

# Examining Language Assessment Literacy for Saudi Pre-service EFL Teachers

Faiza Abdalla Elhussien<sup>1</sup>, & Safaa Moustafa Khalil<sup>2</sup>

<sup>1</sup>Department of English, College of Education, Majmaah University, Al-Majmaah, 11952, Saudi Arabia & Omdurman Islamic University, Sudan

<sup>2</sup>Department of English, College of Education, Majmaah University, Al-Majmaah, 11952, Saudi Arabia & Sadat Academy for Management Sciences, Egypt

Correspondence: Faiza Abdalla Elhussien, Department of English, College of Education, Majmaah University, Al-Majmaah, 11952, Saudi Arabia & Omdurman Islamic University, Sudan.

Received: April 12, 2023      Accepted: May 24, 2023      Online Published: June 14, 2023

doi:10.5430/wjel.v13n6p431

URL: <https://doi.org/10.5430/wjel.v13n6p431>

## Abstract

Teacher assessment skills include knowledge of assessment methods and techniques, the ability to interpret assessment results, and the ability to use assessment to tailor teaching strategies to student needs. The present study examines non-native Teaching English as a Foreign Language (TEFL) pre-service teachers' awareness of students' assessment. The study also investigates whether studying a course in assessment significantly impacts pre-service teachers' assessment literacy. To answer the research questions, an online test was administered to 52 pre-service teachers who are seniors at Majmaah University. The collected data were analyzed descriptively and inferentially with the aid of IBM SPSS version 22. The results revealed that non-native pre-service TEFL teachers have a low level of assessment literacy. Despite the overall weak performance in the test, the participants showed strength in some of the standards for teachers' competence. They revealed a high capability of choosing and developing appropriate assessment methods. On the other hand, pre-service teachers possessed little awareness about their ability to implement students' test results to adjust the curriculum or make better instructional decisions. Based on the study results, institutions may consider adding more assessment courses to the TEFL study plan or adopting some of the existing curricula and extending it to include all the assessment skills needed for creating qualified teachers. Designing a careful professional development training program for EFL faculty members is another important recommendation of the present study. Finally, future research in the field is needed for planning and implementing training programs to improve teachers' assessment skills.

**Keywords:** assessment literacy, assessments' criteria, standards for teacher competence, pre-service TEFL teachers, curriculum development, course learning outcomes (CLOS)

## 1. Introduction

Language assessment literacy, referring to the ability to design, conduct, and interpret assessments effectively, is an essential component of language education. As the importance of language proficiency in today's globalized world continues to grow, the evaluation of language competency has become increasingly important. However, the quality and reliability of language assessments depend heavily on the competence of those who design, administer, and interpret the tests. Language assessment literacy has thus emerged as a crucial area of focus in language education, to equip language educators with the necessary knowledge, skills, and tools to develop and conduct effective language assessments. These issues have been addressed by metrology experts (Cronbach, 1960; Messick, 1984; Thorndike, 1904); and (Plake et al., 1993) and Assessing Language Inventory frameworks (ALI) by Mertler (2003) and Mertler and Campbell (2005). A recent study by Wise (2020) examined the impact of test participation on the Program for International Student Assessment (PISA) to provide essential knowledge contributions to the power of testing measures and how students interact with test results. It assessed six key ideas and their implications for assessment programs over the past 15 years. The impact of test results on both students and teachers which consequently affects teaching practices was revealed by Gashaye (2020).

### 1.1 Statement of the Problem

The present investigation endeavors to assess the efficacy of pre-service educators regarding identifying their strengths and limitations in the development of assessment measures. This is undertaken to ensure their aptitude in crucial areas such as abilities, knowledge, skills, as well as literacy values, specifically before their graduation.

### 1.2 The Purposes

The research problem deserves to be investigated because it aims primarily to raise the level of awareness among pre-service educators about exploring the fundamental categories of evaluations, their suitable application, and the process of utilizing evaluation results to ascertain student progress. The study draws inspiration from Webb's (2002) work on the subject matter. Furthermore, the investigation

endeavors to appraise the proficiency of the participants in the various procedural stages required to create assessments that cater to measuring language learning outcomes.

