AKADÉMIAI KIADÓ

Teachers' assessment literacy: A descriptive literature review

PISETH HULL[1]* and TIBOR VÍGH[2]

[1] Doctoral School of Education, University of Szeged, Hungary

[2] Institute of Education and MTA-SZTE Reading and Motivation Research Group, University of Szeged, Hungary

RESEARCH ARTICLE

© 2024 The Author(s)

ABSTRACT

This study aims to synthesize teachers' assessment literacy (AL) reviews of past studies to find how AL is conceptualized, measured, and developed and formulate recommendations for further empirical studies. As AL is important for teachers to contribute to teaching and learning improvement, a growing body of teacher AL research has been established in the literature. The present study identifies, assesses, synthesizes, and describes the characteristics of reviews on AL from 2014 to 2023. The results show that studies used different terms and definitions for AL. Researchers apply different forms of instruments to measure AL in various aspects of assessment. AL is influenced by mediating factors, including assessment conceptions and efficacy. It can be effectively developed through training workshops, assessment courses, and teachers' self-reflection on assessment practices. The review suggests that further AL studies are needed among in- and pre-service teachers at different levels and contexts and directional relationships between assessment-related constructs that encompass AL.

* Corresponding author. E-mail: piseth.hull@edu.u-szeged.hu

AKJournals

INTRODUCTION

Assessment literacy (AL) is important for teachers to handle their responsibilities in using assessment to improve education quality, i.e., improving their instruction and students' learning outcomes, and to be accountable for the students' learning results (G. T. L. Brown, 2006). They use their AL to select, construct, and use different types of assessment to help improve their students' learning and communicate the assessment results effectively and ethically (Brookhart, 2011; Khadijeh & Amir, 2015; Popham, 2017). However, research often confirms that teachers have low AL levels (Boothroyd, Robert, & Robert, 1992; DeLuca & Bellara, 2013; Yamtim & Wongwanich, 2014). For example, Williams (2015) reported that teachers have weaknesses in designing paper-and-pencil tests, and their skills in grading students' achievement need further development. Others found that teachers had limited knowledge about some assessment aspects, for instance, regarding applying reliability and validity (Tao, 2014) and assessing students' reflections (Chan & Luo, 2020). They need AL that is necessary for their practical use in their context (DeLuca & Klinger, 2010).

Inquiries on teachers' AL focused mostly on collecting evidence from various processes of assessment issues, including (1) how assessment was provided in instruction (Kremmel, Eberharter, Holzknecht, & Konrad, 2018; Nier, Anne, & Margaret, 2009; Oo, Alonzo, & Asih, 2022), (2) how AL was developed (Gan & Lam, 2020; Hasselgreen & Carlsen, 2004; Lan & Fan, 2019), (3) what AL level teachers had (DeLuca, Valiquette, Coombs, LaPointe-McEwan, & Luhanga, 2018; Ogan-Bekiroglu & Suzuk, 2014), (4) how they enacted their assessment knowledge into practices (Ayalon & Wilkie, 2020; Ogan-Bekiroglu & Suzuk, 2014), and (5) how teacher education programs related to assessment contribute to teacher candidates' AL development (Ayalon & Wilkie, 2020; DeLuca & Klinger, 2010; Ogan-Bekiroglu & Suzuk, 2014). As a large and growing body of AL studies has been established in the literature, it is worth conducting a review study about teachers' AL based on previous review studies to find feasible gaps for further empirical investigations.

RESEARCH QUESTIONS

This study aimed to review previous AL reviews to explore how teachers' AL was conceptualized, measured, and developed in past studies and to formulate recommendations for further empirical studies. To guide the review, the following questions were formulated:

1. What concepts were used in previous studies to define AL?
2. What methods were used in previous studies to measure teachers' AL?
3. What background factors were identified as influencing teachers' AL levels?
4. How can teachers' AL be effectively developed?

METHODS

During the identification process of this narrative literature review, first, we used keywords, 'assessment literacy' AND 'review,' to search for the literature reviews on teachers' AL in the

Scopus database; we limited our search only to review studies in English. Second, we used the back-reference searching method for the related literature we found. Third, we performed an additional search in Google Scholar for the related reviews and referenced articles. During the screening process, based on the title, abstract, and keywords of the records, we excluded non-review articles on teachers' AL and selected studies that are related to teachers' AL and professional development. We assessed the review articles that met the selection criteria for eligibility to answer the research questions.

The literature search yielded fifteen studies that were published between 2014 and 2023. As shown in Table 1, these studies are classified into four categories according to their common themes. The first group of studies focuses on how AL is conceptualized at different educational levels. The second category contains review studies about teachers' AL measurements that focus on how researchers measure this construct by using tests and questionnaires. The third group involves studies that review teachers' AL development, focusing on how teachers' AL develops through assessment training and courses in teacher education programs. The last group contains reviews that explore trends in AL research in past studies.

We assess the quality of reviews based on the methodology, number of reviewed studies, journal ranking, and number of studies that cite the study as shown in Table 2. Eleven studies were published in high-quality journals and have been cited by many other studies except for the study by Juanjuan and Mohd Yusoff (2022). Although studies by Pastor's (2022), Puspawati (2022), Khalid et al. (2021) and SeviMel-ŞahiN and Subaşi (2019) were not published in a high-quality journals, they were included because of their relevance to our research questions.

