… give them reprimand. If… you didn't get the average so you just have to repeat the class or find another school.

11.Perceived Student Mathematics Efficacy (PSME)	PSME:DW	Do Well (1)	Teachers' belief in their students' ability to pass mathematics national examinations.	Yeah for me I do believe that my students they are going to do better.
	PSME:NDW	Not Do Well (2)		I believe they can do well if they can change their attitude ...everything is possible if they work hard.
11.1 Reasons For PSME	PSME:DW-RE-GTS	Do Well-Remedial Classes & My Good Teaching Strategies (1)	Potential sources of teacher believe about their students' mathematics efficacy. For example, believing student will do well based on their previous performances, teaching approaches, or believing that students will not do well because of students' negative attitude towards mathematics	Sometimes I conduct remedial classes. Because I use a lot of good teaching methods.
	PSME:DW-APE	Do Well-As Previous Examinations (2)		Exactly, because they did it in Form two, why not Form four?
	PSME:NDW-VE PERM	Not Do Well-Negative Attitude (3)		I believe they can do well if they can change their attitude ...everything is possible if they work hard.
12:Perceived Student Feedback Treatment (SPFT)	PSFT:MOST-UFB	Most-Use Feedback to Improve (1)	Teachers' perception of their students' treatment of mathematics feedback. Positive perceived student use of feedback exemplified by using feedback to correct mistakes (errors) and asking more questions from the feedback. Negative perception of relates to not using feedback, rejecting/throwing away feedback. Highlight some conditions fostering feedback acceptance/rejection (good performance vs. failure).	It depends if the feedback will be good they will be able to increase more efforts in order to do well in next test.
	PSFT:IGFB-GU	Ignore Feedback-Given Up (2)		Yeah, most of them they make use of that feedback but some of them they just ignore it.
13:Response to Students' Feedback Reactions (TRFR)	TRFR:DIFB_RT	Discuss Feedback & Provide-Remedial Teaching (1)	Depicting teacher response to students' feedback reactions. A clear identification of possible strategies done by teachers to increase chances of feedback acceptance and application by students such as explaining the feedback, encouraging students. Barriers for student feedback use such as reprimand, or not taking further steps after providing feedback.	-I conduct remedial classes so that to make them improve their mathematics performance. -When I am going to provide feedback to my students first of all I try to explain how the feedback is...
	TRFR:NW-SGU	No Way-Some Given Up (2)		Actually if what she has on the paper is what she deserves, there are no more marks.
	TRFR: COUN&ENCO-PR	Counsel & Encourage Students to Practice More (3)		First of all am encouraging them that even if you got poor performance I am sure next test/exam you will perform well.

Appendix L. Chapter 4 and 5 (Study 2) - Overview of the feedback and error handling training.

Block	*Instruction*	*Activities*	*Duration*	*Materials by instructor*	*Materials by teachers*
1: 9:00-10:30	Performance trends of mathematics education in Tanzanian secondary schools.	Teachers were asked for their views on current mathematics education.	20 minutes	PPT presentation of students' previous results in secondary mathematics (2003-2013) (link to Continuous Assessment).	Brainstorming on possible causes and how to reverse the trend (teacher side).
	Formative Assessment (theory).	Teachers were asked questions about their previous FA practices.	20 minutes	PPT presentation on FA theory, principles and practices.	Reflect on how formative their assessment practices are.
	Feedback (theory).	Teachers were asked questions about their experience with FB practices.	20 minutes	PPT presentation on Hattie and Timperley feedback model.	Examples of their own feedback on student tests.
	Levels of mathematics written (and/or oral) feedback on student tests.	Teachers were asked to reflect on feedback levels on their student mathematics tests.	30 minutes	PPT presentation on samples of typical mathematics teacher feedback on student tests.	Use samples of feedback on student tests to reflect on the levels of their own feedback on student tests.
10:30-11:00	BREAK				
2:11:30-12:30	Learning from errors in mathematics & negative knowledge (theory).	Types (sources) of student errors in mathematics classes.	20 minutes	PPT presentation on types of student errors on student tests.	Using sample feedback on student tests to reflect on typical the types of student errors in tests.
	Negative knowledge (theory).	Role of negative knowledge in student learning.	20 minutes	PPT presentation on negative knowledge and its roles in learning.	Using samples of feedback on student tests to analyse possible reasons behind student errors in tests.
	Opportunities in learning from errors	Watch and reflect on a six minute video on effective use of errors in instruction.	20 minutes	Video clip on using errors 'my favourite no'	Reflect on the presented video in relation to how they (can) support students in error situations.
12:30-13:30	BREAK				