### 1.3 Questions

To tackle the aforementioned concerns, the research seeks to respond to the subsequent questions:

1. What are the assessment criteria that EFL teachers are aware of and unaware of?
2. To what extent does enrolment in an English language testing course significantly affect pre-service EFL teachers' assessment literacy?

### 1.4 The Importance of the Study

This research has a positive impact on the development of educational assessment. Furthermore, the importance of this research lies in its role to boost EFL teachers' performance who need a range of assessment skills and knowledge to effectively improve their teaching quality, fairness, and equity. It also gives insights to curriculum designers to enhance course specifications including assessment tools.

### 1.5 Relevant Scholarship

#### 1.5.1 Assessment Literacy

Assessment is generally considered a systematic basis for inferring student progress and success and gathering data about student learning. Its goal is to obtain a piece of evidence of student learning (Cheng and Fox, 2017). Scores depend on certain specifications that define the content of a test and describe how the content indicates that an item should appear in the test evaluation (Kenji & Lee, 2009). There is an increasing need for the language evaluator profession to precisely examine what literacy assessment means and articulate the evaluator's role in creating assessment educational materials and programs. In addition, they provide new assessment tools of language to meet teachers' changing needs for a new era (Fulcher, 2012)

Giraldo (2018) stated that the field of applied linguistics that addresses language assessment literacy considers examining the knowledge, skills, and principles needed for assessment. In this field, the two most important concerns are the definition of language assessment literacy and its different interpretations among teachers and students. White (2015) believed that assessment should be viewed as a training tool and a reliable tool to ensure that training has been conducted. He added: "For educators, the challenge is to use pre-graded assessment to deepen learning and disclose developmental skills to students" (p. 18). Educators need to develop and select assessments to fit the context that reflects specific performance goals. Therefore, the need for the profession of language testing is highly growing.

#### 1.5.2 Assessments' Criteria

##### *Reliability and Validity of Assessment*

Measuring reliability and validity is important in assessment because they determine how accurately a particular question can be answered. Formal and summative assessments are more valuable and trustworthy because those who develop and implement them will know what they are doing, and regarding the validity, he stated that validity determines how accurately a particular question can be answered. Formal and summative assessments are more valuable and trustworthy (Stiggins, 2000). Validity is the most fundamental concept in testing, as it is the essential factor that leads to meaningful conclusions in the test, they added Validity is the most basic and broad in testing. Without validity, the test and the conclusions drawn from it are meaningless (Hubley & Zumbo, 1996). Messick (1989) regarded validity as a tool for judgment that integrates empirical evidence and theoretical justification to support reliable conclusions.

##### *Fairness & Ethics*

Fairness is usually practiced to ensure consistency and improve the reliability of results and the validity of conclusions (Kunnan 2003). Assessments must provide equal opportunities for all students, an equal chance of success, and respond to individual and unique needs. Therefore, assessment tools should always include a wide range of parameters to properly assess each student in the class and allow them to bring out the best in themselves. He must also introduce a fair and appropriate system, equitable and free from the prejudices of the prevailing culture. When an assessment system is proper, it positively impacts students, minimizes adverse impacts to the extreme (Smith et al., 2004), and progressively leads to social equity in the classroom.

##### *Assessment Standards for Teacher Competence (TALQ & ALI Frameworks)*

It represents teachers' awareness and skills in assessing students' performance. Two well-known instruments were developed and widely used; the "Teacher Assessment Literacy Questionnaire" (TALQ) (Plake et al. 1993) and the Assessment Literacy Inventory (ALI) (Mertler 2003 & Mertler and Campbell 2005). Both tools were generated to comply with "Standards for Teacher Competence in the Educational Assessment of Students." These standards are coined as a conceptual framework for what teachers should consider when assessing students (Webb, 2002):

*Standard 1.* (Teachers' skill in choosing suitable assessment methods)

*Standard 2.* (Teachers' skill in developing suitable assessment methods)

*Standard 3.* (Teachers' skill in conducting, scoring, and explaining the results of the assessment method)

