University of Groningen

Students' formative assessment perceptions, feedback use and mathematics performance in secondary schools in Tanzania

Kyaruzi, Florence; Strijbos, J. W. ; Ufer, Stefan; Brown, Gavin T. L.

Published in:
Assessment in Education: Principles, Policy & Practice

DOI:
[10.1080/0969594X.2019.1593103](https://doi.org/10.1080/0969594X.2019.1593103)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Kyaruzi, F., Strijbos, J. W., Ufer, S., & Brown, G. T. L. (2019). Students' formative assessment perceptions, feedback use and mathematics performance in secondary schools in Tanzania. *Assessment in Education: Principles, Policy & Practice*, 26(3), 278-302. <https://doi.org/10.1080/0969594X.2019.1593103>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Students' formative assessment perceptions, feedback use and mathematics performance in secondary schools in Tanzania

Florence Kyaruzi, Jan-Willem Strijbos, Stefan Ufer & Gavin T. L. Brown

To cite this article: Florence Kyaruzi, Jan-Willem Strijbos, Stefan Ufer & Gavin T. L. Brown (2019) Students' formative assessment perceptions, feedback use and mathematics performance in secondary schools in Tanzania, *Assessment in Education: Principles, Policy & Practice*, 26:3, 278-302, DOI: [10.1080/0969594X.2019.1593103](https://doi.org/10.1080/0969594X.2019.1593103)

To link to this article: <https://doi.org/10.1080/0969594X.2019.1593103>

Published online: 21 Mar 2019.

Submit your article to this journal [↗](#)

Article views: 535

View related articles [↗](#)

View Crossmark data [↗](#)

Citing articles: 1 View citing articles [↗](#)

Students' formative assessment perceptions, feedback use and mathematics performance in secondary schools in Tanzania

Florence Kyaruzi ^a, Jan-Willem Strijbos ^b, Stefan Ufer^c and Gavin T. L. Brown ^d

^aLudwig-Maximilians-Universität München, Department of Psychology, Germany & Dar es Salaam University College of Education, University of Dar es Salaam, Tanzania; ^bLudwig-Maximilians-Universität München, Department of Psychology, Germany & University of Groningen, Department of Educational Sciences, The Netherlands; ^cLudwig-Maximilians-Universität München, Chair of Mathematics Education, Germany; ^dUniversity of Auckland, Faculty of Education and Social Work, New Zealand

ABSTRACT

This study investigates the impact of secondary school students' perceptions of mathematics teachers' formative assessment practices and feedback delivery on their feedback use, and mathematics performance. The sample consisted of 2767 Form 3 (Grade 11) students from 48 secondary schools in Tanzania. Surveys and focus group discussions were used to measure students' perceptions of formative assessment, feedback delivery, and feedback use. Structural Equation Modelling (SEM) of survey data showed that students' perceptions of the quality of teacher feedback delivery and perceived scaffolding positively predicted students' feedback use whereas perceived monitoring negatively predicted feedback use. In turn, students' feedback use positively predicted their mathematics performance to a small extent. Content analysis of focus group discussions illustrated that most students valued their mathematics teacher's assessment practices. The findings imply that in Sub-Saharan African educational systems the quality of teacher feedback delivery and promotion of student feedback use can improve students' mathematics performance.

ARTICLE HISTORY

Received 13 July 2017
Accepted 17 February 2019

KEYWORDS

Formative assessment;
mathematics education;
secondary education; mixed
methods; structural
equation modelling

Introduction

Formative Assessment (FA) and Assessment for Learning (AfL) have received increased attention over the past three decades (Black & Wiliam, 1998; Crooks, 1988). 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. 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).

CONTACT Florence Kyaruzi sakyaruzi@gmail.com Faculty of Education, Dar es Salaam University College of Education, P. O. Box 2329, Dar es Salaam, Tanzania

© 2019 Informa UK Limited, trading as Taylor & Francis Group

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). Supporting student learning (scaffolding) is positively related to students’ use of deep-level learning strategies and learning outcomes (Baas, Castelijns, Vermeulen, Martens, & Segers, 2015; Pinger, Rakoczy, Besser, & Klieme, 2018a). FA and AfL literature provides extensive evidence that, if well implemented and well perceived by students, FA and AfL have the potential to contribute to improved student learning (Black & Wiliam, 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.

Since Black and Wiliam (1998) the field has used both FA and AfL to point to the use of assessment as a mechanism for improving student learning outcomes. However, while the terms might be synonymous, it is clear that AfL tends to be associated with a more interactionist and pedagogic approach (Stobart, 2006) that might be synonymous with good teaching (Brown, 2013). In contrast, FA, following from Scriven (1991), refers to all assessment practices, including testing, that are conducted before the end of teaching processes and used to inform changes to both teaching and learning. In this paper, we prefer FA because it includes more formal approaches to assessment, rather than the more pedagogical approach advocated by a pure AfL perspective (Swaffield, 2011). Feedback is a key component of FA (Lipnevich, Berg, & Smith, 2016; Shute, 2008; Wiliam, 2011) and the added value of FA – 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). Feedback is the information provided by an agent (e.g., teacher, peer, etc.) regarding aspects of one’s performance or understanding (Hattie & Timperley, 2007). It refers to all post-response information that is provided to a learner to inform the learner on his or her actual state of learning or performance (Narciss, 2008). Importantly, Sadler (1989) and Shute (2008) refer to formative feedback as the information communicated to the learner that is intended to modify his or her thinking or behaviour for improving learning. Feedback is an important instructional practice that informs students about the gap between expected learning goals and their status. FA involves giving students feedback, which is found to be a key predictor of student achievement (Furtak et al., 2016). Briggs, Woodfield, Martin, and Swatton (2006) maintain that useful feedback provides precise information on what is wrong and how it can be corrected. It is widely acknowledged that effective feedback should promote self-regulated learning and allow the learner to interact with the feedback to confirm, add, overwrite, tune, or restructure their knowledge (Butler & Winne, 1995; Hattie & Timperley, 2007; Jonsson, 2013). Kollar and Fischer (2010) showed that the feedback process involves various activities such as feedback provision by a teacher (or peer), feedback reception by a student, and application of feedback to improve their work. It is argued by some that feedback is formative only if the information is used by the learner to improve his/her performance (Black & Wiliam, 2009).