3:13:30-14:30	What to do when giving written (and/or oral) feedback on student tests.	Discussion on effective feedback focus using Hattie and Timperley model (FT, FP, FR, FS).	45 minutes	PPT presentations on examples of feedback focusing on FT, FP and FR levels.	Using the feedback examples on student tests to provide feedback focusing on FT, FP and FR.
		Discussion on other features of feedback.	15 minutes	PPT presentation on other features of effective (elaborated vs. specific) feedback.	Discuss the possibility and obstacles for providing such feedback.
4:14:30-15:30	What to do when discussing a test in a plenary discussion.	Discussion on how to support students in error situations using Heemsoth and Heinze (2016) model.	45 minutes	PPT with concrete examples on: (i) describe the error, (ii) explain the error, (iii) correct the error, and (iv) generate a new similar task in which the same error might occur.	Reflect on how to implement each aspect of the error handling strategy stages in their mathematics classes.
		Sample of good/bad student experience of teacher support in error situations (from Study 1).	15 minutes	PPT presentation on some student experiences of teacher support in error situations.	Reflection on their practices of (affective or cognitive) support in relation to student errors.
15:30-15:45	BREAK				
15:45-16:00	Reflection and feedback on the FBT.	Answer a short reflection questionnaire.	15 minutes	Answer a short reflection questionnaire.	Answer a short reflection questionnaire.

Note. FA = Formative assessment; FT = Task level feedback; FP = Process level feedback; FR =Self-regulation feedback; FS =Self-level feedback; FBT = Feedback training; PPT = PowerPoint

Appendix M. Chapter 4 and *5* (Study 2) - Student general questionnaire

Through this questionnaire, you are invited to provide information regarding how you experience mathematics assessment and feedback practices in your class. The questionnaire covers some general information (such as your age) and specific questions regarding how you perceive your mathematics teacher's feedback practices on your mathematics tests/assignments. There are no right or wrong answers. Your answers will remain anonymous.

1. Date: ______ (day) ___________ (month) ______ (year)

2. Answer the following questions. Together they form an anonymous code

	The last letter of your first name
	The first letter of your surname
	The first letter of your month of birth
	The first letter of your favourite food

3. Gender: ☐ Male ☐ Female

4. Year of birth ______

5. School name: ______________________

6. Stream ________________

7. Form/Class: ______________________

8. What was your **Mathematics** score in Form Two national examination: _______________

To be completed by the researcher

Schoolnumber		District		Region	
Respondentgroup					
School ownership		Lettercode ST		School category	

Read each statement carefully and place a tick (√) inside the box alongside your best option among: *Completely disagree, mostly disagree, somewhat disagree, somewhat agree, mostly agree or completely agree.* There are no right or wrong answers. Your answers will remain anonymous