*Standard 4.* (Teachers' skill in employing assessment results in instructional and administrative improvement decisions)

*Standard 5.* (Teachers' skill in developing accurate grading systems)

*Standard 6.* (Teachers' skill in identifying immoral, illegal, and other unethical practices regarding assessment information)

### 1.5.3 Classroom Evaluation

Classroom assessment should go beyond simply measuring a student's knowledge and instead of supporting their learning by providing new opportunities for formal testing (Poehner et al., 2017). It is claimed to be a collaborative effort to measure what learners can do independently and their level of performance after giving support (Lantolf and Pöhner, 2004). Therefore, it is always aimed at promoting student growth and collaboration. According to Angelo and Cross (1993), in-class assessments help individual university professors get helpful feedback on what and how good students are. Educators use this information to guide learning and enable students to learn more effectively and efficiently. Walker (2012) believed that some of the challenges in higher education are to ensure that assessments are meaningful, and feedback is timely to facilitate learning. Another challenge is how to provide teachers with quick and impartial feedback on their effectiveness.

### 1.6 Related Studies

In a recent study by Umer, Zaharia, and Alshara (2018), they examined how Saudi EFL teachers develop assessment tasks, how those tasks impact student learning, and how well the teachers' assessment practices align with recommended practices. The researchers analyzed summative assessment tasks and conducted a student survey with both open and closed-ended questions. Quantitative analysis was used in addition to survey responses. The results revealed that teachers' assessment tasks did not align well with course learning outcomes, with many tasks not assessing higher-order learning outcomes. Most tasks comprised selected-response questions, leading students to rely on memorization strategies. The study suggests that professional development for Saudi university teachers, with an emphasis on assessment literacy, must be implemented for the efficacy of the (higher) education system.

Another study done by Alhareth and Al Dighrir (2014) examined the assessment criteria used in Saudi's higher education sector, they discovered that the assessment processes that are implemented at Saudi's secondary and university levels do not evaluate student abilities. The study provides some standardized forms of evaluation to be employed. They recommend that: a) teacher training has to be reviewed, informed, and standardized in Saudi Arabia, b) rethinking may be needed to develop attitudes both by the teacher and a continuous assessment system to identify the differences between attitudes and practices.

Alkharusi (2011) employed the Teacher Assessment Literacy Questionnaire (TALQ) developed by Plake and Impara (1993). The questionnaire is composed of 35 multiple-choice items in an attempt to examine 359 preservice teachers' knowledge and skills in educational evaluation in Oman. Item analysis showed that the TALQ effectively discriminated between the different preservice teachers' assessment literacy levels. Furthermore, it had an excellent internal consistency reliability coefficient. The confirmatory factor and regression analysis are collectively and significantly correlated with evaluation ability in educational measurements. There was also a significant positive correlation between the TALQ score and the academic performance score. These results show that TALQ provides a valid and reliable interpretation of the evaluation.

However, further research was done by Herrera and Macias (2015) to raise awareness of the relevance of Assessment Literacy (AL) in English education. They reviewed some definitions for developing AL with future teachers and in-training teachers for students' and institutions' sake. The findings showed that EFL teachers in Colombia usually focus on the accuracy of exams to evaluate the student's learning without realizing that the accuracy of exams is only a part of the overall picture of language evaluation. Therefore, they recommend that education programs include alternatives to help EFL teachers develop a wide range of practical knowledge of evaluation skills. At this point, it should be emphasized that training in assessment literacy should not be limited to a single course in teachers' education programs.

In a recent study in China, Xu & Brown (2017) administered an adaptive version of the Teacher Evaluation Literacy Questionnaire to Chinese university English teachers (n = 891) to explore their levels of assessment literacy. The results showed the base level of AL in a particular dimension with limited demographic impact. The discussion focuses on the verification of AL equipment, the causes of the AL's capacity decline, and the significant factors that affected AL. This study concludes with a discussion of contextual AL preparation and its impact on teacher evaluation training principles, policies, and practices. This finding was supported by Ping (2013) who believed that quality training could better qualify pre-service teachers.