Table 1. Themes of reviews on AL from 2014 to 2023

Themes of the studies	n	Reference
Conceptualization of AL	6	SeviMel-ŞahiN and Subaşi (2019)
		Singh, Singh, Singh, Moneyam, and Abdullah (2022)
		Weng and Shen (2022)
		Xu and Brown (2016)
		Coombs and DeLuca (2022)
		Pastore (2022)
AL measurements	3	DeLuca, LaPointe-McEwan, and Luhanga (2016b)
		Gotch and French (2014)
		Juanjuan and Mohd Yusoff (2022)
AL development	2	Giraldo (2021)
		Oo et al. (2022)
Trends of AL studies	4	Khalid, Latif, and Yusof (2021)
		Puspawati (2022)
		Qotboddin, Hossein, and Khalil (2020)
		Wang, Zuo, Liu, and Sun (2023)
Total	15	

Table 2. Quality assessment of included reviews

Studies	Methodology	Studies reviewed	Journal ranking	Studies cite
Xu and Brown (2016)	Scoping review	100	Q1	781
Coombs and DeLuca (2022)	Scoping review	190	Q1	16
DeLuca et al. (2016b)	Systematic review	23	Q1	324
Gotch and French (2014)	Systematic review	36	Q1	173
Oo et al. (2022)	Systematic review	12	Q2	12
Wang et al. (2023)	Scientometric review	163	Q2	9
Weng and Shen (2022)	Systematic review	NA	Q2	15
Giraldo (2021)	Content analysis	14	Q2	40
Juanjuan and Mohd Yusoff (2022)	Systematic review	18	Q3	1
Qotboddin et al. (2020)	Systematic review	37	Q3	11
Singh et al. (2022)	Systematic review	69	Q4	9
Pastore (2022)	Systematic review	8	NA	1
Puspawati (2022)	Systematic review	21	NA	2
Khalid et al. (2021)	Systematic review	4	NA	4
SeviMel-ŞahiN and Subaşi (2019)	Systematic review	21	NA	12

Note. NA stands for not applicable.

RESULTS AND DISCUSSION

Concepts of AL

Variations of AL. Four variations of AL were found in the previous studies that were defined differently at distinct education levels. The variations include assessment competence, AL, assessment capability, and assessment identity (Coombs & DeLuca, 2022). *Assessment competence* represents teachers' cognitive dimension in conducting high-quality assessment practices to improve their students' learning and their own instruction. It includes teachers' knowledge of assessment purpose, understanding of what and how to assess, attributes of high-quality assessment, skills in using assessment, understanding of assessment communication, and feedback ability (Stiggins & Conklin, 1992, as cited in Coombs & DeLuca, 2022). *AL* appears in the Standards for Teacher Competence in Educational Assessment of Students (American Federation of Teachers et al., 1990). According to Stiggins (1991), assessment-literate teachers should (1) understand what students' achievement targets they have to assess (e.g., the contents, cognitive levels, and behavior) and (2) have the ability to use the assessment methods (e.g., different types of assessment) to assess their students' achievements. *Assessment capability* stresses the importance of how assessment-competent teachers enable students to attain the curriculum through assessment, and they should possess the knowledge to construct, modify, and use assessment as a tool to motivate students to learn through self-regulation (Absolum, Flockton, Hattie, Hipkins, & Reid, 2009). *Assessment identity* is defined as a teacher's identification of their personal and professional work of assessment, including their knowledge and skills about curriculum, assessment practice knowledge, confidence in practicing effective assessment, belief in assessment process effectiveness, and how prior experience and context influence their assessment identification (Wyatt-Smith & Adie, 2021).

As explained by Coombs and DeLuca (2022), although all these terms conceptualize similar constructs of teachers' knowledge and skills of assessment, they differ to some extent regarding the trend, location, and scope. While AL tends to be abundant in the literature, the terms assessment competence, assessment capability, and assessment identity are found to be used in fewer studies. While AL is widely used in the literature across regional contexts, assessment competence is used mainly in the research context of Europe, and assessment capability and assessment identity are mainly studied in New Zealand and Australia. Furthermore, while assessment competence, AL, and assessment identity are related to teachers' knowledge about assessment, assessment capability expands its conceptualization to focus on students' involvement in the self-regulated aspect of learning in assessment.

Definition of AL at the compulsory education level. Regarding the definition of AL, Pastore (2022) found that researchers define it differently according to education levels. At the general education levels (both primary and secondary education), AL is defined as teachers' understanding of assessment theory and appropriate assessment practices to improve students' learning (DeLuca & Klinger, 2010; Willis, Adie, & Klenowski, 2013). It is context- and experience-driven (Looney, Cumming, Van Der Kleij, & Harris, 2018; Willis et al., 2013) and is a part of teachers' data literacy covering teachers' assessment knowledge and the related factors enable them to use various data (Cowie & Cooper, 2017). Assessment-literate teachers know a wide range of assessment situations (Herppich et al., 2018) and balance their assessment practices depending on assessment-related factors (e.g., students' cognitive levels and assessment policy) to ensure their effectiveness (Looney et al., 2018). Xu and Brown (2016) propose that teachers' AL consists of three mastery levels: educational assessment knowledge based on principles and theories, knowledge of assessment based on perception of assessment experience, and self-directed awareness of assessment processes and one's own identity as an assessor. These three components work together to help teachers practice and reflect on their assessment practices and gain better insights.