	Statement	Completely disagree	Mostly disagree	Somewhat disagree	Somewhat agree	Mostly agree	Completely agree
1_SCoA_ VI_1_SIS	I pay attention to my assessment results in order to focus on what I could do better next time	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_2_SSC	Assessment encourages my class to work together and help each other.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_5_SIT	Assessment helps teachers track my progress.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_8_SIT	Assessment is a way to determine how much I have learned from teaching.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_9_SIT	Assessment is checking off my progress against achievement objectives or standards.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_10_SIS	I make use of the feedback I get to improve my learning.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_12_SS C	Assessment motivates me and my classmates to help each other.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_14_SIS	I look at what I got wrong or did poorly on to guide what I should learn next.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_15_SIS	I use assessments to take responsibility for my next learning steps.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_17_SS C	Our class becomes more supportive when we are assessed.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_19_SIS	I use assessments to identify what I need to study next.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_21_SS C	When we do assessments, there is a good atmosphere in our class.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_23_SI T	My teachers use assessment to help me improve.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_25_SS C	Assessment makes our class cooperate more with each other.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_27_SI T	Teachers use my assessment results to see what they need to teach me next.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_28_SS C	When we are assessed, our class becomes more motivated to learn.	☐	☐	☐	☐	☐	☐
1_SCoA_ VI_30_SI T	Assessment shows whether I can analyse and think critically about a topic.	☐	☐	☐	☐	☐	☐

	Statement	Completely disagree	Mostly disagree	Somewhat disagree	Somewhat agree	Mostly agree	Completely agree
1_SAFL_Q_1_SPM	My mathematics teacher encourages me to reflect on how I can improve my assignments.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_2_SPM	After examining my test results, my mathematics teacher discusses the answers I gave to the test with me.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_3_SPM	Whilst working on my mathematics assignments, my mathematics teacher asks me how I think I am doing.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_4_SPM	My mathematics teacher allows me to think about what I want to learn in school.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_5_SPM	My mathematics teacher gives me the opportunity to decide on my own learning objectives.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_6_SPM	My mathematics teacher inquires what went well and what went badly in my work.	☐	☐	☐	☐	☐	☐
1_PISA_2004_1_SSE	I am sure I can understand even the most difficult topic in mathematics.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_7Q_SPM	My mathematics teacher encourages me to reflect on my learning process and to think about how to improve next time.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_8_SPM	My mathematics teacher stresses my strengths concerning learning.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_9_SPM	My mathematics teacher identifies my weaknesses concerning learning.	☐	☐	☐	☐	☐	☐
1_MSC5(R)	Mathematics just is not my subject.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_10_SPM	I am encouraged by my mathematics teacher to improve my learning process.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_11_SPM	My mathematics teacher gives me guidance to assist my learning.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_12_SPM	My mathematics teacher discusses assignments with me to help me understand the subject matter better.	☐	☐	☐	☐	☐	☐
1_PISA_2004_2_SSE	I am convinced that I can understand even the most difficult topic taught by my mathematics teacher.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_13_SPM	My mathematics teacher discusses with me the progress I make.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_14_SPM	After each assessment my mathematics teacher informs me how to improve the next time.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_15_SPM	My mathematics teacher discusses with me how to exploit my strengths to improve my assignment.	☐	☐	☐	☐	☐	☐

	Statement	Completely disagree	Mostly disagree	Somewhat disagree	Somewhat agree	Mostly agree	Completely agree
1_SAFL_Q_16_SPM	My mathematics teacher and I consider ways to improve my weak points.	☐	☐	☐	☐	☐	☐
1_MSC1(R)	Mathematics is too hard to like the subject.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_17_SPS	When I do not understand a topic, my mathematics teacher tries to explain it in a different way.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_18_SPS	My mathematics teacher provides me with hints to help understand the subject matter.	☐	☐	☐	☐	☐	☐
1_PISA_2004_3_SSE	I am convinced that I can achieve good results in mathematics homework and exams.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_19_SPS	During mathematics class I have an opportunity to show what I have learned.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_20_SPS	My mathematics teacher asks questions in a way I understand.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_21_SPS	My mathematics teacher asks questions that help me gain understanding of the subject matter.	☐	☐	☐	☐	☐	☐
1_MSC3(R)	I am just not good at mathematics.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_22_SPS	My mathematics teacher allows for my contribution during the lesson.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_23_SPS	I have the opportunity to ask my classmates questions during the mathematics lesson.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_24_SPS	My mathematics teacher makes me aware of the areas I need to work on to improve my results.	☐	☐	☐	☐	☐	☐
1_PISA_2004_4_SSE	I know exactly what to do at home in order to understand the mathematics instruction.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_25_SPS	There is an opportunity to ask questions during the mathematics lesson.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_26_SPS	I am aware of the criteria by which my math assignment will be evaluated.	☐	☐	☐	☐	☐	☐
1_MSC2(R)	Although I make a real effort, mathematics seems to be harder for me than for my fellow students.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_27_SPS	When I receive a mathematics assignment it is clear to me what I can learn from it.	☐	☐	☐	☐	☐	☐

