
LANGUAGE ASSESSMENT LITERACY: INVESTIGATING ENGLISH LANGUAGE TEACHERS' ASSESSMENT LITERACY IN MOZAMBIQUE

Felizardo Muianga

Department of English Language and Literature, Faculty of Education, Masaryk University in Brno

Abstract:

This quantitative study constitutes the first phase of an explanatory sequential mixed-methods study on English language teachers' assessment literacy in Mozambique. This phase of the study explored (n=72) English language teachers' experience with formal training in language assessment and their self-perceived levels of classroom-based language assessment literacy. The findings suggest that the training received by the respondents in four dimensions of classroom-based language assessment literacy sits between moderate and advanced. The respondents seem to have attained the recommended levels of classroom-based language assessment literacy in two dimensions out of four, accentuating the need for more training in language assessment.

Key words:

Language assessment, formal training, assessment literacy, Mozambique

Introduction

Teachers dedicate about a third of their professional time to various assessment-related activities (Stiggins, 2014, p. 68). Assessments are categorised as summative and formative. Summative assessments are often conducted at the end of a unit, semester, or school year to measure student achievement. The assessment results are often used to make high-stakes decisions about the students, such as retention and promotion (Cizek, 2010). Conversely, formative assessment is continuous and intended to promote student learning (Stiggins, 2002; Wiliam, 2010). The term formative assessment tends to be used as a synonymy of classroom-based assessment (Davison & Leung, 2009). However, the term classroom-based assessment is broader, encompassing both formative and summative assessments (Hill & McNamara, 2012). Despite the difference between the two terms, they both refer to teacher-mediated, context-based, and classroom-embedded assessment-related activities (Davison & Leung, 2009, p. 395).

Classroom-based assessment has aroused researchers' interest due to its importance to instruction (Popham, 2008, 2011, 2014, 2018; Stiggins, 2014). Classroom teachers are expected to develop assessment literacy to meet the demands of classroom-based assessment (Popham, 2014, 2018) and continuously update their assessment knowledge to keep pace with changes in the assessment field (Tsagari & Vogt, 2017). Although assessment is central to instruction, empirical evidence shows that teacher assessment literacy remains underdeveloped (Kvasova & Kavytska, 2014; Lam, 2015, 2019; Sultana, 2019; Xu & Brown, 2017). Tsagari and Vogt's (2017) study involving foreign language teachers in different European countries found that participants'

perceived level of language assessment literacy was inadequate for engaging in various assessment activities. Vogt and Tsigari (2014) found that some teachers even showed a limited understanding of concepts related to language assessment. Lan and Fan (2019) uncovered that English language teachers' perceived language assessment literacy level was at around the functional level—based on the assessment literacy continuum scale proposed by Pill and Harding (2013). Overall, despite the importance of assessment to instruction and calls for teachers to develop assessment literacy (e.g., Popham, 2018; Stiggins, 1991; Tsigari & Vogt, 2017), research shows that teacher assessment literacy remains underdeveloped in different contexts worldwide, and some teachers are even unfamiliar with some concepts related to language assessment. These findings underline the need for teacher professional development.

In 2008, the government of Mozambique introduced a new English language curriculum in secondary education, which emphasises formative assessment rather than summative assessment. According to the guiding documents (e.g., INDE/MINED, 2010a, 2010b, 2010c, 2010d, 2010e), teachers are expected to (1) use assessment to cater to different students' needs; (2) use assessment to measure student achievement; (3) use formal and informal tools to gather evidence of students' learning; (4) engage students in the assessment process through self- and peer-assessment; and (5) assess both productive and receptive skills. These assessment activities include both summative and formative assessments, which means that teachers have to play a dual role—that of “facilitator and monitor of language development and that of assessor and judge of language performance as achievement” (Rea-Dickins, 2004, p. 253). To fulfil these assessment responsibilities, teachers need expertise in language assessment, which could be attained through pre-service teacher education or professional training and development. Although there is a vast literature on language assessment and testing, the Mozambican context remains under-researched. This research study intends to explore English language teachers' assessment literacy in Mozambique to fill this gap.

Classroom-Based Language Assessment

Assessment refers to all methods used to gather information about student knowledge, skills, and abilities (Purpura, 2016, p. 191). Currently, it is widely agreed that classroom assessment is an integral part of effective instruction (Lan & Fan, 2019; Rea-Dickins, 2004). Hill and McNamara (2012, p. 396) define classroom assessment as “any reflection by teachers (and/or learners) on the qualities of a learner's (or group of learners') work and the use of that information by teachers (and/or learners) for teaching, learning (feedback), reporting, management or socialisation purposes.” Classroom-based assessment fulfils both summative and formative purposes (Boraie, 2018; Hill & McNamara, 2012; Mathew & Poehner, 2013). However, some experts contend that it should focus more on later purposes than the former (Green, 2018; Migliacci, 2018). Although classroom-based assessment should concentrate more on improvement, Rea-Dickins (2004, p. 249) noted that when teachers talk about classroom assessment, they have “a tendency to prioritise the ‘formal’ and the ‘procedural’ and to underplay the observation-driven approaches to assessment.” These findings suggest that some teachers view classroom assessment from the summative standpoint.

