

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

Florence Kyaruzi,  
LMU Munich, Department of Psychology & Faculty of Education, Dar es Salaam University  
College of Education, University of Dar es Salaam [sakyaruzi@gmail.com](mailto:sakyaruzi@gmail.com)

Jan-Willem Strijbos,  
LMU Munich, Department of Psychology & University of Groningen, Department of  
Educational Sciences, [j.w.strijbos@rug.nl](mailto:j.w.strijbos@rug.nl)

Stefan Ufer,  
LMU Munich, Department of Mathematics, [ufer@math.lmu.de](mailto:ufer@math.lmu.de)

Gavin T. L. Brown,  
University of Auckland, Faculty of Education and Social Work, [gt.brown@auckland.ac.nz](mailto:gt.brown@auckland.ac.nz)

Correspondence concerning the article should be addressed to Dr. Florence Kyaruzi, Faculty of Education, Dar es Salaam University College of Education, P. O. Box 2329, Dar es Salaam, Tanzania. Email: [sakyaruzi@gmail.com](mailto:sakyaruzi@gmail.com)

This paper has been accepted for publication in *Studies in Educational Evaluation*.

**Current Recommended citation (26 Jan. 18)**

Kyaruzi, F., Strijbos, J. W., Ufer, S., & Brown, G. T. L. (in press). Teacher AfL perceptions and feedback practices in mathematics education among secondary schools in Tanzania. *Studies in Educational Evaluation*.

### Abstract

Feedback that monitors and scaffolds student learning has been shown to support learning. This study investigates the effect of mathematics teachers' perceptions of Formative Assessment (FA) and Assessment for Learning (AfL) and their conceptions of assessment on the quality of their feedback practices. The study was conducted in 48 secondary schools in Tanzania with 54 experienced mathematics teachers teaching Form three (Grade 11). Validated questionnaires were combined with interviews to investigate mathematics teachers' perceptions, conceptions, and feedback practices. Data were analysed by structural equation modeling and content analysis techniques. Results from the structural equation model indicated that mathematics teachers' perceptions of FA and AfL and their conceptions of assessment purposes positively predicted the quality of their feedback practices. Interview results illustrated that mathematics teachers used their students' assessment information for both formative and summative purposes. Future interventions for improving the quality of mathematics teacher's feedback practices are proposed.

*Keywords:* Formative assessment; perceptions of FA and AfL; teacher feedback practices; mathematics education; mixed methods.

### Highlights

- 54 Year 11 mathematics teachers in 48 Tanzanian schools surveyed and interviewed.
- Teachers evaluated positively their own FA, AfL, and feedback practices.
- Assessment conceptions and practice perceptions predicted 58% of variance of feedback practices.
- Teachers reported to use students' assessment information for formative and summative purposes.

### Introduction

Formative assessment (FA) and Assessment for Learning (AfL) are 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 formatively for improving the teaching and learning process, certifying students, placement of students in tracks, or for curriculum improvement (Brown, 2008; Pellegrino, 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).

FA and AfL practices are assumed to serve two core functions: monitoring to track student progress and scaffolding to help students improve their learning (Pat-El, Tillema, Segers, & Vedder, 2013; Stiggins, 2005). Monitoring allows teachers to know the direction, speed, and quality of students' learning progress so that supportive interventions can be put in place. Scaffolding is feedback that explicitly provides task, process, or self-regulatory information (Hattie & Timperley, 2007) so that students know how to proceed. However, recent research has shown that the implementation of FA and AfL is far from straightforward. For example, peer and self-assessment can be biased due to students' intra- and interpersonal factors (Panadero, 2016), feedback is often ineffective (Lipnevich, Berg, & Smith, 2016), teachers do not always ask good questions (Airasian, 1997; Barnette, Orletsky, & Sattes,

1994), or actively promote feedback seeking (Winstone, Nash, Parker, & Rowntree, 2017). As a result, the nature of FA and AfL and how it leads to improved outcomes has been debated in FA and AfL literature (Bennett, 2011; Black & Wiliam, 2003).

### **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 provided early enough in the process so as to inform improvements. Bloom (1969) shifted the initial focus of formative evaluation from ‘program evaluation’ to ‘student evaluation’. The purpose of formative evaluation was to provide feedback and corrections at each stage in the teaching and learning process (Bloom, 1969; Bloom, Hastings, & Madaus, 1971). Later on, ‘formative evaluation’ which was about testing and assessment, evolved into what is now referred to as ‘formative assessment’ (FA).

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). Tests can be used to collect information that helps teachers or students to adjust their actions accordingly; however, tests themselves are neither formative nor summative (Popham, 2014). Nonetheless, rather than focusing on the evaluative or grading function of assessment, the present study focuses on the pedagogical use of FA and AfL as an instructional strategy to regulate both the teacher’s teaching processes and the students’ learning tactics and active use of feedback to improve outcomes. In this framework, the decisions made by the teacher and their students’ regarding assessment-elicited evidence make assessment formative. Decision-making depends in part upon beliefs that frame and guide thinking about a phenomenon (Fives & Buehl, 2012). Beliefs about assessment, which are rooted in their experience of assessment, influence the degree to which teacher assessment practices act formatively (Barnes, Fives, & Dacey, 2016; Fulmer, Lee, & Tan, 2015).

