



Preservice teachers' assessment literacy: A systematic review

Piseth Hull^{a,1,2}, Tibor Víg^{b,*,1,3}

^a Doctoral School of Education, University of Szeged, Hungary

^b Institute of Education and MTA-SZTE Reading and Motivation Research Group, University of Szeged, Hungary

ARTICLE INFO

Keywords:

Assessment literacy
Assessment conceptions
Assessment efficacy
Assessment capability
Preservice teachers
Systematic review

ABSTRACT

Assessment literacy is essential for improving teaching, learning, and accountability. While teacher education is crucial in developing preservice teachers' assessment literacy, previous reviews have not systematically examined how it has been operationalized and conceptualized in empirical studies. This study addresses this gap through a systematic review of 39 empirical studies, conducted in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses protocol. Results indicate that researchers relied on tests and questionnaires grounded in assessment standards, guidelines, or contextual curricula. This review proposes a conceptual framework demonstrating that preservice teachers' assessment literacy development is a dynamic process shaped by the interplay of personal (beliefs, efficacy, prior experience), contextual (training opportunities, institutional culture), and external (supervising agencies, policy) factors. The findings underscore the need for context-sensitive methodological approaches and reflect the increasing integration of information and communication technology knowledge, while offering implications for teacher education and research in educational assessment.

1. Introduction

Assessment literacy (AL) is critical for teachers because it helps them reflect on their teaching and students' learning needs (Abell & Siegel, 2011; Khadijeh & Amir, 2015; Pastore & Andrade, 2019). Teachers must develop their assessment knowledge and skills (DeLuca & Klinger, 2010) and become assessment literate to enhance their assessment for learning (AFL) practices (Engelsen & Smith, 2014). Thus, AL should be included in content areas for current and future teacher development programs (Fulcher, 2012; Popham, 2009). The importance of AL in education has attracted many researchers (DeLuca et al., 2016a; Lan & Fan, 2019; Lo & Leung, 2022; Starck et al., 2018; Sun & Zhang, 2022; Vogel et al., 2006; Yan & Pastore, 2022) to examine which dimensions of AL are necessary for teachers, how it develops, and how it relates to other constructs, including assessment conceptions, efficacy, and practices.

Despite the growing body of research on teachers' AL, there is still limited systematically synthesized evidence on how prospective teachers' AL has been examined and conceptualized in relation to assessment-related constructs in empirical studies. Although Oo et al. (2022) reviewed research on preservice teachers' (PTs') AL, they primarily

focused on how AL content is developed within teacher education programs. To date, no systematic review has comprehensively examined the methodological approaches and instruments used to operationalize PTs' AL or synthesized how AL has been conceptualized in relation to constructs such as assessment conceptions, efficacy, and practices. Accordingly, this study aims to systematically review how PTs' AL has been investigated and conceptualized in empirical research to inform the design and development of teacher education programs and identify possible directions for future research.

This study focuses on PTs' AL, as teacher education represents the primary training phase that lays the foundation for the development of AL and may shape future assessment practices. However, since the investigation of PTs' AL is fundamentally based on research conducted among practicing teachers, we first discuss the characteristics of the construct in this context, and then present the specific characteristics that affect PTs, thus establishing the relevance of our review.

1.1. Definitions of AL

Teachers' AL has been defined differently on the basis of educational

* Corresponding author.

E-mail addresses: piseth.hull@edu.u-szeged.hu (P. Hull), vigh.tibor@edpsy.u-szeged.hu (T. Víg).

¹ Postal address: Petőfi Sándor sgt. 32–34, 6722 Szeged, Hungary

² ORCID: <https://orcid.org/0009-0006-5078-7835>

³ ORCID: <https://orcid.org/0000-0002-7201-9187>

levels. At the general education level, this construct is defined as teachers' knowledge about assessment and ability to appropriately assess their students' learning according to the assessment's theoretical and philosophical stance (DeLuca & Klinger, 2010). AL is also interpreted as teachers' practice of probing the knowledge of the classroom and culture with learners to develop and use assessment to help achieve students' learning (Willis et al., 2013). To become assessment literate, Xu and Brown (2016) proposed that three levels of assessment mastery should be met: (1) educational assessment knowledge based on principles and theories; (2) knowledge of assessment based on the perception of assessment experience and self-directed awareness of assessment processes; and (3) one's own identity as an assessor. These levels help teachers enhance their assessment practices and reflection, resulting in deeper insights into assessment and highlighting that AL is a dynamic and contextual construct (Xu & Brown, 2016). At the higher education level, Deeley and Bovill (2017) explained that assessment-literate instructors are familiar with subject matter and assessment language, which emphasizes their knowledge of subject matter and assessment principles.

1.2. Variations and competing terminologies in AL research

Building on a comprehensive scoping review of 190 studies, Coombs and DeLuca (2022) identified three other variations of AL: assessment competence, capability, and identity. Researchers have used these terms to refer to related concepts of AL, including teachers' assessment knowledge and skills to construct and assess students' learning. However, they differ in terms of scopes, geographical areas, and emerging frequency in studies (Coombs & DeLuca, 2022). While assessment competence focuses on teachers' assessment knowledge and skills, assessment identity extends its conceptualization by focusing on students' self-regulated learning aspects. In addition, while AL is found across most regional contexts, assessment capability and assessment identity appear mostly in studies in Australia and New Zealand. Coombs and DeLuca (2022) also indicated that AL is the most frequently used term among the four.

1.3. Conceptual scope of AL

Teachers' AL has been widely discussed since the 1990s after the emergence of the Standards for Teacher Competence in Educational Assessment of Students (AFT et al., 1990) and seminal works by Stiggins (1991) on educational assessment. According to these standards (AFT et al., 1990), teachers should be skilled at seven components of assessment competence: (1) choosing appropriate methods; (2) developing appropriate methods; (3) administering, scoring, and interpreting results; (4) using assessment results appropriately; (5) developing valid grading; (6) communicating assessment results; and (7) recognizing the unethical, illegal, and inappropriate use of assessment information. These assessment competences were used to operationalize and measure teachers' AL in earlier research (Mertler, 2004; Plake et al., 2005).

Besides the seven assessment competences above, current research has suggested that to capture a better picture of teachers' AL, investigation should cover multiple dimensions of the AL construct, including assessment conceptions/beliefs, assessment practices, and assessment confidence, as teachers' AL is driven by contexts, affective factors, cognition, assessment learning, and assessment identity (DeLuca et al., 2020; Xu & Brown, 2016). Drawing on Xu and Brown's (2016) conceptualization, assessment knowledge is treated as a core component of AL, while assessment conceptions, beliefs, practices, attitudes, and values are conceptualized as interrelated elements that shape how assessment knowledge is enacted in practice. Empirical evidence supports this overarching conceptualization by verifying the relationships between assessment knowledge, practices, and attitudes (Tajeddin et al., 2022; Quilter & Gallini, 2000) and revealing that teachers' reflection on educational goals and their beliefs about assessment, collaborative

practice, and peer feedback might foster higher mastery levels in their AL (Bijsterbosch et al., 2019).