Definition of AL at the higher education level. At the higher education level, Pastore (2022) revealed that assessment-literate teachers are familiar with both subject matter language (i.e., the knowledge of the subject matter) and the assessment language (i.e., the knowledge of assessment principles) (Deeley & Bovill, 2017). In addition, Medland (2019) provides six characteristics of AL: community (e.g., stakeholders), dialogue (i.e., interaction between assessment-related participants), knowledge and understanding, program-wide approach (i.e., relationships between curriculum and student learning), self-regulation (i.e., teachers' prior judgement ability of the students' work), and standards (i.e., shared understanding of assessment).

There is a consensus among researchers in defining AL as teachers' assessment knowledge and skills that teachers use in assessment practices at all education levels (Pastore, 2022). However, at the higher education level, some researchers tend to extend their focus on AL beyond the common factors by elaborating extensively on aspects of assessment as learning, such as self-regulation (Medland, 2019). At the compulsory education level, some researchers, like Looney et al. (2018), define AL in a way that goes beyond what is considered the knowledge of the teachers about assessment; that is, not only what they know and can do about assessment but also what they perceive to do and not to do about it. Their definition seems to reflect not only teachers' assessment knowledge, skills, and use of assessment but also their perceptions of assessment.

Measurements of teachers' AL

In measuring teachers' AL, researchers constructed several tools to assess this construct in different forms (e.g., multiple choice tests and Likert-scale questionnaires). Regarding measurements of overall AL proficiency, DeLuca et al. (2016b) and Juanjuan and Mohd Yusoff (2022) reviewed three common forms of teachers' AL measurements that covered the overall teachers' AL competency: tests, questionnaires, and a combination of tests and questionnaires.

Measurements as a test. Plake et al.'s (2005) Teacher Assessment Literacy Questionnaire (TALQ) and Mertler and Campbell's (2005) Revised Assessment Literacy Inventory (ALI) are measurements of teachers' overall AL, both of which consisted of 35 items based on the 1990 Standards of teachers' competence in assessment. They measured the seven standards (i.e., choosing assessment methods; developing assessment; administering, scoring, and interpreting the results; using assessment results; developing valid pupil grading procedures; communicating assessment results; and recognizing unethical, illegal, and inappropriate assessment methods), each of which consisted of five multiple-choice items. Mertler and Campbell's (2005) instrument is the revised form of Plake et al.'s (2005) TALQ. While Plake et al.'s (2005) measure of 555 in-service teachers' AL yielded a low reliability coefficient ($\alpha = 0.54$), Mertler and Campbell (2005) measured the AL of 250 pre-service teachers and found the reliability of the instrument to be satisfactory ($\alpha = 0.74$). Similarly, Talib, Ghafar, and Naim (2014) constructed a test to measure secondary school teachers' AL ($n = 465$) not only based on the 1990 Standards but also the Malaysian Teacher Training Syllabus and Stiggins's (1999) Competency Assessment Model. This test consisted of 45 multiple-choice items covering five dimensions of AL: assessment concepts, measurement methods, testing, scoring, and grading, and statistics and reporting, and had a high level of internal consistency ($\alpha = 0.85$). Dunlai and You (2017) adapted the AL Test from Coombe et al.'s (2016) work, as cited in Juanjuan and Mohd Yusoff (2022). It measures the AL of 39 middle school English teachers in China before they took the language assessment course. There were 10 tasks covering both the language-specific testing knowledge (i.e., the four macro-skills: listening, speaking, reading, and writing) and the overall test-related knowledge (i.e., the test preparation process and the statistical interpretation). Its estimated reliability of the internal consistency was high ($\alpha = 0.828$).

Measurements as a questionnaire. Volante and Fazio (2007), Esfandiari and Nouri (2016), DeLuca et al. (2016a), Nikmard and Mohamadi Zenouzagh (2020), and Kelly, Feistman, Dodge, St. Rose, and Littenberg-Tobias (2020) took a different approach to measuring teachers' overall and perceived AL by constructing questionnaires in a Likert-scale form based on different sources that measure different aspects of assessment. Volante and Fazio's (2007) questionnaire, which is based on Earl (2003) and Earl and Katz's (2006) work, as cited in Juanjuan and Mohd Yusoff (2022), focused on three conceptions of assessment, i.e., assessment of/for/as learning. It consisted of seven open-ended questions about four main areas, including the self-described level of AL, perception of assessment purposes, frequency of using diverse assessment methods, and improvement suggestions for further training and the pre-service education program. It was distributed to 69 Canadian pre-service teachers. Esfandiari and Nouri (2016) designed an instrument based on literature on AL and frameworks specializing in assessment (Bachman, 1990; Fulcher & Davidson, 2009; Hughes, 2003). Participants self-rated their AL proficiency on a five-point Likert scale of all 50 items, ranging from 1 for the lowest to 5 for the highest. It focused on

several aspects of AL, including knowledge of statistics, testing, and interpretation of the test results. It was validated by experts and piloted with 310 in-service teachers. The reliability coefficient of the questionnaire was excellent ($\alpha = 0.97$).

DeLuca, LaPointe-McEwan, and Luhanga (2016a) constructed the Approaches to Classroom Assessment Inventory based on the latest revised classroom assessment standards in response to the call for instrument development that reflected the current assessment standards and the increase in accountability issues in education. Three sections–approaches to classroom assessment, perceived skill in classroom assessment, and assessment professional learning priorities and preferences–were framed in the instrument. In the first part of the instrument, they introduced 20 scenario-based items to measure four aspects of the approach to assessment, including assessment purposes, processes, fairness, and measurement theory. In the second part of the questionnaire, they included 12 items that asked the teachers to self-rate their assessment proficiency levels on a five-point scale ranging from 1 = novice to 5 = expert. The final part of the questionnaire included 21 items to measure teachers' professional learning priorities and preferences in assessment. The expert panel was asked to review the content of the questionnaire, and it was piloted with both in-service and pre-service teachers; the reliability coefficients ranged from.74 to.92 across all sub-scales.