	Statement	Completely disagree	Mostly disagree	Somewhat disagree	Somewhat agree	Mostly agree	Completely agree
1_MSC4(R)	Some topics in mathematics are just so hard that I know from the start I will never understand them.	☐	☐	☐	☐	☐	☐
1_SAFL_Q_28_SPS_	My mathematics assignments allow me to show what I am capable of.	☐	☐	☐	☐	☐	☐
1_IFOS_Q_1_SFU	I think feedback from my mathematics teacher is important in improving my mathematics performance.	☐	☐	☐	☐	☐	☐
1_FES_Q_1_SFDTE	My mathematics teacher is supportive when giving me feedback about my mathematics performance.	☐	☐	☐	☐	☐	☐
1_IFOS_Q_5_SFU	I think that feedback from my mathematics teacher provides clear direction on how to improve my performance.	☐	☐	☐	☐	☐	☐
1_IFOS_Q_6_SFU	Feedback from my mathematics teacher can be a valuable form of praise.	☐	☐	☐	☐	☐	☐
1_FES_Q_5_SFDTE	My mathematics teacher is tactful when giving me feedback about my mathematics performance.	☐	☐	☐	☐	☐	☐
1_IFOS_Q_8_SFU	Feedback from my mathematics teacher motivates me to improve my mathematics performance.	☐	☐	☐	☐	☐	☐
1_IFOS_Q_7_SFU	I pay careful attention to instructional feedback from my mathematics teacher.	☐	☐	☐	☐	☐	☐
1_FES_Q_2_SFDTE	My mathematics teacher considers my feelings when giving me feedback about my mathematics performance.	☐	☐	☐	☐	☐	☐
1_IFOS_Q_3_SFU	I listen carefully when my mathematics teacher provides oral feedback.	☐	☐	☐	☐	☐	☐
1_IFOS_Q_2_SFU	I usually reflect on my mathematics teacher’s feedback.	☐	☐	☐	☐	☐	☐
1_FES_Q_3_SFDTE	My mathematics teacher generally provides feedback in a thoughtful manner.	☐	☐	☐	☐	☐	☐
1_IFOS_Q_4_SFU	I am extremely encouraged by positive feedback from my mathematics teacher.	☐	☐	☐	☐	☐	☐
1_FES_Q_4_SFD (R)	My mathematics teacher provides feedback in an intimidating manner.	☐	☐	☐	☐	☐	☐
1_IFOS_Q_10_SFU	I feel relieved when I receive positive feedback from my mathematics teacher.	☐	☐	☐	☐	☐	☐