Recent empirical research conducted among PTs also indicates the multidimensional nature of AL and suggests that the development of PTs' AL is shaped by personal, contextual, and external factors. Personal factors, including prior experiences as students, influence PTs' initial assessment conceptions, often leading them to believe that assessment is mainly used for evaluating students rather than for improving teaching and learning (Xu & Brown, 2016). However, targeted training can shift these conceptions toward more positive, formative-oriented views after being exposed to AL training (Levy-Vered & Alhija, 2018; Levy-Vered et al., 2022; Xu & He, 2019). Therefore, contextual factors within teacher education programs, including specific assessment courses and AL training, are critical for enhancing assessment knowledge and shaping assessment conceptions (Atjonen et al., 2022). However, the investigation of the relationships between AL training and assessment conceptions (Xu & Brown, 2016) revealed that conceptions of assessment can either hinder or enhance AL training: if assessment conceptions are treated as independent from AL, participants tend to resist the training. The integration of assessment practices with the development of assessment knowledge and conceptions is essential to support the understanding of different assessment purposes, like assessment of, for, and as learning (Lam, 2019). Finally, external factors, such as supervising agency, school context, and available resources, can influence PTs' classroom assessment decision-making processes (Oo et al., 2021). Therefore, considering these factors is crucial to provide an integrated synthesis of how PTs' AL is operationalized, conceptualized, and developed.

1.4. Research gap and relevance of the present review

While the review literature on AL is substantial, it has mainly focused on in-service teachers. Existing reviews on teachers' AL focused on five broad AL areas: (1) how AL is measured in different contexts (DeLuca et al., 2016b; Gotch & French, 2014; Juanjuan & Mohd Yusoff, 2022), (2) how it develops (Giraldo, 2021; Oo et al., 2022), (3) how it is conceptualized (Singh et al., 2022; Weng & Shen, 2022; Xu & Brown, 2016), (4) how it is defined in different areas and at different levels of education (Coombs & DeLuca, 2022; Pastore, 2022), and (5) what study area trends impact teachers' AL (Puspawati, 2022; Qotboddin et al., 2020; Wang et al., 2023). Previous review studies focused on language teachers' AL (Giraldo, 2021; SeviMel-ŞahiN & Subaşı, 2019; Wang et al., 2023) and general schoolteachers' AL (Coombs & DeLuca, 2022; Khalid et al., 2021; Oo et al., 2022). Past review studies on AL have targeted mainly in-service teachers, yet only a few have focused on PTs' AL (Oo et al., 2022), leading to less consideration of the unique characteristics of PTs.

The decision to focus specifically on PTs' AL in this review is grounded in the professional differences between PTs and in-service teachers due to their assessment training, experience, and competency levels. As Juanjuan and Mohd Yusoff (2022) argue, AL acquired during teacher education might differ from that of in-service teachers, as PTs are in the process of reflecting on the experiences they gained as former students, learning and practicing competencies required in their future daily teaching practices. Synthesizing AL at this stage is crucial to addressing the gap between assessment theory and practice that PTs often experience during their training (SeviMel-ŞahiN & Subaşı, 2019). Furthermore, teacher education programs aimed at improving teachers' assessment abilities vary according to contexts, objectives, and professional development approaches (Giraldo, 2021). Additionally, most training programs were for in-service teachers (e.g., Baker & Riches, 2018; Giraldo, 2021; Nier et al., 2009), but fewer programs were dedicated to student-teachers (Walters, 2010) and PTs (Bolivar, 2020; Giraldo & Murcia, 2019; Jaramillo Delgado & Gil Bedoya, 2019; Walters, 2010). While Oo et al. (2022) provided a valuable review of AL content in teacher education, systematic evidence on how PTs' AL is

operationalized and conceptualized remains limited. Therefore, this systematic review aims to fill this gap by establishing guiding principles for the development of future PT AL instruments and by providing a conceptual framework that synthesizes factors influencing AL development in teacher education.

2. Research questions

Building on the identified conceptual and empirical gaps, the present review addresses the following research questions:

1. What methodological approaches and instruments have been employed to operationalize and examine PTs' AL in past studies?
2. How do personal, contextual, and external factors contribute to the conceptualization and development of PTs' AL?

3. Research methodology

This study reports a systematic literature review synthesizing empirical research investigating the operationalization and

development of PTs' AL. The integrative conceptualization of AL outlined above informed the search strategy, coding scheme, and synthesis procedures of the present review. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol (Page et al., 2021) was followed to identify empirical studies, describe their characteristics, summarize findings, and formulate recommendations for future research. Fig. 1 presents the PRISMA flow diagram illustrating the identification, screening, eligibility assessment, and inclusion of studies, with detailed reasons for full-text exclusions.

3.1. Identification

Empirical studies were included if they were published in English in peer-reviewed journals indexed in one of the Web of Science Core Collections, such as the Science Citation Index Expanded, the Social Science Citation Index, or the Emerging Sources Citation Index, and/or listed in the Scimago journal ranking. Scopus in combination with the EBSCO-host platform, Academic Search Complete, Education Resources Information Center, and Complementary Index were used to identify relevant sources. These databases were last searched on July 25, 2025.