Rather than perceiving feedback as the information of which learners are mere recipients, the contemporary view in the feedback literature emphasizes the learner’s active engagement with the feedback (Carless & Boud, 2018; Winstone, Nash, Parker, &

Rowntree, 2017). When learners play an active role in the feedback process and engage with the comments received, they not only improve their performance (Hattie & Timperley, 2007) but also develop self-regulation skills (Butler & Winne, 1995). Learners' active engagement with the feedback and their subsequent use of it is therefore crucial for effective learning. Although most literature advocating this conception of feedback is grounded in higher education (Carless & Boud, 2018; Jonsson, 2013; Winstone et al., 2017), this conception is likely equally applicable to secondary schools. However, given contextual constraints in African systems such as high-stakes examinations (Akyeampong, Lussier, Pryor, & Westbrook, 2013) such a student-oriented conception of feedback, emphasizing agency and engagement, is less likely to be realised. In this paper, the focus of interest is on how students think and feel about their mathematics teachers' FA and feedback practices. There is little agreement on the distinction between perception, conception, belief, and attitude, with Brown (2008) arguing that these terms are fundamentally synonymous. 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.

Perceptions of formative assessment and feedback practices

Unfortunately, studies show that not all feedback provided to learners contributes to improved 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 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 perceived as supportive.

The effectiveness of formative feedback depends on a student's motivation, opportunity to receive timely feedback, and a student's means to use such feedback (Lipnevich et al., 2016; Shute, 2008). Furthermore, feedback is related to learning only 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 (Crichton & McDaid, 2016; Poulos & Mahony, 2008; Strijbos, Narciss, & Dünnebier, 2010). For example, King, Schrodtt, and Weisel (2009) showed that students' perceptions that their teacher's feedback is useful were related to feedback retention (use), self-efficacy, and academic performance. However, among university students, reliance on peer or teacher feedback was associated with lower academic performance (Brown et al., 2016). Thus, the impact of feedback may depend on a number of contextual factors.

From a mathematics education perspective, formative assessment occurs naturally in the 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. Furthermore, meta-analyses indicate that

substantial positive effects of FA in terms of students' mathematics performance are not easily achieved (Bennett, 2011; Veldhuis & Van den Heuvel-Panhuizen, 2014). Student perceptions and application of teacher feedback warrant further investigation in mathematics education. Although several studies in FA have investigated students' perceptions of feedback (Brown, 2007; Carnell, 2000; Harris et al., 2014; Peterson & Irving, 2008), few studies have investigated the role of perceptions of assessment in mathematics education (Adams & Hsu, 1998; Al Duwairi, 2013). In particular, research on students' perceptions of mathematics teacher's FA 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.

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 for 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 any overreliance on national summative examinations, Tanzania introduced a Continuous Assessment (CA) program in 1976 in secondary schools, which was designed so 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). While potentially formative in that it provides early signals of success or difficulty, CA is in effect summative.

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). The Diploma in Teacher Education 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 in Tanzania comprise specific subject courses and didactical courses including a teaching practicum. Despite school improvement programmes such as the Secondary Education Development Programme (Ministry of Education and Vocational Training (MoEVT), 2008), the Basic Education Statistics in Tanzania (BEST, 2014) show that mathematics education in secondary schools in Tanzania has suffered from low passing rates. 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 (Qorro, 2013; Vuzo, 2007), (b) large class sizes (BEST, 2014), (c) curriculum content overload (Kitta & Tilya, 2010), and (d) lack of in-service teacher professional development (Komba, 2007). Further challenges include the lack of assessment skills to implement effective school based assessment (Ottevanger, Akker, & Feiter, 2007).

The present study

Despite having CA in Tanzanian secondary schools, students consistently underperform in national mathematics examinations. The Basic Education Statistics in Tanzania (BEST, 2014) indicate that for ten consecutive years (2004–2013), the majority of secondary schools students failed their mathematics national examinations. Students' poor performance among other factors raises the concern that CA is not as formative as intended to be. The Tanzanian educational system with its CA program in place has a certain infrastructure for assessment interventions; however, less is known about the formative nature and effectiveness of this system. An analysis of the Tanzanian system may allow insights into possible problems of such systems and identify potential interventions. 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 practices and feedback delivery and 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 providing formative monitoring and scaffolding?
- (2) To what extent do students' perceptions of their mathematics teachers' feedback practices and quality of delivery predict their feedback use?
- (3) To what extent do students' perceptions of their own feedback use predict their mathematics performance?

