

RESEARCH

Open Access



Exploring tertiary EFL practitioners' knowledge base component of assessment literacy: implications for teacher professional development

Muhammad Wasim Latif 

Correspondence: ml463@exeter.ac.uk; mwaseemlatif21@gmail.com
Department of English, Jubail University College, Jubail Industrial City, Al-Jubail, Kingdom of Saudi Arabia

Abstract

Language teachers' assessment literacy has been a much debated subject in the educational arena recently. Teachers' assessment knowledge base and skills, important aspects of their assessment literacy, have been extensively investigated at the school level; however, the research on this subject in the tertiary EFL context remains underdeveloped. This study reports findings of tertiary EFL teachers' assessment literacy in terms of their assessment knowledge base in the context of Saudi Arabia. The study is informed by sociocultural theoretical background, with pragmatism as the philosophical underpinning. It uses an adapted instrument called Classroom Assessment Literacy Questionnaire, and the data are obtained from 80 questionnaire respondents. The statistical analysis of the data revealed that tertiary EFL teachers' current assessment knowledge base is very limited and not consistent with the contemporary trends and approaches to educational assessment in terms of their preparation and readiness for the continuously mounting challenges posed by classroom-based assessments. These results indicating gaps and inadequacies in tertiary EFL practitioners' assessment literacy have strong implications for teacher development in the area of assessment and testing at the level of policy, practice and professional development programmes.

Keywords: Assessment literacy, Classroom assessment, Language assessment and testing, Knowledge base, Professional development

Introduction

Since there is an indispensable link between assessment quality and student accomplishment (Mertler, 2002; Stiggins, 1999), teachers essentially need to be assessment literate so that they can carry out effective results-oriented pedagogical practice. Research on teacher assessment in different contexts shows that a typical teacher spends as much as a third to a half (30 to 50%) of their total professional time on assessment or testing-related activities (Bachman, 2014; Cheng, 2001; Coombe, Davidson, Sullivan, & Stoyhoff, 2012; Coombe, Troudi, & Al-Hamly, 2012; Stiggins, 2007); however, this time is usually spent without the advantage of having appropriate knowledge and skills

relative to assessment literacy. A language teacher's understanding and knowledge regarding language learning theories, classroom assessment practices, and the effective use of this knowledge to gauge and improve student learning by employing various assessment methods and strategies is often referred to as language assessment literacy (Inbar-Lourie, 2008; Taylor, 2009).

Adequate assessment literacy enables educators to distinguish between comprehensive and flawed assessment. Moreover, it makes them cognizant of the negative effects of tests (Shohamy, 2001). Their understanding of issues such as pretest validity, posttest validity, test reliability, authenticity, and inauthenticity helps them enhance learning inside and outside the classroom. While there is no denying the fact that one of the major responsibilities of teachers is to measure and evaluate student achievement (National Education Association, 1983; Schafer, 1993), the research on teachers' knowledge of assessment has revealed that many teachers' knowledge and understanding of the essentials of assessment and evaluation is inadequate (Al-Bahlani, 2019; Coombe, Davidson, Sullivan, & Stoyloff, 2012; Coombe, Davidson, Gebriel, Boraie, & Hidri, 2017; Jannati, 2015; Jawhar & Subahi, 2020; Lam, 2015; Latif, 2017; Popham, 2009). Despite the great significance of an educational culture that recognizes and promotes teacher assessment literacy, the progress in this regard, in any EFL/ESL context in general and in the study context—the Kingdom of Saudi Arabia—in particular, has been either very slow or not up to the required standards (Coombe, Davidson, et al., 2012; Coombe, Troudi, & Al-Hamly, 2012; Rajab, 2013; Umer, Zakaria, & Alshara, 2018). Popham (2009) warns that teachers' assessment shortcomings due to insufficient knowledge and skills can ruin the quality of any education system.

The present study is an attempt to explore tertiary EFL practitioners' understanding of the various dynamics of assessment literacy and test development in terms of their knowledge base. The study also aims to explore tertiary teachers' assessment knowledge base in relation to variables such as teachers' academic qualifications, teaching experience, and professional training in assessment and testing. It is hoped that the study will have implications in terms of creating awareness and better understanding amongst policy makers, administrators, and teacher educators regarding tertiary EFL teachers' professional development needs in the area of assessment, testing, and evaluation in the pursuit of effective pedagogical practices in any EFL/ESL context in general and the Saudi higher education context in particular.

Literature review

The concept of language assessment literacy—the evolution

The term 'language assessment literacy' (LAL) is often used as a subordinate category to assessment literacy (AL), which has been a focus in the assessment and testing-related general educational literature for the past two decades. Since Stiggins (1991) influential publication on 'assessment literacy', a term he identifies as an individual's ability to evaluate the strengths and weaknesses of an assessment and apply such knowledge in decision-making related to student achievement, the language assessment literacy phenomenon has been a much debated subject. Research indicates an agreement among language assessment scholars that LAL is unique on account of the intricacies that are involved in the assessment of linguistic and communicative competence,

knowledge, and skills (Levi & Inbar-Lourie, 2020; Vogt, Tsagari, Csépes, Green, & Sifakis, 2020). To Brindley (2001), the first language assessor to address the notion of assessment literacy, a language assessment literate is one who is trained in and capable of handling various aspects of curriculum-based classroom assessments in a given social context. Davies (2008), based on a review of the past five decades of language testing literature, defines language assessment literacy as a concept comprising three elements (i.e., skills, knowledge, and principles). Here, “skills” refers to teachers’ expertise in test development and results analysis methods; “knowledge” relates to teachers’ background knowledge about language learning theories, assessment, and classroom pedagogies; and “principles” refers to teachers’ conceptual and practical understanding of assessment concepts such as ‘test ethics’, ‘fairness’, ‘impact’, and ‘professionalism’ (p. 335). Brown and Abeywickrama (2010) add ‘practicality’ and ‘authenticity’ to this list of the principles of assessment together with validity, reliability, and washback. Davies’ *skills, knowledge, and principles* model is similar to Inbar-Lourie’s (2008) *how, what, and why* classification of Brindley’s (2001) LAL professional development model (Harding & Kremmel, 2016). For Fulcher (2012), teacher assessment literacy means having adequate knowledge and understanding of the basic principles of assessment literacy being placed “within wider historical, social, political, and philosophical frameworks” (p. 125). O’Loughlin (2013) endorses these views, arguing that language teachers need to have the ability to critically evaluate the roles of assessment and testing in a given context from both educational and societal angles. In a study published in the same year, Scarino (2013) emphasizes that language teachers need to be capable of exploring and appraising their own assessment-related beliefs, personal theories, and preconceptions in order for them to be self-aware as assessors, which is an essential element of their assessment literacy. In her later article, Scarino stresses the need for language teachers to reconceptualize their assessment practices in alignment with innovative approaches that respect cultural and linguistic diversity against the background of the ever more complex linguistically and culturally diverse classroom contexts arising from globalization (Scarino, 2017). More recently, agreeing with these arguments, Coombe, Vafadar, and Mohebbi (2020) emphasize the importance of language teachers broadening their “contextual-related knowledge and inter-related competencies”, taking into account multiple inter-linked dynamics such as “teacher independence, identity as assessor, and critical perspectives” as important features of their assessment literacy (p. 12). These views are further endorsed by Yan and Fan (2020) and Levi and Inbar-Lourie (2020), who state that the development of LAL as a social and co-constructed phenomenon should be viewed as a dynamic learning process facilitated by a variety of experiential and contextual factors instead of being considered as a passive accumulation of skills and knowledge. The above discussion highlights that the concept of LAL has been continuously evolving over time, though there is no commonly acknowledged LAL framework thus far (Harding & Kremmel, 2016).