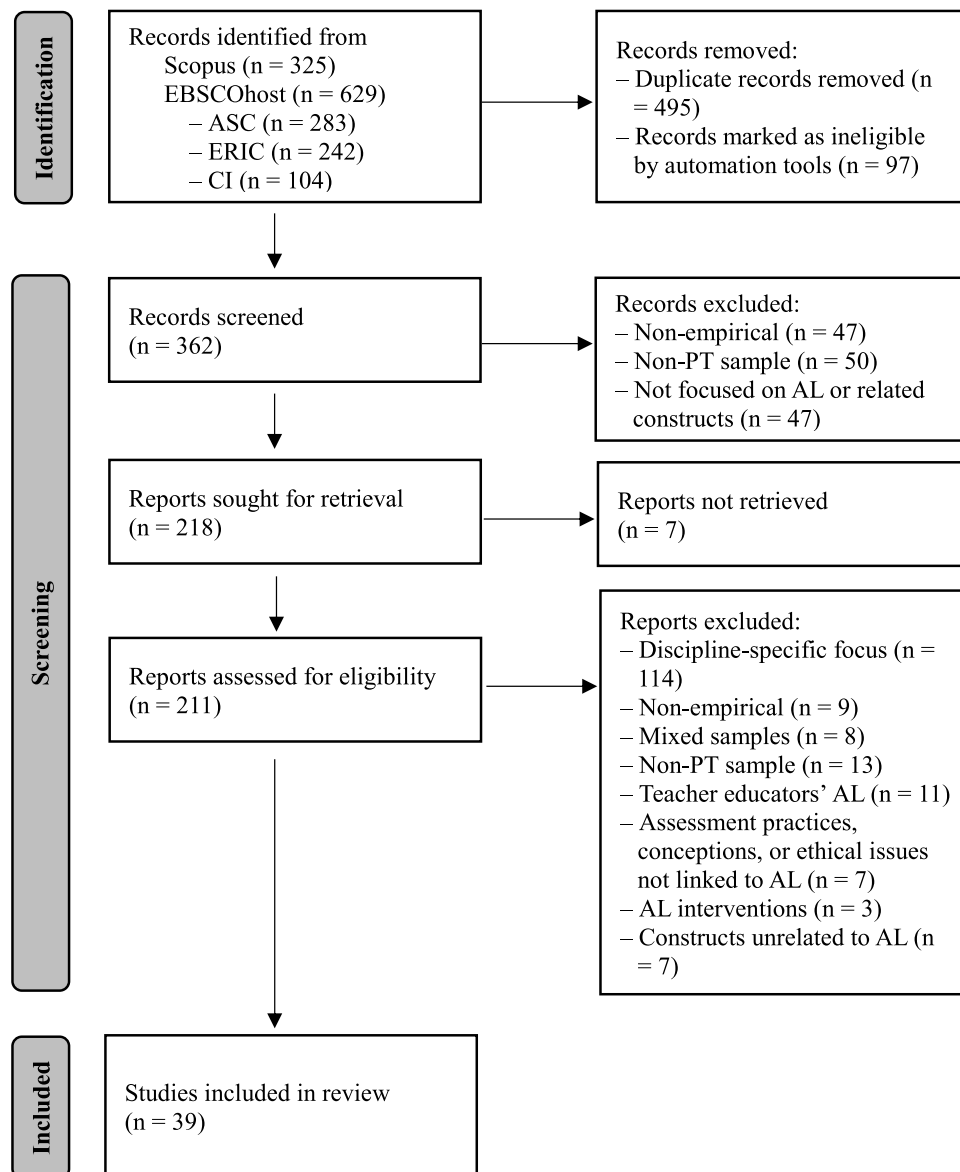


Fig. 1. Summary of the Literature Search and Review Process. Note. EBSCO = Elton B. Stephens Company; ASC = Academic Search Complete; ERIC = Education Resources Information Center; CI = Complementary Index; AL = assessment literacy; PT = preservice teacher.

Table 1 includes the search terms organized into two categories related to PTs as participants and AL or other similar relevant constructs. Boolean operators were used to link the two inclusion categories with AND and the keywords within each category with OR. The search equations are shown in Table 2. No year limitation was used during the database search, which was conducted by title, abstract, and keywords. As shown in Fig. 1, the initial search resulted in 954 records.

3.2. Screening

After removing duplicates and records marked ineligible by automation filters, 362 records remained for screening based on titles, abstracts, and keywords. Screening was conducted in two steps. First, the study characteristics were checked, and records were excluded if they were nonempirical or if PTs were not involved. Second, the construct was examined, and the record was excluded if the article did not focus on AL or another relevant construct included in Table 1. The record was also excluded if only PTs' assessment practice was examined without connecting it with AL or with another similar construct.

Using this decision process, two reviewers first independently screened the titles and abstracts of the first 100 records to determine which records to exclude or include from the further process. After that, their decisions were compared, and discrepancies were solved via discussion. Second, the reviewers independently screened the rest of the records and calculated Cohen's kappa, indicating substantial agreement between them ($\kappa = .813$; raw agreement = 88.9%). At this stage, 144 records were excluded for the reasons presented in Fig. 1, leaving 218 potentially eligible articles, of which seven could not be retrieved.

Eligibility assessment was conducted through full-text review using a coding scheme collaboratively developed by the reviewers through both deductive and inductive procedures. Deductive categories were determined based on conceptualizations of AL (see Section 1.2), while inductive coding through iterative comparison ensured the identification of recurring constructs across the included studies. As shown in the coding framework (Table 3), extracted information included study characteristics, sample features (participant group, educational level, subject focus, country, and sample size), and research design. To address the first research question regarding methodological operationalization of PTs' AL, studies were coded according to their focus, instrument type, and key characteristics. To address the second research question on AL conceptualization, studies were coded for focal AL constructs, related assessment constructs, development orientation and context, and stage of investigation. Using the coding scheme, a multi-stage coding process was employed. First, the two reviewers piloted the scheme on fifteen articles to ensure alignment in their decisions. Second, the reviewers independently assessed the articles for eligibility based on subject focus, research design, sample type, focal AL construct, and related assessment constructs; their inclusion and exclusion decisions were compared ($\kappa = .829$; raw agreement = 91.4%), and disagreements were resolved through discussion. Third, the first author extracted data from the included studies using the deductively and inductively developed

Table 1
Search Terms and Limits Used During the Search Phase.

Category	Related terms or limits
Inclusion categories	
Constructs	assessment literacy, assessment competency, assessment competence, assessment knowledge, assessment identity, assessment capability, assessment conception, assessment practice, assessment efficacy
Stakeholders	preservice teachers, teacher candidates, prospective teachers, teacher trainees, student teachers
Filters	
Language	English
Publications	Journal articles

Table 2
Search Equations.

Platform	Equation
Scopus	(TITLE-ABS-KEY ("assessment literacy" OR "assessment competenc*" OR "assessment knowledge" OR "assessment identity" OR "assessment capability" OR "assessment conception*" OR "conception* of assessment" OR "assessment practice*" OR "assessment efficacy") AND TITLE-ABS-KEY ("preservice teacher*" OR "pre-service teacher*" OR "teacher candidate*" OR "prospective teacher*" OR "teacher trainee*" OR "student teacher*")) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (SRCTYPE, "j"))
EBSCO	(TI ("assessment literacy" OR "assessment competenc*" OR "assessment knowledge" OR "assessment identity" OR "assessment capability" OR "assessment conception*" OR "conception* of assessment" OR "assessment practice*" OR "assessment efficacy") OR AB ("assessment literacy" OR "assessment competenc*" OR "assessment knowledge" OR "assessment identity" OR "assessment capability" OR "assessment conception*" OR "conception* of assessment" OR "assessment practice*" OR "assessment efficacy") OR KW ("assessment literacy" OR "assessment competenc*" OR "assessment knowledge" OR "assessment identity" OR "assessment capability" OR "assessment conception*" OR "conception* of assessment" OR "assessment practice*" OR "assessment efficacy")) AND (TI ("preservice teacher*" OR "pre-service teacher*" OR "teacher candidate*" OR "prospective teacher*" OR "teacher trainee*" OR "student teacher*") OR AB ("preservice teacher*" OR "pre-service teacher*" OR "teacher candidate*" OR "prospective teacher*" OR "teacher trainee*" OR "student teacher*") OR KW ("preservice teacher*" OR "pre-service teacher*" OR "teacher candidate*" OR "prospective teacher*" OR "teacher trainee*" OR "student teacher*"))

Table 3
Coding Framework for Eligibility Assessment, Data Extraction, and Synthesis.