	Statement	Completely disagree	Mostly disagree	Somewhat disagree	Somewhat agree	Mostly agree	Completely agree
1_F1	My mathematics teacher is patient and does not tell me off when something does not work out.	☐	☐	☐	☐	☐	☐
1_L8	When I am at home after school, I check the mistakes that I made during the mathematics class.	☐	☐	☐	☐	☐	☐
1_A1 (R)	I get anxious when I make a mistake in the mathematics class.	☐	☐	☐	☐	☐	☐
1_L1	I willingly correct mistakes in my assignments, even if my mathematics teacher does not explicitly ask for it.	☐	☐	☐	☐	☐	☐
1_F7	My mathematics teacher admits when he or she made a mistake.	☐	☐	☐	☐	☐	☐
1_L4	If I do something wrong in mathematics class I perceive this as an opportunity to learn.	☐	☐	☐	☐	☐	☐
1_F3	My teacher assists and discusses with me the mistake(s) I made in a mathematics assignment.	☐	☐	☐	☐	☐	☐
1_A5 (R)	If I make a mistake in mathematics class, I blame myself for not paying enough attention.	☐	☐	☐	☐	☐	☐
1_L2	Sometimes it helps me to keep a mistake in mind that I made during the mathematics class, so that I do not make it again.	☐	☐	☐	☐	☐	☐
1_F5	If I make a mistake in mathematics class, my teacher discusses it with me in a way that I really learn from it.	☐	☐	☐	☐	☐	☐
1_L7	I enjoy that I can learn from making mistakes.	☐	☐	☐	☐	☐	☐
1_L6	I reconsider mistakes to mathematics tasks multiple times.	☐	☐	☐	☐	☐	☐
1_A2 (R)	I am afraid of my mathematics teacher when I made many mistakes in my homework.	☐	☐	☐	☐	☐	☐
1_F6 (R)	My mathematics teacher tries to cover up when making a mistake.	☐	☐	☐	☐	☐	☐
1_A3 (R)	I feel ashamed when I make a mistake in front of the class in mathematics.	☐	☐	☐	☐	☐	☐
1_F4	It is not a problem if I make a mistake in front of my mathematics teacher.	☐	☐	☐	☐	☐	☐
1_L5	The mistakes that I make in mathematics help me to improve my performance.	☐	☐	☐	☐	☐	☐

	Statement	Completely disagree	Mostly disagree	Somewhat disagree	Somewhat agree	Mostly agree	Completely agree
1_A4 (R)	Before the start of a mathematics class I am sometimes afraid that I will make mistakes during the class.	☐	☐	☐	☐	☐	☐
1_L3	I enjoy trying out different solutions to solve mathematics assignments	☐	☐	☐	☐	☐	☐
1_F2	My mathematics teacher is patient when I do not understand something.	☐	☐	☐	☐	☐	☐

******Thank you for answering this questionnaire******

Appendix N. Chapter 4(Study 2) -Students' perceptions of feedback on mathematics test

Through this questionnaire, you are invited to provide information regarding how you experience feedback practices in your mathematics class. The questionnaire covers some general information (such as your age) and specific questions regarding how you perceive your mathematics teacher's feedback on your mathematics test/ assignment. There are no right or wrong answers. Your answers will remain anonymous.

1. Date: ______ (day) ___________ (month) ______ (year)

2. Answer the following questions. Together they form an anonymous code

	The last letter of your first name
	The first letter of your surname
	The first letter of your month of birth
	The first letter of your favourite food

3. Gender: ☐ Male ☐ Female 4. Year of birth ___________________

5. School name: ______________________ 6. Stream: ___________________

7. What is your score on the mathematics test/assignment you just received? ___________

To be completed by the researcher

Schoolnumber		District		Region	
Respondentgroup					
School ownership		Lettercode ST		School category	

Please answer the following statements in relation to the test/assignment feedback you just received.

Read each statement carefully and place a tick (√) inside the box alongside your best option among: *Completely disagree, mostly disagree, somewhat disagree, somewhat agree, mostly agree or completely agree.* There are no right or wrong answers. Your answers will remain anonymous.