The ‘knowledge base’ construct in language assessment literacy and related research

Teachers’ assessment knowledge base is a fundamental component of their assessment literacy (Giraldo, 2021; Inbar-Lourie, 2008; Popham, 2009). Since the publication of the 1990 Standards for Teacher Competence in Educational Assessment of Students

(American Federation of Teachers, National Council on Measurement in Education, National Education Association, 1990) and the seminal publication of Black and Wiliam (1998), which led to the formation of the “Assessment Reform Group (2002)” based on constructivist assessment philosophy (Levi & Inbar-Lourie, 2020, p. 169), the scholarly work in many dimensions of the language assessment field has made a substantial progress. Some research projects have focused on determining which particular aspects of knowledge and skills might be necessary for teachers to be considered assessment literate, conceptualizing teacher assessment literacy in the general and language education fields (e.g., American Federation of Teachers, National Council on Measurement in Education, National Education Association, 1990; Levi & Inbar-Lourie, 2020; Taylor, 2013). Taylor (2013) identifies eight dimensions of teacher assessment knowledge, skills, and principles that constitute their language assessment literacy. These are knowledge of theory, technical skills, principles and concepts, language pedagogy, sociocultural values, local practices, personal beliefs/attitudes, and scores and decision making. She stresses that the level of competence and expertise in these areas of required assessment-related knowledge varies according to the level of responsibility and role in the assessment process. Since different stake holders such as language test developers, language assessment researchers, and language teachers have different roles and responsibilities, they are expected to have different assessment knowledge base profiles based on their needs, interests, and opportunities in the assessment process (Kim, Chapman, Kondo, M., Kondo, A., & Wilmes, 2020; Kremmel & Harding, 2020; Yan & Fan, 2020). Another area that has received scholarly attention has been the connection between teacher assessment quality and student accomplishment (e.g., Coombe, Davidson, et al., 2012; Coombe, Troudi, & Al-Hamly, 2012; Djoub, 2017; Earl, 2013; Popham, 2009). Studies on this topic indicate that the implementation of quality classroom assessments has a strong impact on the quality of classroom pedagogical practices, resulting in higher levels of learner achievement. Linked to these conclusions, there have been some studies in different contexts across the globe that have endeavoured to examine teachers’ assessment knowledge base and level of assessment literacy using different assessment literacy measures at the pre-service and the in-service level (e.g., Al-Bahlani, 2019; Al-kharusi, Aldhafri, Alnabhani, & Alkalbani, 2012; Davidheiser, 2013; Djoub, 2017; Jawhar & Subahi, 2020; Levi & Inbar-Lourie, 2020; Mertler, 2003; Volante & Fazio, 2007). Most of the research on pre-service teachers’ preparation in the area of assessment and testing (e.g. Al-kharusi et al., 2012; Beziat & Coleman, 2015; Campbell & Mertler, 2004; Graham, 2005; Volante & Fazio, 2007) focus on future school teachers. A common finding of most of these studies is that pre-service teachers lack a sufficient assessment knowledge base and the skills necessary to carry out various assessment tasks indicating the need for assessment literacy training based on practice-based, assessment-related experience activities to bridge the gap between theory and practice. Similarly, investigations into teachers’ assessment knowledge base at the in-service level have also addressed school teachers in general, indicating a scarcity of research on tertiary EFL teachers’ assessment-related competence; this is particularly true in the EFL context of the Middle East. The research on school teachers’ assessment knowledge and competence can be grouped into two categories: those that employ self-described measures and those that use objective measures (i.e., a variety of test instruments) to assess teachers’ assessment-related skills and capabilities. Among the studies

that are based on self-report measures (i.e., the questionnaires and surveys through which the language teachers describe their assessment competence) are Al-kharusi et al. (2012), Fulcher (2012), Plake, 1993; Lam (2019), and Tsagari and Vogt (2017). The common finding of all these studies is EFL teachers' self-described inadequate assessment knowledge base, indicating a strong need for in-service teacher professional training in the area of language assessment, testing, and evaluation. The researchers reported that the teachers involved in the studies were cognizant of the fact that their assessment-related competence and knowledge base was inadequate and deficient and that they needed professional preparation to fulfil their roles. For instance, some of the common areas of professional training need were assessment purposes; assessment methods; classroom-based assessments; ethical aspects of assessment use; and issues related to assessment reliability, validity, and wash back. The other category of studies on school teachers' assessment knowledge base includes investigations based on objective measures, that is, some kind of assessment knowledge test or teacher assessment literacy questionnaire (e.g. Al-kharusi et al., 2012; Davidheiser, 2013; Djoub, 2017; Kiomrs, Abdolmehdi, & Naser, 2011; Muhammad, Hama, & Bardakçi, 2019; Plake, Impara, & Fager, 1993). Like the self-reported measures-based studies, these more objective measures-based investigations report that teachers lack sufficient assessment knowledge and understanding.

In the context of the Middle East, Al-kharusi et al. (2012), employing a descriptive survey as the research design, measured the assessment competence of 165 school teachers teaching various subjects in different parts of the Sultanate of Oman. The results of the study indicated that teachers' average score on the Teacher Assessment Literacy Questionnaire (Plake et al., 1993) was 12.42 on the 32-item test, suggesting poor and insufficient assessment competence. In the same context, another study that examined English language teachers' assessment knowledge base was conducted by Al-Bahlani (2019). The findings of the study suggest that although tertiary EFL teachers perceive themselves as having moderate assessment-related competence and training, their actual knowledge and understanding of assessment and testing is very limited, as is obvious from their average performance on the assessment knowledge test ($M = 56.34\%$), which is based on Brookhart's (2011) formative assessment principles. The findings also reveal that teachers have gaps in their understanding of assessment principles used to evaluate assessment tasks such as clarity, authenticity, practicality, validity, and reliability. In the context of Saudi Arabia, to my knowledge, the only two studies conducted on this subject in the tertiary EFL context are by Umer et al. (2018) and Jawhar and Subahi (2020). Umer et al. (2018), in their mixed-method study, investigated university teachers' assessment practices in terms of their alignment with the recommended good practice in the literature, whereas Jawhar and Subahi (2020) studied the assessment literacy level of tertiary instructors working at a public sector university based on questionnaire data obtained via the Classroom Assessment Literacy Inventory developed by Mertler and Campbell (2005) and a demographic questionnaire. The results of both these small-scale studies reveal participants' low level of assessment literacy.

The literature reviewed above clearly suggests that there is dearth of scholarship that explores tertiary language teachers' assessment literacy in terms of their knowledge base in the EFL context of the Middle East in general and the context of the present

study, Saudi Arabia, in particular. In addition, it is important to note that the assessment literacy instruments used in most of the studies reviewed above whether concerning pre-service or in-service teachers' assessment knowledge base (e.g., ALI, Mertler, 2002; CALI, Mertler & Campbell, 2005; TALQ, Plake et al., 1993), were developed in an English-as-a-first-language context, the USA, and they were used to assess school teachers' understanding of the general concepts of assessment and testing applicable to the general education field and not necessarily the English language field. The adapted instrument (the Classroom Assessment Literacy Questionnaire—CALQ) employed in the present large-scale study, however, reflects tertiary EFL contextual requirements. It includes two additional standards, one on the importance of keeping assessment records and the second on the importance of assessment quality assurance and control issues.

Considering the fact that the number of studies on tertiary EFL teachers' assessment literacy is limited, the present study is an attempt to bridge this gap in the literature. Since language assessment literacy is viewed as “a dynamic context-dependent social practice” (Willis, Adie, & Klenowski, 2013, p. 242), investigating teacher assessment literacy in a specific contextualized setting is critical in terms of further exploring the various dynamics of the concept of language assessment literacy. Given the massive socioeconomic reforms that Saudi Arabian society is currently undergoing and their recognition of the significance of promoting knowledge-based economy by developing skilled and educated human resources as the main target of the National Transformation Program (NTP), the importance of such a study cannot be underestimated. Some of the ambitious strategic goals of NTP, an important pillar of the Kingdom's Vision 2030, include teacher training and professional development, creative and innovative learning environment, improved curricula and effective classroom pedagogies aligned with successful contemporary practices, and improved recruitment policy. Against such a background, the recognition of the importance of assessment and teachers' assessment literacy in order to better understand the implications of educational outcomes is inevitable. Research indicates that an individual teacher's assessment knowledge base is an essential component of their academic and professional repertoire and plays a vital role in the overall success and effective implementation of the assessment system in any educational context (Fishbein & Ajzen, 2010; Xu & Brown, 2016). Taking this into account, this study investigates EFL teachers' assessment knowledge base in the Saudi higher education context in the pursuit of teacher development in the area of assessment and testing.