To explore the challenges and prospects of foreign language teacher exams at Ukrainian universities, Vogt & Tsugari (2014) duplicated a European study on the need for language testing and assessment and designed a survey to identify particular strengths and weaknesses of their evaluation. The results are similar to those in Europe, with slight differences. This is mainly due to the newly introduced evaluation format. According to their research, foreign language teachers at Ukrainian universities are very test-oriented and ready to do excellent language testing and evaluation. Lessons learned from the data can be used to develop a curriculum for the education and training of teachers in this area.

Further study on Assessment Literacy is that of Deluca, LaPointe-McEwan & Luhanga (2014). They carried out a systematic review study over vast areas as the aim of the study is twofold. First, the study aimed to explore evaluation literacy requirements from 5 English-speaking countries to recognize shifts with inside the evaluation panorama over the years rough out areas. Secondly, researchers examined distinguished evaluation literacy measures that evolved after 1990 through a magic evaluation of (n=15) evaluation requirements and an exam of (n=8) evaluation literacy measures. Findings imply significant shifts in conditions over the years. However, most measures remain

primarily based on early concepts of evaluation literacy. Additionally, results serve to outline more than one dimension of evaluation literacy and yield critical suggestions for measuring it.

According to Hailaya et al. (2014), teachers utilize assessments to identify and enhance their students' learning, underscoring the significance of having proficient evaluation skills. The Evaluation Literacy Inventory (Mertler and Campbell, 2005) is one of the instruments employed in studying these skills. After being administered to 582 teaching staff and scrutinized using traditional statistical methods, the Evaluation Literacy Inventory was found to be effective in measuring teacher proficiency in educational assessment. However, the evaluation of students still poses a challenge for new psychometric techniques. The article concludes by discussing the implications for tool development, educational evaluation research, policies and practices, and teacher professional development.

Kemmel & Harding (2020) developed a Language Evaluation Literacy Survey (LEL) devoted to all researchers. They believed that many scholars had proposed different theoretical models that represent the perspective of the linguistic evaluation researcher rather than the stakeholders themselves. Therefore, they conducted an empirical study that started with a discussion of the review and pre-testing phases by research and development experts. Then, they displayed the results of a factor analysis of data collected from a large cohort ( $n = 1086$ ). The results of the study shed light on the dimensions of LEL and create a needs profile across these dimensions.

## 2. Methods

### 2.1 Participants Characteristics

The study was conducted with a purposive group of undergraduate English Major pre-service teachers, enrolled in the "English Language Testing course" in 2022. The groups included ( $n=52$ ) females at Zulfi College of Education Majmaah University, Saudi Arabia. They were senior students majoring in Teaching English as a foreign language (TEFL).

### 2.2 Sampling Procedures

The study covered the complete set of knowledge necessary for how pre-service teachers could test, assess, and avoid the harmful backwash of testing and unethical phases through the online tests' items. For these goals, the online test items have been coded using themes listed in the "English Language Testing" course's specifications besides (TALQ) and (ALI) frameworks. The test was administered through a link sent to the preservice teachers' Telegram and WhatsApp groups. The test was conducted in a mono attempt. The instructions were written for all the questions, and the responses were automatically saved. The participants were asked to indicate their level of understanding of the inventory questions by choosing the best choice from the multiple-choice (see Appendix). The data were carefully examined using both descriptive and inferential statistical analyses to understand its significance, recognize any trends or correlations, and isolate the pertinent information that relates to the research query.

#### 2.2.1 The Data Collection Tool

The current study used an assessment literacy test adopted from two commonly used assessment literacy surveys TALQ (Plake, 1993) and ALI (Mertler & Campbell, 2005). The test consists of six scenarios with 24 questions, covering six of the teachers' standards. However, the original structure of the TALQ and ALI elements methods was retained to maintain equipment integrity while the parts were fitted. The study chose an online testing tool that employed different operations and procedures to prepare representative assessments to assess English language skills using both standard and specific test techniques. Moreover, examine how pre-service teachers use different kinds of testing and how to construct suitable assessment criteria (rubrics) for assessing standard and specific test techniques.