More recently, Nikmard and Zenouzagh (2020) and Kelly et al. (2020) measured teachers' AL with different tools on distinct aspects of assessment. Nikmard and Zenouzagh (2020) investigated the assessment knowledge of 150 English teachers on four dimensions based on their self-assessment, including validity, reliability, interpretability, and efficiency, with a 30-item instrument. The reliability estimate of the scale was high ($\alpha = 0.78$). Kelly et al. (2020) measured teachers' self-perceived AL proficiency with their questionnaire that was based on Brown and Mednick's (2012) Quality Performance Assessment Framework. It consisted of 27 items, measuring five dimensions of performance assessment practices, including valid assessment design, reliable scoring, assessment data analysis, fairness in assessment, and student voice and choice. A total of 1,080 teachers were asked to rate their confidence on a 6-point Likert-scale from 0 (not confident) to 6 (extremely confident). A five-factor hierarchical model confirmed that this instrument measured teachers' AL appropriately.

Measurements as both test and questionnaire. Daniel and King (1998) designed an instrument based on the literature to measure teachers' AL by combining both the test and questionnaire forms, consisting of 30 items in the test and 30 items in the questionnaire. The true/false test was used to probe 90 teachers' knowledge of testing and measurement, while the five-point Likert scale questionnaire ranging from the least frequency to the greatest frequency elicited the use of assessment techniques among elementary and middle school teachers. It was reported to have reliability coefficients of.60 for the first section and.93 for the second.

Background factors influencing AL

Teachers' AL is influenced by individual and contextual factors, which have relationships with other constructs, including assessment conceptions and efficacy, and several background factors influence AL levels. The following section describes these factors and relationships.

Weng and Shen (2022) found that teachers' AL is influenced by individual factors including linguistic backgrounds and years of teaching (Crusan, Plakans, & Gebril, 2016), teachers' academic degrees, training experiences, and fields of study (e.g., different subjects taught in English)

(Soodmand Afshar & Ranjbar, 2021). Some other studies found that the contextual factors that influence teachers' language AL include assessment cultures (Sultana, 2019; Tsagari, 2021), educational landscapes (e.g., the cultural and contextual dimensions), policies at the national and local level (Gu, 2014), school policies (Mansouri, Molana, & Nazari, 2021), institutional mandates (Yan, Zhang, & Fan, 2018), and infrastructure provided by institutions (Firoozi, Razavipour, & Ahmadi, 2019). Xu and Brown (2016) claimed that contextual factors might influence teachers' AL at three different levels. At the national level, AL could be influenced by policies, i.e., they shape professional development activities and teacher autonomy (Forsberg & Wermke, 2012), and policies influence AL by obligating teachers to conform to curriculum standards, textbook use, and large-scale tests (Gu, 2014). At the institutional level, teachers' personal practical knowledge of assessment might be shaped by structural conditions (e.g., workplace conditions) (Xu & Liu, 2009), and at the personal level, teachers' awareness of their role and identity as assessors could influence their AL (Adie, 2013; Scarino, 2013).

According to Xu and Brown (2016), assessment training has relationships with various mediating factors for both pre-service and in-service teachers, including assessment conceptions and efficacy. Teachers' AL is deemed to have relationships with conceptions of assessment, as Xu and Brown (2016) proposed that if teachers receive sufficient training on assessment, they tend to improve their AL, but if their AL does not improve well, there must be some mediating factors hindering its development. However, the results of the studies are rather diverse. While some studies found that assessment training might not have an impact on teachers' beliefs about assessment purposes (e.g., Brown, 2008, as cited in Xu & Brown, 2016), some other studies (Levy-Vered & Alhija, 2018; Quilter & Gallini, 2000) found relationships between assessment training and teachers' AL. Yet, researchers tend to agree that experience in assessing students for summative purposes is hindering their conceptions of assessment for improvement purposes (Gunn & Gilmore, 2014), and this experience also has impacts on pre-service teachers' AL development (Quilter & Gallini, 2000). Teachers' conceptions of assessment tend to be reflected on and modified after gaining knowledge about assessment (DeLuca & Lam, 2014). Xu and Brown (2016) also found that teachers' training needs have a relationship with their assessment efficacy. For example, Gullickson (1993) found a mismatch between what the teachers wanted to learn and what the professors wanted them to learn. DeLuca and Klinger (2010) found that during the assessment training, educators intended to teach everything about assessment rather than specifically what the teachers wanted to learn. Volante and Fazio (2007) reported that teachers lack confidence in assessing their students' learning, and they need more assessment training, support, and time.

Some background factors evidently have impacts on teachers' AL, such as teachers' experience, education levels, training experience, and fields of study (Weng & Shen, 2022). Thus, it is important to consider these factors when discussing AL for teachers' professional development and training programs since teachers might want to learn in line with their needs, and their different backgrounds might require them to learn certain assessment areas aligning with their teaching fields. Their conceptions of assessment might also make them value assessment differently, and they probably have an impact on their attitudes towards the assessment training in the programs.