The study

Paradigmatically, the study is based on pragmatism as a philosophical approach that opposes the incompatibility thesis stance, the term used by Howe (1988) to discourage involvement in paradigm wars between positivist and interpretivist philosophies (Tashakkori & Teddlie, 1998). Pragmatism primarily focuses on what is best suited to solve the research problem being investigated. Theoretically, the study is informed by Vygotsky's sociocultural theory (1978) recognizing the place and role of sociocultural dynamics in the process of language and language assessment literacy development. Hill and McNamara (2011) point out that classroom assessment being entrenched in

multifarious sociocultural subtleties results in meaning-making being a social and constructive phenomenon. McNamara (2000) holds that there is an inseparable connection between assessment and its social context. Every context has its own distinct institutional and educational policy dynamics, which influence pre-assessment, during-assessment, and post-assessment processes and procedures. Such recognition of the significant role of the dynamics of sociocultural milieu in the whole process of language learning, its assessment, and teachers’ assessment knowledge base has led to calls for multidimensional professional expertise on the part of language teacher–assessors. Hence, it is important to explore if the assessment-related decisions and practices of language teachers are founded on sound theoretical as well as practical knowledge or if there are gaps in their understanding of the dynamics of assessment literacy.

Research questions

1. What is the tertiary EFL practitioners’ knowledge base in assessment and testing?
2. Which areas of assessment literacy and test development do EFL practitioners in the KSA find challenging?
3. What is the relationship between tertiary EFL teachers’ assessment knowledge base and their academic profile, teaching experience, and PD training background?

Methodology

The participants and the context of the study

A randomly selected large probabilistic sample of 80 tertiary-level EFL practitioners participated in the research. The majority of the respondents were male (71%), as shown in Table 1 below. Regarding their educational background, more than half of the respondents (45) held a master’s degree, 21 were bachelor’s degree holders, and the remaining 14 had a doctoral degree. Most of the respondents were experienced EFL practitioners. Half of them (40) reported having between 16 and 20 years of teaching experience.

Data were collected from EFL teachers working in five public and private sector (4 public; 1 private) higher education institutes located in three main cities of the Eastern

Table 1 Participants’ demographic information

	Variable	N = 80	Percentage
Gender	Male	57	71%
	Female	23	29%
Educational background	Bachelor	21	26%
	Master’s	45	56%
	Doctorate	14	18%
Teaching experience	0–5 Years	11	14%
	6–10 Years	14	18%
	11–15 Years	15	19%
	16–20 Years	40	50%
PD training/educational background	Yes	44	55%
	No	36	45%

Province of Saudi Arabia. This was done purposely taking into consideration the variety of backgrounds of the English teacher population. Most of the universities in the kingdom offer a Preparatory Year English Language Program that comprises two semesters and sometimes a third (summer semester) if needed. A large number of tertiary-level students intending to pursue undergraduate degrees in diverse disciplines enrol in this programme. These students have to pass the 1-year English (EAP) programme, which is, in most cases, an outcomes-based curriculum, to be able to transfer to various university departments where the medium of instruction is English. In addition to a PYP, some of these universities offer bachelor's degrees in English language. The teachers who teach these programmes come from multi-cultural backgrounds and have wide-ranging academic and professional achievements.

Method

Instruments and data collection

An assessment literacy questionnaire (adapted from the Assessment Literacy Inventory, Mertler & Campbell, 2005 and Classroom Assessment Knowledge Test, Tao, 2014) was used in order to understand the dynamics of teachers' knowledge base in assessment and testing. Both the research instruments (i.e., the Assessment Literacy Inventory of Mertler and Campbell (2005) and the Classroom Assessment Knowledge Test by Tao (2014)) went through a rigorous development process before they were used for data collection. The Assessment Literacy Inventory used by Mertler and Campbell (2005) was based on an original instrument named the Teacher Assessment Literacy Questionnaire introduced by Plake et al. (1993). The standards for Teacher Competence in the Educational Assessment of Students (American Federation of Teachers, National Council on Measurement in Education, National Education Association, 1990) serve as the basis of this questionnaire, which later went through in-depth content validation and detailed item analysis by the members of the National Council on Measurement in Education. The modified form of this questionnaire used by Mertler and Campbell (2005) was pilot tested in two stages before it was used. The overall reliability of the instrument based on pilot data was measured as Cronbach's alpha of .74.

This assessment literacy instrument was further modified by Tao (2014) to develop his Classroom Assessment Knowledge Test, which was also subjected to thorough content, construct, and face validation as well as instrument reliability check procedures using the pilot stage data and findings. The assessment literacy questionnaire used in the present study is an adapted form of the above-mentioned instruments. Minor modifications were made to the original instruments. These modifications include changing names to better reflect the context of the study, changing in the order of distractors, and changing a few words and phrases in both the stem and the distractors. However, it was ensured that the content and the construct validity of the original instruments were not compromised. The overall scale reliability of the instrument used in the study was measured as a Cronbach's alpha of .713.

The assessment literacy questionnaire has four scenarios; each one is based on nine multiple-choice questions aligned with the nine total standards. Of these, seven are "Standards for Teacher Competence in the Educational Assessment of Students" (American Federation of Teachers, National Council on Measurement in Education,

National Education Association, 1990), and the other two were introduced by Tao (2014). All nine standards are based on the following principles of assessment and testing:

- a. Teachers should have skills to *select and choose appropriate assessment methods* that are suitable for making apt instructional decisions.
- b. Teachers should have skills to *develop appropriate assessment methods* that inform effective pedagogical practices.
- c. Teachers should have skills to *administer, mark, and interpret the assessment results* of assessments that they produce themselves and those that are developed and produced by others.
- d. Teachers should have skills to *make important decisions regarding individual learners, classroom pedagogy, curriculum development, and institutional progress* based on the assessment results.
- e. Teachers should have skills to *plan and develop appropriate marking standards/ rubrics and procedures* to assess learner progress.
- f. Teachers should have skills to *communicate assessment results to various stake holders such as learners, school management, and parents*.
- g. Teachers should have skills to *recognize and appreciate the principles of test ethics*; they should have appropriate understanding of issues such as test fairness, test impact, and the use of illegal and inappropriate assessment methods and information.
- h. Teachers should have knowledge and skills to *record assessment-related data professionally and use it when needed or required*.
- i. Teachers should have appropriate *knowledge regarding assessment quality assurance and quality control issues in assessment and testing*.

Before, during, and after data collection, all ethical research procedures were followed. These included getting data collection approval from the Institutional Research Committee heads at the research setting and informing participants through verbal as well as written communication about the research purpose; the intended methods and procedures to be employed; the importance of their participation in the research process; the use of data analysis; and the protection of their anonymity, confidentiality, and data.

Data analysis

The assessment literacy questionnaire data was analysed statistically using SPSS version 24. After uploading the questionnaire data into SPSS, a statistical analysis was carried out for the descriptive and inferential statistics. First, scale reliability check (Cronbach's alpha) was carried out. Then, the descriptive statistics were calculated, which included the frequency for nominal/categorical variables (gender, educational background, and professional development in assessment and testing) and descriptives for scale variables (questionnaire scores) for overall mean, range, standard deviation, minimum, and maximum. Before running the inferential statistics, first, it was ensured that the data met the parametric assumptions regarding a randomized sample, normal distribution of

data, independence (data collected from individuals and not groups), higher-level data (ratio scale), and equal variance of the variables. Then, an independent samples t test was run to compare the mean scores. To determine the statistical difference in the mean score between and within groups, one-way ANOVA was run. This was followed by post hoc tests for multiple comparisons to identify the statistical difference between and within groups.

Results

The statistical analysis of the questionnaire data involved calculating descriptive and inferential statistics in SPSS (version 24). The descriptive analysis produced frequencies relating to the respondents’ gender, educational background, and teaching experience and descriptives relating to their performance on the questionnaire in terms of the overall mean, standard deviation, range, minimum, and maximum.