### **Teacher perceptions of FA and AfL practices and conceptions of assessment**

While multiple terms are used to refer to the mental representations humans have for a phenomenon (e.g., perception, conception, belief, attitude, etc.), it seems likely these terms refer to the same thing (Brown, 2008). Hattie (2015) suggested that while Australian scholars use the term ‘beliefs’, those in the United States commonly use the term ‘epistemology’, while those in Europe prefer the term ‘conception’. In the present study, we consider the terms perceptions and conceptions separately—in particular because of the terminology used in survey instruments to refer to different content. However, it is acknowledged that others may consider perceptions and conceptions as largely synonymous.

In this study, 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). According to Wood, Bruner and Ross (1976) scaffolding permits learners to concentrate upon and complete only those elements that are within their range of competence. Kyaruzi, Strijbos and Ufer (2016) showed that students’ perceptions of their mathematics

teacher's monitoring and scaffolding practices were significantly related to their mathematics achievement.

In contrast, conceptions of assessment refer to the more general representations teachers have concerning the 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). Three conceptions of assessment have been generally attested to, including (1) assessment improves teaching and learning, (2) assessment evaluates and holds accountable students, schools, and teachers, and (3) assessment is irrelevant (Bonner, 2016).

The conception that assessment improves teaching and learning is the central argument for FA and AfL (Black & Wiliam, 1998; Popham, 2014) and requires teachers to use evidence about student progress to support their learning (Brown, 2004). The conception that assessment makes schools and/or teachers accountable is the rationale behind accountability policies that use student assessment results to judge and reward or punish schools and teachers (Brookhart, 1994; Nichols & Harris, 2016). Student accountability is evidenced in the 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 into further opportunities (Brown, 2008). The conception that assessment is irrelevant regards assessment as a negative practice, which, because it is unfair to students or inaccurate, can be ignored—even if it is imposed. If assessment cannot help teachers improve student learning, then teachers may choose to ignore it (Deneen & Brown, 2016).

Although studies into the role of teacher conceptions of assessment and assessment perceptions on teaching and learning processes have proliferated in the last two decades (e.g., Brown, 2004; Brown, Chaudhry & Dhamija, 2015; Gibbs & Simpson, 2003; MacLellan, 2001; Pat-El et al., 2013), few studies have examined 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 feedback practices**

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 (Njabili, 1999; Wiliam, Lee, Harrison, & Black, 2004; Wiliam, 2011), and especially for struggling learners (Black & Wiliam, 1998). Specifically, the quality of how teachers deliver feedback and how it promotes students to seek feedback (i.e., 'feedback delivery' and 'promote feedback seeking') is essential in contributing to both student outcomes and increased student regulation of their own learning. In fact, the more considerate a feedback source is when providing feedback, the more likely an individual is to accept and respond to the feedback provided (Duijnhouwer, Prins, & Stokking, 2012; Gregory & Levy, 2015; King, Schrodt, & Weisel, 2009; Strijbos, Pat-El, & Narciss, 2010). Considerate feedback, among other things, maximises clarity of information (Winstone et al., 2017).

An important goal of FA and AfL is to have students become active participants in assessment and active seekers of feedback. Feedback seeking requires students to identify areas in which they need help and seek feedback that aligns with their learning needs (Carless, Salter, Yang, & Lam, 2010). However, students are likely to seek feedback only if the social dynamics of the classroom or the teacher promotes feedback-seeking behaviours (Neitzel & Davis, 2014). Unfortunately, there is limited information on how teachers can

promote students' becoming active feedback seekers instead of being passive feedback receivers (Winstone et al., 2017). Thus, teacher feedback delivery and promoting students to seek feedback are important aspects of feedback practices that require further scrutiny.

It is also noteworthy that previous research has examined teacher perceptions of FA and AfL independent of their conceptions of assessment. While perceptions and conceptions may overlap considerably, there is a possibility that teacher perceptions of feedback practices may vary systematically with their conception of assessment. For example, Brown and Harris (2009) showed that when New Zealand primary teachers conceived of assessment as a measure of student accountability they perceived assessments as formal tests and measures of surface cognitive processes. Thus, it may be that teachers, who have a strong conception of assessment as an evaluation of students (perhaps consistent with systemic use of external examinations), would perceive feedback as accuracy on task-focused skills rather than supportive of self-regulation skills. It may also be that such a conception would be associated with more self-oriented feedback practices in which unsuccessful students are blamed and successful ones praised. This would stand in contrast to a teacher who conceives of assessment as oriented towards improving teaching and thus uses student failure as a way to reflect on pedagogy and curriculum rather than the student. The present study is an early attempt at bringing together perceptions of feedback with conceptions of assessment.