Teachers' AL development

Assessment training and professional development. As AL training is important for teachers, understanding what assessment knowledge they need is necessary for the assessment training

program to plan what should be taught to meet their needs (Gan & Lam, 2020; Ölmezer-Öztürk & Aydin, 2018). Weng and Shen (2022) found that language teachers' AL was developed in response to their contextual needs through language assessment courses and teachers' reflections. Teachers need assessment knowledge differently depending on their contexts. For example, teachers in Europe preferred different types of assessments, such as portfolios, peer assessments, and self-assessments (Hasselgreen & Carlsen, 2004). In China, language teachers at secondary schools preferred practical assessment training rather than theoretical assessment courses (Lan & Fan, 2019; Yan et al., 2018), while college teachers wanted to advance their competence on assessment theories and concepts (Gan & Lam, 2020). Regarding the AL topics, research shows common areas included in the training programs, such as the construction and evaluation of assessment tools (Giraldo, 2021; Kremmel et al., 2018; Levi & Inbar-Lourie, 2020), the principles of assessment knowledge (e.g., assessment reliability and validity) (Bolivar, 2020; Levi & Inbar-Lourie, 2020; Walters, 2010), ethics, fairness, democracy, and transparency (Levi & Inbar-Lourie, 2020).

Assessment training and teacher education. Oo et al. (2022) found that assessment content in the initial teacher education program includes four dimensions of assessment orientation, depending on the nature of the program. While an audience-oriented assessment program provides assessment training in response to pre-service teachers' needs (Childs & Lawson, 2003), a theory-driven assessment program provides content that is influenced by the philosophical theory of learning, like the Vygotskian sociocultural approach to learning (Brevik, Blikstad-Balas, & Engelien, 2017) or other scholars' frameworks (Ogan-Bekiroglu & Suzuk, 2014). Another dimension is embedded in a policy-driven program that develops assessment training content by conforming to the policy of educational assessment (Izci & Caliskan, 2017; Levy-Vered & Alhija, 2018), and a classroom practice-driven assessment program focuses on assessment practice at the school level, such as an assessment scenario, an assessment task, and a scoring rubric (DeLuca & Klinger, 2010; Deneen & Brown, 2016; Giraldo & Murcia, 2019; Izci & Caliskan, 2017; McGee & Colby, 2014). Oo et al. (2022) indicated that initial teacher education programs on assessment use a stand-alone course, an integrated curriculum unit, and an intervention or workshop to empower pre-service teachers' AL. The duration of the course is often one semester or shorter (Levi & Inbar-Lourie, 2020; Walters, 2010), and only a few programs last several years (Kremmel et al., 2018).

AL development has impacts on teachers' assessment conceptions, their abilities to design assessment, and the construct of assessment, i.e., they tend to look at language assessment more broadly in terms of testing that just focuses on macro-skills rather than wholistic assessment (Giraldo, 2021). Tian, Louw, and Khan (2021) found that self-reflection and seeking help from others who have the same or similar problems may help teachers improve their AL. Some others show that teachers' reflections can help them become more knowledgeable about assessment and more independent in enacting the power of being an assessor (Babaii & Asadnia, 2019).

CONCLUSION

Although research has defined AL with different terms, the term 'AL' is still the dominant one found in the literature. It is defined differently across studies, but its common themes are about teachers' assessment knowledge and skills in designing and using assessment to gauge the

students' learning. While AL measurements are basically based on the Standards developed in 1990, they tend to embed the dimensions of assessment for learning in contemporary assessment studies due to the revised assessment guidelines. AL measurements need validation across contexts with different stakeholders based on the characteristics among these participants. Training in assessment needs to consider background factors that encompass AL development and that might hinder or enhance teachers' assessment development and practice. In assessment training programs, consideration of teachers' assessment conceptions, efficacy, and training needs is necessary to help align the content of assessment training to meet their assessment needs and practices.

Currently, reviews on AL focus mainly on in-service teachers and teachers in general. Only very few review studies have been conducted about pre-service teachers' AL. Similarly, empirical studies on teachers' AL are basically about teachers in general and language teachers, but for pre-service teachers and other subject teachers' studies, only a few studies are found available (e.g., about physic education teachers). Therefore, more studies on these groups of teachers should be conducted, especially focusing on validations of AL measurement for these groups of teachers because, based on assessment experience, fields, and needs, they might obtain different assessment knowledge and practices. Another feasible research area is the relationship between teachers' AL and other constructs. This is also very applicable to the studies with pre-service teachers, so future studies should be about how teachers' AL is related to other psychological constructs that support teachers' AL, such as assessment conceptions, practices, and efficacy. Future studies should focus on the direction of impacts between constructs rather than on the association between constructs. Finally, investigation on more specific types of teachers' assessment competence, both general and specific subject teachers and pre-service teachers, such as teachers' literacy of assessment of/for/as learning.

In this study, we limited our inclusion criteria to only previous AL review studies, and the search results showed that these review studies were published between 2014 and 2023; thus, more current, useful empirical studies might be missing from our reviews. Although back-referenced studies in the included reviews were used, they might include only studies that were published before 2023, and it might impact the generalization of our findings.

This study has four implications. First, teacher trainers should consider the teachers' assessment needs and practices. This would help them align their assessment competence with practices. Second, training curriculum developers should consider trainees' needs, course contents, and approaches that might benefit assessment training. Third, assessment policymakers should encourage teachers to learn and use innovative assessments by aligning assessment policy with teachers' innovative classroom practice of assessment for learning. Finally, researchers should carefully consider the validity of assessment measurements in different contexts with different stakeholders in further studies.

Declaration of conflict interests: No potential conflict of interest was reported by the authors.