The increasing focus on classroom-based language assessment accentuates the need for language teachers to achieve and maintain adequate levels of language assessment literacy (Rea-Dickins, 2007; Stoyonoff, 2012; Tsigari & Vogt, 2017). As Green (2018) points out, language teachers need to be able to develop assessment

instruments that accurately gauge students' language skills and use the evidence to adjust instruction.

Classroom-Based Language Assessment Literacy

Stiggins (1991) coined the term *assessment literacy* to refer to knowledge and skills teachers need to engage in good classroom-based assessment practices. For Stiggins, an assessment literate teacher can distinguish between high-quality and low-quality assessment and can use their knowledge about assessment to make informed decisions about students' learning. The language assessment literacy (henceforth LAL) field derives from general assessment literacy, and its conceptualisation has also been influenced by general assessment literacy (Stabler-Havener, 2018). What distinguishes the two fields is that the assessment construct in LAL is language (Giraldo, 2018).

Despite the latest developments in language assessment, the question surrounding the conceptualisation of LAL and differentiated LAL for each assessment stakeholder (e.g., test developers and language teachers) remains partially answered (Kremmel & Harding, 2020; Levi & Inbar-Lourie, 2020). However, several experts have attempted to spell out LAL components (e.g., Brindley, 2001; Davies, 2008; Fulcher, 2012; Inbar-Lourie, 2008a; Kremmel & Harding, 2020; Pill & Harding, 2013; Taylor, 2013). Generally, LAL is conceptualised as consisting of knowledge, skills, and principles. The tripartite view of LAL is embraced by several authors, such as Davies (2008), Fulcher (2012), Giraldo (2018), and Inbar-Lourie (2008). Regarding differentiated LAL profiles for different assessment stakeholders, Taylor (2013) used the assessment literacy continuum scale adapted by Pill and Harding (2013)¹ to build four assessment profiles. According to the model (see figure 1), the level of LAL that language teachers should attain varies depending on the dimension: knowledge of theory (2), principles and concepts (2), technical skills (3), language pedagogy (4), socio-cultural values (3), local practices (3), personal beliefs and attitudes (3), and scores and decision making (2). Kremmel and Harding (2020) criticise Taylor's profiles for being speculative. Notwithstanding the criticism, this model has been used to explore teachers' LAL in different contexts. Using Taylor' (2013) model as a reference, this study explored English language teachers' assessment literacy in Mozambique. The study sought to answer the following research questions:

1. How well does Lan and Fan's (2019) classroom-based language assessment literacy model fit the new dataset?
2. How much formal training did teachers receive in different dimensions of classroom-based language assessment literacy?
3. What is the teachers' self-perceived level of classroom-based language assessment literacy?

¹ Pill and Harding's (2013) assessment literacy continuum scale: *0-illiteracy* (unfamiliarity with concepts and methods of language assessment), *1-nominal literacy* (understanding that a particular term is related to language assessment but with some misconceptions), *2-functional literacy* (good understanding of basic terms and concepts related to language assessment), *3-procedural and conceptual literacy* (understanding of the important concepts in the field, and the ability to translate assessment knowledge into practice), and *4-multidimensional literacy* (knowledge expanded beyond the understanding of ordinary concepts, philosophical, historical and social components).

Methodology

Research Design

An explanatory sequential mixed-methods design was adopted (Creswell & Creswell, 2018). This design consists of collecting quantitative data first, analysing the data, and then using the results to plan the qualitative phase of the study. The quantitative phase of the study explored teachers' experience with formal training in language assessment and their self-perceived level of language assessment literacy. The next phase of the study will focus on how individual teachers use classroom-based language assessments formatively, thereby shedding more light on the quantitative part of the study.

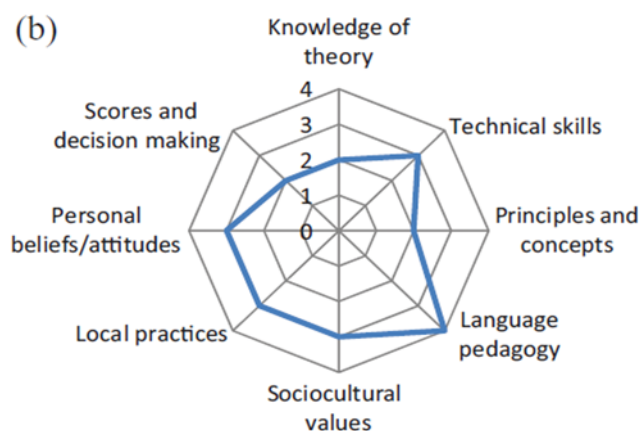


Figure 1: LAL Profile for classroom teachers
Transcribed from Taylor (2013, p. 410)

Participants

The respondents of the English language teacher questionnaire (ELTQ) were secondary school teachers of English as a foreign language based in Maputo, Mozambique. The proportional stratified random sampling technique was used to collect the data (Johnson & Christensen, 2016). This technique consists of dividing the target population into strata and subsequently sampling each group randomly. The stratification variables were gender and school location (Maputo City and Maputo Province). There are about four hundred teachers of English as a foreign language in Maputo, where about 60% work in Maputo City and about 40% in Maputo Province (MINED, 2018). While the male teachers are about 89%, the female teachers are roughly 11%.