Conceptual framework

The underlying assumptions of this study were that perceptions influence behaviour and outcomes (Ajzen, 1991; 2005). Based on the theoretical conceptualizations of studies on formative assessment (Pat-El et al., 2013, 2015; Stiggins, 2005), 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 student perceptions of the nature and quality of FA practices and feedback would predict their self-reported use of the feedback, which in turn would influence performance in mathematics. Figure 1 hypothesizes that the more students perceive their teacher assessment practices as monitoring and scaffolding, rather than evaluative, and the more they perceive the delivery of feedback as supportive, the more they would use the feedback. Greater use of such feedback would contribute to greater mathematics performance.

Method

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

Figure 1. Conceptual model.

educational statistics (Ministry of Education and Vocational Training (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 school ranking (Ministry of Education and Vocational Training (MoEVT), 2013), class-size (<40 , ≥ 40), and school-type (private, government). Within the 48 randomly sampled schools there were 2767 Form 3 (Grade 11) students (53.3% female, 46.2% male, 0.5% missing) from schools varying in mathematics performance ($N_{\text{high}} = 421$, $N_{\text{middle}} = 997$, and $N_{\text{low}} = 1349$). 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 3 (Grade 11) class was selected for this study because it contains more teacher-based assessment practices compared to Form 1 and 2 (Grade 9 and 10). 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).

Design

A mixed-method research approach with quantitative (survey) and qualitative (focus group discussions) methods was applied (Creswell, 2009; Dingyloudi & Strijbos, 2018). 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.

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 Form 3 (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 (1994) 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 perceptual meaning (Dunham & Davison, 1991; Kulas & Stachowski, 2009).

Questionnaires

First, to measure student perceptions of their teachers’ FA 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 five items of the feedback delivery subscale of the ‘Feedback Environment Scale’ (Steelman, Levy, & Snell, 2004). We used the threshold of 0.70 for Cronbach’s alpha as the minimum acceptable value for each scale’s internal consistency. For the present study, all scales registered above this criterion. 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 the present study.

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

Table 1. Scales, sample items and estimates of reliability.

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

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

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 (75.7%, $N = 2094$) of the students opted for the Swahili version compared to 673 students (24.3%) who answered in English.

Focus groups discussions

Six focus group discussions (FGDs) were randomly sampled from six secondary schools. Each focus group consisted of six students, resulting in a sub-sample of 36 Form 3 students (Female = 20, Male = 16). The questions explored students' perceptions on the 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 discussions was 62 minutes.

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; students' mathematics teacher was not present while they filled in the questionnaire. The researcher or assistant first demonstrated how to use the rating scales and while filling-in the questionnaire students could ask for clarification of items. The students needed approximately 15–25 minutes to complete the questionnaire.

Analyses

Data inspection

Prior to data analysis, all negatively phrased items were recoded and data inspection was carried out to account for outliers, as well as missing value analysis. Only 39 out of the initial 2806 respondents (approx. 1%) had more than 10% missing values and 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, $\chi^2_{(56611)} = 48,876.79$, $p = 1.00$ (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. We also assessed the normality of the data obtained with an ordinal 4-point scale. While no standardised skewness was outside the range of -3 to $+3$, only 4 out of 43 manifest variables had a standardised kurtosis outside the range of -3 and $+3$. Given the large sample ($N = 2767$) and the fact that we used the

MLR estimator that produces robust standard errors, we concluded that the variables were reasonably normal (Curran, West, & Finch, 1996).

Language measurement invariance

Multigroup invariance testing determines whether responses to questionnaires are statistically equivalent between groups. If groups have statistically similar characteristics, then it can be concluded that 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 (Brown, Harris, O'Quin, & Lane, 2017; 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 *d*MACS procedure which determines the size of differences for each item and then allows calculation of the average factor *d*MACS (Nye & Drasgow, 2011). The *d*MACS effect size values can be interpreted as trivial when $|dMACS| < .20$ and small when $|dMACS| < .40$ (Cohen, 1987; Hattie, 2009; Rosenthal, Rosnow, & Rubin, 2000).

The measurement invariance tests (see Appendix A) showed that the two language versions were configural and metric invariant, but lacked scalar equivalence. However, the average *d*MACS for each factor was trivial to small; that is, perceived monitoring (*d*MACS = .10), perceived scaffolding (*d*MACS = .07), feedback delivery (*d*MACS = .17), and feedback use (*d*MACS = .10). Given the average *d*MACS effect sizes for all scales across language versions were trivial, it was decided to treat the students as members of the same population and combine the data from the two language versions.

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 (Muthén, 1994). Because the Chi-square statistic is overly sensitive in large sample sizes above 250 (Byrne, 2010), we report multiple fit indices. The comparative fit index (CFI) and the root mean square error of approximation (RMSEA) are not stable estimators because CFI rewards simple models while RMSEA rewards complex models (Fan & Sivo, 2007). The Gamma hat statistic and standardized root mean residual (SRMR) have been shown to be stable estimators (Fan & Sivo, 2007). 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). Importantly, as Steiger (1990) recommends, we report the 90%

confidence interval (90%CI) for the RMSEA. Finally, intraclass correlations (ICC) were estimated for each scale to show how much variation in the students' score is attributable to membership of the same classroom (Cress, 2008; Field, 2013; Raudenbush & Bryk, 1986). ICC values above .15 suggest that significant variation at the student-level is due to these students' membership of a classroom taught by the same mathematics teacher (Hox, 2010).