#### 2.2.2 Research Design

To answer the study questions, decide on its objectives, and draw more precise conclusions, the study adopted a case study research design to gain a rich, detailed, understanding of a specific context and phenomenon (Könür, 2018). The e-test was conducted with the purposive group of undergraduate English Major pre-service teachers, enrolled in the "English Language Testing course" in 2022.

### 2.3 Limitations of the Study

The researchers explored the strengths and weaknesses of this study in terms of the accessibility of information, time constraints, and conflict of bias. It was observed that the study was limited to female participants, although it has not led to any statistical limitations, or poor data management that could lead to deviations in the accuracy of the data compromised or limitations in generalizability. Therefore, it is advisable to conduct future studies including males.

## 3. Results

Data analysis encompassed descriptive and inferential analyses. Descriptive research, including frequencies and reliability tests, was conducted at three levels: the level of the items, the level of the six standards, and the level of total test scores. Each standard level was developed by grouping the relevant test items (based on *the Standards for Teacher Competence*). Inferential data analysis involved a one-sample t-test comparing pre-service teachers' mean scores to the proposed level of awareness in the actual test. All statistical analyses were conducted using IBM SPSS version 22 and presented to answer research questions.

### 3.1 RQ1: What Assessment Criteria Are EFL Pre-Service Most and Least Aware of?

Data analysis revealed the instruments' high level of internal consistency reliability,  $\alpha = .85$ . Results showed that the overall average score was  $M = 12/24$  (50%). Detailed examination of the results indicated that only 32.7% ( $n = 17$ ) of the participants scored  $\geq 14$ , whereas 67.3%

of them (n = 35) scored less than 14. This reflected a low level of assessment literacy among non-native EFL pre-service teachers.

When comparing the means of the six standards, results indicated the gaps in the participants' literacy level in each measure, as displayed in Table 1. Participants were found to score the highest means in Standard 1 (i.e., *The skill of choosing suitable assessment methods*) as the standard overall mean was  $M = 3.87$ . The following highest means ( $M = 2.87$ ) was that of standard 2, which is concerned with teachers' skills in developing assessment methods suitable for instructional decisions. Standard 6, which highlights teachers' ethics regarding assessment methods, came in third place among the highest means, with a total score of 1.83.

Table 1. Total mean for each of the Teacher Competence Standards.

	N	Minimum	Maximum	Mean	Std. Deviation
Standard 1	52	1.00	7.00	3.87	1.69230
Standard 2	52	00	6.00	2.87	1.83676
Standard 3	52	00	3.00	1.29	1.09072
Standard 4	52	.00	2.00	.61	.52966
Standard 5	52	00	3.00	1.23	.7577
Standard 6	52	1.00	3.00	1.83	.83363
Valid N (listwise)	52				

When comparing participants' low total scores, standards 3 and 5 scored very similar low means ( $M = 1.29$  &  $1.23$  respectively). Participants' lowest mean ( $M = 0.62$ ) was found in Standard 4 (e.g., *The skill of employing assessment results in making industrial and administrative improvement decisions*).

On the individual item level, the most correctly answered two items belong to Standard 1 and Standard 6, as the percentages of the correct responses were 78.8% and 76.9%, respectively. In contrast, the items with the lowest rate of correct answers were found in Standard 4 as the percentage was 0.04%, and in Standard 5 with the rate of 17.3%.

3.2 RQ2: To What Extent Does Enrollment in an English Language Testing Course Significantly Affect Pre-Service EFL Teachers' Assessment Literacy?

To answer this question, a One-sample t-test was used to compare participants' total scores in the assessment literacy test with the proposed score of 14 (which constitutes 60% of the entire test score).

Table 2 One-sample t-test of the participants' total scores in the assessment literacy test

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Total Score	-436,031	51	.000	-13,51282	-13,5750	-13,4506

As displayed in Table 2, results revealed a significant difference between the participants' total scores in the test and the test value as  $p = .003 < .01$ . All participants were seniors who had just passed a 42-hour course about English language testing. Therefore, Results indicated that studying such courses as undergraduate students was not helpful enough to elevate preservice teachers' awareness of language assessment.