ABOUT THE AUTHORS

Piseth Hull is currently a PhD student in Educational Science at the Doctoral School of Education, University of Szeged, Hungary. He is also a teacher trainer at the Regional Teacher Training Centre of Kampong Cham, Cambodia. His research interest is teacher' assessment literacy and related psychological factors of assessment.

Tibor Vígh holds a PhD in Education from the University of Szeged, Hungary. He is an assistant professor at the Institute of Education and a member of the MTA-SZTE Reading and Motivation Research Group, University of Szeged. His research covers foreign language assessment and the examination of teacher candidates' competencies.

ACKNOWLEDGEMENTS

REFERENCES

Absolum, M., Flockton, L., Hattie, J., Hipkins, R., & Reid, I. (2009). Directions for assessment in New Zealand. *Directions for assessment in New Zealand (DANZ): Developing students' assessment capability*. Wellington: Ministry of Education. http://assessment.tki.org.nz/Assessment-inthe-classroom/Assessment-position-papers.

Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 91–106. https://doi.org/10.1080/0969594X.2011.650150.

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), 30–30. https://doi.org/10.1111/j.1745-3992.1990.tb00391.x.

Ayalon, M., & Wilkie, K. J. (2020). Developing assessment literacy through approximations of practice: Exploring secondary mathematics pre-service teachers developing criteria for a rich quadratics task. *Teaching and Teacher Education*, *89*, 103011. https://doi.org/10.1016/j.tate.2019.103011.

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. (1990). *Fundamental considerations in language testing* (1. publ., 11. [print.]). Oxford University Press.

Bolivar, E. M. R. (2020). *Monitoring preservice teachers' language assessment literacy development through journal writing* (Vol. 17).

Boothroyd, R. A., Robert, M. F., & Robert, P. M. (1992). *What do teachers know about measurement and how did they find out? Paper presented at the annual meeting of the national council on measurement in education*. Sanfriscisco, CA.

Brevik, L. M., Blikstad-Balas, M., & Engelien, K. L. (2017). Integrating assessment for learning in the teacher education programme at the University of Oslo. *Assessment in Education: Principles, Policy & Practice*, *2*(24), 164–184.

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. (2006). Teachers' conceptions of assessment: Validation of an abridged version. *Psychological Reports*, *99*, 166–170. https://doi.org/10.2466/pr0.99.1.166-170.

Brown, C., & Mednick, A. (2012). *Quality performance assessment: A guide for schools and districts*. Center for Collaborative Education.

Chan, C. K. Y., & Luo, J. (2020). An exploratory study on teacher assessment literacy: Do novice university teachers know how to assess students' written reflection? *Teachers and Teaching*, *26*(2), 214–228. https://doi.org/10.1080/13540602.2020.1787375.

Childs, R. A., & Lawson, A. (2003). What do teacher candidates know about large-scale assessments? What should they know? *Alberta Journal of Educational Research*, *4*(49), 354–367.

Coombs, A., & DeLuca, C. (2022). Mapping the constellation of assessment discourses: A scoping review study on assessment competence, literacy, capability, and identity. *Educational Assessment, Evaluation and Accountability*, *34*(3), 279–301. https://doi.org/10.1007/s11092-022-09389-9.

Cowie, B., & Cooper, B. (2017). Exploring the challenge of developing student teacher data literacy. *Assessment in Education: Principles, Policy & Practice*, *24*(2), 147–163. https://doi.org/10.1080/0969594X.2016.1225668.

Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing*, *28*, 43–56. https://doi.org/10.1016/j.asw.2016.03.001.

Daniel, L. G., & King, D. A. (1998). Knowledge and use of testing and measurement literacy of elementary and secondary teachers. *The Journal of Educational Research*, *91*(6), 331–344. https://doi.org/10.1080/00220679809597563.

Deeley, S. J., & Bovill, C. (2017). Staff student partnership in assessment: Enhancing assessment literacy through democratic practices. *Assessment & Evaluation in Higher Education*, *42*(3), 463–477. https://doi.org/10.1080/02602938.2015.1126551.

DeLuca, C., & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, *64*(4), 356–372. https://doi.org/10.1177/0022487113488144.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 419–438. https://doi.org/10.1080/0969594X.2010.516643.

DeLuca, C., & Lam, C. Y. (2014). Preparing teachers for assessment within diverse classrooms: An analysis of teacher candidates' conceptualizations. *Teacher Education Quarterly*, *3*(41), 3–24.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, *21*(4), 248–266. https://doi.org/10.1080/10627197.2016.1236677.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016b). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*(3), 251–272. https://doi.org/10.1007/s11092-015-9233-6.

DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2018). Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice*, *25*(4), 355–375. https://doi.org/10.1080/0969594X.2016.1244514.

Deneen, C. C., & Brown, G. T. L. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education*, *3*(1), 1225380. https://doi.org/10.1080/2331186X.2016.1225380.

Dunlai, L., & You, S. (2017). An investigation of Chinese middle school in-service English teachers' assessment literacy. *Indonesian EFL Journal*, *1*(1), 1. https://doi.org/10.25134/ieflj.v1i1.607.

Esfandiari, R., & Nouri, R. (2016). A mixed-methods, cross-sectional study of assessment literacy of Iranian university instructors: Implications for teachers' professional development. *Iranian Journal of Applied Linguistics*, *19*(2), 115–154. https://doi.org/10.29252/ijal.19.2.115.