Results of the descriptive statistics

A total of 80 respondents ($n = 80$) completed the questionnaire. Of them, the majority (57) are male. As regards their academic background, more than half of them (45) are master’s degree holders. Half of them (40) reported having between 16 and 20 years of teaching experience or more. Of the other half, most (29) respondents had between 6 and 15 years of teaching experience. As far as academic qualifications or in-service professional development training in the area of assessment and testing are concerned, 44 respondents reported some academic qualification or in-service training, and 36 reported having no such assessment-related training or qualification. The questionnaire data resulting from the analysis of the performance of tertiary EFL practitioners ($n = 80$) revealed an adequate internal consistency level (i.e., $\alpha = .713$). This accords with the recommendations in the literature regarding reliability coefficients (e.g., Kehoe, 1995; Nitko, 2001).

The respondents’ overall mean score on the questionnaire was $M = 17.25/36$ (47.92 %) as shown in Table 2 below, which indicates a very limited level of classroom assessment literacy among the tertiary EFL teachers who participated in the research.

The descriptive statistics revealed gaps in the assessment knowledge base of the study sample for each individual standard, too, as shown in Table 3 below.

The figures indicate that the participants performed the highest on standard 7 (i.e., skills/ability to recognize and appreciate the principles of test ethics), with a total mean score of $M = 2.56/4$ (64%). The poorest performance was on standard 3 (i.e., skills to administer, mark, and interpret the assessment results of assessments), with total mean score of $M = 1.09/4$ (27%).

Overall, for five standards, the total mean score is above 50%, with 64% for one standard (i.e., standard 7), and a little above 50% for the other four standards: standard 1 (i.e., skills in choosing assessment methods appropriate for instructional decisions), with a total mean score of 52%; standard 4 (i.e., skills in using assessment results when

Table 2 Total questionnaire respondents and overall mean

	N	Range	Minimum	Maximum	Mean	Std. deviation
Overall	80	22.00	5.00	27.00	17.25	5.27

Table 3 Total questionnaire mean for each standard

	N	Range	Minimum	Maximum	Mean	Std. deviation
Standard 1 ALQ items 1, 10, 19, 28	80	4.00	.00	4.00	2.10	1.11
Standard 2 ALQ items 2, 11, 20, 29	80	3.00	.00	3.00	1.71	.81
Standard 3 ALQ items 3, 12, 21, 30	80	3.00	.00	3.00	1.09	.86
Standard 4 ALQ items 4, 13, 22, 31	80	4.00	.00	4.00	2.04	1.04
Standard 5 ALQ items 5, 14, 23, 32	80	4.00	.00	4.00	2.14	1.19
Standard 6 ALQ items 6, 15, 24, 33	80	4.00	.00	4.00	1.71	1.39
Standard 7 ALQ items 7, 16, 25, 34	80	4.00	.00	4.00	2.56	1.07
Standard 8 ALQ items 8, 17, 26, 35	80	4.00	.00	4.00	1.83	.98
Standard 9 ALQ items 9, 18, 27, 36	80	4.00	.00	4.00	2.08	1.09
Valid N (listwise)	80					

making decisions about individual students, planning teaching, developing curriculum, and school improvement), with a total mean score of 51%; standard 5 (i.e., developing valid grading procedures that use student assessments), with a total mean score of 53%; and standard 9 (i.e., knowledge regarding quality assurance and quality control issues in assessment and testing), with a total mean score of 52%.

The areas wherein the total mean score is below 50% are standard 2 (i.e., skills in developing assessment methods appropriate for instructional decisions), with a mean score of 42%; standard 3 (i.e., skills in administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods), with a mean score of 27%; standard 6 (i.e., skills in communicating assessment results to different stakeholders such as students, parents, administrators, and other educators), with a mean score 42%; and standard 8 (i.e., knowledge and skills in recording assessment-related data professionally and using it as needed), with a mean score of 45%.

The results based on the descriptive statistics were used to answer the research questions one and two. The first research question is related to tertiary EFL practitioners’ knowledge base in assessment and testing. The findings indicate that the knowledge base of tertiary EFL practitioners who participated in the research by completing the questionnaire is very limited (overall $M = 17.25/36$; $SD = 5.27$). This is evident in their performance on each individual standard, too. There is just one area where their mean score is above 60%.

The second research question relates to the areas of assessment literacy and test development that the EFL practitioners in the KSA find challenging. The results of the study show that generally the teachers have difficulties in almost all areas delineated as “Standards for Teacher Competence in the Educational Assessment of Students” by the American Federation of Teachers, National Council on Measurement in Education, National Education Association (1990) and Tao (2014). Although there is only one area, standard 7 (i.e., skills/ability to recognize and appreciate the principles of test ethics),

where the mean score is above 60%, we consider those areas as challenging where the participants’ performance is below 50%, indicating very limited understanding and knowledge relating to the various aspects of classroom assessment. These areas encompass standards 2, 3, 6, and 8. These standards relate to the teachers’ skills in developing assessment methods appropriate for instructional decisions; skills in administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods; skills in communicating assessment results to different stakeholders such as students, parents, administrators, and other educators; and skills in recording assessment-related data professionally and using it as needed. On all these areas, the most challenging one, as reflected by the questionnaire results, is standard 3 (i.e., skills in administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods), which had an overall mean score of $M = 1.09/4$; $SD = .86$ (27%).

Results of the inferential statistics

The inferential statistics were based on an independent samples t test and one-way between-groups analysis of variance (ANOVA), which were used to determine if there was any significant difference in the respondents’ mean score in terms of gender, educational background, teaching experience, and professional development profile in the area of assessment and testing. According to the results of the independent samples t test, there was no statistically significant difference between the overall mean scores for the male ($M = 17.14$; $SD = 5.54$) and female respondents ($M = 17.52$; $SD = 4.61$). There was also no statistically significant difference between the overall mean scores of the respondents who had some assessment and testing-related qualification or professional development training background and those who did not, as shown in Table 4 below.

On the other hand, the results of one-way between-groups analysis of variance (ANOVA), which was conducted in order to explore the impact of respondents’ educational background and teaching experience on their performance on the assessment literacy questionnaire, indicate a statistically significant difference at $p \leq .05$. For instance, as shown in the table below (Table 5), the overall mean score of those who had a doctoral degree ($M = 19.36$; $SD = 5.26$) was significantly higher than that of those who had a bachelor’s degree ($M = 15.10$; $SD = 6.16$). However, there was no significant difference in the overall mean score between master’s degree holders ($M = 17.60$; $SD = 4.54$) and doctoral degree holders ($M = 19.36$; $SD = 5.26$).

As regards the impact of respondents’ teaching experience on their mean score on the questionnaire, the results shown in Table 6 below indicate that the respondents with teaching experience ranging between 11 and 20 or more years scored higher than those whose teaching experience ranged between 0 and 10 years.

The third research question was answered using the results of the above inferential statistics. This question focused on exploring teachers’ assessment knowledge base in relation to demographic variables such as gender, educational background, teaching

Table 4 Total questionnaire mean according to background in assessment & testing

		Academic qualification or PD training in assessment and testing	N	Mean	Std. deviation
Overall	Yes		44	16.93	5.40
	No		36	17.64	5.15

Table 5 Total questionnaire mean according to educational background

Educational background		N	Mean	Std. deviation
Overall	BA	21	15.10	6.16
	MA	45	17.60	4.54
	PhD	14	19.36	5.26
	Total	80	17.25	5.27

experience, and professional training profile related to assessment education. The results indicate better performance on the questionnaire of respondents with a higher academic and work experience background. However, the findings also suggest that there was no significant difference in performance between the respondents with and without any assessment-related professional training background.

Discussion

Teachers’ assessment knowledge base is an integral component of their assessment literacy (Fulcher, 2012; Giraldo, 2021; Inbar-Lourie, 2008; Popham, 2009; Taylor, 2013).