Coding category	Description
Study characteristics	References, purpose, journal quartile (Q1–Q4), WoS CC indexing
Sample type	Participant group (e.g., preservice teachers, prospective teachers, student teachers, teacher candidates)
Educational level	Teaching level targeted (primary, secondary, or mixed)
Subject focus	General (cross-disciplinary or non-specified subject focus) or discipline-specific focus
Sample size	Number of participants included in the study
Research country	Country or region where data were collected
Research design	Quantitative, qualitative, or mixed-methods design
Focal AL construct	Primary construct investigated (e.g., assessment literacy, competence, capability, or identity)
Related assessment constructs	Constructs examined in relation to AL (e.g., assessment conceptions, beliefs, efficacy, practices, attitudes, learning, culture, supervising agency)
Methodological focus	Overall AL or specific AL subcomponents
Instrument type	Data collection method (e.g., questionnaire, test, content analysis)
Instrumentation features	Underlying framework, AL components, number of items, scale type, and reported psychometric properties
Development orientation	Product-oriented (investigation conducted at a single time point) or process-oriented (investigating development through cross-sectional or longitudinal designs)
Development context	Course-based or program-based AL development
Stage of investigation	Timing (before, during, or after course or program)

Note. AL = assessment literacy; WoS CC = Web of Science Core Collection

categories to generate a descriptive overview and address the research questions. To enhance coding reliability, the second author systematically reviewed the extracted data against the full texts. This approach was adopted because many coding categories (Table 3) relied on explicitly reported study characteristics. For most categories, no discrepancies were identified. Discrepancies occurred in the more interpretive categories, including development orientation and stage of investigation, particularly when relevant study details were not explicitly reported. The initial aggregate raw agreement between reviewers was 94.6%, with lower agreement observed in these interpretive categories. Discrepancies in the categorization were resolved through

consensus-based discussion of the coding decisions, resulting in full agreement on the final dataset.

Of the 211 full-text articles, 172 studies were excluded from further analysis. The primary reasons for exclusion included a discipline-specific focus, nonempirical designs, and mixed or non-PT samples. A detailed list of exclusion reasons and corresponding frequencies is presented in Fig. 1. Studies focusing exclusively on discipline-specific AL (e.g., language, social studies, chemistry, physics, and physical education) were excluded because AL is conceptualized and operationalized differently across subject domains, often using domain-specific frameworks and instruments (Puspawati, 2022; Xu & Liu, 2025). Including such studies would have limited the comparability of findings and the coherence of synthesis; therefore, they fall beyond the scope of the present review. Studies addressing assessment practices, conceptions, or ethical issues in assessment were included only when these constructs were explicitly situated within an AL framework to ensure alignment with the second research question. Therefore, seven studies examining these constructs in isolation were excluded to maintain alignment with the research questions. For similar reasons, a further seven studies were excluded because they focused on constructs unrelated to AL, such as PTs' perceptions of pedagogical courses, microteaching effects, data literacy development, social loafing, and competency-based assessment in PTs' learning.

3.3. Synthesis of included studies

The 39 included studies were synthesized using two approaches: analyzing construct definition and research methodology and summarizing common themes. For each study, extracted data were systematically compared across studies to identify similarities and differences in methodological approaches and conceptual linkages, which were then aggregated into higher-order themes aligned with the research questions. Accordingly, studies on various instruments for PTs' AL were first summarized and synthesized, followed by studies conceptualizing PTs'

AL in relation to other assessment-related constructs.

The complete coding dataset for the included studies, covering all categories presented in the coding framework (Table 3), is provided as Supplementary Material. Tables 4 and 5 synthesize results related to the instruments of the overall AL and its sub-components, including their scope, content, and psychometric properties. Table 6 summarizes studies (n = 35) examining relationships between AL and related constructs, categorized across cognitive, affective, and contextual domains. The synthesis yielded two overarching themes: (1) the methodological approaches and instruments employed in PTs' AL research, and (2) the conceptual integration of AL through the interplay of personal, contextual, and external factors.

4. Results

4.1. Overview of included studies

The final synthesis comprised 39 empirical studies. A substantial portion focused on assessment conceptions (n = 16) and AL (n = 13), while others examined related constructs, including assessment competence (n = 3), practices (n = 2), capability (n = 1), knowledge (n = 1), beliefs (n = 2), and attitudes (n = 1).

Participants were predominantly referred to as PTs (n = 20), and other synonymous terms, including teacher candidates (n = 9), prospective teachers (n = 5), student-teachers (n = 4), and teacher trainees (n = 1), were also employed, reflecting a consistent focus on the initial phase of teacher development. Regarding research design, most studies employed quantitative methods (n = 29), followed by qualitative (n = 6) and mixed-method approaches (n = 4).

4.2. Methodological approaches and instruments for examining PTs' AL

To address the first research question, we synthesize the methodological approaches and instruments used to examine PTs' AL. Across the

Table 4
Instruments and Studies Examining Overall AL among PTs.

Instrument	Frameworks and Components	Type (Scale)	Items	References	Country	N	Reliability (α/KR20)
Teacher Assessment Literacy Questionnaire (TALQ)	1990 Standards: (1) Choosing ass. methods, (2) Sound design, (3) Scoring, (4) Using results, (5) Grading, (6) Communicating results, (7) Ethical ass.	Test (MC)	35	Alkharusi (2011a, b)	Oman	259	.84
Assessment Literacy Inventory (ALI)	1990 Standards (see above for components)	Test (MC)	35	McGee and Colby (2014)	USA	190	.75
			32	Sondergeld et al. (2020)	USA	96	.75
Educational Assessment part of KOSTA	KOSTA Standards: (1) Ass. techniques, (2) Recognition of heterogeneity, (3) Awareness of observational bias	Quest. (6-pt rating)	13	Schneider and Bodensohn (2017)	Germany	931	.70-.81
Assessment Literacy Questionnaire	(1) Beliefs about and understanding of ass., (2) Planning ass. confidence, (3) Confidence in conducting ass., (4) Confidence in using ass., (5) Ethical issues in ass.	Quest. (5-pt Likert)	35	Oo et al. (2023)	Myanmar	335	Not reported
Teacher Educational Assessment Competencies (TEAC)	(1) ACK, (2) Practical ass. skills, (3) ICT competencies for educational ass., (4) Competencies related to attitudes, behaviors, and ethics	Quest. (5-pt Likert)	43	Hoang et al. (2025)	Vietnam	314	.91-.99
Assessment Literacy Survey	(1) Self-described level of AL, (2) Main purposes of ass., (3) Utilization of different ass. methods, (4) Need for further training, and suggested methods for promoting AL	Quest. (Open and 10-pt rating)	9	Volante and Fazio (2007)	Canada	69	Not reported
Approach to Classroom Assessment Inventory (ACAI)	Contemporary ass. standards: (1) Ass. Purpose, (2) Process, (3) Fairness, (4) Theory	Scenario	60	DeLuca et al. (2020)	Canada	206	Not reported
				Coombs et al. (2020)	Germany	182	> .85
				DeLuca et al. (2019)	Canada	457	> .82
				Schneider et al. (2020)	USA	396	> .82
				Schneider et al. (2020)	Germany	182	Not reported
					Canada	206	