### **Tanzanian Context**

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). It is implied that even school based-assessment have impact on students' final summative examination. Ottevanger, Akker and Feiter (2007) pointed out that although most Sub-Saharan African countries – including Tanzania – have integrated school-based continuous assessment, testing at the school level was mainly summative and hardly used formatively, that is, for instructional purposes or to provide feedback to students. Zembazemba (2017) argued that most assessment practices in Tanzanian secondary schools are teacher-centred with low student involvement.

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. Teachers who are qualified to teach in secondary schools in Tanzania are supposed to possess a Diploma or Degree in Education or above (Ministry of Education and Vocational Training, MoEVT, 2014). Diploma in Teacher Education course comprises two years of pre-service teacher education and a Bachelor degree in Education comprises at least three years of pre-service teacher education. Teacher training colleges prepare Diploma teachers while Universities prepare Degree teachers. Teacher education programmes provide pre-service teachers with subject related courses and didactical courses including a teaching practicum.

Despite school improvement programmes such as the Secondary Education Development Programme (MoEVT, 2008), mathematics education in secondary schools in Tanzania has suffered from low passing rates (BEST, 2014; Kitta, 2004). Several studies have examined general educational challenges in Tanzania that might explain this: (a) the

transition from Swahili as the language of instruction in primary schools to English in secondary schools (Brock-Utne, 2007; Qorro, 2013; Vuzo, 2007), (b) large class (BEST, 2014) curriculum content overload (Kitta & Tilya, 2010), and (d) 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). Since it seems logical to consider that teacher conceptions and/or perceptions will be consistent with the contextual demands in which the teacher is operating (Brown & Harris, 2009), these challenges may influence their conceptions and perceptions of their own teaching and assessment practices, including the quality of feedback practices.

### **The present study**

Formative assessment practices ought to be well perceived by teachers because conceptions and perceptions of assessment frame and guide teacher assessment practices (Ajzen, 1991; Fives & Buehl, 2012). Hence, the present study investigates the relationship of Tanzanian secondary school mathematics teachers' perceptions of FA and AfL, their conceptions of the purposes of assessment, and the quality of their feedback practices. Four research questions are posed:

- 1) To what extent do Tanzanian secondary school mathematics teachers perceive their own assessment practice in terms of the monitoring and scaffolding functions?
- 2) What conception of assessment do Tanzanian secondary school mathematics teachers report?
- 3) To what extent do Tanzanian 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) relate to the quality of their self-reported feedback practices?
- 4) For what purposes do Tanzanian secondary school mathematics teachers typically use students' assessment information (such as student's scores in terminal and mid-term tests)?

### **Method**

#### **Participants**

The study was conducted in 48 Tanzanian secondary schools. Just over half ( $n = 25$ ) of the schools were in the mostly urban Dar es Salaam region and the remaining schools ( $n = 23$ ) were in the mostly rural Kilimanjaro region. Based on national educational statistics (MoEVT, 2013) the mean GPA for the sampled schools in mathematics ( $M = 4.63$ ,  $SD = 0.69$ ) did not deviate statistically from the overall schools' mean GPA ( $M = 4.85$ ,  $SD = 0.70$ ). Three criteria were used to achieve a representative sample of teachers from these 48 schools: school mathematics performance (high, medium, low) according to school ranking (MoEVT, 2013), average class-size ( $< 40$ ,  $\geq 40$ ), and school-type (private, government). Within the 48 randomly sampled schools, which were drawn from schools of varying mathematics performance ( $N_{\text{high}} = 8$ ,  $N_{\text{middle}} = 19$ ,  $N_{\text{low}} = 27$ ), 54 mathematics teachers in Grade 11 (Form three in the Tanzanian system) participated. The participants included nine women and 45 men, whose overall mean age was 37.26 ( $SD = 10.96$ ; range 23 to 66). Teachers had on average close to 11 years of teaching experience ( $M = 10.87$ ,  $SD = 10.39$ , range 1-36 years). In terms of highest qualification: 19 held a Diploma, 33 held a Bachelor degree and two held a Master degree. The average class size for the participating teachers was large ( $M = 49$  students,  $SD = 20.49$ ) and the average number of 40-minute teaching periods per week ( $M = 22.05$ ,  $SD = 7.33$ , range = 6-38 periods) meant that the average work-load was about 15 hours per week. Grade 11 was selected because it should contain a wide variety of assessment and feedback practices relatively unconstrained by the official public examination.

## Design

A mixed-method research approach was used, 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 analysed 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.