	Statement	**Completely disagree**	**Mostly disagree**	**Somewhat disagree**	**Somewhat agree**	**Mostly agree**	**Completely agree**
1_FA1	I am satisfied with this feedback.	☐	☐	☐	☐	☐	☐
1_US1	I consider this feedback useful.	☐	☐	☐	☐	☐	☐
1_AC1	I accept this feedback.	☐	☐	☐	☐	☐	☐
1_FA2	I consider this feedback fair.	☐	☐	☐	☐	☐	☐
1_WI2	I am willing to invest a lot of effort in my mathematics performance.	☐	☐	☐	☐	☐	☐
1_FA3	I consider this feedback justified.	☐	☐	☐	☐	☐	☐
1_WI1	I am willing to improve my mathematics performance.	☐	☐	☐	☐	☐	☐
1_AC3 (R)	I reject this feedback.	☐	☐	☐	☐	☐	☐
1_US2	I consider this feedback helpful.	☐	☐	☐	☐	☐	☐
1_WI3	I am willing to work on further mathematics assignments.	☐	☐	☐	☐	☐	☐
1_AC2 (R)	I dispute this feedback.	☐	☐	☐	☐	☐	☐
1_US3	This feedback provides me a lot of support.	☐	☐	☐	☐	☐	☐
1_AF1 (R)	I felt offended after this feedback on my mathematics test.	☐	☐	☐	☐	☐	☐
1_AF2	I felt satisfied after this feedback on my mathematics test.	☐	☐	☐	☐	☐	☐
1_AF3 (R)	I felt angry after this feedback on my mathematics test.	☐	☐	☐	☐	☐	☐
1_AF4	I felt confident after this feedback on my mathematics test.	☐	☐	☐	☐	☐	☐
1_AF5 (R)	I felt frustrated after this feedback on my mathematics test.	☐	☐	☐	☐	☐	☐
1_AF6	I felt successful after this feedback on my mathematics test.	☐	☐	☐	☐	☐	☐

******Thank you for answering this questionnaire******

Appendix O. Chapter 4 and 5 (Study 2) - Student perceptions of feedback plenary discussion

Through this questionnaire, you are invited to provide information regarding how you experience mathematics assessment and feedback practices in your class. The questionnaire covers some general information (such as your age) and specific questions regarding how you perceive your mathematics teacher feedback practices on your mathematics tests/assignments. There are no right or wrong answers. Your answers will remain anonymous.

1. Date: ______ (day) ___________ (month) ______ (year)

2. Answer the following questions. Together they form an anonymous code

	The last letter of your first name
	The first letter of your surname
	The first letter of your month of birth
	The first letter of your favourite food

3. Gender: ☐ Male ☐ Female

4. Year of birth ______

5. School name: ______________________

6. Stream_____________

To be completed by the researcher

Schoolnumber		District		Region	
Respondentgroup					
School ownership		Lettercode ST		School category	

Read each statement carefully and place a tick (√) inside the box alongside your best option among: *Completely-disagree, mostly disagree, somewhat disagree, somewhat agree, mostly agree or completely agree.* There are no right or wrong answers. Your answers will remain anonymous.

	Statement	Completely disagree	Mostly disagree	Somewhat disagree	Somewhat agree	Mostly agree	Completely agree
1_SP P1	After this plenary feedback discussion, I now understand what I should do to improve my mathematics performance.	☐	☐	☐	☐	☐	☐
1_SP P2	After this plenary feedback discussion, I now understand how I can use the feedback to improve my mathematics performance.	☐	☐	☐	☐	☐	☐
1_PPr 3	Even after this feedback plenary discussion, I still do not understand most of the test questions I failed.	☐	☐	☐	☐	☐	☐
1_SP P4	After this plenary feedback discussion, I now know how I can correct most of my mistakes.	☐	☐	☐	☐	☐	☐
1_SP P5	After this plenary feedback discussion, I now realize why I made some of my mistakes.	☐	☐	☐	☐	☐	☐

Read each statement carefully and tick (√) inside the box alongside your best option