Initially, the teachers were invited to participate in the study through email and SMS. However, only a handful of them responded to the invitation. Therefore, we decided to visit 38 secondary schools (22 in Maputo City and 16 in Maputo Province) to invite teachers to participate in the study. Although 142 teachers (126 males and 16 females) were invited to respond to the questionnaire, only 72 participants (84.7% males and

15.3% females) submitted their responses. All the prospective respondents received either a soft or a hard copy of the ELTQ (see appendices). For more information about the participants, see table 1.

Data Collection

This quantitative study partly replicated Lan and Fan's (2019) research instrument, which was adapted from Vogt and Tsagari (2014), and previously used by Hasselgreen et al. (2004). Lan and Fan (2019) used the instrument to investigate teachers' LAL in China. The authors concentrated on the English teachers' self-perceived level of LAL and their training needs in four dimensions of LAL. The research instrument encompasses 18 items, which focus on four dimensions of classroom-based language assessment: *technical skills* (ELTQ1 to ELTQ6), *scores and decision making* (ELTQ7 to ELTQ10), *language pedagogy* (ELTQ11 to ELTQ15), and *theories and principles* (ELTQ16 to ELTQ18). The authors' factor analysis work found four factors.

The first part of the questionnaire focused on the sociodemographic characteristics of the respondents. It was changed to reflect the Mozambican context, emphasising respondents' age, gender, years of experience, academic qualifications, university attended, and school location. The second part is divided into two sections. The first section focuses on teachers' experience with formal training in classroom-based LAL. Using Lan and Fan's (2019) questionnaire and a five-point Likert scale², the respondents were asked to specify the amount of training they received in each aspect of language assessment. The second section deals with teachers' self-perceived level LAL. Using the same questionnaire and the five-pointed Likert scale adapted by Pill and Harding (2013), the respondents were asked to specify their perceived level of LAL.

Table 1: Participants

School Location	Qualifications	Years of experience	Gender	Age	Institution
Maputo City 40	Diploma ³ 3	1-5 27	Males 61	20-25 12	UEM 31
Maputo Province 32	B.A 66	6-10 20	Females 11	26-30 10	UP 32
	M.A 3	11-15 16		31-35 25	Others 6
	PhD 0	16-20 7		36-40 17	
		21 and more 2		41 and more 8	

Data Analysis

Three statistical procedures were performed in this study. Firstly, a measure of internal consistency of the questionnaire was performed using Cronbach's alpha. Secondly, confirmatory factor analysis was performed to test measurement invariance (Brown, 2015; Harrington, 2008). This test helped determine whether the four-factor structure in Lan and Fan's (2019) could be replicated in the new dataset from 72 respondents from Mozambique. In reporting the goodness-of-fit of a model with a new dataset, Kline (2015) recommends selecting chi-square (χ^2), root mean square error of approximation (RMSEA), comparative fit index (CFI), and standardised root mean square residual (SRMS). However, Hurley et al. (1997) recommend using various goodness-of-fit

² Likert scale: 0-no training at all; 1- little training (few days/ a month); 2- moderate training (a semester); 3- advance training (a year); 4-very advanced training (more than a year)

³ Diploma in teaching: non-degree programmes offered by various Teacher Training Institutes in Mozambique.

indices simultaneously, indexing distinct aspects of the model. Therefore, besides the indices recommended by Kline (2015), the study reported other indices: goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), normed fit index (NFI), and Tucker-Lewis Index (TLI). Thirdly, descriptive statistics were performed to determine the formal training teachers received in four dimensions of language assessment and their self-perceived level of LAL. While the confirmatory factor analysis was performed using IBM SPSS Amos 26.0 software, descriptive statistics and the measure of internal consistency were done using IBM SPSS Statistics 27.0 software.

Results

Confirmatory Factor Analysis

The Cronbach's alpha of the first and the second sections of the questionnaire is .935 and .943, respectively, suggesting that the items in the questionnaire in the two sections are internally consistent. Regarding the assessment of normality, Skewness and Kurtosis indices are presented in table 2 in appendices. According to Brown (2006), when using structural equation modelling, appropriate values for Skewness range between - 3 and + 3 and Kurtosis from - 10 and + 10. Therefore, the ELTQ can be considered normally distributed.

Model Fit

Lan and Fan's (2019) model is a four-factor model of classroom-based language assessment literacy. The four latent variables were named TS (technical skills), SDM (scores and decision making), LP (language pedagogy), and TP (theories and principles). Initially, with $\chi^2 = 226.122$, $df = 129$, and $p < .05$ and with $GFI = .763$, $AGFI = .685$, $NFI = .762$, $TLI = .863$, $CFI = .884$, $RMSEA = .098$, and $SRMS = .105$, the four-factor model of classroom-based language assessment literacy did not fit the new dataset well. For more information, see figure 2 in the appendices. Following Brown's (2006) criteria for model fit indices, CFI should be close to 0.95 or greater, TLI close to 0.95, and RMSEA close to 0.06. The rule of thumb for $NFI = .95$, $GFI = .95$, $AGFI = .90$, and $SRMS < .05$ (Schermele-Engel et al., 2003).