## Instruments

**Questionnaires.** We used 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. First, we used the Teacher Assessment for Learning Questionnaire (TAFLQ) to measure teacher’s perceptions of their AfL practice in terms of ‘perceived monitoring’ and ‘perceived scaffolding’ (Pat-El et al., 2013). Second, mathematics teachers’ conceptions of assessment were measured using the Teacher Conceptions of Assessment survey (TCoA-III) (Brown, 2004). This measure had nine 3-item first order factors which aggregate into four different conceptions: school accountability, improvement, irrelevant, and student accountability. Third, we adopted the ‘feedback delivery’ and ‘promoting feedback seeking’ subscales from the Feedback Environment Scale (FES) to measure the quality of teachers’ feedback practices (Steelman, Levy, & Snell, 2004). The various scales differed in response options (i.e., 5, 6, or 7) and were adapted to a common balanced 4-point scale: fully disagree (1), somewhat disagree (2), somewhat agree (3) and fully agree (4). We deliberately refrained from a middle category due to its ambiguous perceptual meaning (Dunham & Davison, 1991; Kulas & Stachowski, 2009). It should be noted that the ‘teacher conceptions of assessment subscales’ and ‘feedback environment’ subscales were below the .70 threshold for Cronbach’s alpha, which could be due to low sample size, suggesting that robust analyses methods would be required to draw valid conclusions—therefore our results should be interpreted cautiously. Table 1 summarises the scales, number of items per scale, a sample item per scale, and the Cronbach’s  $\alpha$  from the original studies and for this study.

Table 1. Scales, Sample items and estimates of Reliability

Scale	k	Sample item	Cronbach’s $\alpha$	
			Original Study	Present Study
<i>Teacher Assessment for Learning Questionnaire</i>				
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
<i>Teacher Conceptions of Assessment</i>				
School accountability	3	Assessment is a good way to evaluate a school.	.77	.54
Improvement	9	Assessment helps students improve their learning.	.83	.68
Assessment is Irrelevant	3	Assessment interferes with teaching.	.69	.52
Student Accountability	3	Assessment places students into categories.	.55	.48
Ignore Assessment	3	Assessment results are filed and	.69	.40

Scale	k	Sample item	Cronbach's $\alpha$	
			Original Study	Present Study
		ignored.		
<i>Feedback Environment Scale</i>				
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

*Note.*  $k$  = number of items per scale

**Teacher Interviews.** The interview questions were specifically developed for the present study to gain in-depth understanding of the topics covered in questionnaire scales. The interview focused on two main goals: (a) teachers' teaching, assessment, and feedback practices such as: teacher reactions to student errors, teacher perceptions of FA and AfL practices, and (b) teachers' perception of student experiences with teaching, assessment, and feedback such as: perceptions of student reactions to teacher feedback. For this study, teachers' responses to the question "For what purposes do you typically use students' assessment information (e.g., student's scores in terminal and mid-term tests)" were analysed. The average duration of the interviews was 27 minutes ranging from 17 to 54 minutes.

### Procedure

The study was conducted with ethics clearance from the University of Dar es Salaam. All participating teachers signed an informed consent form. Questionnaires were administered by the researcher or by one of two research assistants. The researcher or assistant demonstrated how to use the rating scales prior to the teachers filling-in the questionnaire. The teachers needed approximately 20 minutes to do so.

Prior to data analyses, data screening was carried out to identify and address 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 missing values using expectation maximization (EM), which is an effective imputation method when data are MCAR (Musil, Warner, Yobas, & Jones, 2002). Investigation of the EM estimated statistics such as item means showed minimal differences to the un-estimated data (i.e., differences in descriptive statistics were at the third decimal point).

### Analyses

**Questionnaire analyses.** Ideally, confirmatory factor analysis should be used to ensure that the factor structure of an existing measurement applies equally to a new sample (Brown, Harris, O'Quinn, & Lane, 2017). However, with just 54 teachers and 64 items the study did not meet the expectation of 5 to 10 cases per variable (Costello & Osborne, 2005). Hence, the existing scale structure was tested with this sample using Cronbach's alpha values to identify groups of items that replicated the original scales. Where the scale (e.g., TCoA Irrelevance) did not generate a plausible alpha value (i.e.,  $\alpha > .40$ ), information about the possible alpha value after deletion of an item was used to create defensible scales.

To identify the relationship of instruments to each other it was decided to parcel (Little, Cunningham, Shahar, & Widaman, 2002) the manifest item responses into factor mean scores, according to the CFA results, so that the ratio of cases to variables was 54:9 (i.e., 6:1) meeting conventional standards. This approach meant that structural equation modeling (SEM) of the relationship of factors to each other with a relatively small sample could be conducted. The application of SEM with small samples reduces the chance of correlated residuals and sampling errors (Marsh, Hau, Balla, & Grayson, 1998).

SEM 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 regression because it provides a stronger framework to account for response bias and takes into account non-random measurement errors (Comşa, 2010). Furthermore, SEM is preferred to path analysis since latent constructs can be retained in a model, rather than only manifest variables.

The evaluation of CFA and SEM model fit requires reporting multiple indicators and determining whether fit is acceptable or good (Hu & Bentler, 1999). Unfortunately, not all indicators have been shown to be stable as model complexity, sample size, or model misspecification occurs (Fan & Sivo, 2007). The chi-square statistic is overly-sensitive in sample sizes above 250 (Byrne, 2010), the comparative fit index (CFI) rewards models with three or fewer factors, while the root mean square error of approximation (RMSEA) rewards complex models with more than three factors (Fan & Sivo, 2007). In contrast, the gamma hat and the standardised root mean residual (SRMR) have been shown to be stable estimators (Fan & Sivo, 2007). Acceptable fit is indicated when CFI and gamma hat  $>.90$ , RMSEA and SRMR  $<.08$ , and the ratio of  $\chi^2/df$  has  $p >.05$ . Good fit is indicated when CFI and gamma hat  $>.95$ , RMSEA and SRMR  $<.05$  and  $\leq .06$  respectively. As recommended by Steiger (1990), the 90% confidence interval for RMSEA can be used to determine whether the observed value may be  $<.05$ .