Note. Quest. = questionnaire; AL = assessment literacy; ass. = assessment; N = Sample size; MC = Multiple choice; PCK = Pedagogical content knowledge; ACK = Assessment content knowledge; AFL = Assessment for learning; AoL = Assessment of learning; ICT = Information and communication technology; 1990 Standards = Standards for Teacher Competence in the Educational Assessment of Students (AFT et al., 1990); KOSTA = Educational Assessment Part of Competence and Standard Orientation in Teacher Education; USA = United States of America.

Table 5
Questionnaires and Studies Examining AL Subcomponents among PTs.

Instrument	Frameworks and Components	Scale Type	Items	References	Country	N	Reliability (α /KR20)
Beliefs about Assessment Questionnaire	(1) Principles of AfL and AoL, (2) Ass. rigor, (3) Ass. methods	5-pt Likert + Open	46 + 5	Hill et al. (2017)	NZ	224	Not reported
		5-pt Likert	22	Chen and Cowie (2016)	China	531	.68-.76
Assessment Belief Questionnaire	Revised National Curriculum in South Africa: (1) Clarity of focus, (2) High expectations for learner success, (3) Providing expanded learning opportunities, (4) Designing down, (5) Aligning outcomes, teaching, and ass.	5-pt Likert + Open	32	Vandeyar and Killen (2006)	South Africa	48	Not reported
Teacher Conceptions of Assessment Inventory (TCoA)	(1) Improvement, (2) School accountability, (3) Student accountability, (4) Irrelevance	6-pt Likert	25	Izci and Caliskan (2017)	Turkey	118	.69-.95
		6-pt Likert	22	Tessema et al. (2024)	Ethiopia	197	.84
TCoA-IIIa (Abridged version)	Adapted (TCoA-C): (1) Improvement, (2) Accountability, (3) Irrelevance	6-pt Likert	53	Pastore (2020)	Italy	409	.39-.89
		6-pt Likert	31	Chen and Brown (2013)	China	765	Not reported
	Adapted: (1) Ass. for teaching and learning, (2) AoL, (3) Ass. as harmful action	7-pt Likert	20	Kyttälä et al. (2022)	Finland	287	.69-.91
		6-pt Likert	27	Levy-Vered and Alhija (2018)	Israel	297	.75-.78
	(1) Improvement, (2) School accountability, (3) Student accountability, (4) Irrelevance	6-pt Likert	27	Brown and Remesal (2012)	Spain NZ	672	.18-.83
		6-pt Likert	27	Reeves et al. (2023)	USA	554	.60-.82
		6-pt Likert	27	Daniels et al. (2014)	Canada	436	.60-.73
5-pt Likert	24	Daniels and Poth (2017)	Canada	388	.34-.70		
6-pt Likert	27	Rasooli et al. (2023)	Canada	248	.50-.85		
Students' Conceptions of Assessment (SCoA) Questionnaire	(1) Student accountability, (2) School accountability, (3) Ass. is fun, and (4) Ass. is ignored	6-pt Likert	11	Al-Haddad et al. (2023)	Bahrain	278	.77-.85
Teachers' Classroom Assessment Scale (TCAS)	(1) Making learning explicit, (2) Promoting autonomy, (3) Performance orientation	5-pt Likert	21	Eren (2010)	Turkey	304	.78-.86
		5-pt Likert	21	Eren (2013)	Turkey	423	.82-.88
Confidence Levels in Educational Assessment Knowledge	(1) Ass. theory, (2) Ass. practice, (3) Ass. philosophy	5-pt Likert	45	DeLuca and Klinger (2010)	Canada	288	.55-.93
Educational Assessment Confidence Inventory	(1) Confidence in using practical ass. approaches to measure student learning, (2) Confidence in engaging in ass. praxis	5-pt Likert	19	DeLuca, Chavez, and Cao (2013)	USA	48	.50-.89

Note. Quest. = questionnaire; AL = assessment literacy; ass. = assessment; AfL = Assessment for Learning; AoL = Assessment of Learning; N = Sample size; USA = United States of America; NZ = New Zealand.

reviewed studies, researchers have operationalized AL through two distinct approaches: comprehensive instruments targeting overall AL and tools focusing on AL subcomponents. Comprehensive instruments were used to examine PTs' competence in several aspects of assessment, including their ability to design, implement, and communicate assessment results (Juanjuan & Mohd Yusoff, 2022), often covering assessment knowledge and other assessment-related constructs such as assessment practices, conceptions, and efficacy. In contrast, targeted instruments focus on assessment-related constructs encompassed within AL, such as assessment efficacy, conceptions, and identity (Xu & Brown, 2016), and examine one or more constructs beyond core assessment competence (e.g., an assessment conception or assessment practices). We analyze all AL tools on the basis of three main criteria: AL content, investigation mode (e.g., through testing or self-reporting), and reported psychometric properties.

4.2.1. Instruments targeting overall AL

As shown in Table 4, seven instruments were used to examine PTs' overall AL. These instruments can be categorized according to their methodological approaches and formats.

The Teacher AL Questionnaire (TALQ), developed by Plake and Impara (1992), and its revised version, the AL Inventory (ALI) (Mertler & Campbell, 2005), consist of multiple-choice items to measure AL with correct and incorrect responses grounded in the AFT et al. (1990) assessment competency standards. These tests demonstrated varying

reliability across different samples; while Plake et al. (2005) reported an average reliability for in-service teachers (KR20 = .54), studies involving PTs generally yielded higher estimates, ranging from .75 to .84 (Alkharusi, 2011a, 2011b; Campbell et al., 2002; Mertler & Campbell, 2005). Consequently, their application in empirical research (McGee & Colby, 2014; Sondergeld et al., 2020) often required adaptations in item wording and descriptions to improve content validity and align with the specific content taught to PTs.