4. Discussion and Conclusion

Teachers' assessment literacy test reflects their knowledge about assessment methods and techniques. They also reveal teachers' ability to implement assessment methods appropriately and ethically to get results that are critical to making instructional and administrative decisions. (Mertler, 2005)

The significant finding of the current study is that most of the non-native pre-service TEFL teachers demonstrated weakness in their level of assessment literacy. The results of the study highlight the low level of assessment literacy among non-native EFL pre-service teachers. The finding that only 32.7% of the participants scored 14 or higher and 67.3% scored less than 14 is a cause for concern. This shows that the majority of the participants lack the necessary skills to use assessment as a tool for effective teaching.

Furthermore, the study revealed the voids in the participants' literacy level in each measure, with Standard 1 being the highest score and Standard 4 being the lowest. Standard 1 focuses on the skill of choosing suitable assessment methods, and Standard 4 is concerned with the skill of employing assessment results in making industrial and administrative improvement decisions. The third highest score was in Standard 6, which assesses the participants' ethics in assessment methods. This shows that the participants had some understanding of the importance of ethical practices in assessment methods.

The results suggest that there is a need for improving assessment literacy among non-native EFL pre-service teachers. This result coincides

with Latif's study (2021) which was conducted on 80 tertiary EFL teachers and concluded that they have a very limited assessment knowledge base. He also called for teachers' professional development programs that address the need for raising awareness in teachers' assessment literacy. Several other studies correlate with the same result (Alkharusi, 2011; Malone, 2013; Xu and Brown, 2017).

Moreover, the preservice TEFL teachers showed little awareness when tested about their ability to implement their students' test results to make improvement decisions and to adjust the curriculum to enhance students' performance, therefore, achieving better learning outcomes (Al-Bahlani, 2019, Xu and Brown, 2017). Accordingly, in his study on 33 teachers, Gashaye (2020) concluded that exams affect instructional decisions and teaching practices.

When assessing the impact of studying a language testing course on pre-service teachers' awareness of assessment, results revealed no significant effect. An interpretation of this finding can be ascribed to two factors the teacher level, and the educational institutions level that impact assessment literacy as summarized by Al-Bahlani (2019). After investigating literature concerned with teacher assessment literacy, he summed up those factors based on three levels: the teacher level, the educational institutions level and the assessment literacy research level. This finding relates to the first two factors as it depends on the students themselves and the compulsory course they studied as TEFL undergraduates. The nature of the course content they studied as seniors are targeted only toward designing and implementing EFL tests with no focus on the importance of reflecting on assessment results when planning or making instructional decisions.

Overall, the findings suggest that EFL pre-service teachers need to be offered training in the areas where they scored poorly to improve their assessment literacy skills. Additionally, the study discovered individual items within the standards where strong or weak performance was demonstrated, highlighting areas where additional focus is necessary for pre-service teachers to enhance their assessment practices. Training and workshops should be provided to help them develop the necessary skills to use assessment as an effective tool for teaching. Future research could explore the factors that contribute to low assessment literacy among pre-service teachers and the effectiveness of training programs in improving their assessment literacy.

This study will help both pre-service and in-service teachers in addition to educational institutions realize the skills non-native TEFL teachers need to elevate their awareness of students' assessment. Based on the study results, institutions may consider adding assessment courses to the TEFL study plan if they are not included. Institutions that already teach the course may think of adapting the existing curriculum and extending it to include all the assessment skills needed for equipping qualified teachers. Carefully designed professional development training programs for EFL teachers is another important recommendation of the present study. The programs should include skills in designing and implementing creative assessment tools and techniques that will in turn foster students' thinking abilities, particularly critical thinking skills (Bankole- Minaflinou, 2018)

This empirical research study has significant ramifications across multiple domains - theoretical, practical, and social. In terms of theoretical significance, this study contributes to the extant literature related to the assessment literacy of TEFL (Teaching English as a Foreign Language) instructors. Specifically, this study serves to raise awareness among non-native TEFL teachers, both pre-service and in-service, about assessment literacy. In practical terms, the results of this research may be used by educational institutions to re-evaluate their plans and teacher development programs, to cultivate highly skilled TEFL instructors. Additionally, the social implications of this study are noteworthy, as the findings contribute to improvements in the overall quality of pre-service teacher training, which in turn positively affects students' educational outcomes, as well as society more broadly. The potential implications of this research extend to the level of individual students, families, and society as a whole.