**Interview analyses.** The interviews were content analysed (Braun, & Clarke, 2006; Strijbos, Martens, Prins, & Jochems, 2006). A data-derived coding scheme was developed using about ten percent of all interviews. The threshold for segmentation agreement was 80% (Strijbos et al., 2006) and a Krippendorff alpha of 0.80 or higher indicated good coding reliability (Krippendorff, 2013). Two independent coders were involved in all data analysis after a 50 minutes training session on the study rationale and the coding scheme. Four iterations of independent coding were performed. After these rounds, it was deemed that codes had been appropriately and reliably assigned to interview utterances.

Table 2: *Interview segmentation agreement and Krippendorff alpha*

Trial	Size	Segmentation (%)		Krippendorff alpha	
		Coder 1	Coder 2	Value	90% C.I
1	7 interviews	67	70	—	—
2	6 interviews	83	86	.63	[.51, .75]
3	6 interviews	71	74	—	—
4	6 interviews	87	89	.88	[.78, .96]

## Results

### Mathematics teachers' perceptions of their FA and AfL practices

The first research question sought to describe mathematics teachers' FA and AfL perceptions, conceptions of assessment, and feedback environment (Table 3). Scale analysis showed that the two long scales for TAFLQ had strong alpha values, the two shorter FES

scales had moderate estimates, and three of the TCoA scales had moderate estimate values. The TCoA Irrelevance factor had to be split into Ignore and Irrelevant scales, each with three items and moderate estimate values. Hence, in total, 58 items were used to create nine scales.

Table 3. *Descriptive statistics per scale and their inter-correlations*

	Descriptive			Scale inter-correlations							
	Min.	M	SD	I	II	III	IV	V	VI	VII	VIII
<i>AfL Perceptions</i>											
I. Monitoring	2.44	3.61	0.31	--							
II. Scaffolding	2.75	3.64	0.32	<b>.72**</b>	--						
<i>Conceptions of Assessment</i>											
III. Improvement	2.33	3.40	0.40	.43**	.34*	--					
IV. Student Accountability	2.33	3.45	0.49	.37**	.26	.46**	--				
V. School Accountability	2.00	3.49	0.51	.34*	.09	.48**	<b>.58**</b>	--			
VI. Irrelevant	1.33	2.55	0.70	.24	.08	.42**	<b>.35**</b>	<b>.32*</b>	--		
VII. Ignore Assessment	1.00	2.00	0.61	-.32*	-.24	-.49**	<b>-.13</b>	<b>.00</b>	<b>-.19</b>	--	
<i>Feedback Environment Perceptions</i>											
VIII. Feedback delivery	2.50	3.49	0.43	.49**	.56**	.20	.08	.12	.22	-.27*	--
IX. Promote Feedback seeking	2.50	3.54	0.44	.30*	.30*	.18	.13	.21	.02	-.26	<b>.37**</b>

*Note.* Values in bold are within inventory correlations;  $N=54$ ; Maximum score =4.00; \*\*  $p < .01$ , \*  $p < .05$

Table 3 reveals that mathematics teachers' assessment perceptions were above 'somewhat agree' (3.00) for all scales except for conceptions of assessment as irrelevant and ignore assessment, suggesting that the mathematics teachers evaluated positively their own FA and AfL practices, conceptions of assessment, and the quality of feedback practices. The lower mean scores for irrelevant and ignore assessment conceptions indicates that teachers did not ignore assessment nor did they consider assessment to be irrelevant. Furthermore, the inter-correlations indicate that mathematics teachers' FA and AfL perceptions had moderate positive correlations with feedback environment and conceptions of assessment scales (improvement, student accountability and school accountability), except for the relationship between perceived scaffolding, school accountability and student accountability—which was not significant. The association of AfL perceptions (monitoring and scaffolding) with the conception of assessment for improvement, suggests reasonable convergence around learning-oriented assessment practices and philosophy. However, the ignore assessment conception had moderate negative significant correlation with monitoring, and a small negative correlation with feedback delivery. This relationship indicates that for teachers to effectively monitor student learning and deliver feedback in a thoughtful manner they should not ignore assessment information.