Model Modification

The proposed model did not fit the dataset well, so the researcher decided to modify it. The first noticeable modification proposed by modification indices (MI) involves item e15 (testing aspects of culture). The MI suggest adding paths from e15 to two latent variables, namely language pedagogy (LP) and theories and principles (TP). Additionally, the MI suggest adding an error covariance between e15 and e16, e15 and e8, e15 and e10, and e15 and e13. It seems that e15 is a very complex item, so the researcher decided to remove it. Its removal resulted in significant improvements to the model. The second most noticeable modification proposed by MI involves adding a covariance between e9 (placing students) and e10 (awarding certificates). Both items belong to the same latent variable—scores and decision-making factor (SDM). The two items are related to student classification based on their performance; thus, adding the covariance between the two items seems reasonable. The last modification involved e8 (finding out what needs to be taught/learned). The MI propose adding paths between e8 and two latent variables: language pedagogy (LP) and theories and principles (TP). The MI also suggest adding a covariance between e8 and e11. The items e8 and e11 do not belong to the same latent variable, so akin to item e15, the researcher decided to remove item e8, which resulted in considerable improvements to the model.

The modified model (see figure 3 in appendices) preserves the four-factor structure of the original model, but it comprises only 16 items rather than 18. Following Brown's

(2006) criteria for model fit, with a $\chi^2 = 115.132$, $df = 97$, $p = .101$, $RMSEA = .051$, $TLI = .964$, and $CFI = .971$, the modified model fits the dataset well. Although the SRMR should be less than .05, values smaller than .10 (e.g., $SRMR = .065$) are considered acceptable (Schermelleh-Engel et al., 2003). Other indices such as $GFI = .84$, $AGFI = .78$, $NFI = .845$ have improved significantly. Nevertheless, they still suggest that the modified model does not fit the dataset well (Schermelleh-Engel et al., 2003).

Teachers' Training in Language Assessment

The second research question aims at uncovering how much formal training teachers received in four dimensions of classroom-based language assessment literacy. Figure 4 (see appendices) presents the mean of each item of the questionnaire. The first dimension, technical skills, involves six items (ELTQ1 to ELTQ6). The six items deal with preparing classroom tests, using ready-made tests, giving feedback, using self- and peer-assessment, informal assessment, and portfolio assessment. The mean of all six items but one (ELTQ6) is above 2.5, suggesting that the participants' training in technical skills is between moderate (a semester) and advanced (a year). The second dimension, scores and decision making, comprises three items that deal with giving grades, placing students, and awarding certificates. While the mean of giving grades is above 2.5, placing students and awarding certificates is below this figure, suggesting that teachers might have received more formal training in giving grades than placing students and awarding certificates. This is unsurprising because classrooms are barely streamed, so teachers hardly conduct placement tests in Mozambique. Also, the Mozambican education system tends to emphasise grading. All in all, the training received by the participants in this dimension also seems to lie between moderate and advanced.

Language pedagogy involves four items that concentrate on testing receptive skills, productive skills, grammar/vocabulary, and integrated skills. The mean suggests that teachers might have received more formal training in this dimension of classroom-based language assessment than others. The mean of testing productive skills and testing grammar/vocabulary is 3 and 3.03, suggesting that the respondents received advanced training in these aspects. This is not surprising given that language tests tend to concentrate on vocabulary, grammar, and reading. The mean of the other two items is above 2.5, suggesting that the training received lies between moderate and advanced. The last dimension, principles of language assessment, focuses on reliability, validity, and statistics. The mean for reliability and validity is above 2.5, while for using statistics is below this figure, suggesting that the received training in this dimension sits between moderate to advanced. Overall, the formal training received by the respondents in the four dimensions of classroom-based language assessment is between moderate and advanced.

Teachers' Self-Perceived Language Assessment Literacy

The third research question focuses on the teachers' self-perceived level of LAL. Figure 5 (see appendices) illustrates the mean of each item of the questionnaire. The mean of each item of technical skills (ELTQ1 to ELTQ6) ranges from 2.38 to 2.99, suggesting that teachers' self-perceived level of LAL sits between functional literacy and procedural and conceptual literacy. Following Taylor's (2013) LAL Profile for classroom teachers, it seems that participants have not attained the recommended level of LAL, which is procedural and conceptual literacy. Scores and decision making comprises three items (ELTQ7 to ELTQ10). The mean of the items ranges between 2.21 and 2.96. The perceived level of LAL in scores and decision making is between functional literacy and procedural and conceptual literacy, which is slightly above the recommended (Taylor, 2013).

Language pedagogy involves four items (ELTQ11 to ELTQ14). The teachers' perceived level of LAL is marginally higher in this dimension than others. While the perceived level of LAL is slightly above procedural and conceptual literacy level in testing receptive skills, productive skills, and grammar/vocabulary, it is marginally below this level in testing integrated language skills. One of the possible explanations is that assessments tend to be discrete point rather than integrative in Mozambique. Based on Taylor's (2013) LAL Profile for classroom teachers, the respondents' perceived level of LAL is below the recommended. Theories and principles encompass the last three items (ELTQ16 to ELTQ18). The mean of the items varies between 2.54 and 2.89, meaning that the teacher perceived level of LAL is between functional literacy and procedural and conceptual literacy, which is slightly above the recommended level (Taylor, 2013).