		Completely untypical	Mostly untypical	Somewhat untypical	Somewhat typical	Mostly typical	Completely typical
1_SAPD 1	Was the videotaped plenary feedback discussion typical/ representative for the lessons your teacher normally teaches?	☐	☐	☐	☐	☐	☐
		Completely not similar	**Mostly not similar**	**Somewhat not similar**	**Somewhat similar**	**Mostly similar**	**Completely similar**
1_SAPD 2	How would you describe your teachers' behaviour during the taped lesson? Compared to usual lessons, the teacher behaved ….	☐	☐	☐	☐	☐	☐
		Completely not nervous	**Mostly not nervous**	**Somewhat not nervous**	**Somewhat nervous**	**Mostly nervous**	**Completely nervous**
1_SAPD 3r	How nervous did you feel because of the filming?	☐	☐	☐	☐	☐	☐
		Completely unhelpful	**Mostly unhelpful**	**Somewhat unhelpful**	**Somewhat helpful**	**Mostly helpful**	**Completely helpful**
1_SAPD 4	What is your overall impression of the lesson?	☐	☐	☐	☐	☐	☐

Appendix P. Chapter 5 (Study 2). Multigroup measurement invariances for interaction groups (Pretest-Control, Pretest-Experimental, Posttest-Control, and Posttest-Experimental).

Models	Comparison	SRMR	RMSEA	X^2	df	ΔX^2	Δdf	p-value	CFI	ΔCFI
1. Student emotions and perceived support in errors										
Unconstrained (configural)-A	B vs. A	0.089	0.071	408.742	204	25.706	27	**0.535**	0.871	
Measurement weights (metric)-B	C vs. A	0.084	0.081	434.449	231	67.105	63	**0.338**	0.872	**-0.001**
Measurement intercepts (scalar)-C	B vs. C	0.079	0.091	475.847	267	41.399	36	**0.247**	0.869	**0.002**
2. Authenticity and perception of feedback plenary										
Unconstrained (configural)-A	B vs. A	0.075	0.143	185.799	52	26.784	15	**0.031**	0.912	
Measurement weights (metric)-B	C vs. A	0.107	0.132	212.583	67	46.898	36	**0.106**	0.904	**0.008**
Measurement intercepts (scalar)-C	B vs. C	0.112	0.114	232.697	88	20.114	21	**0.514**	0.905	**0.007**

Note. SRMR = Standardized Root Mean Residual; RMSEA = Root Mean Squared Error of Approximation; CFI = Comparative Fit Index; values marked in bold indicate that the statistical equivalence assumption holds (difference in χ^2, taking into account the difference in *df*, should not be statistically significant and/or the difference in CFI should be Δ CFI ≤ 0.

Eidesstattliche Versicherung

Statement of Scientific Integrity

Kyaruzi, Florence

Name, Vorname

Last name, first name

Ich versichere, dass ich die an der Fakultät für Psychologie und Pädagogik der Ludwig-Maximilians-Universität München zur Dissertation eingereichte Arbeit mit dem Titel:

I assert that the thesis I submitted to the Faculty of Psychology and Pedagogy of the Ludwig-Maximilian-Universität München under the title:

Formative assessment practices in mathematics education among secondary schools in Tanzania

selbst verfasst, alle Teile eigenständig formuliert und keine fremden Textteile übernommen habe, die nicht als solche gekennzeichnet sind. Kein Abschnitt der Doktorarbeit wurde von einer anderen Person formuliert, und bei der Abfassung wurden keine anderen als die in der Abhandlung aufgeführten Hilfsmittel benutzt.

is written by myself, I have formulated all parts independently and I have not taken any texts components of others without indicating them. No formulation has been made by someone else and I have not used any sources other than indicated in the thesis.

Ich erkläre, das ich habe an keiner anderen Stelle einen Antrag auf Zulassung zur Promotion gestellt oder bereits einen Doktortitel auf der Grundlage des vorgelegten Studienabschlusses erworben und mich auch nicht einer Doktorprüfung erfolglos unterzogen.

I assert I have not applied anywhere else for a doctoral degree nor have I obtained a doctor title on the basis of my present studies or failed a doctoral examination.

Munich, 24.7.2017

Ort, Datum

Place, Date

Florence Kyaruzi

Unterschrift Doktorandin/Doktorand

Signature of the doctoral candidate