### **Teachers' FA and AfL perceptions, conceptions of assessment, and quality of feedback practices**

Since both the TAFLQ and TCoA inventories related to perceptions or conceptions of assessment respectively, it was decided to model these self-reported beliefs as correlated predictors of the FES as the dependent variable. The structural model was motivated by the

theoretical evidence underpinning that conceptions and perceptions of assessment have the potential to influence teacher assessment practices (Ajzen, 2005; Fives & Buehl, 2012). Hence, this model was tested in SEM (Figure 1). The model had good fit:  $\chi^2 = 40.86$ ;  $df = 25$ ;  $\chi^2/df = 1.63$  ( $p = .024$ ); CFI = .884, Gamma hat = .938, SRMR = .085, and RMSEA = .109 [.040, .168]. Mathematics teachers' conceptions of assessment purposes were moderately and positively correlated with their perceptions of AfL ( $r = .55$ ). The regression from AfL practices to Feedback Environment was strong ( $\beta = .76$ ), whereas the path from TCoA to FES was statistically not significant (and therefore not shown in Figure 1). Combined, teacher's conceptions of assessment and perceptions of AfL accounted for a large portion of variance in the Feedback Environment factor (SMC = .57,  $f^2 = 1.32$ ; Cohen, 1992).

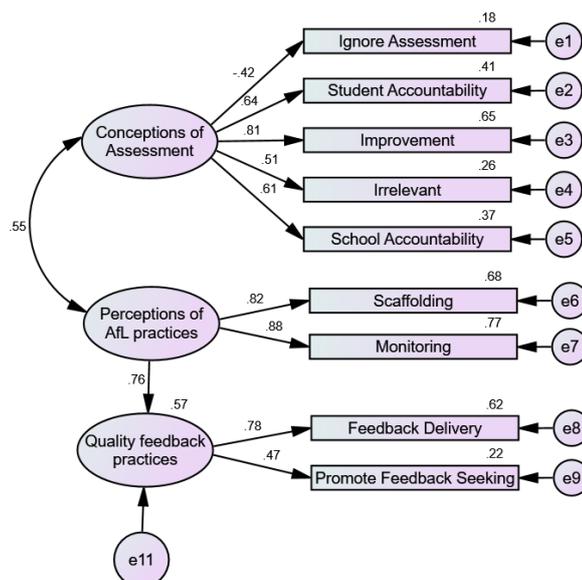


Figure 1. Teacher's FA and AfL perceptions and assessment conceptions as predictors of the quality of their feedback practices.

### Mathematics teachers' use of students' assessment information

With the help of the interview, we aimed to investigate how the mathematics teachers used students' assessment information. 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. Analyses of interviews resulted in 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 (see Table 4).

These six main themes reflect a mix of formative and summative purposes. Teachers reported 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. They also reported a summative use of student assessment information such as ability grouping (if no specific support was provided to each ability group), accountability reports, and using assessment to reprimand low achievers.

Table 4. *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).

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 the remark "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 as evidenced in the remark, "We normally use student's scores first of all in ranking students" (Teacher 53).

In general, the interview results support the questionnaire results in Table 3 in various ways. First, teachers used assessment information to show students how to improve and to devise their teaching approaches, which is supported by their high rating in the conception that assessment is for 'improvement' (teaching and learning). Second, the observed teacher use of assessment for ability grouping supports that the 'assessment is for student accountability' conception. Third, teacher use of student assessment information for accountability reports (such as student reports to inform parents) supports their high agreement with the conception that 'assessment is for school accountability'. In general, the mathematics teachers self-reported uses of their students' assessment information in the interviews corresponds to their conceptions of assessment in the questionnaires.

### Discussion

This study investigated the effect of mathematics teachers' FA and AfL perceptions and conceptions of assessment on the quality of their feedback practices. The first research question sought to investigate the extent to which Tanzanian secondary school 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 as 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, Tillema, Segers, & Vedder, 2015; Rach et al., 2013; Veldhuis & Van den Heuvel-Panhuizen, 2014). The second research question investigated the conceptions of assessment reported by Tanzanian secondary school mathematics teachers. Our results indicate that mathematics teachers had positive conceptions of assessment, i.e. that the purpose of assessment was to improve student learning, promote student accountability and school accountability. However, teachers did not agree that assessment is irrelevant nor ignored assessment information. 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 third research question investigated the extent to which Tanzanian secondary school mathematics teachers' perceptions of their own FA and AfL practices and conceptions of assessment 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 the quality of their feedback practices. These findings support previous studies that perceptions of assessment are related to teacher assessment practices (Fives & Buehl, 2012). Moreover, 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. Furthermore, the findings on the impact of teacher perceptions of FA and AfL on their assessment practices is in line with recent findings that teachers' perceptions of peer assessment (part of AfL) predicted how often teachers used peer assessment (Panadero & Brown, 2017). Likewise, our results are in line with recent findings indicating that teachers' positive perceptions and/or experiences with self-assessment (also part of AfL) influence their assessment practices (Panadero, Brown & Courtney, 2014). Hence, consistent with previous studies, our findings support that teacher feedback practices (feedback delivery and promoting feedback seeking) are shaped by their conceptions about assessment purposes and perceptions of FA and AfL.