Overall, the study looked at the teachers' self-perceived level of language assessment literacy in four dimensions of classroom-based language assessment literacy. The findings suggest that the respondents have attained the recommended levels of LAL in two dimensions (scores and decision making and theories and principles), whereas they have not in the other two dimensions (technical skills and language pedagogy).

Conclusion and Limitations

This quantitative study explored English language teachers' language assessment literacy in Mozambique. The research study concentrated on teachers' experience with formal training in language assessment and their self-perceived level of classroom-based language assessment literacy. Regarding teachers' experience with formal training in language assessment, the findings suggest that the respondents received training that sits between moderate and advanced in the four dimensions of classroom-based assessment. Regarding the teachers' self-perceived level of classroom-based language assessment, the data suggest that the respondents have attained the recommended levels of classroom-based language assessment literacy in two out of the four dimensions of classroom-based language assessment literacy. These results seem to accentuate the need for more training in language assessment.

This research study is not devoid of limitations. Therefore, we recommend interpreting these findings with caution. The main limitation of the study derives from the sample size. Due to the COVID-19 pandemic, which led to school closures for more than a year, we could only sample ($n=72$) teachers from the initial target of 150. According to Kline (2015), sample sizes less than 100 are considered small and only suitable for simple models. Due to the sample size, the modifications made to Lan and Fan's (2019) model of classroom-based language assessment literacy should not be interpreted as an indication of the weakness of the model. The last limitation is related to the self-report methods of data collection. Self-report data may suffer from several problems, including socially desired responding and extreme responding (Robins et al., 2009).

References

- Boraie, D. (2018). Types of Assessment. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1–7). American Cancer Society. <https://doi.org/10.1002/9781118784235.eelt0350>
- Brindley, G. (2001). Language assessment and professional development. *Experimenting with Uncertainty: Essays in Honour of Alan Davies*, 11, 137–143.
- Brown, T. (2006). *Confirmatory Factor Analysis for Applied Research*. Guilford Press.
- Brown, T. (2015). *Confirmatory Factor Analysis for Applied Research, Second Edition*. Guilford Publications.

- Cizek, G. (2010). An introduction to formative assessment: History, characteristics, and challenges. In *Handbook of formative assessment* (pp. 15–29). Routledge.
- Creswell, J., & Creswell, J. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.
- Davies, A. (2008a). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/0265532208090156>
- Davies, A. (2008b). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/0265532208090156>
- Davison, C., & Leung, C. (2009). Current Issues in English Language Teacher-Based Assessment. *TESOL Quarterly*, 43(3), 393–415. JSTOR.
- Fulcher, G. (2012). Assessment Literacy for the Language Classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- Giraldo, F. (2018). Language Assessment Literacy: Implications for Language Teachers. *Profile Issues in Teachers' Professional Development*, 20(1), 179–195. <https://doi.org/10.15446/profile.v20n1.62089>
- Green, A. (2018). Assessment of Learning and Assessment for Learning. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1–6). American Cancer Society. <https://doi.org/10.1002/9781118784235.eelt0352>
- Harrington, D. (2008). *Confirmatory Factor Analysis*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195339888.001.0001>
- Hasselgreen, A., Carlsen, C., & Helness, H. (2004). European survey of language testing and assessment needs. Report: Part one—general findings. Retrieved March, 15, 2006.
- Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395–420. <https://doi.org/10.1177/0265532211428317>
- Hurley, A., Scandura, T., Schriesheim, C., Brannick, M., Seers, A., Vandenberg, R., & Williams, L. (1997). Exploratory and Confirmatory Factor Analysis: Guidelines, Issues, and Alternatives. *Journal of Organizational Behavior*, 18(6), 667–683.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402. <https://doi.org/10.1177/0265532208090158>
- INDE/MINED. (2010a). *Inglês, Programa da 8ª Classe*. DINAME, Moçambique.
- INDE/MINED. (2010b). *Inglês, Programa da 9ª Classe*. DINAME, Moçambique.
- INDE/MINED. (2010c). *Inglês, Programa da 10ª Classe*. DINAME, Moçambique.
- INDE/MINED. (2010d). *Inglês, Programa da 11ª Classe*. DINAME, Moçambique.
- INDE/MINED. (2010e). *Inglês, Programa da 12ª Classe*. DINAME, Moçambique.
- Johnson, R. B., & Christensen, L. (2016). *Educational Research: Quantitative, Qualitative, and Mixed Approaches*. SAGE Publications.
- Kline, R. (2015). *Principles and Practice of Structural Equation Modeling*, Fourth Edition. Guilford Publications.
- Kremmel, B., & Harding, L. (2020). Towards a Comprehensive, Empirical Model of Language Assessment Literacy across Stakeholder Groups: Developing the Language Assessment Literacy Survey. *Language Assessment Quarterly*, 17(1), 100–120. <https://doi.org/10.1080/15434303.2019.1674855>