Firoozi, T., Razavipour, K., & Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, *9*(1), 2. https://doi.org/10.1186/s40468-019-0078-7.

Forsberg, E., & Wermke, W. (2012). Knowledge sources and autonomy: German and Swedish teachers' continuing professional development of assessment knowledge. *Professional Development in Education*, *38*(5), 741–758. https://doi.org/10.1080/19415257.2012.694369.

Fulcher, G., & Davidson, F. (2009). *Language testing and assessment: An advanced resource book* (Repr). Routledge.

Gan, L., & Lam, R. (2020). Understanding university English instructors' assessment training needs in the Chinese context. *Language Testing in Asia*, *10*(1), 11. https://doi.org/10.1186/s40468-020-00109-y.

Giraldo, F. (2021). Language assessment literacy and teachers' professional development: A review of the literature. *Profile: Issues in Teachers' Professional Development*, *23*(2), 265–279. https://doi.org/10.15446/profile.v23n2.90533.

Giraldo, F., & Murcia, D. (2019). Language assessment literacy and the professional development of pre-service language teachers. *Colombian Applied Linguistics Journal*, *2*(21), 243–259. https://doi.org/10.14483/22487085.14514.

Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, *33*(2), 14–18. https://doi.org/10.1111/emip.12030.

Gu, P. Y. (2014). The unbearable lightness of the curriculum: What drives the assessment practices of a teacher of English as a Foreign Language in a Chinese secondary school? *Assessment in Education: Principles, Policy & Practice*, *21*(3), 286–305. https://doi.org/10.1080/0969594X.2013.836076.

Gullickson, A. R. (1993). 1. Matching measurement instruction to classroom-based evaluation: Perceived discrepancies, needs, and challenges. In *Teacher training in measurement and assessment skills*. Lincoln, Neb: Buro institute of mental measurement, University of Nebraska-Lincoln.

Gunn, A. C., & Gilmore, A. (2014). Early childhood initial teacher education students' learning about assessment. *Assessment Matters*, *7*. https://doi.org/10.18296/am.0124.

Hasselgreen, A., & Carlsen, C. (2004). *European survey of language testing and assessment needs*. http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf.

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, *76*, 181–193. https://doi.org/10.1016/j.tate.2017.12.001.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge University Press.

Izci, K., & Caliskan, G. (2017). Development of prospective teachers' conceptions of assessment and choices of assessment tasks. *International Journal of Research in Education and Science*, 464–464. https://doi.org/10.21890/ijres.327906.

Juanjuan, G., & Mohd Yusoff, N. (2022). A review of teachers' assessment literacy proficiency measures. *International Journal of Evaluation and Research in Education (IJERE)*, *11*(3), 1143. https://doi.org/10.11591/ijere.v11i3.22300.

Kelly, M. P., Feistman, R., Dodge, E., St. Rose, A., & Littenberg-Tobias, J. (2020). Exploring the dimensionality of self-perceived performance assessment literacy (PAL). *Educational Assessment, Evaluation and Accountability*, 32(4), 499–517. https://doi.org/10.1007/s11092-020-09343-7.

Khadijeh, B., & Amir, R. (2015). Importance of teachers' assessment literacy. *International Journal of English Language Education*, 3(1), 139. https://doi.org/10.5296/ijele.v3i1.6887.

Khalid, N. H. M., Latif, A. A., & Yusof, I. J. (2021). Research focusing on assessment literacy: A systematic literature review. *Studies of Applied Economics*, 39(10). https://doi.org/10.25115/eea.v39i10.5816.

Kremmel, B., Eberharter, K., Holzknecht, F., & Konrad, E. (2018). Fostering language assessment literacy through teacher involvement in high-stakes test development. In D. Xerri, & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 173–194). Springer International Publishing. https://doi.org/10.1007/978-3-319-77177-9_10.

Lan, C., & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation*, 61, 112–122. https://doi.org/10.1016/j.stueduc.2019.03.003.

Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, 17(2), 168–182. https://doi.org/10.1080/15434303.2019.1692347.

Levy-Vered, A., & Alhija, F. N.-A. (2018). The power of a basic assessment course in changing preservice teachers' conceptions of assessment. *Studies in Educational Evaluation*, 59, 84–93. https://doi.org/10.1016/j.stueduc.2018.04.003.

Looney, A., Cumming, J., Van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442–467. https://doi.org/10.1080/0969594X.2016.1268090.

Mansouri, B., Molana, K., & Nazari, M. (2021). The interconnection between second language teachers' language assessment literacy and professional agency: The mediating role of institutional policies. *System*, 103, 102674. https://doi.org/10.1016/j.system.2021.102674.

McGee, J., & Colby, S. (2014). Impact of an assessment course on teacher candidates' assessment literacy. *Action in Teacher Education*, 36(5–6), 522–532. https://doi.org/10.1080/01626620.2014.977753.

Medland, E. (2019). I'm an assessment illiterate': Towards a shared discourse of assessment literacy for external examiners. *Assessment & Evaluation in Higher Education*, 44(4), 565–580. https://doi.org/10.1080/02602938.2018.1523363.

Mertler, C. A., & Campbell, C. (2005). Measuring teachers' knowledge & application of classroom assessment concepts: Development of the assessment literacy inventory. In *Annual meeting of the American educational research association*. Montreal.

Nier, C. V., Anne, E. D., & Margaret, E. M. (2009). Increasing assessment literacy among LCTL instructors through blended learning. *The Journal of the National Council of Less Commonly Taught Languages*, 9, 105–136.