#### **Abbreviation and their Meaning:**

TEFL: Teaching English as a Foreign Language.

TALQ: Teacher Assessment Literacy Questionnaire

ALI: Assessment Literacy Inventory

#### **Acknowledgments**

The authors would like to thank the Deanship of Scientific Research at Majmaah University for supporting this work under Project Number No. R-2023-443

#### **References**

- Al Kharanai, H. (2011). An analysis of the internal and external structure of the teacher assessment literacy questionnaire. *The International Journal of Learning Annual Review*, 18(1), 515-528. <https://doi.org/10.18848/1447-9494/CGP/v18i01/47461>
- AlBahlani, S. M. (2019). *Assessment literacy: A study of EFL teachers' Assessment knowledge, perspectives and classroom behavior*. Unpublished doctoral dissertation. The University of Arizona, Arizona.
- Alhareth, Y., & Al Dighrir, I. (2014). The assessment process of pupils' learning in Saudi education system: A literature review. *American Journal of Educational Research*, 2(10), 883-891. <https://doi.org/10.12691/education-2-10-6>
- Angelo, T. A., & Cross, K. P. (1993). *Classroom Assessment Techniques: A Handbook for College Teachers* (2nd ed.). Jossey-Bass Publishers 94 San Francis.
- Bankole-Minaflinou, E. (2018). Promoting critical thinking skills in efl university students in Benin. *International Journal of English*

*Language and Literature Studies*, 8(1), 1-13. <https://doi.org/10.18488/journal.23.2019.81.1.13>