- Kvasova, O., & Kavvytska, T. (2014). The assessment competence of university foreign language teachers: A Ukrainian perspective. *Language Learning in Higher Education*, 4(1), 159–177.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197. <https://doi.org/10.1177/0265532214554321>
- Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System*, 81, 78–89. <https://doi.org/10.1016/j.system.2019.01.006>
- Lan, C., & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation*, 61, 112–122. <https://doi.org/10.1016/j.stueduc.2019.03.003>
- Levi, T., & Inbar-Lourie, O. (2020). Assessment Literacy or Language Assessment Literacy: Learning from the Teachers. *Language Assessment Quarterly*, 17(2), 168–182. <https://doi.org/10.1080/15434303.2019.1692347>
- Mathew, R., & Poehner, M. (2013). Monitoring Progress in the Classroom. In *The Companion to Language Assessment* (pp. 631–645). American Cancer Society. <https://doi.org/10.1002/9781118411360.wbcla073>
- Migliacci, N. (2018). Classroom Testing. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1–7). American Cancer Society. <https://doi.org/10.1002/9781118784235.eelt0374>
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381–402. <https://doi.org/10.1177/0265532213480337>
- Popham, W. (2008). *Transformative Assessment*. ASCD.
- Popham, W. (2011). *Transformative Assessment in Action: An Inside Look at Applying the Process*. ASCD.
- Popham, W. (2014). The right test for the wrong reason. *Phi Delta Kappan*, 96(1), 46–52. <https://doi.org/10.1177/0031721714547862>
- Popham, W. (2018). *Classroom assessment: What teachers need to know* (8th Edition). Pearson.
- Purpura, J. (2016). Second and Foreign Language Assessment. *The Modern Language Journal*, 100(S1), 190–208. <https://doi.org/10.1111/modl.12308>
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249–258. <https://doi.org/10.1191/0265532204lt283ed>
- Rea-Dickins, P. (2007). Classroom-Based Assessment: Possibilities and Pitfalls. In J. Cummins & C. Davison (Eds.), *International Handbook of English Language Teaching* (pp. 505–520). Springer US. https://doi.org/10.1007/978-0-387-46301-8_36
- Rea-Dickins, P. (2008). Classroom-based language assessment. *Encyclopedia of Language and Education*, 7, 257–272.
- Robins, R., Fraley, R., & Krueger, R. (2009). *Handbook of Research Methods in Personality Psychology*. Guilford Press.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.

- Stabler-Havener, M. (2018). Defining, conceptualising, problematising, and assessing language teacher assessment literacy. *Studies in Applied Linguistics and TESOL*, 18(1).
- Stiggins, R. (2014). Improve assessment literacy outside of schools too. *The Phi Delta Kappan*, 96(2), 67–72. JSTOR.
- Stiggins, R. (1991). Assessment Literacy. *The Phi Delta Kappan*, 72(7), 534–539. JSTOR.
- Stiggins, R. (2002). Assessment Crisis: The Absence of Assessment for Learning. *Phi Delta Kappan*, 83(10), 758–765. <https://doi.org/10.1177/003172170208301010>
- Stoyhoff, S. (2012). Looking backward and forward at classroom-based language assessment. *ELT Journal*, 66(4), 523–532. <https://doi.org/10.1093/elt/ccs041>
- Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9(1), 1.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412. <https://doi.org/10.1177/0265532213480338>
- Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, 6(1), 41–63.
- Vogt, K., & Tsagari, D. (2014). Assessment Literacy of Foreign Language Teachers: Findings of a European Study. *Language Assessment Quarterly*, 11(4), 374–402. <https://doi.org/10.1080/15434303.2014.960046>
- William, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. *Handbook of Formative Assessment*, 18–40.
- Xu, Y., & Brown, G. (2017). *University English Teacher Assessment Literacy: A survey-test report from China*. <https://researchspace.auckland.ac.nz/handle/2292/34245>

Author's address

Felizardo Muianga

Department of English Language and Literature, Faculty of Education, Masaryk University in Brno

Žerotínovo nám. 617/9, 601 77 Brno

452888@mail.muni.cz

Appendices

Table 2: English language teacher questionnaire (ELTQ+ number of the item in the questionnaire). Skewness and Kurtosis (n =72)

<i>Code and Item</i>	<i>Skewness</i>	<i>Kurtosis</i>
Technical skills		
ELTQ1 Preparing classroom tests	-.570	.203
ELTQ2 Using ready-made tests	-.042	-.625
ELTQ3 Giving feedback based on	-.818	.216

assessment

ELTQ4 Using self-/Peer-assessment	-.506	-.253
ELTQ5 Using informal assessment	.395	-.512
ELTQ6 Using language Portfolio	.183	-.169

or Portfolio

Scores and Decision Making

ELTQ7 Giving grades Scores and decision making	-1.011	.940
ELTQ8 Finding out what needs to be taught/learned	-.485	-.743
ELTQ9 Placing students	-.369	-.677
ELTQ10 Awarding final certificates	.421	-.626

Language pedagogy

ELTQ11 Testing receptive skills (reading/listening)	-.542	-.515
ELTQ12 Testing productive skills (writing/speaking)	-.671	-.298
ETQ13 Testing grammar/vocabulary	-.645	-.338
ELTQ14 Testing integrated language skills	.076	-.564
ELTQ15 Testing aspects of culture	-.381	-.823