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 practices (i.e., perceived monitoring with 16 items, and perceived scaffolding with 12 items). This model had good fit (CFI = .95, Gamma hat = .95, SRMR = .035, RMSEA = .044 [90% CI = .042, .046]). Inspection of modification indices showed that two pairs of items – one pair in each factor – had strong inter-correlations and overlapping content. Hence, one item from each pair was removed resulting in 15 items for monitoring and 11 items for scaffolding. The re-estimated model had improved fit (CFI = .96, Gamma hat = .96, SRMR = .031, RMSEA = .040 [90% CI = .038, .042]).

Perception of feedback delivery, with five items, had good fit (CFI = .97, Gamma hat = .99, SRMR = .034, RMSEA = .076 [90% CI = .062, .091]). Elimination of one item, which was negatively phrased and had a low factor loading ($\beta = .14$), improved the model fit (CFI = .999, Gamma hat = 1.00, SRMR = .010, RMSEA = .026 [90% CI = .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 ($r = .98$); the items were combined into one new scale entitled 'feedback use'. The combined scale had acceptable fit (CFI = .90, SRMR = .041, RMSEA = .051 [90% CI = .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 [90% CI = .037, .046]).

Focus group discussion analyses

Content analysis was used to analyse focus group discussions. A data-derived coding scheme was developed using about ten percent of the data. 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 rationale and the segmentation and coding procedures. Two independent coding trials were performed. The first trial analysed all questions from two randomly sampled focus group discussions (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 [95%CI = .80–.94]. 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 analysed 40% of randomly sampled questions from the remaining four focus group discussions (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 [95%CI = .82–.94].

Results

Student perceptions of their mathematics teacher assessment practices

The first research question sought to describe students' perceptions of their mathematics teachers' assessment practice as providing formative monitoring and scaffolding. 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).

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 in the Form 3 (Grade 11) terminal examination ($M = 43.13$, $SD = 18.58$) was below the passing threshold of 50%. The large standard deviation is most likely attributable to the inclusion of students from a wide range of schools. The ICCs were large ($ICC > .15$) for all scales except 'feedback use'. This provides evidence that students' judgements depended on the practices of the individual teacher who taught them. This is taken into account in further analysis by including the grouping effect in the structural modelling.

FA perceptions, feedback use, and mathematics performance

The full structural model (Figure 2) consisted of 42 manifest variables and an additional manifest variable for mathematics performance, representing the five factors of interest. The model had acceptable to good fit ($CFI = .92$, $\Gamma = .95$, $SRMR = .038$, $RMSEA = .037$ [90% CI = .036, .038]). Trimming out 23 low loading items did improve the fit ($CFI = .98$, $\Gamma = .99$, $SRMR = .025$, $RMSEA = .030$ [90% CI = .027, .033]), but did not change the meaning of the model, so the full structural model is reported. The three predictor scales of perceived monitoring, perceived scaffolding, and feedback delivery were highly inter-correlated (i.e., $r > .80$). These factors had statistically significant regression weights onto students' feedback use, which subsequently predicted students' mathematics performance. More specifically, student perception of feedback delivery strongly positively predicted their feedback use ($\beta = 0.73$, $p < .001$). Although perceived monitoring

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

Variables	<i>M</i>	<i>SD</i>	Scale inter-correlations					ICC
			I	II	III	IV	V	
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

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

Figure 2. Structural equation model illustrating the relations between perceived monitoring, perceived scaffolding, feedback use, feedback delivery and mathematics performance.

and scaffolding were highly positively correlated, as in the original study (Pat-El et al., 2013), each factor had an inverse prediction to feedback use; that is, monitoring negatively predicted use ($\beta = -0.43$, $p < .001$), while scaffolding positively predicted use ($\beta = 0.41$, $p < .001$). The negative coefficient implies that for students who perceived feedback as monitoring, the less intensively they claimed to use feedback. Figure 2 represents the structural equation model for the relations under investigation.

Combined, the four predictor scales explained almost 60% of the variance in students' self-reported feedback use and only 4% 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 indicated that all relations between perceived monitoring, scaffolding, feedback delivery, and mathematics performance were fully mediated via feedback use. Table 3 represents the scales and items for each scale that were included in the present study.

Focus group discussions

Analysis of focus group discussions showed that a majority of students (19 of 36, 53%) valued their mathematics teachers' assessment and feedback practices. This occurred for two major reasons; that is, being told how to fix mistakes and receiving feedback in a supportive and friendly manner. First, ten (28%) students were satisfied with their mathematics teachers' feedback only if the teacher told them what to improve their work (e.g., *I am satisfied because when we make mistakes she corrects us and advises what to do*; Focus group 5). Second, nine (25%) students were satisfied with the

Table 3. Items for each scale.