The participants’ average performance on the Assessment Literacy Questionnaire suggests their limited assessment knowledge base. Their total mean score on the questionnaire of either below 50% or a little above 50% on 8 of the 9 standards indicates their limited understanding, skills, and abilities relating to the various assessment competencies or standards as delineated by American Federation of Teachers, National Council on Measurement in Education, National Education Association (1990) and Tao (2014). These assessment competencies or standards are related to teachers’ skills and abilities in several key areas discussed earlier in the section on Instruments and data collection. The area where the participants performed the best is related to recognizing test ethics-related issues ($M = 2.56/4$), which is consistent with the results of Davidheiser’s (2013) and Tao’s (2014) studies. On the other hand, they performed worst in the area related to administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods.

These results are somewhat consistent with the literature in that most of the researchers who have used either CALI (Campbell & Mertler, 2004) or the Teacher Assessment Literacy Questionnaire—TALQ (Plake et al., 1993) have reached a consensus that teachers’ assessment-related knowledge is by and large insufficient with regard to assessment standards and expectations, with some variations in terms of strengths and weaknesses in different areas with different samples (e.g. Al-kharusi, Kazem, & Al-Musawai, 2011; Campbell, Murphy, & Holt, 2002; Davidheiser, 2013; Jawhar & Subahi, 2020; Kiomrs et al., 2011; Muhammad et al., 2019; Plake et al., 1993; Tao, 2014). The variation in the results in different areas may be due to the difference in assessment-

Table 6 Total questionnaire mean according to teaching experience

		N	Mean	Std. deviation
Overall	0–5 years	11	14.91	5.01
	6–10 years	14	15.43	5.32
	11–15 years	15	18.40	3.81
	16–20 years	40	18.10	5.58
	Total	80	17.25	5.27

related courses in teacher education programmes, educational policies, and curriculum in different contexts. These results suggest that teacher assessment illiteracy is a serious educational concern internationally. The literature reveals that the factors that conspire against teacher assessment knowledge base development, in general, are inappropriate assessment policies and lack of professional standards; institutional and/or contextual power dynamics; the absence of assessment literacy benchmarks in recruitment criteria; leadership styles; teacher agency; general beliefs and misconceptions about teacher assessment; teacher attitudes; complexity of assessment language and concepts; the intimidating nature of tests; insufficient resources and facilities provided for assessment development; difficult work conditions; and lack of innovation aligned with contemporary trends in the delivery of assessment education courses/training or perhaps a complete lack of any such training at all (Coombe, Davidson, et al., 2012; Coombe, Troudi, & Al-Hamly, 2012; Davidson & Coombe, 2019; Inbar-Lourie, 2008; Islam, Hasan, Sultana, Karim, & Rahman, 2021; Lam, 2015; Latif, 2017; Rea-Dickens, 2004; Shohamy, 2001; Tao, 2014; Troudi, Coombe, & Al-Hamly, 2009; Xu & Liu, 2009; Xu & Brown, 2017).

In the present study, a combination of most of the above-mentioned factors, in general, could explain the gaps in teachers' assessment literacy in terms of their assessment knowledge base and skills especially if we take into account the varied academic, socio-cultural, assessment, and teaching contextualized background of the respondents. Among the dominant contributing factors, however, may be the lack of strong assessment policies and professional standards, institutional and/or contextual power dynamics, and lack of innovation and appropriateness in the delivery of both pre-service teacher education and in-service assessment-related teacher professional development programmes. Saudi Arabia, like many other EFL contexts in the Middle East and Asia, is an examination-driven society, where tremendous emphasis is laid on objective-type norm-referenced testing at both the school and university level (Alsamaani, 2014; Umer et al., 2018). The examination system at the school level is centrally managed and controlled by the Ministry of Education, whereas at the tertiary higher education institutes and universities, it is either the local test committees (which is the case in most of the settings), or the individual teachers who are responsible for preparing, administering, and controlling exams, especially in undergraduate programmes (Umer et al., 2018). In a typical Saudi tertiary educational institute, the assessment system is based on both summative and formative assessment types; however, the maximum weight is given to the summative assessments. In educational systems where assessment is centrally controlled, either externally, usually by district- or national-level educational bodies or internally by local testing committees, the assessment is generally used to establish control or impose policy decisions, working on the principle of 'selection' instead of 'democratization' (Shohamy, 2001, p. 29). In such centralized systems, teachers lack opportunities to develop and practice assessment literacy at the classroom level. Externally mandated high-stakes achievement tests or internally mandated large-scale summative tests are generally preferred to on-going formative classroom-based teacher assessments. Here, the main focus is on meeting set standards for accountability reasons serving the interests of departmental, institutional, or governmental authorities instead of designing assessments for learning purposes catering to the specific needs of students and teachers (Camp, 1993; Huot, 2002). In such contexts, teachers' role in

assessment is extremely limited and is restricted to just administering a centrally developed test, with no significant voice in the assessment process, or where it is somewhat active but it constantly suffers from institutional summative–formative assessment conflict (Inbar-Lourie & Donitsa-Schmidt, 2009; Troudi et al., 2009). The limited role of teachers in the assessment process might be one of the reasons behind respondents' low performance on the CALQ.

Another shortcoming or gap in the educational assessment policies that may explain the inadequate assessment knowledge base of the respondents could be the lack of some professional standards at the institutional level serving as quality assurance mechanisms for monitoring assessment practices and evaluating teacher assessment literacy. This certainly functions against the teacher assessment literacy development process.

The literature reveals that traditional styles and approaches to institutional leadership also influence and hamper teacher assessment practices and learning. Pounder (2006) claims that traditional leadership models are based on control, power, and remote management of teachers at the level of either school principals, department heads, or testing committees, and generally at the level of all three at the same time. Research indicates that in centrally controlled examination-driven systems, teachers lack empowerment, so their assessment knowledge base and skills do not have sufficient opportunity to develop (Davison, 2004; Inbar-Lourie & Donitsa-Schmidt, 2009; Shohamy, Donitsa-Schmidt, & Ferman, 1996; Troudi et al., 2009). Teachers' lack of appropriate opportunities with regard to assessment development and implementation activities might be another reason for the inadequate assessment knowledge base of the respondents. The third major factor explaining the insufficient teacher assessment knowledge base is the lack of innovation and appropriateness in the delivery of both pre and in-service assessment-related teacher development programmes aligned with the requirements of the field and contemporary trends. This neglect of teacher assessment literacy development in teacher education and training programmes in the Saudi educational context were highlighted in a recent study by Latif and Wasim (2021). As a matter of fact, since these institutional and contextual assessment-related policy conundrums are, unfortunately, beyond teachers' authority and power, it is the responsibility of policy makers, assessment specialists, teacher educators, and institutional administrators to work collaboratively towards devising a system that ensures teacher professional development in the area of assessment and testing.

Another area that was investigated in the study was teachers' assessment knowledge base in relation to demographic variables such as gender, academic qualifications, teaching experience, and professional development training background. The results of the study indicated better performance on the Assessment Literacy Questionnaire by the more experienced and academically sound participants (i.e., those holding a master's or doctoral degree), suggesting a strong connection between relevant teaching experience, a strong educational background, and the assessment knowledge base, which is in line with the results of several previous empirical studies (e.g. Hoover, 2009; King, 2010; Mertler, 2004). On the other hand, the results indicating no statistical difference between the scores of those with some professional development training in the area of assessment and testing and those with no such training background generally point to some of the shortcomings, difficulties, and challenges facing professional development educational and training programmes. Among these challenges are teachers' belief that