Theories and principles

ELTQ16 Establishing Reliability	-.534	-.373
ELTQ 17 Establishing Validity	-.592	-.097
ELTQ 18 Using statistics	-.336	-.801

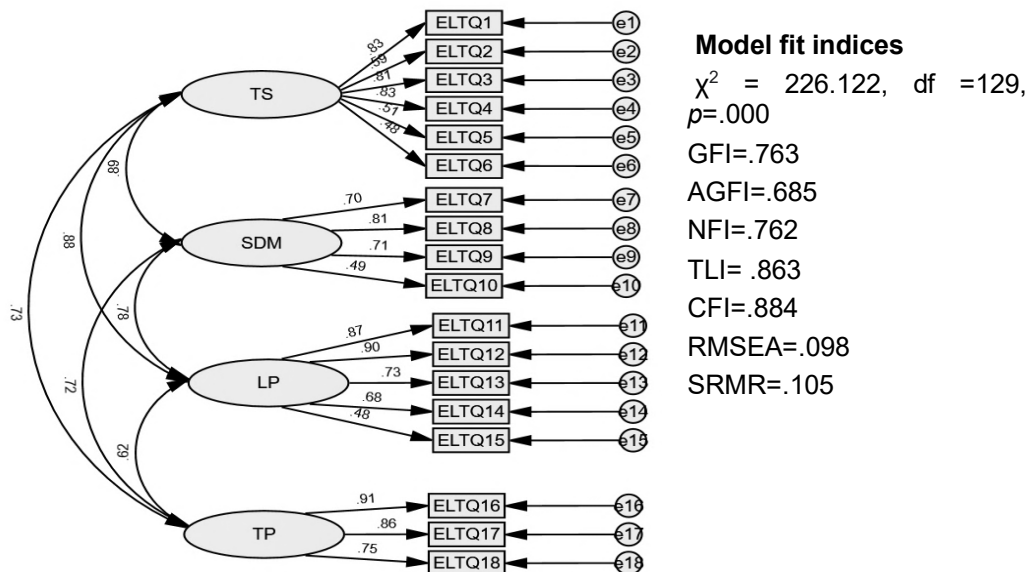


Figure 2: Results of confirmatory factor analysis of the ELTQ. Standardised Estimates and Model Fit Indices (n=72)

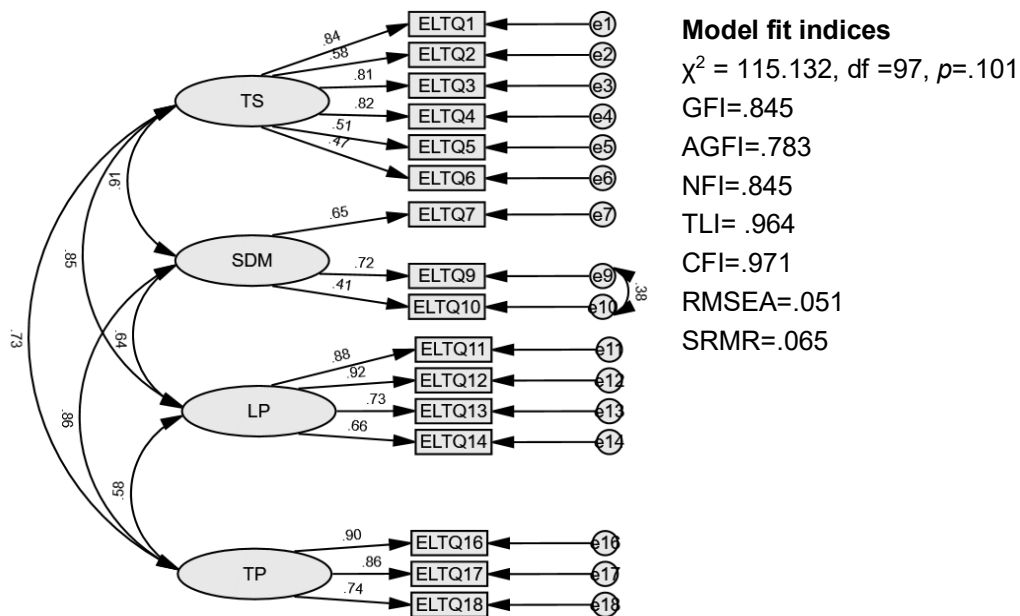


Figure 3: Modified Model. Standardised Estimates and Model Fit Indices (n=72)