Other researchers took a different methodological approach and examined PTs' perceived and/or self-reported AL with questionnaires. Schneider and Bodensohn (2017) developed the Educational Assessment Part of Competence and Standard Orientation in Teacher Education (KOSTA), targeting German PTs' self-reported assessment competence across three factors (see Table 4); these subscales' reliability estimates were reported to be .81, .70, and .74, respectively. The following three questionnaires have been developed in Asian countries. The AL Questionnaire, originally developed by Nguyễn (2016) for preservice English as a foreign language teachers, focuses on assessment beliefs and confidence levels, including ethical issues in assessment practices as an expanded AL component. Because the instrument demonstrated high reliability (Cronbach's $\alpha = .85$), Oo et al. (2023) adapted it to investigate PTs' AfL literacy improvement in a Myanmar professional development program, focusing on five dimensions of assessment confidence and ethics (see Table 4). A recent questionnaire, developed by Hoang et al. (2025), was used to examine Vietnam PTs' assessment competencies in

Table 6
Studies on PTs' AL and Related Constructs.

References	Investigation purpose	Country	N	Focal AL construct	Development orientation	Development context	Investigation stage	Related constructs
I. Personal factors								
<i>Prior experiences and epistemological beliefs</i>								
1. Brown and Remesal (2012)	To examines prospective teachers' assessment conceptions in Spain and New Zealand	New Zealand, Spain	996	Assessment conceptions	Process-oriented	Course-related	Beginning and during the course	Training and prior experience
2. McGee and Colby (2014)	To investigate the impact of a required assessment course on the teacher candidates' AL	USA	190	AL	Process-oriented	Course-related	Before and after the course	Training and prior experience
3. Tessema et al. (2024)	To examine the relationship between assessment conceptions and epistemological beliefs	Ethiopia	197	Assessment conceptions	Product-oriented	Program-related	During the program	Epistemological beliefs
4. Chen and Cowie (2016)	To investigate how PTs in China perceive assessment	China	531	Assessment beliefs	Process-oriented	Program-related	During the program	Experience and culture
5. Atjonen et al. (2022)	To examine Finnish PTs' AL through assessment module	Finland	168	AL	Product-oriented	Course-related	During the course	Training and experience
<i>Psychological factors: mindset, attitudes, and efficacy</i>								
6. DeLuca et al. (2019)	To explore the relationship between teacher candidates' approaches to assessment and their mindset	Canada	396	AL	Product-oriented	Program-related	End of the program	Mindset
7. Alkharusi (2011a)	To explore the internal and external structure of TALQ	Oman	259	AL	Product-oriented	Course-related	End of the course	Attitude, self-confidence
8. Ng et al. (2020)	To explore if deep learning approach of peer assessment on peer assessment improves PTs' AL	China	9	AL	Process-oriented	Course-related	During the course	Attitude
9. Schneider et al. (2020)	To examine the relationship between personality traits and AL	Canada Germany	388	AL	Process-oriented	Program-related	During the program	Efficacy
10. Volante and Fazio (2007)	To examine the AL of primary/junior teacher candidates in all four years of their concurrent program	Canada	69	AL	Product-oriented	Program-related	End of the program	Efficacy
11. Alkharusi (2010)	To investigate whether self-efficacy and self-concept relate to attitude toward educational measurement	Oman	120	Assessment attitudes	Process-oriented	Course-related	Before and after the course	Self-efficacy and self-concept
<i>Assessment and teaching conceptions</i>								
12. Daniels and Poth (2017)	To examine relationships between assessment conceptions, approaches to instruction and approaches to assessment in PTs	Canada	388	Assessment conceptions	Process-oriented	Course-related	Beginning of the course	Approach to assessment
13. Chen and Brown (2013)	To investigate how prospective teachers conceive of teaching excellence and assessment purposes	China	765	Assessment conceptions	Process-oriented	Program-related	During the program	Conceptions of teaching excellence
14. Eren (2010)	To examine the consonance and dissonance between prospective teachers' values and practices of their conceptions about teaching/ learning and assessment	Turkey	304	Assessment conceptions	Process-oriented	Program-related	During the program	Conceptions of teaching and learning
15. Al-Haddad et al. (2023)	To examine assessment conceptions and their relationship with self-regulation and self-confidence	Bahrain	278	Assessment conceptions	Product-oriented	Program-related	During the program	Self-regulation and self-confidence
16. Reeves et al. (2023)	To examine the validity of Conceptions of Assessment III Abridged instrument with US PTs and in-service teachers	USA	554	Assessment conceptions	Product-oriented	Course-related	During the course	Cultural differences in assessment conceptions
17. Levy-Vered and Alhija (2018)	To explore PTs' assessment conceptions and examine whether and how they change during an e-learning basic assessment course	Israel	297	Assessment conceptions	Process-oriented	Course-related	Before and after the course	Assessment learning
II. Contextual factors								
<i>Assessment training and professional development</i>								
18. Gallego-Noche et al. (2022)	To investigate the impact of PTs' ideas and knowledge construction of assessment	Spain	24	Assessment knowledge	Process-oriented	Course-related	Before and after the course	Training

(continued on next page)

Table 6 (continued)

References	Investigation purpose	Country	N	Focal AL construct	Development orientation	Development context	Investigation stage	Related constructs
19. Hill et al. (2017)	To investigate teacher candidates' AL in their student assessment practices	New Zealand	224	Assessment capability	Process-oriented	Program-related	Beginning and during the program	Training
20. Lomax (1996)	To examine the level of PTs' AL before and after their assessment course and their assessment use	USA	6	AL	Process-oriented	Course-related	Before and after the course	Training
21. Sondergeld et al. (2020)	To examine the establishment of AL growth in PTs' achievement	USA	96	AL	Process-oriented	Course-related	Before and after the course	Training
22. Oo et al. (2023)	To explore the utilization of a needs-based professional development program to build PTs' assessment for learning literacy	Myanmar	335	AL	Process-oriented	Program-related	Before and after the program	Professional development
23. DeLuca, Chavez, Bellara, and Cao (2013)	To examine pedagogies that support positive changes in teacher candidates' assessment conceptions	USA	97	Assessment conceptions	Process-oriented	Course-related	Beginning, in the middle, and at the end of the course	Pedagogical constructs and training methods
24. DeLuca and Klinger (2010)	To examine the effect of training on teacher candidates' AL	Canada	288	AL	Product-oriented	Course-related	End of the course	Training
25. Izci and Caliskan (2017)	To explore the impact of an assessment course on prospective teachers' assessment conceptions and choices of assessment tasks	Turkey	118	Assessment conceptions	Process-oriented	Course-related	Before and after the course	Assessment training
26. Schneider and Bodensohn (2017)	To explore views on how student teachers perceive assessment	Germany	931	Assessment competence	Product-oriented	Program-related	During the program	Training
<i>Learning culture and assessment practices</i>								
27. DeLuca et al. (2020)	To investigate the influence of education culture on student teachers' developing assessment competence	Canada, Germany	388	Assessment competence	Process-oriented	Course-related	During the course	Culture
28. Coombs et al. (2020)	To investigate factors that influence different groups of teacher candidates' approaches to classroom assessment	Canada	457	Assessment competence	Product-oriented	Program-related	End of the program	Learning experience and context
29. Eren (2013)	To examine whether prospective teachers' efficacy beliefs mediate the relationships between their assessment values and practices	Turkey	423	Assessment conceptions	Product-oriented	Program-related	During the program	Efficacy and practices
30. Rasooli et al. (2023)	To explore beginning teacher candidates' approaches to grading in relation to their broader assessment conceptions	Canada	248	Assessment conceptions	Product-oriented	Program-related	Beginning of the program	Approach to grading
III. External factors								
<i>Educational policy and supervising agencies</i>								
31. Daniels et al. (2014)	To examine relationships between PTs' assessment conceptions, approaches to instruction and approaches to assessment	Canada	436	Assessment conceptions	Process-oriented	Course-related	Beginning of the course	Context
32. Pastore (2020)	To investigate how teacher trainees value assessment	Italy	409	Assessment conceptions	Process-oriented	Course-related	During the course	Context and accountability
33. Vandeyar and Killen (2006)	To identify the beliefs, perceptions and attitudes about assessment prior to the assessment course	South Africa	48	Assessment beliefs	Product-oriented	Course-related	Before the course	Context and principles
34. Liu et al. (2016)	To investigate the perceptions of PTs in the US and China about the ethics of classroom assessment practices	USA, China	347	Assessment practice ethics	Product-oriented	Course-related	During the course	Culture
35. Oo et al. (2021)	To explore how PTs implement their assessment and their decision making	Myanmar	30	Assessment practice beliefs	Product-oriented	Program-related	End of the program	Supervising teachers