Nikmard, F., & Mohamadi Zenouzagh, Z. (2020). Designing and validating a potential assessment inventory for assessing ELTs' assessment literacy. *Language Testing in Asia*, 10(1), 8. https://doi.org/10.1186/s40468-020-00106-1.

Ogan-Bekiroglu, F., & Suzuk, E. (2014). Pre-service teachers' assessment literacy and its implementation into practice. *The Curriculum Journal*, 25(3), 344–371. https://doi.org/10.1080/09585176.2014.899916.

Ölmezer-Öztürk, E., & Aydin, B. (2018). Toward measuring language teachers' assessment knowledge: Development and validation of Language Assessment Knowledge Scale (LAKS). *Language Testing in Asia*, 8(1), 20. https://doi.org/10.1186/s40468-018-0075-2.

Oo, C. Z., Alonzo, D., & Asih, R. (2022). Acquisition of teacher assessment literacy by pre-service teachers: A review of practices and program designs. *Issues in Educational Research*, *32*(1).

Pastore, S. (2022). Assessment literacy in the higher education context: A systematic review. *Intersection: A Journal at the Intersection of Assessment and Learning*, *4*(1).

Plake, B. S., Impara, J. C., & Fager, J. J. (2005). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, *12*(4), 10–12. https://doi.org/10.1111/j.1745-3992.1993.tb00548.x.

Popham, W. J. (2017). *Classroom assessment: What teachers need to know* (8th ed.). Pearson Education.

Puspawati, I. (2022). Teachers' language assessment literacy: A systematic review. *Indonesian EFL Journal*, *8*(1).

Qotboddin, J. M., Hossein, K., & Khalil, M. (2020). Assessment literacy of Iranian EFL teachers: A review of recent studies. *The Journal of AsiaTEFL*, *17*(2), 689–698. https://doi.org/10.18823/asiatefl.2020.17.2.27.689.

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.

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.

SeviMel-ŞahiN, A., & Subaşi, G. (2019). An overview of language assessment literacy research within English language education context. *Journal of Theoretical Educational Science*, *12*(4), 1340–1364. https://doi.org/10.30831/akukeg.501817.

Singh, C. K. S., Singh, H. K. J., Singh, T. S. M., Moneyam, S., & Abdullah, N. Y. (2022). ESL teachers' assessment literacy in classroom: A review of past studies. *Journal of Language and Linguistic Studies*, *18*(1), 1–17.

Soodmand Afshar, H., & Ranjbar, N. (2021). EAP teachers' assessment literacy: From theory to practice. *Studies in Educational Evaluation*, *70*, 101042. https://doi.org/10.1016/j.stueduc.2021.101042.

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*.

Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, *18*(1), 23–27. https://doi.org/10.1111/j.1745-3992.1999.tb00004.x.

Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, *9*(1), 1. https://doi.org/10.1186/s40468-019-0077-8.

Talib, R., Ghafar, M. N. A., & Naim, H. A. (2014). Assessment literacy: A catalyst to the success of school-based assessment in Malaysian schools. In A. Kasim, W. S. A. Wan Omar, N. H. Abdul Razak, N. L. Wahidah Musa, R. Ab. Halim, & S. R. Mohamed (Eds.), *Proceedings of the international conference on science, technology and social Sciences (ICSTSS) 2012* (pp. 197–202). Singapore: Springer. https://doi.org/10.1007/978-981-287-077-3_23.

Tao, N. (2014). *Development and validation of classroom assessment literacy scales: English as a foreign language (EFL) instructors in a Cambodian higher education setting* (Vol. 326).

Tian, W., Louw, S., & Khan, M. K. (2021). Covid-19 as a critical incident: Reflection on language assessment literacy and the need for radical changes. *System*, *103*, 102682. https://doi.org/10.1016/j.system.2021.102682.

Tsagari, D. (2021). Gauging the assessment literacy levels of English teachers in Norway. *European Journal of Applied Linguistics TEFL*, *10*, 162–192.

Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education/Revue Canadienne de l'éducation*, *30*(3), 749. https://doi.org/10.2307/20466661.

Walters, F. S. (2010). Cultivating assessment literacy: Standards evaluation through language-test specification reverse engineering. *Language Assessment Quarterly*, *7*(4), 317–342. https://doi.org/10.1080/15434303.2010.516042.

Wang, X., Zuo, J., Liu, F., & Sun, Z. (2023). A scientometric review of research Trends in language assessment literacy. *Education Sciences*, *13*(2), 190. https://doi.org/10.3390/educsci13020190.

Weng, F., & Shen, B. (2022). Language assessment literacy of teachers. *Frontiers in Psychology*, *13*, 864582. https://doi.org/10.3389/fpsyg.2022.864582.

Williams, J. C. (2015). "Assessing without levels": Preliminary research on assessment literacy in one primary school. *Educational Studies*, *41*(3), 341–346. https://doi.org/10.1080/03055698.2015.1007926.

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising 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.

Wyatt-Smith, C., & Adie, L. (2021). The development of students' evaluative expertise: Enabling conditions for integrating criteria into pedagogic practice. *Journal of Curriculum Studies*, *53*(4), 399–419. https://doi.org/10.1080/00220272.2019.1624831.

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010.

Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, *43*(3), 493–513.

Yamtim, V., & Wongwanich, S. (2014). A study of classroom assessment literacy of primary school teachers. *Procedia - Social and Behavioral Sciences*, *116*, 2998–3004. https://doi.org/10.1016/j.sbspro.2014.01.696.

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.