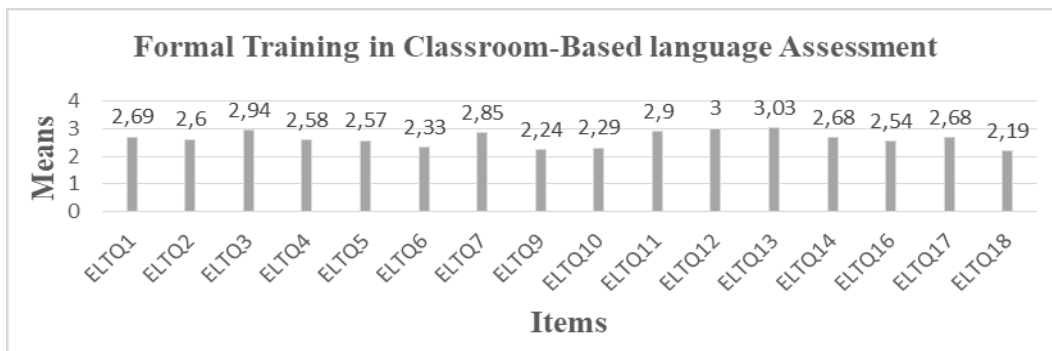


Figure 4

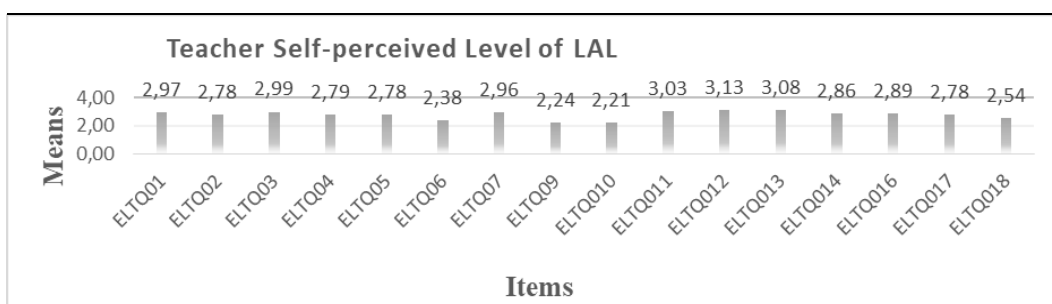


Figure 5

English Language Teacher Questionnaire (ELTQ)

Part I

Socio-demographic Characteristics of the Participants

1. Age _____

2. Gender

Male Female

3. Years of Experience _____

4. Highest Academic Qualifications

Diploma Bachelor's degree master's degree PhD

5. University Attended

UEM UP Other universities

6. School Location

Maputo Cidade

Maputo Provincia

Part II
2.1. Using the scale below, indicate how much formal training you received on each aspect of language assessment/testing

0- No training at all

1- Little training (few days/ a month)

2- Moderate training (a semester)

3- Advance training (a year)

4-Very advanced training (more than a year)

Items	0	1	2	3	4
ELTQ1. Preparing classroom tests.					
ELTQ2. Using ready-made tests from textbook packages or from other sources.					
ELTQ3. Giving feedback to students based on information from tests/assessment.					
ELTQ4. Using self- or peer-assessment.					
ELTQ5. Using informal, continuous, non-test type of assessment.					
ELTQ6. Using the Language Portfolio, an adaptation of it or some other portfolio.					
ELTQ7. Giving grades.					
ELTQ8. Finding out what needs to be taught/learned.					
ELTQ9. Placing students onto courses, programs, etc.					
ELTQ10. Awarding final certificates (from school/program; local, regional, or national level).					
ELTQ11. Testing/Assessing Receptive skills (reading/listening).					
ELTQ12. Testing/Assessing Productive skills (speaking/writing).					
ELTQ13. Testing/Assessing Microlinguistic aspects (grammar/vocabulary).					

ELTQ14. Testing/Assessing Integrated language skills.					
ELTQ15. Testing/Assessing Aspects of culture.					
ELTQ16. Establishing reliability of tests/assessment.					
ELTQ17. Establishing validity of tests/assessment.					
ELTQ18. Using statistics to study the quality of tests/assessment.					

2.2. Using the scale below, indicate your level of knowledge on each aspect. (0 is the lowest and 4 is the highest).

0-Illiteracy (the individual is not familiar with concepts and methods to language assessment and testing).

1-Nominal literacy (the individual understands that a term is related to language assessment and testing but may indicate a misconception).

2-Functional literacy (the individual has a good understanding of basic terms and concepts related to language assessment and testing).

3-Procedural and conceptual literacy (the individual understands the important concepts related to language assessment and testing and can put their knowledge into practice).

4-Multidimensional literacy (the individual's knowledge about language assessment and testing extends beyond ordinary concepts including philosophical, historical and social dimension of assessment).

Items	0	1	2	3	4
ELTQ1. Preparing classroom tests.					
ELTQ2. Using ready-made tests from textbook packages or from other sources.					
ELTQ3. Giving feedback to students based on information from tests/assessment.					
ELTQ4. Using self- or peer-assessment.					

ELTQ5. Using informal, continuous, non-test type of assessment.					
ELTQ6. Using the Language Portfolio, an adaptation of it or some other portfolio.					
ELTQ7. Giving grades.					
ELTQ8. Finding out what needs to be taught/learned.					
ELTQ9. Placing students onto courses, programs, etc.					
ELTQ10. Awarding final certificates (from school/program; local, regional or national level).					
ELTQ11. Testing/Assessing Receptive skills (reading/listening).					
ELTQ12. Testing/Assessing Productive skills (speaking/writing).					
ELTQ13. Testing/Assessing Microlinguistic aspects (grammar/vocabulary).					
ELTQ14. Testing/Assessing Integrated language skills.					
ELTQ15. Testing/Assessing Aspects of culture.					
ELTQ16. Establishing reliability of tests/assessment.					
ELTQ17. Establishing validity of tests/assessment.					
ELTQ18. Using statistics to study the quality of tests/assessment.					