- Cronbach, L. J. (1984). *Essentials of Psychological Testing* (2nd ed.). Harper & Row. New York, USA.
- DeLuca, C., LaPointe-McEwan, D. & Luhanga, U. (2014). Teacher assessment literacy: a review of international standards and measures. *Educ Asse Eval Acc* 28, 251–272 (2016). <https://doi.org/10.1007/s11092-015-9233-6>
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9, 113-132. <https://doi.org/10.1080/15434303.2011.642041>
- Gashaye, S. (2020). The washback of the English National Examination (ENE) for grades 9 and 10 on teaching practices in Ethiopia: A case study of Debre Markos Secondary Schools. *International Journal of English Language and Literature Studies*, 9(2), 106-120. <https://doi.org/10.18488/journal.23.2020.92.106.120>
- 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>
- Hailaya, W., Alagumalai, S. & Ben, F. (2014). Examining the utility of Assessment Literacy Inventory and its portability to education systems in the Asia Pacific region. *Australian Journal of Education*, 58(3), 297-317. <https://doi.org/10.1177/0004944114542984>
- Herrera, L., & Macias, D. F. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, 17(2), 302-312. <https://doi.org/10.14483/udistrital.jour.calj.2015.2.a09>
- Hubley, A. M., & Zumbo, B. D. (1996). A dialectic on validity: where we have been and where we are going. *The Journal of General Psychology*, 123, 207-215. <https://doi.org/10.1080/00221309.1996.9921273>
- Impara, J. C., Plake, B. S., & Fager, J. J. (1993). Educational administrators' and teachers' knowledge of classroom assessment. *Journal of School Leadership*, 3, 510-521. <https://doi.org/10.1177/105268469300300504>
- Kenji, H., & Lee L. J. (2009). *Guidelines for the Assessment of English Language Learners*.
- Kömür, S. (2018). Preservice English teachers' assessment awareness: level of readiness for classroom practice. *Journal of Language Teaching and Learning*, 8(1), 109-121. Retrieved from <https://www.jltl.com.tr/index.php/jltl/article/view/60>
- 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), 100-12. <https://doi.org/10.1080/15434303.2019.1674855>
- Kunnan, A. J. (2003). *Test fairness*. In M. Milanovic & C. Weir (Eds.), *Europe language testing in a global context: Selected Papers from the ALTE conference in Barcelona*. Cambridge, England: Cambridge University Press.
- Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment: Bringing the past into the future. *Journal of Applied Linguistics*, 1, 49-74. <https://doi.org/10.1558/japl.1.1.49.55872>
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329-344. <https://doi.org/10.1177/0265532213480129>
- Mertens, D. M. (2010). *Research and Evaluation in Education and Psychology* (3rd ed.). Los Angeles, CA: SAGE Publications, Inc
- Mertler, C. A. (2003). *Pre-service versus in-service teachers' assessment literacy: does classroom experience make a difference?* In the Annual Meeting of the Mid-Western Educational Research Association, Columbus.
- Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the Assessment Literacy Inventory*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada. (ERIC Document Reproduction Service No. 490 355)
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13– 103). New York, NY: American Council on Education/Macmillan.
- Ping, W. (2013). Micro-Teaching: A Powerful Tool to Embedding the English Teacher Certification Testing in the Development of English Teaching Methodologies. *International Journal of English Language and Literature Studies*, 2(3), 163-175. Retrieved from <https://archive.aessweb.com/index.php/5019/article/view/695>
- Plake, B., Impara, J., & Fager, J. (1993). Assessment competencies of teachers: a national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-39. <https://doi.org/10.1111/j.1745-3992.1993.tb00548.x>
- Poehner, M. E., Davin, K. J., & Lantolf, J. P. (2017). Dynamic Assessment. In Shohamy, E., Lair, G., & May, S. (Eds.). *Language, Testing and Assessment*, 243-256. Switzerland: Springer. [https://doi.org/10.1007/978-3-319-02261-1\\_18](https://doi.org/10.1007/978-3-319-02261-1_18)
- Smith, M. E., Teemant, A., & Pinnegar, S. (2004). Principles and practices of sociocultural assessment: Foundations for effective strategies for linguistically diverse classrooms. *Multicultural Perspectives*, 6(2), 38-46. [https://doi.org/10.1207/s15327892mcp0602\\_8](https://doi.org/10.1207/s15327892mcp0602_8)
- Stiggins, R. (2000). *Specifications for a Performance-Based Assessment System for Teacher Preparation*. National Council for Accreditation of Teacher Education, Washington, D.C. Retrieved June 15, 2004, from

<http://www.ncate.org/resources/commissioned%20papers/stiggins.pdf>

- Umer, M., Zakaria, M. & Alshara, M. (2018). Investigating Saudi University EFL Teachers' Assessment Literacy: Theory and Practice. *International Journal of English Linguistic*, 8(3), 345-356. <https://doi.org/10.5539/ijel.v8n3p345>
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11, 374-402. <https://doi.org/10.1080/15434303.2014.960046>
- Walker, D. (2012). Classroom Assessment Techniques: An Assessment and Student Evaluation Method. *Creative Education*, 3, 903-907. <https://doi.org/10.4236/ce.2012.326136>
- Webb, N. L. (2002). *Assessment Literacy in a Standards-Based Urban Education Setting*. A paper presented at the American Educational Research Association Annual Meeting in New Orleans, Louisiana, April 1-5, 2002. Retrieved from Assessment Literacy in a Standards-Based Urban Education Setting (wceruw.org)
- White, E. (2015). Synergizing Formative and Summative Assessment of Presentation Slideshows. *Arab World English Journal (AWEJ)*, 6(3), 3-29. <https://doi.org/10.24093/awej/vol6no3.1>
- Wise, S. L. (2020). Six insights regarding test-taking disengagement. *Educational Research, and Evaluation*, 26(5-6), 328-338. <https://doi.org/10.1080/13803611.2021.1963942>
- Xu, Y., and Brown, G.T. L. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, 6(1), 133-156. <https://doi.org/10.58379/UZON5145>

### Appendix

The link to the e.test: <https://forms.gle/WCZNVx2bZzQMe6mX6>

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).