assessment knowledge gained during teacher educational and training programmes is not pedagogically effective, being only theoretical and thus irrelevant to normal classroom assessment practices; that the knowledge is not contextually needs-based; that a cookie-cutter approach is employed to gain this knowledge; and that the focus of most teacher training initiatives is to provide assessment training generically and superficially and not holistically considering various contextual needs and current requirements (Coombe et al., 2020; Davidson & Coombe, 2019; Leung, 2014; Yan, Zhang, & Fan, 2018). The results of the present study showing the ineffectiveness of assessment training exposure and experience in terms of strengthening teachers' assessment-related knowledge and expertise are consistent with those of several previous studies such as Brown (2008), Casale (2011), Koh (2011), Lam (2019), Quilter and Gallini (2000), Vogt and Tzagari (2014), Xu and Brown (2016), Yurtsever (2013), and Zhang and Burry-Stock (2003). These investigations call for some necessary reforms in assessment educational and professional development training programmes. These reforms are related to focus more on teacher attitudes, perceptions, and conceptions regarding assessment and testing in such educational and training programmes (Brown, 2008; Quilter & Gallini, 2000); teacher training based on rigorous needs analysis aligned with contextualized demands in terms of assessment policy and practice (Vogt & Tzagari, 2014; Xu & Brown, 2016; Zhang & Burry-Stock, 2003); on-going and sustained training in line with a constructivist professional development model entailing self-directed learning, peer-coaching, and mentoring, as opposed to the traditional/transmission training model; and finally, appropriate follow-up of and feedback on assessment training (Koh, 2011; Lam, 2019; Yurtsever, 2013). The inadequate assessment-related competence and knowledge base, as evidenced by the results of this study, supports Popham's (2009) argument that the teachers today possess only a little knowledge regarding educational assessment. These results should be a wake-up call for teacher educators in the Saudi higher education sector and have implications for language teachers' assessment literacy development at the level of both policy and practice. At the level of policy and administration, teacher assessment knowledge base development based on context-specific needs and requirements ensuring teachers' voice and inclusion, a broadened assessment purpose recognizing teacher agency, collaborative action research, and quality reflective language assessment research and practices empowering teachers through increased autonomy need to be encouraged. As transformative agents, teachers need to be provided with an assessment system where their identities as reflective practitioners, teacher assessors, and critical pedagogues can be recognized and established. Moreover, it needs to be recognized that no assessment policy or training initiative, whether it is at the pre-service or in-service level, could be effective unless the complexity of teachers' assessment conceptions and beliefs is challenged and addressed (G. T. L. Brown, 2004). Furthermore, it is equally important that the policy makers, administrators, and other assessment-related decision makers also have a good understanding of the various aspects of language, assessment practices, and teacher professional development models because the literature reveals that policy makers' lack of appropriate understanding and misconceptions about language assessment practices result in misinformed and imprudent decision-making causing harm to the cause of education (Kremmel & Harding, 2020; Pill & Harding, 2013). Most importantly, the content of the assessment education or training programmes must be updated and aligned with

recent policy innovations and research practices. There is a need to recognize the significant role of the language assessment learning groups and communities that function as dynamic platforms for sharing suitable resources; having discussions about the latest assessment theories, practices, and trends; establishing small-scale research-based Special Interest Groups (SIGs); and identifying and promoting the latest virtual language assessment trends and tools (Babaii & Asadnia, 2019). Last but not the least, since teachers' professional development is a "continual intellectual, experiential, and attitudinal growth of teachers" (Farrell, 2013, p. 22), teacher assessment literacy development initiatives should be carried out on a continuous, long-term and sustainable basis. The role of appropriately trained assessment literate teachers and educators is fundamental in terms of producing academically sound individuals equipped with lifelong skills necessary to meet the needs and challenges of the modern era (Coombe, Davidson, et al., 2012; Coombe, Troudi, & Al-Hamly, 2012; Popham, 2014). Without this, it will be a serious challenge to realize the target aims of Saudi Arabia's Vision 2030, which pursues diversified business, economic, social, cultural, and educational reforms.

As with any other research investigation, the present study also has several limitations. The use of a single instrument to study the complex subject of LAL could be the first likely limitation. Since the aim of the study was to investigate teachers' assessment knowledge base as a standalone construct, the use of the CALQ was considered sufficient. In future studies, however, the triangulation of both quantitative and qualitative data using instruments such as interviews, focus group interviews, and classroom observation along with the assessment literacy questionnaire should be considered so that the complexity of language teachers' assessment literacy can be investigated from various angles. Another possible limitation could be related to the certain variables, such as the length of the questionnaire, the number of questions, and the somewhat technical nature of the content, which admittedly might have affected some of the respondents' motivation to some extent, resulting in some impact on the findings. Of the nine assessment literacy domains that the CALQ was based on, each was assessed through four items. This was considered important to gain a detailed understanding of the respondents' knowledge in each domain. Another important point to consider is that the adapted CALQ used in the study is not context-specific; it is an adapted version of instruments developed in two different contexts, the USA and Cambodia. Future studies on teachers' LAL need to consider using an instrument that includes generic assessment literacy principles and concepts applicable to all contexts and those that are contextually grounded and aligned with contextual policy, cultural values, and traditions. Moreover, the LAL instrument should be based on language-specific elements catering to the needs in specific educational settings. The instrument should also address each assessment literacy competency area separately, and it should be measured based on sufficient number of items. This is important as a close look at the CALQ shows that Standard 3, wherein the respondents performed the lowest, includes three competency areas (i.e., skills related to administering, scoring and interpreting the results), and these are measured under one domain. This needs to be addressed in future studies. The development of such an instrument must take into account educational assessment policies and teachers' assessment beliefs, conceptions, and practices. Despite these limitations, the results of the study have strong implications for teacher development in the area of language assessment literacy and test development. The study unravels the

complexity of language assessment literacy in terms of teachers' assessment knowledge base in a particular contextualized setting; however, the phenomenon needs to be further explored and addressed in terms of teacher conceptions, beliefs, attitudes, practices, and teacher identity issues in this context and other EFL/ESL contexts with multiple methods and varied stake holder groups so that the multi-faceted aspects of the concept of language assessment literacy can be further explored and better understood.

Conclusion

The overall purpose of the study was to explore tertiary EFL practitioners' assessment literacy and test development in terms of their knowledge base. The findings suggest that tertiary EFL practitioners' assessment knowledge base is very limited, as indicated by their overall average performance on the Classroom Assessment Literacy Questionnaire (CALQ) comprising nine areas of assessment-related competencies or standards as delineated by American Federation of Teachers, National Council on Measurement in Education, National Education Association (1990) and Tao (2014). The participants performed best with regard to recognizing test ethics-related issues and performed worst with regard to administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods. Moreover, the findings revealed a positive relationship between variables such as sound academic and professional background in terms of pertinent work experience and the participants' assessment-related knowledge base. However, in terms of the impact of professional development education or training background on the participants' assessment knowledge base as reflected in their performance on the CALQ, no marked difference was noted indicating concerns related to the shortcomings, difficulties, and challenges facing professional development educational and training programmes. These findings clearly indicate that language teachers need to be equipped with an adequate level of assessment literacy. The literature reveals that the success of an educational system heavily relies on properly trained assessment literate educators who are capable of preparing individuals for the challenges and needs of the 21st century (Coombe, Davidson, et al., 2012; Coombe, Troudi, & Al-Hamly, 2012; Popham, 2014). The findings of the study suggesting gaps and inadequacies in tertiary EFL practitioners' knowledge base of assessment literacy have strong implications for teacher professional development in the area of language assessment and testing in the context of Saudi Arabia and other EFL contexts around the world.

Regarding the theoretical knowledge that teachers are expected to have mastery of for their assessment literacy development, it is expected that teachers, in both their pre-service teacher education and in-service teacher professional development, be sufficiently exposed to and trained in the fundamentals of assessment (i.e., assessment standards (American Federation of Teachers, National Council on Measurement in Education, National Education Association, 1990)). The greatest emphasis should be placed on classroom assessment along with the pedagogic and discipline-oriented content knowledge by establishing a link between theoretical concepts and their actual implementation. Some of the essential components of an assessment literate educator's repertoire are knowledge of the assessment purpose (formative/summative/accountability), knowledge of assessment methods (selected/constructed/personal), knowledge of

assessment marking/grading (holistic/analytic), knowledge of feedback (direct/indirect/descriptive/evaluative/supportive), knowledge of assessment results' interpretation (norm/criterion-referenced-based) and communication, knowledge of alternative assessments (self and peer assessment; teacher questioning; reflective journals; portfolio assessments; projects; authentic assessments; presentations), and knowledge of the ethical aspects of assessment. In addition to these key elements of an assessment knowledge base, an assessment literate educator is also expected to have appropriate theoretical as well as practical understanding of classroom assessment principles such as validity, reliability, practicality, washback, and authenticity (Brown & Abeywickrama, 2010; Malone, 2011). Coombe et al. (2020) emphasize that assessment courses should be a mandatory part of teachers' professional credentials and requirements. They further accentuate that language teachers' assessment knowledge base needs to be kept constantly updated based on research and innovations at the level of policy and practice. Moreover, in order for teachers to have a profound learning experience, they need to be engaged in extensive and sustained professional training.