Scale and Items
<i>Perceived scaffolding</i> When I do not understand a topic, my mathematics teacher tries to explain it in a different way. My mathematics teacher provides me with hints to help understand the subject matter. During mathematics class I have an opportunity to show what I have learned. My mathematics teacher asks questions in a way I understand. My mathematics teacher asks questions that help me gain understanding of the subject matter. My mathematics teacher allows for my contribution during the lesson. I have the opportunity to ask my classmates questions during the mathematics lesson. My mathematics teacher makes me aware of the areas I need to work on to improve my results. There is an opportunity to ask questions during the mathematics lesson. I am aware of the criteria by which my math assignment will be evaluated. When I receive a mathematics assignment it is clear to me what I can learn from it.
<i>Perceived monitoring</i> My mathematics teacher encourages me to reflect on how I can improve my assignments. After examining my test results, my mathematics teacher discusses the answers I gave to the test with me. Whilst working on my mathematics assignments, my mathematics teacher asks me how I think I am doing. My mathematics teacher allows me to think about what I want to learn in school. My mathematics teacher inquires what went well and what went badly in my work. My mathematics teacher encourages me to reflect on my learning process and to think about how to improve next time. My mathematics teacher stresses my strengths concerning learning. My mathematics teacher identifies my weaknesses concerning learning. I am encouraged by my mathematics teacher to improve my learning process. My mathematics teacher gives me guidance to assist my learning. My mathematics teacher discusses assignments with me to help me understand the subject matter better. My mathematics teacher discusses with me the progress I make. After each assessment my mathematics teacher informs me how to improve the next time. My mathematics teacher discusses with me how to exploit my strengths to improve my assignment. My mathematics teacher and I consider ways to improve my weak points.
<i>Feedback delivery</i> My mathematics teacher is supportive when giving me feedback about my mathematics performance. My mathematics teacher is tactful when giving me feedback about my mathematics performance. My mathematics teacher considers my feelings when giving me feedback about my mathematics performance. My mathematics teacher generally provides feedback in a thoughtful manner.
<i>Feedback use</i> The feedback on my mathematics assignments prompts me to revise instructional material covered earlier in the course. The feedback on my mathematics assignments helps me with subsequent assignments. I read written feedback on my mathematics assignments carefully to understand what the feedback is saying. I use feedback on my mathematics assignments for revising. I use the feedback on my mathematics assignments to review what I have done. I think feedback from my mathematics teacher is vitally important in improving my mathematics performance. I think that feedback from my mathematics teacher provides clear direction on how to improve my performance. Feedback from my mathematics teacher motivates me to improve my mathematics performance. I pay careful attention to instructional feedback from my mathematics teacher. I listen carefully when my mathematics teacher provides oral feedback. I usually reflect on my mathematics teacher's feedback. I am extremely encouraged by positive feedback from my mathematics teacher.

feedback if they perceived the teacher as delivering the feedback in a friendly manner (e. g., *I am satisfied because she gives us feedback very friendly and convinces us to continue studying and do more exercises*; Focus group 2). This is consistent with recent reviews on students' feedback preferences and feedback use (Jonsson, 2013; Lipnevich et al., 2016). These approaches to providing feedback also seem consistent with the idea that feedback should scaffold learning, as opposed to monitor it. It seems that the students expected to gain more teacher directed feedback rather than to generate their own feedback that might improve or regulate their learning. Our findings do support earlier

work by Rach et al. (2013) who showed that secondary school students need teacher support to learn from their mistakes. In particular, given that most of the students in Tanzania fail their national mathematics exam, students are less likely to regulate their learning, relying solely on their own perceptions or the feedback of their peers. Teacher scaffolding support seems a logical source for learning.

In contrast, dissatisfaction arose for 12 of 36 (33%) students because teachers did not give the test results confidentially and personally to students and sometimes did not even provide corrections. For example, six (17%) students were dissatisfied when teachers did not correct student errors (e.g., *I am not satisfied because when he gives the test papers he tells us to do corrections of our errors ourselves*; Focus group 4). Furthermore, some students were dissatisfied with feedback when the teacher did not maintain privacy in the feedback they provided (e.g., *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*; Focus group 6). 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 & Walker, 2011; Lizzio & Wilson, 2008; Weaver, 2006). Dissatisfaction also came from three (8%) students who reported being reprimanded by their teachers for making errors or getting low scores (e.g., *I am not satisfied because if you get low marks you become reprimanded*; Focus group 6). 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 those students who participated in the focus group discussions used various methods to inform students how to reduce the discrepancy between the desired goal and their current state. In response to the question ‘Could you please give examples of what your mathematics teacher does 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, almost half of the students (17 of 36, 47%) indicated their mathematics teacher provided them with cognitive support showing and/or correcting their errors. Cognitive support involved teacher scaffolding practices that explicitly showed students how to fix their errors (e.g., *Our teacher is very friendly to us and when you do some mistakes, she is ready to do corrections and teach you even personally*; Focus group 2). Secondly, one-third of the students (12 of 36) said their teacher gave them positive affective support in error situations. Affective support involved scaffolding practices such as encouraging students and reducing fear resulting from making errors (e.g., *When I do errors in a test my mathematics teacher advises me and tells me to pull up my socks [work hard]*; Focus group 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 (e.g., *When I make errors in a mathematics test as my fellow student said, my teacher reprimands me*; Focus group 6).

In general, the focus group discussions support the relationships revealed by the structural equation modelling of the survey data; that is, perceived monitoring is negative to feedback use. In other words, students did not perceive all monitoring practices as supportive, while supportive scaffolding and teacher feedback delivery predicted greater feedback use. The mechanisms teachers used to provide (formative)

feedback clearly matter to students' sense of (dis)satisfaction with teacher's feedback practices and to whether or not they make use of the feedback.

Discussion

Student perceptions of FA 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, only when it was warm and friendly. This replicates results from previous studies (e.g., Pat-El et al., 2015; Weaver, 2006) that students value warm, supportive teacher feedback. However, according to Rach et al. (2013) students' positive perceptions of teacher assessment practices do not necessarily result into effective performance.