Note. AL = Assessment literacy; PT = Preservice teacher; N = Total sample size. The table contains data from 35 studies, as four studies (Alkharusi, 2011b; DeLuca, Chavez, & Cao, 2013; Hoang et al., 2025; Kyttälä et al., 2022) were excluded because they focused exclusively on the development or validation of AL instruments.

the context of information and communication technology (ICT). The instrument comprises four main components (Table 4), including ICT competencies in assessment, as a novel component absent from previous standards. The questionnaire assesses PTs' competence levels using a five-point Likert scale (1 = novice, 5 = expert), and it demonstrated strong internal consistency (Cronbach's $\alpha = .99$) and sampling adequacy

(KMO =.98), with subscale reliabilities ranging from .91 to .97.

A smaller subset of studies used open-ended questionnaires to investigate self-reported AL. Volante and Fazio (2007) designed a questionnaire in Canada based on the works of Earl (2003) and Earl and Katz (2005), covering three assessment conceptions in relation to AfL, assessment as learning (AaL), and assessment of learning (AoL).

However, the questionnaire's psychometric properties were not reported due to its open-ended format.

Rather than relying on self-reported AL instruments, scenario-based methodologies were also used to examine PTs' assessment decisions in authentic contexts. The Approach to Classroom Assessment Inventory (ACAI), developed by DeLuca et al. (2016a), contains three parts. The first part presents participants with practice-relevant scenarios, covering four aspects of the approach to assessment: assessment purposes, processes, fairness, and measurement theory (Table 4). In the second part, 12 items asked PTs to self-rate their assessment proficiency on a five-point scale ranging from 1 (novice) to 5 (expert). In the final part, 21 items targeting professional learning priorities and preferences in assessment were included in the questionnaire. In DeLuca et al.'s (2016a) study, the instrument was validated by experts, and the reliability estimates ranged from .74 to .92 across all subscales. The ACAI has been widely used to compare PTs' AL in different cultural contexts (DeLuca et al., 2020; Schneider et al., 2020), to examine AL in relation to PTs' mindset (DeLuca et al., 2019), personality traits (Schneider et al., 2020), as well as in relation to methods of assessment, motivation for teacher education, and their self-perception of assessment competence (Coombs et al., 2020).

The results of the present review (see Table 4) indicate that methodological approaches to operationalizing PTs' AL have evolved from a predominant focus on examining assessment knowledge or competence through testing toward broader approaches that include instructional knowledge, perceived assessment practices, and ICT-related assessment knowledge, often relying on self-reported and self-perceived questionnaires. While this shift reflects a more comprehensive operationalization of AL, it is also associated with greater variability in how AL is implemented. Notably, many instruments appear to be closely tied to specific curriculum content and contextual practices, which may limit their comparability across studies.

4.2.2. Instruments targeting AL subcomponents

As presented in Table 5, this review identified seven instruments focusing on specific AL subcomponents, including assessment beliefs, conceptions, and efficacy or confidence. As all tools employ a questionnaire format, the following analysis organizes them according to their specific focus to highlight how these AL subcomponents are operationalized across research contexts.

The investigation of assessment beliefs mainly focuses on how PTs perceive the purpose and impact of assessment, often contrasting formative and summative assessment. Hill et al. (2014) developed the Beliefs about Assessment Questionnaire to examine primary and early childhood PTs' assessment beliefs related to principles of AfL and AoL, assessment rigor, and assessment methods. When Hill et al. (2017) used this instrument with PTs in New Zealand, they identified four interrelated factors: assessment as irrelevant/negative, assessment as formal/summative, the role of teachers in AaL, and the role of students in AaL. However, the cultural differences of these beliefs are demonstrated by Chen and Cowie (2016), who reported that the original model did not fit their sample of Chinese PTs. Instead, the original questionnaire required item reduction and factor restructuring to identify a 22-item, four-factor structure (assessment to trace development and inform action, assessment for various accountability purposes, teacher assessment knowledge, and negative aspects of assessment) with acceptable model fit indices ($\chi^2/df = 1.96$, $p > .05$; CFI = .90; RMSEA = .042) and reliability (Cronbach's $\alpha = .68-.76$). These findings highlight that PTs' beliefs may depend on their local educational context. Consequently, instrument development was grounded on national curriculum standards, as evident in Vandeyar and Killen's (2006) questionnaire, which was based on the revised national curriculum of South Africa. Although it covers five core assessment beliefs (see Table 5), its psychometric properties remained unreported, suggesting a need for its further validation.