Abbreviations

CALQ: Classroom Assessment Literacy Questionnaire; AFT: The American Federation of Teachers; NCME: The National Council on Measurement in Education; NEA: The National Education Association

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40468-021-00130-9>.

Additional file 1. Classroom Assessment Literacy Questionnaire (<https://doi.org/10.6084/m9.figshare.13757653.v1>).

Additional file 2. The Source of Assessment Standards and their Associated Items within the Classroom Assessment Literacy Questionnaire (<https://doi.org/10.6084/m9.figshare.13757653.v1>).

Additional file 3. Data analysis files (<https://doi.org/10.6084/m9.figshare.13757653.v1>).

Acknowledgements

I would like to thank Dr. Susan Riley and Dr. Shirley Larkin, the senior faculty members at the Graduate School of Education, University of Exeter, UK for their insightful comments on the earlier draft of this paper.

Author's contributions

MWL is the sole contributor to this research paper. The author(s) read and approved the final manuscript.

Author's information

Muhammad Wasim Latif has a doctoral degree (EdD TESOL) from the University of Exeter, UK. He has a wide range of teaching experience at tertiary level in both EFL & ESL contexts. He is currently associated with Royal Commission Colleges and Institutes, Jubail Industrial city, Saudi Arabia. His research interests include teacher education, language pedagogy, teacher assessment literacy and language testing and assessment.

Funding

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Declaration

Competing interests

The author declares that he has no competing interests.

Received: 14 February 2021 Accepted: 17 June 2021

Published online: 02 September 2021

References

Al-Bahlani, S. M. (2019). *Assessment literacy: A study of EFL teachers' assessment knowledge, perspectives, and classroom behaviors* [Published doctoral dissertation, The University of Arizona]. UA Campus Repository, University Libraries

- Al-kharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2012). Educational assessment attitudes, competence, knowledge, and practices: An exploratory study of Muscat teachers in the Sultanate of Oman. *Journal of Education and Learning*, 1(2). <https://doi.org/10.5539/jel.v1n2p217>.
- Al-kharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of pre-service and in-service teachers in educational measurement. *Asia-Pacific Journal of Teacher Education*, 39(2), 113–123. <https://doi.org/10.1080/1359866X.2011.560649>.
- Alsamaani, A. (2014). Evaluating classroom assessment techniques of novice Saudi EFL teachers. *Journal of Arabic and Human Sciences*, 7(2), 63–81. <https://doi.org/10.12816/0009602>.
- American Federation of Teachers, National Council on Measurement in Education, National Education Association (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30–32.
- Assessment Reform Group (2002). Assessment for learning. Retrieved from http://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/Eng_DVD/doc/Afl_principles.pdf.
- Babaii, E., & Asadnia, F. (2019). A long walk to language assessment literacy: EFL teachers' reflection on language assessment research and practice. *Reflective Practice*, 20(6), 745–760. <https://doi.org/10.1080/14623943.2019.1688779>.
- Bachman, L. F. (2014). Ongoing challenges in language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment*, (vol. 3, 1st ed., pp. 1586–1603). Wiley.
- Beziat, T. L. R., & Coleman, B. K. (2015). Classroom assessment literacy: Evaluating pre-service teachers. *The Researcher*, 27(1), 25–30.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7–74.
- Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing*, 18(4), 393–407.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy and Practice*, 11(3), 301–318.
- Brown, G. T. L. (2008). Assessment literacy training and teachers' conceptions of assessment. In C. Rubie-Davies, & C. Rawlinson (Eds.), *Challenging thinking about teaching and learning*, (pp. 285–302). Nova Science.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and practices*. Pearson.
- Camp, R. (1993). Changing the model for the direct assessment of writing. In M. M. Williamson, & B. Huot (Eds.), *Validating holistic scoring for writing assessments: Theoretical and empirical foundations*, (pp. 45–78). Hampton.
- Campbell, C., & Mertler, C. A. (2004). Measuring teachers' knowledge & application of classroom assessment concepts: Development of the assessment literacy inventory. In *Paper presented at the annual meeting of the American Educational Research Association, Montréal, Quebec, Canada*.
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002, October). Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers. In *Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH*.
- Casale, M. (2011). *Teachers' perceptions of professional development: An exploration of delivery models [Unpublished doctoral dissertation]*. Johnson and Wales University.
- Cheng, L. (2001). An investigation of ESL/EFL teachers' classroom assessment practices. *Language Testing Update*, 29, 53–83.
- Coombe, C., Davidson, P., Gebriil, A., Boraie, D., & Hidri, S. (2017). *Language assessment in the Middle East and North Africa: Theory, practice and future trends*. UAE, TESOL Arabia.
- Coombe, C., Davidson, P., Sullivan, B., & Stoyoff, S. (2012). *The Cambridge guide to second language assessment*. Cambridge University Press.
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language assessment literacy: Issues, challenges and recommendations. In *The Cambridge guide to second language assessment*. Cambridge University Press.
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn. *Language Testing in Asia*, 10(3). <https://doi.org/10.1186/s40468-020-00101-6>.
- Davidheiser, S. A. (2013). *Identifying areas for high school teacher development: A study of assessment literacy in the Central Bucks School District [Unpublished doctoral thesis]*. The Drexel University.
- Davidson, P., & Coombe, C. (2019). Language assessment literacy in the Middle East and North Africa (MENA) region. *Arab Journal of Applied Linguistics*, 4(2), 1–23.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/0265532208090156>.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practice in Australia and Hong Kong secondary schools. *Language Testing*, 21, 305–334.
- Djoub, Z. (2017). Assessment literacy: Beyond teacher practice. In *Revisiting EFL assessment, second language learning and teaching*. Springer International Publishing. https://doi.org/10.1007/978-3-319-32601-6_2.
- Earl, L. M. (2013). *Assessment as learning: Using classroom assessment to maximize student learning*, (2nd ed.,). Corwin Press.
- Farrell, T. S. C. (2013). *Reflective practice in ESL teacher development groups: From practices to principles*. Palgrave MacMillan. <https://doi.org/10.1057/9781137317193>.
- Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. Psychology Press, Taylor & Francis.
- 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. (2021). A reflection on initiatives for teachers' professional development through language assessment literacy. *Profile: Issues in Teachers' Professional Development*, 23(1), 197–213. <https://doi.org/10.15446/profile.v23n1.83094>.
- Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, 21(6), 607–621. <https://doi.org/10.1016/j.tate.2005.05.001>.
- Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari, & J. Banerjee (Eds.), *Handbook of second language assessment*, (pp. 413–427). De Gruyter. <https://doi.org/10.1515/9781614513827-027>.