Given the sample size of this study, we might expect this result to generalize to all Tanzanian secondary school students. However, the large ICCs indicated that students' perceptions of their mathematics teacher assessment practices varied considerably between classrooms (Hox, 2010; Kyaruzi, Strijbos, & Ufer, 2016). While the ICCs indicate that there is clearly a significant teacher effect, there were insufficient teachers in this study to establish the teacher or school characteristics that account for those effects. Furthermore, due to our sampling of students from private and government owned schools, socioeconomic factors might also explain the large ICCs. Nonetheless, the current results do provide insights into how Tanzanian students experience and perceive feedback; most likely in accordance with the warmth and supportive nature of teacher practices.

Student use of feedback

Student perceptions of teacher FA 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 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 when teachers are also expected to monitor and evaluate student learning.

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 formative assessment ought to enhance student learning rather than merely

monitor student learning and calls for a closer examination of typical monitoring practices by mathematics teachers in Tanzania. First, 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 own individual work reduces the perception that feedback is scaffolding. Similarly, practices such as reprimanding low achievement increased students' anxiety in error situations and reduced the likelihood of students' using the feedback. Generally, in focus group discussions, students indeed reported that they did not perceive all monitoring practices as supportive; this applies in particular to those practices which are not used for scaffolding purposes, such as reprimanding for errors. These findings support Rach et al. (2013) who showed that, in mathematics, students did not always benefit from teacher assessment practices. Secondly, 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 ...'). This implies 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).

In light of these results, it seems that mathematics teachers ought to consider student perceptions of their feedback practices more extensively when implementing FA in mathematics classes. For example, greater individual customisation of assessment feedback would probably increase student use of feedback, although this could be challenging in large classes. Furthermore, greater respect for the privacy and self-worth of students might also enhance student use of feedback.

The impact of FA perceptions on mathematics performance

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 contribute to greater learning outcomes. It may be, however, that when students are functioning well below the expected standard of the test, their use of helpful feedback may help them progress toward, but not achieve the expectations of the examination. 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. Making use of feedback may be necessary for improvement, but that does not necessarily equate with meeting the standards. The findings support recent intervention studies in mathematics education that positive effects in terms of mathematics achievement are found when feedback is embedded in instruction with the emphasis on feedback utilization (Pinger, Rakoczy, Besser, & Klieme, 2018b).

A crucial finding given the examination-centred Tanzanian education system was that all relations from student perceptions of teacher FA practices and feedback delivery to mathematics performance were fully mediated via 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 and relationships. It may also be that the quality of feedback depends in part on the quality of instruction, which has not been evaluated in this study. 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 & Hirschfeld, 2008), and demonstrates how to correct mistakes (Shute, 2008).

Methodological limitations

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. Secondly, while the model depends on student self-report ratings of teacher practices, there is no observational evidence that teachers' practices are consistent with these perceptions. Third, although our analyses show evidence for the feasibility of using self-reported student mathematics performance, other objective measures of student performance are to be encouraged. Nevertheless, the current results signal the kinds of practices that need to be systematically observed in future studies.

Theoretical contribution

Student perceptions of their teachers' formative assessment 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 practices regulate students' feedback use. These results support previous work that related student perceptions of assessment to their performance (Brown & Hirschfeld, 2007; 2008), but it further extended these results by showing that feedback use contributed to students' mathematics performance. These results support the planned behaviour theory that perceptions (beliefs) influence behaviour, which subsequently predicts outcomes (Ajzen, 1991). Students are less likely to apply feedback when they perceive it to be monitoring 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 very important in FA since feedback is formative when it can be used by learners to improve their learning (Black & Wiliam, 2009; Lizzio & Wilson, 2008; Poulos & Mahony, 2008).

Practical contribution

Findings in this study highlight the need for increasing teacher awareness of the impact of their assessment and feedback practices on student learning. Teachers as key facilitators of FA in schools are supposed to provide feedback that helps students to bridge the gap between their current performance and the desired standard (Kyaruza, Strijbos,

Ufer, & Brown, 2018) as well as regulate their learning. Teacher's efforts towards teaching, assessing, and providing feedback to students might be less productive if those practices are not positively perceived by students and subsequently used to improve learning. However, improving teacher assessment and feedback practices is far from trivial as teachers might lack knowledge in the form of assessment and feedback literacy, 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 professional development aimed at improving mathematics teachers' 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 personally 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). Our results might be transferable to educational systems that apply similar high-stakes examination assessment programmes such as India, Egypt, China, or Hong Kong. For example, Carless (2011) suggested that, in the high-stakes examination culture of Hong Kong (and Confucian-Heritage settings in general), an important and initial step toward interactive formative assessment involved developing teachers' formative use of summative test information. In India, also a high-stakes examination society, it was found that use of diagnostic tests was associated strongly with conceptions of assessment as improvement; hence, for mathematics testing to contribute effectively to feedback, it has to be done for the purpose of improvement, not evaluation (Brown, Chaudhry, & Dhamija, 2015). Given that Tanzania has a similar focus on high-stakes examination, our findings might be extended to such contexts and that students in those settings are more likely to prefer teacher-scaffolding feedback practices rather than those seen as monitoring. This signifies the necessity for further research into the role of monitoring practices and to consider the specific educational and assessment context when conceptualizing teacher professional development for formative assessment.