The fourth research question sought to identify typical uses of student's assessment information by Tanzanian secondary school mathematics teachers. In line with Al Duwairi (2013), our interviews showed that mathematics teachers used students' assessment information for both formative and summative purposes. The reported formative uses of student assessment information (i.e., improving student learning and instruction) are in line with how Canadian teachers in various instructional domains conceive the functions of assessment (Hunter, Mayenga, & Gambell, 2006). Furthermore, the interviews indicate that mathematics teachers' 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). This is 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. Finally, the reported summative uses of assessment information such as accountability reports to parents and students' ability grouping, provide further support that conceptions of assessment promote student and school accountability (Brown, 2006; Firestone, Mayrowetz, & Fairman, 1998). Nevertheless, how mathematics teachers balance summative and formative uses of student assessment information is an important issue for future research.

### Methodological limitations

Our results should be interpreted in light of some limitations. Firstly, we mainly used self-report data from surveys and teacher interviews, which might be limited in their scope—although our data provide evidence for their validity. Hence, future research could further substantiate our findings with other measures such as observational data. Secondly, the reliability of the ‘conceptions of assessment scales’ and ‘feedback environment scales’ were below the typical threshold (i.e., Cronbach’s  $\alpha > .70$ ), which indicates that our results should be interpreted cautiously. However, we applied structural equation modeling with parcelling which is a robust technique for small sample sizes and takes into account random and non-random measurement errors. Thirdly, based on the relatively small sample, the models’ multiple fit indicators were acceptable to good; hence, we cannot generalise our findings beyond this sample. These results may be substantiated by future studies using observational and/or longitudinal data to examine other potential factors that might influence the quality of teacher feedback practices. Fourthly, we adopted the scales with different response options (i.e., 5, 6 or 7) to a common balanced 4-point scale which might have influenced item variance, resulting in low reliability for some scales.

### Theoretical and practical implications

Our results showed that the mathematics teachers were 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, the interviews showed that the mathematics teachers reported various uses of student assessment information (Gronlund & Linn, 1990). These uses of assessment information aligned with established teacher conceptions of assessment purposes that assessment improves teaching and learning that promotes school and student accountability (Black & Wiliam, 1998; Brown, 2006, 2011). Surprisingly, a few mathematics teachers used their student’s assessment information to reprimand low achievers, which might be attributed to the high-stakes assessment culture in Tanzania. Hence, future research could investigate how teachers can be encouraged to use student errors in mathematics tests or assignments to inform students on how to improve (Rach et al., 2013), or provide educational counseling instead of reprimanding low achieving students (Yaghambe & Tshabangu, 2013). Likewise, interventions could be designed and implemented to improve the quality of teacher feedback practices to capitalize on teacher conceptions of assessment and perceptions of their FA and AfL 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. doi:10.1111/j.1949-8594.1998.tb17413.x
- 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.

- Al Duwairi, A. (2013). Secondary teachers' conceptions and practices of assessment models: The case for mathematics teachers in Jordan. *Education, 134*(1), 124-133.
- Assessment Reform Group (ARG) (2002). *Assessment for Learning: 10 Principles*. Retrieved from <http://www.assessment-reform-group.org.uk>
- 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, USA.
- 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 (2014). *Basic Education Statistics in Tanzania (National data)*. Prime Minister's Office Regional Administration and Local Government: Dodoma, Tanzania.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. doi:10.1080/0969595980050102
- Black, P., & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal, 29*(5), 623-637. doi:10.1080/0141192032000133721
- 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
- 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.
- Bloom, B., Hastings, J., & Madaus, G. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw Hill.
- Bonner, S. M. (2016). Teachers' perceptions about assessment: Competing narratives. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 21-39). New York: Routledge.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa
- 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
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education, 7*(4), 279-301. doi:10.1207/s15324818ame07042
- 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.
- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York: Nova Science.

- Brown, G. T. L., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources heighten conceptions of assessment as school accountability. *Journal of Multidisciplinary Evaluation*, 6(12), 68-91.
- 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., Chaudhry, H., & Dhamija, R. (2015). The impact of an assessment policy upon teachers' self-reported assessment beliefs and practices: A quasi-experimental study of Indian teachers in private schools. *International Journal of Educational Research*, 71, 50-64. doi:10.1016/j.ijer.2015.03.001
- Brown, G. T. L., Harris, L. R., O'Quin, C., & Lane, K. E. (2017). Using multi-group confirmatory factor analysis to evaluate cross-cultural research: Identifying and understanding non-invariance. *International Journal of Research & Method in Education*, 40(1), 66-90. doi:10.1080/1743727X.2015.1070823
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications and programming*. Routledge: London.
- Carless, D., Salter, D., Yang, M., & Lam, J. (2010). Developing sustainable feedback practices. *Studies in Higher Education*, 36(4), 395-407. doi:10.1080/03075071003642449
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159. doi:10.1037/0033-2909.112.1.155
- 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.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, 10(7), <http://www.pareonline.net/pdf/v10n17.pdf>
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative and mixed approaches*. Los Angeles: Sage.
- Deneen, C. C., & Brown, G. T. L. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education*, 3, 1225380. doi:10.1080/2331186X.2016.1225380
- Dingyloudi, F., & Strijbos, J. W. (in press). 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
- Dunham, T. C., & Davison, M. L. (1991). Effects of scale anchors on student ratings of instructors. *Applied Measurement in Education*, 4(1), 23-35. doi:10.1207/s15324818ame0401\_3
- 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
- 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