- Hill, K., & McNamara, T. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395–420. <https://doi.org/10.1177/0265532211428317>.
- Hoover, N. R. (2009). *A descriptive study of teachers' instructional use of student assessment data* [Unpublished doctoral dissertation, Virginia Commonwealth University].
- Howe, K. R. (1988). Against the quantitative-qualitative incompatibility thesis, or, Dogmas die hard. *Educational Researcher*, 17(8), 10–16. <https://doi.org/10.3102/0013189X017008010>.
- Huot, B. (2002). *(Re)-articulating writing assessment for teaching and learning*. Utah State University Press.
- 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>.
- Inbar-Lourie, O., & Donitsa-Schmidt, S. (2009). Exploring classroom assessment practices: The case of teachers of English as a foreign language. *Assessment in Education: Principles, Policy & Practice*, 16(2), 185–204.
- Islam, M. S., Hasan, M. K., Sultana, S., Karim, A., & Rahman, M. M. (2021). English language assessment in Bangladesh today: Principles, practices, and problems. *Language Testing in Asia*, 1(1). <https://doi.org/10.1186/s40468-020-00116-z>.
- Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *The International Journal of Research in Teacher Education*, 6(2), 26–37.
- Jawhar, S. S., & Subahi, A. M. (2020). The impact of specialty, sex, qualification, and experience on teachers' assessment literacy at Saudi higher education. *International Journal of Learning, Teaching and Educational Research*, 19(5), 200–216. <https://doi.org/10.26803/ijlter.19.5.12>.
- Kehoe, J. (1995). Basic item analysis for multiple-choice tests. *Practical Assessment, Research & Evaluation*, 4(10) Retrieved from <http://PAREonline.net/getvn.asp?v=4&n=10>.
- Kim, A. A., Chapman, M., Kondo, M., Kondo, A., & Wilmes, C. (2020). Examining the assessment literacy required for interpreting score reports: A focus on educators of K–12 English learners. *Language Testing*, 37(1), 54–75. <https://doi.org/10.1177/0265532219859881>.
- King, J. D. (2010). *Criterion-referenced assessment literacy of educators* [Unpublished doctoral dissertation, University of Southern Mississippi], Hattiesburg, MS, United States.
- Kiomrs, R., Abdolmehdi, R., & Naser, R. (2011). On the interaction of test washback and teacher assessment literacy: The case of Iranian EFL secondary school teachers. *English Language Teaching*. <https://doi.org/10.5539/elt.v4n1p156>.
- Koh, H. K. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, 22(3), 255–276. <https://doi.org/10.1080/10476210.2011.593164>.
- 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>.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2) SAGE. <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>.
- Latif, M. W. (2017). Tertiary level EFL practitioners' understanding of the dynamics of assessment literacy: An exploratory study. In C. Coombe, P. Davidson, A. Gebril, D. Boraie, & S. Hidri (Eds.), *Language Assessment in the Middle East and North Africa: Theory, Practice and Future Trends*, (pp. 71–87). TESOL Arabia Publications.
- Latif, M. W., & Wasim, A. (2021). Tertiary EFL practitioners' beliefs and attitudes towards professional development in the area of assessment and testing – An exploratory study. *Language Testing in Focus: An International Journal*, 3, 17–32. <https://doi.org/10.32038/ltf.2021.03.02>.
- Leung, C. (2014). Classroom-based assessment issues for language teacher education. In A. J. Kunnan (Ed.), *The companion to language assessment*, (vol. 3, pp. 1510–1519). Wiley Blackwell.
- 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>.
- Malone, M. (2011). Assessment literacy for language educators. *CALDigest*, 1–2. Retrieved from <http://www.cal.org/resources/digest/assessmentliteracy-for-language-educators.htm>
- McNamara, T. (2000). *Language testing*. Oxford University Press.
- Mertler, C. A. (2002). *Classroom assessment literacy inventory*. (Adapted from the Teacher Assessment Literacy Questionnaire (1993), by Barbara S. Plake & James C. Impara, University of Nebraska-Lincoln, in cooperation with The National Council on Measurement in Education & the W. K. Kellogg Foundation.
- Mertler, C. A. (2003, October 15–18). Pre-service versus in-service teachers' assessment literacy: Does classroom experience make a difference. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, Ohio.
- Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference. *American Secondary Education*, 33, 49–64.
- Mertler, C. A., & Campbell, C. (2005, April 11–15), *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory* [Paper presentation]. Annual meeting of the AERA, Montreal, Quebec, Canada.
- Muhammad, N., Hama, F., & Bardakçı, M. (2019). Iraqi EFL teachers' assessment literacy: Perceptions and practices. *Arab World English Journal (AWEJ)*, 10(2). <https://doi.org/10.2139/ssrn.3418574>.
- National Education Association (1983). *Teachers' views about student assessment*. Author.
- Nitko, A. (2001). *Educational assessment of students*, (3rd ed.,). Merrill Prentice Hall.
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363–380. <https://doi.org/10.1177/0265532213480336>.
- 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>.
- Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(1), 2–27.
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10–12.

- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental. *Theory into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>.
- Popham, W. J. (2014). *Classroom assessment: What teachers need to know*, (7th ed.,). Pearson Education.
- Pounder, J. S. (2006). Transformational classroom leadership: The fourth wave of teacher leadership. *Educational Management Administration & Leadership*, 34(4), 533–545. <https://doi.org/10.1177/1741143206068216>.
- Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator*, 36(2), 115–131. <https://doi.org/10.1080/08878730009555257>.
- Rajab, H. (2013). Developing speaking and writing skills of L1 Arabic EFL learners through teaching of IPA phonetic codes. *Theory and Practice in Language Studies*, 3(4), 653–659.
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment [Editorial]. *Language Testing*, 21(3), 251–258. <https://doi.org/10.1191/0265532204lt283ed>.
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309–327. <https://doi.org/10.1177/0265532213480128>.
- Scarino, A. (2017). Developing assessment literacy of teachers of languages: A conceptual and interpretive challenge. *Papers in Language Testing and Assessment*, 6(1), 18–40.
- Schafer, W. D. (1993). Assessment for teachers. *Theory into Practice*, 32(2), 118–126. <https://doi.org/10.1080/00405849309543585>.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Pearson Education.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effects over time. *Language Testing*, 13(3), 298–317. <https://doi.org/10.1177/026553229601300305>.
- Stiggins, R. (2007). Conquering the formative assessment frontier. In J. McMillan (Ed.), *Formative Classroom Assessment*. Colombia University Teachers College Press.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534–539.
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 18(1), 23–27.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323–342. https://doi.org/10.1207/S15324818AME1604_4.
- Tao, N. (2014). *Development and validation of classroom assessment literacy scales: English as a foreign language (EFL) instructors in a Cambodian higher education setting*. [Unpublished doctoral thesis]. Victoria University.
- Tashakkori, A., & Teddlie, C. (1998). *A mixed methodology: Combining quantitative and qualitative approaches*. Sage Publications.
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36. <https://doi.org/10.1017/S0267190509090035>.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412.
- Troudi, S., Coombe, C., & Al-Hamili, M. (2009). EFL teachers' views of English language assessment in higher education in the United Arab Emirates and Kuwait. *TESOL Quarterly*, 43(3), 546–555. <https://doi.org/10.1002/j.1545-7249.2009.tb00252.x>.
- 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.
- Umer, M., Zakaria, H. M., & Alshara, A. M. (2018). Investigating in Saudi university EFL teachers' assessment literacy: Theory and practice. *International Journal of English Linguistics*, 8(3), 345. <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(4), 374–402. <https://doi.org/10.1080/15434303.2014.960046>.
- Vogt, K., Tsagari, D., Csépes, I., Green, A., & Sifakis, N. (2020). Linking learners' perspectives on language assessment practices to teachers' assessment literacy enhancement (TALE): Insights from four European countries. *Language Assessment Quarterly*, 17(4), 410–433. <https://doi.org/10.1080/15434303.2020.1776714>.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749–770. <https://doi.org/10.2307/20466661>.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualizing teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40(2), 241–256. <https://doi.org/10.1007/s13384-013-0089-9>.
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58(2016), 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>.
- Xu, Y., & Brown, G. T. L. (2017). University English teacher assessment literacy: A survey-test report from China. *Language Testing and Assessment*, 6(1), 133–158.
- Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, 3(3) JSTOR URL: <http://www.jstor.org/stable/27785031>.
- Yan, X., & Fan, J. (2020). "Am I qualified to be a language tester?" Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 0(0), 1–28. <https://doi.org/10.1177/0265532220929924>.
- Yan, X., Zhang, C., & Fan, J. J. (2018). "Assessment knowledge is important, but...": How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, 74, 158–168. <https://doi.org/10.1016/j.system.2018.03.003>.
- Yurtsever, G. (2013). English language instructors' beliefs on professional development models and preferences to improve their teaching skills. Akdeniz Language Studies Conference 2012. *Procedia - Social and Behavioural Sciences*, 70, 666–674. <https://doi.org/10.1016/j.sbspro.2013.01.107>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.