Although the contemporary conception of feedback emphasises learner's agency and active engagement with the feedback (Carless & Boud, 2018; Winstone et al., 2017), our findings indicate that mathematics teachers in Tanzania still employ a teacher-centered conception of feedback. Students also expect external guidance from their teacher, and they are content when the feedback specifies what is wrong and how to close the gap between their current state and the desired state. Although current policy in Tanzania emphasizes learner-centered rather than teacher-centered practices (Kitta & Tilya, 2010), several factors can explain why students want to rely on their teachers. Students are aware of how good their peers are and, when they consider their peers are not very good, they focus on the teacher as the 'true' authority in terms of feedback

and correction (Gao, 2009; Peterson & Irving, 2008). Furthermore, there are reasons why teachers might not make use of students as educational resources in assessment. For example, large class sizes (BEST, 2014), lack of assessment and feedback literacy and in particular formative assessment skills (Ottevanger et al., 2007), and the tensions arising from high stake examinations (Black & Wiliam, 2018; Hopfenbeck, 2017). In addition, learners with low levels of prior knowledge may compel teachers to rely on teacher-centred feedback practices, because until learners have developed basic knowledge and skills to regulate their learning, new instruction is more powerful than the feedback information (Hattie & Timperley, 2007). Distinctively, the findings show that when learners lack prerequisite knowledge, they require teacher scaffolding support (Rach et al., 2013) rather than teacher monitoring practices or self-regulation.

Because most of the education systems in the Sub-Saharan African region apply high-stakes examinations (Akyeampong et al., 2013), improvement in performance can be more readily achieved through effective teacher feedback delivery and scaffolding practices at the expense of monitoring practices. The findings imply that, for students to achieve the desired goal (i.e., master competencies and/or pass examinations), teachers should improve how they deliver feedback to their students. More specifically, teachers need to consider student feelings and emotions when delivering or discussing student feedback. We recommend that future studies investigate typical monitoring practices by Tanzanian mathematics teachers, beyond the tests from the Continuous Assessment program. Likewise, we recommend professional development aimed at improving teacher assessment and feedback literacy, feedback delivery and FA practices, because our results show that promoting positive student perceptions of teacher assessment practices enhances feedback use and subsequently student learning gains.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Florence Kyaruzi is a Lecturer in educational psychology and mathematics education in the Department of Educational Psychology and Curriculum Studies at the Dar es Salaam University College of Education (DUCE). He holds Ph.D in the Learning Sciences from the Ludwig-Maximilians-Universität München (LMU), Department of Psychology. His research expertise includes student and teacher assessment perceptions and conceptions, learning from errors in mathematics education, and teacher education.

Jan-Willem Strijbos is a Professor in the Department of Educational Sciences, University of Groningen and a former Professor in the Ludwig-Maximilians-Universität München, Department of Psychology. His research expertise focuses on Assessment and Feedback; Technology-Enhanced Learning (TEL); (Computer-Supported) Collaborative Learning (CS)CL; Peer assessment; Peer feedback; Feedback dialogues; Communities of learners/communities of practice; Instructional design; Analysis methods for (CS)CL; Content analysis of discourse.

Stefan Ufer is a Professor in the Department of Mathematics, Ludwig-Maximilians-Universität München. Being the head of the Chair for Mathematics Education at LMU, his research projects study the teaching and learning of mathematics at different stages of the learning biography and how mathematical learning processes can be supported in an age-appropriate way.

Gavin T. L. Brown is Professor in the Faculty of Education and Social Work, University of Auckland, and the Director of the Quantitative Data Analysis and Research Unit. His research expertise focuses on Educational testing and assessment, Psychometrics and measurement, Social psychology of assessment, attitudes, values, and beliefs and Structural equation modeling.

ORCID

Florence Kyaruzi <http://orcid.org/0000-0003-4660-6764>

Jan-Willem Strijbos <http://orcid.org/0000-0003-4180-688X>

Gavin T. L. Brown <http://orcid.org/0000-0002-8352-2351>

References

- Adams, T. L., & Hsu, J. Y. (1998). Classroom assessment: Teachers' conceptions and practices in mathematics. *School Science and Mathematics*, 98(4), 174–180.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.
- 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.
- 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.
- Baas, D., Castelijns, J., Vermeulen, M., Martens, R., & Segers, M. (2015). The relation between Assessment for Learning and elementary students' cognitive and metacognitive strategy use. *British Journal of Educational Psychology*, 85(1), 33–46.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25.
- BEST. (2014). *Basic education statistics in Tanzania*. (National data). Dodoma, Tanzania: Prime Minister's Office Regional Administration and Local Government.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*. doi:10.1080/0969594X.2018.1441807
- Briggs, M., Woodfield, A., Martin, C., & Swatton, P. (2006). *Assessment for learning and teaching in primary schools*. Glasgow: Bell & Bain.
- 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. (2013). Assessing Assessment for Learning: Reconsidering the policy and practice. In M. East & S. May (Eds.), *Making a difference in education and social policy* (pp. 121–137). Auckland, NZ: Pearson.
- 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.
- 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.