- 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. doi:10.3102/01623737020002095
- 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*, 22(4), 475-494. doi:10.1080/0969594X.2015.1017445
- 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 & Practice*, 21(1), 16-33. doi:10.1080/0969594X.2013.831030
- 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
- Gregory, J. B., & Levy, P. E. (2015). Feedback delivery and the role of the feedback provider. In J. B. Gregory & P. E. Levy (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.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. doi:10.3102/003465430298487
- Hattie, J. (2009). *Visible learning: A synthesis of meta-analyses in education*. London: Routledge.
- 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
- 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
- Hunter, D., Mayenga, C., Gambell, T. (2006). Classroom assessment tools and uses: Canadian English teachers' practices for writing. *Assessing Writing*, 11(1), 42-65. doi:10.1016/j.asw.2005.12.002
- James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *The Curriculum Journal*, 17(2), 109-138.
- King, P. E., Schrod, 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., & 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*, 29, 77-90.
- 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.
- 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
- 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., Strijbos, J. W., & Ufer, S. (2016, August). Students' assessment for learning perceptions and their mathematics achievement in Tanzanian secondary schools. In C. Csíkós, A. Rausch & J. Szitányi (Eds.), *Proceedings of the 40<sup>th</sup> Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 155-162). Szeged, Hungary: PME.
- 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
- 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
- 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.
- Maclellan, E. (2001). Assessment for learning: The differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, 26(4), 307-318. doi:10.1080/02602930120063466
- 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
- Ministry of Education and Vocational Training (MoEVT) (2008). *Education Sector Development Programme Document*. Dar es Salaam: MoEVT. Retrieved from <https://www.globalpartnership.org/content/tanzania-education-sector-development-programme-2008-17>
- Ministry of Education and Vocational Training (MoEVT) (2013). *Certificate of secondary education examination school rank*. Dar es Salaam: MoEVT.
- Ministry of Education and Vocational Training (MoEVT, 2014). *Sera ya elimu na mafunzo* [Education training and policy]. Dar es Salaam: MoEVT.

- 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
- Ndalichako, J. L. (2015). Secondary school teachers' perceptions of assessment. *International Journal of Information and Education Technology, 5*(5), 326-330. doi:10.7763/IJET.2015.V5.524
- 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.
- National Examinations Council of Tanzania (NECTA) (2004). *Thirty years of the national examinations of Tanzania*. Dar es Salaam: NECTA.
- Nichols, S. L., & Harris, L. R. (2016). Accountability assessment's effects on teachers and schools. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 40-56). New York: Routledge.
- Njabili, A. F. (1999). *Public examinations: A tool for curriculum evaluation*. Dar es Salaam: Mture Educational.
- 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.
- Ottevanger, W., Akker, J., & Feiter, L. (2007). *Developing science, mathematics, and ICT education in Sub-Saharan Africa: Patterns and promising practices* (World Bank Working paper number 101). Washington, DC: The World Bank.
- 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
- Panadero, E., Brown, G., & Courtney, M. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in Education: Principles, Policy & Practice, 21*(4), 365-383. doi:10.1080/0969594X.2014.919247
- Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 247-266). New York: Routledge.
- Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology Education, 32*(1), 133-156. doi:10.1007/s10212-015-0282-5
- Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2013). Validation of Assessment for Learning questionnaires for teachers and students. *The British Journal of Educational Psychology, 83*(1), 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 & Practice, 22*(2), 282-298. doi:10.1080/0969594X.2014.975675
- 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
- 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. doi:10.3102/00346543074004525
- Popham, W.J. (2014). *Classroom assessment: What teachers need to know* (7th ed.). Harlow, US: Pearson.

- 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.
- Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Perspectives of curriculum evaluation* (Vol. 1, pp. 39-55). Chicago: Rand McNally.
- Stelman, 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. doi:10.1177/0013164403258440
- 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.
- 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(4), 29-48. doi:10.1016/j.compedu.2005.04.002
- 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.
- United Republic of Tanzania (1974). *The Musoma resolution*. Dar es Salaam: Government Printers.
- 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.
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14. doi:10.1016/j.stueduc.2011.03.001
- William, 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., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100. doi:10.1111/j.1469-7610.1976.tb00381.x
- 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
- 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 38<sup>th</sup> Conference of the International Group for the Psychology of Mathematics Education and the 36<sup>th</sup> Conference of the North American Chapter of*

*the Psychology of Mathematics Education* (Vol. 5, pp. 329-336). Vancouver, Canada: PME.

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. doi:10.5296/jse.v3i4.4167

Zembazemba, R. D. (2017). Educational assessment practices in Tanzania: A critical reflection. *Business Education Journal*, 1(3), 1-8.