- 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.
- Brown, G. T. L., & Hirschfeld, 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., 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.
- Brown, J. (2007). Feedback: The student perspective. *Research in Post-Compulsory Education*, 12(1), 33–51.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245–281.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications and programming*. London: Routledge.
- 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.
- Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. New York: Routledge.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325.
- Carnell, E. (2000). Dialogue, discussion and feedback — Views of secondary school students on how others help their learning. In S. Askew (Ed.), *Feedback for learning* (pp. 46–61). London: Routledge.
- Chang, L. (1994). A psychometric evaluation of 4-point and 6-point likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, 18(3), 205–215.
- 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.
- Cohen, J. W. (1987). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 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.
- Cress, U. (2008). The need for considering multilevel analysis in CSCL research—An appeal for the use of more advanced statistical methods. *International Journal of Computer-supported Collaborative Learning*, 3(1), 69–84. doi:10.1007/s11412-007-9032-2
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative and mixed approaches*. Los Angeles: Sage.
- Crichton, H., & McDaid, A. (2016). Learning intentions and success criteria: Learners' and teachers' views. *Curriculum Journal*, 27(2), 190–203.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438–481.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16–29.
- Dingyloudi, F., & Strijbos, J. W. (2018). 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* (pp. 444–454). New York: Routledge.
- Dunham, T. C., & Davison, M. L. (1991). Effects of scale anchors on student ratings of instructors. *Applied Measurement in Education*, 4(1), 23–35.
- Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, 42(3), 509–529.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.

- Furtak, E. M., Kiemer, K., Circi, R. K., Swanson, R., de León, V., Morrison, D., & Heredia, S. C. (2016). Teachers' formative assessment abilities and their relationship to student learning: Findings from a four-year intervention study. *Instructional Science*, 44(3), 267–291.
- Gao, M. (2009). Students' voices in school-based assessment of Hong Kong: A case study. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 107–130). Charlotte, NC: Information Age Publishing.
- 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.
- 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.
- 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.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Hopfenbeck, T. N. (2017). Balancing the challenges of high-stakes testing, accountability and students' well-being. *Assessment in Education: Principles, Policy & Practice*, 24(1), 1–3.
- 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.
- Jonsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*, 14(1), 63–76.
- 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.
- 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.
- 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.
- Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction*, 20(4), 344–348.
- 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.
- Kyaruzi, F. (2012). Effectiveness of mathematics continuous assessment feedback in secondary schools in Tanzania. *Journal of Education, Humanities and Sciences*, 1(2), 51–63.

- Kyaruzi, F., Strijbos, J. W., & Ufer, S. (2016, August). Students' assessment for learning perceptions and their mathematics achievement in Tanzanian secondary schools. In C. Csikos, A. Rausch, & J. Sztányi (Eds.), *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 155–162). Szeged, Hungary: PME.
- Kyaruzi, F., Strijbos, J. W., Ufer, S., & Brown, G. T. L. (2018). Teacher AfL perceptions and feedback practices in mathematics education among secondary schools in Tanzania. *Studies in Educational Evaluation*, 59, 1–9.
- 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.
- Lizzio, A., & Wilson, K. (2008). Feedback on assessment: Students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education*, 33(3), 263–275.
- 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: National Examination Council of Tanzania.
- 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.
- Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & 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* (pp. 125–143). Mahwah, NJ: Lawrence Erlbaum.
- 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.
- 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). Washington, DC: The World Bank.
- 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.
- 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.
- Peterson, E. R., & Irving, S. E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning & Instruction*, 18(3), 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.
- Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2018a). Interplay of formative assessment and instructional quality—Interactive effects on students' mathematics achievement. *Learning Environments Research*, 21(1), 61–79.
- Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2018b). Implementation of formative assessment—Effects of quality of programme delivery on students' mathematics achievement and interest. *Assessment in Education: Principles, Policy & Practice*, 25(2), 160–182.
- Popham, W. J. (2014). *Classroom assessment: What teachers need to know* (7th ed.). Pearson: Harlow.

- Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. *Assessment & Evaluation in Higher Education*, 33(2), 143–154.
- Qorro, M. (2013). Language of instruction in Tanzania: Why are research findings not heeded? *International Review of Education*, 59(1), 29–45.
- 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 of the United States of America*, 8(1), 21–30.
- Raudenbush, S., & Bryk, S. A. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59(1), 1–17.
- Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge, UK: Cambridge University Press.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144.
- Scriven, M. (1991). Beyond formative and summative evaluation. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation & education: At quarter century* (Vol. 90, Part II, pp. 19–64). Chicago, IL: NSSE.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.
- 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.
- 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*, 64(8), 22–26.
- Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 133–146). London: Sage.
- 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.
- Strijbos, J. W., Narciss, S., & Dünnebie, 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.
- Swaffield, S. (2011). Getting to the heart of authentic Assessment for Learning. *Assessment in Education: Principles, Policy & Practice*, 18(4), 433–449.
- 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.
- United Republic of Tanzania. (1974). *The Musoma resolution*. Dar es Salaam: Government Printers.
- 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.
- 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.
- Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment and Evaluation in Higher Education*, 31(3), 379–394.

- 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. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14.
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy and Practice*, 11(1), 49–65.
- 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.
- Wu, A. D., Li, Z., & Zumbo, B. D. (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), 1–26. Retrieved from: <http://pareonline.net/getvn.asp?v=12%26n=3>

Appendix A. Language measurement invariance tests.

Model (43 variables)	SRMR	RMSEA	Df	χ^2	Δdf	$\Delta \chi^2$	p	CFI	ΔCFI
Unconstrained	.044	.030	1708	5995.484				.915	
Measurement weights	.054	.030	1747	6130.327	39	134.843	.000	.914	-.001
Measurement intercepts	.052	.032	1790	6893.929	43	763.602	.000	.899	-.015