A structurally complex approach is found in the study of assessment

conceptions, which is dominated by the Teacher Conceptions of Assessment Inventory (TCOA), developed by Brown (2002, 2004), examining four interrelated conceptions: assessment as improvement, school accountability, student accountability, and irrelevant. As can be seen in Table 5, this instrument has been used in various studies, often resulting in different functioning. While Izci and Caliskan (2017), Tessema et al. (2024), and Pastore (2020) confirmed its high reliability (average Cronbach's $\alpha = .86$), Chen and Brown (2013) adapted it for the Chinese context by aggregating the original model into three dimensions. Similarly, Kyttälä et al. (2022) adapted this questionnaire for the Finnish context, using 10 TCoA items along with 10 context-specific items, resulting in a three-factor model (Table 5) with acceptable fit indices ($\chi^2/df = 2.46$, CFI = .93, RMSEA = .06) and reliability (Cronbach's $\alpha = .69-.91$). The complexity of cross-cultural validation of these conceptions is further evidenced by studies using the abridged version, the TCoA-IIIa (Brown, 2006). While Levy-Vered and Alhija (2018) reported acceptable fit of the original model for Israeli PTs (pretest: $\chi^2/df = 2.89$, CFI = .96, RMSEA = .08; posttest: $\chi^2/df = 2.89$, CFI = .96, RMSEA = .08) and acceptable reliability (Cronbach's $\alpha = .75-.78$), other studies encountered validation challenges. Brown and Remesal (2012) found unacceptable fit indices in New Zealand and Spain, and Reeves et al. (2023) reported similar validity issues in the U.S. context. Daniels and Poth (2017), after adapting the inventory to 24 items with PTs in Canada, based on a previous validation study by Daniels et al. (2014), also found poor model fit; thus, an exploratory analysis was performed, which identified eight new factors: assessment describes ability, improves learning, improves teaching, is valid, makes schools accountable, is bad, is ignored, and is inaccurate. This procedure was used by Rasooli et al. (2023), allowing them to identify three factors: assessment as (1) improvement and accountability, (2) irrelevant, and (3) inaccurate. These differences across studies indicate that PTs' conceptions of assessment are sensitive to local educational context, highlighting the need to combine deductive and inductive approaches in the validation process to identify context-specific assessment conceptions. Another way of identifying assessment conceptions was demonstrated by Al-Haddad et al. (2023), who investigated PTs' conceptions by adapting the Students' Conceptions of Assessment (SCoA) questionnaire, originally developed by Brown and Hirschfield (2007), in Bahrain and identified a four-factor structure (see Table 5), aligning with the core dimensions of TCoA.

To examine PTs' assessment values and practices in relation to their efficacy beliefs, Eren (2010, 2013) applied an adapted version of the Teachers' Classroom Assessment Scale (TCAS), originally developed by James and Pedder (2006), covering three main assessment areas (see Table 5). In Eren's (2013) validation study involving Turkish PTs, the scale demonstrated acceptable psychometric properties for both values ($\chi^2/df = 1.88$; CFI = .95; RMSEA = .05) and practice ($\chi^2/df = 2.34$; CFI = .94; RMSEA = .06). The high reliability coefficients reported in these studies suggest that the TCAS is an effective tool for examining how PTs' assessment values align with their intended classroom actions.

The final category of subcomponents focuses on assessment confidence and self-efficacy. DeLuca and Klinger (2010) developed a questionnaire targeting confidence levels in educational assessment knowledge to examine PTs' perceived competencies in AoL and AfL across three domains: assessment theory, practice, and philosophy. While internal consistency estimates were high (Cronbach's $\alpha = .82-.93$) across twelve subscales covering item formats, test design, assessment practices of AoL and AfL philosophies of the classroom and large-scale assessment, and reliability and validity issues, specific subscales for assessment type and statistical technique had lower reliability (Cronbach's $\alpha = .55$ and .66, respectively). Building on this study, DeLuca, Chavez, and Cao (2013) developed the Educational Assessment Confidence Inventory, comprising assessment scenarios based on Popham's (2008) work, to examine U.S. PTs' confidence in two areas (see Table 5). Their findings revealed a three-factor confidence structure in applying practical assessment approaches (informal and formal assessments,

communicating growth and achievement) and three further factors related to confidence in assessment praxis (designing standards- and theory-based assessments, rationalizing assessment decisions, and explaining and using assessment theory), with reliability estimates ranging from .50 to .89. These instruments suggest that confidence, similarly to the investigation of assessment decisions in authentic contexts, can be effectively examined through scenario-based methodologies to investigate PTs' confidence in both assessing student learning and in engaging in assessment practices.

In conclusion, the results of this review (Table 5) indicate that, rather than examining AL as an integrated construct, many studies focus on specific components, such as assessment beliefs, conceptions, and confidence. Although such targeted instruments often demonstrate higher psychometric properties and internal consistency within their specific domains, they inherently capture only partial aspects of PTs' AL. Furthermore, recurring validation issues across different countries, as exemplified by TCoA-III, suggest that these subcomponents are embedded in specific research and cultural contexts, which may limit the interpretability and comparability of findings across studies, thereby highlighting the need for greater conceptual coherence in future PTs' AL frameworks.

4.3. Factors contributing to the conceptualization and development of PTs' AL

As AL is defined in context-sensitive ways and is conceptualized through related constructs such as assessment competence, capability, or identity (Coombs & DeLuca, 2022), its conceptualization has been investigated through its relationships with assessment knowledge/competence and other assessment-related constructs, such as assessment conceptions/beliefs/values, efficacy, and practices (Xu & Brown, 2016). Accordingly, to address the second research question, this section synthesizes how personal, contextual, and external factors contribute to the conceptualization and development of PTs' AL.

Table 6 provides an overview of the 35 included studies that specifically examine these relationships. Four studies (Alkharusi, 2011b; DeLuca, Chavez, & Cao, 2013; Hoang et al., 2025; Kytätälä et al., 2022) were excluded from this specific analysis as they focus on the operationalization of AL rather than on factors influencing its conceptualization and development. As shown in Table 6, the literature is balanced between studies focusing on AL-related constructs, including assessment competence, knowledge, and capability ($n = 17$), and those focusing on assessment conceptions, beliefs, and values ($n = 18$). Regarding development orientation and context, the majority of research identified is process-oriented and course-related, typically investigating AL development through cross-sectional comparisons of subgroups or before and after specific interventions. Fewer studies adopt a product-oriented approach at the program level, usually conducted during or at the end of the training period.

Consistent with research on in-service teachers, PTs' AL is grounded in assessment knowledge or competence derived from established standards (AFT et al., 1990; Alkharusi, 2011b; McGee & Colby, 2014; Sondergeld et al., 2020). However, the results of this study indicate an expansion of the AL construct to incorporate personal psychological factors, contextual training environments, and external influences. To provide a synthesis of these factors contributing to the conceptualization and development of PTs' AL, the following section is organized into three main categories. First, personal factors are examined, including prior experiences, psychological factors (mindset, efficacy, attitudes), and assessment and teaching conceptions. Second, the impacts of contextual factors are presented, including assessment training, professional development, and the institutional learning culture. Finally, external factors are addressed, specifically the role of educational policy and supervising agencies in shaping PTs' AL.

4.3.1. Personal factors shaping PTs' AL

PTs' engagement with assessment is found to be significantly shaped by their prior experiences as students and their underlying epistemological beliefs. Brown and Remesal (2012) found that prospective teachers considered the nature and purpose of assessment relevant based on both the knowledge gained from assessment training and their prior student experience. Similarly, McGee and Colby (2014) reported that certain AL aspects, including ethical assessment, scoring, and choosing assessment methods, were already present before the investigated assessment course; nevertheless, exposure to the course potentially increased their AL in some areas, such as the sound design of assessments and the ability to communicate assessment results. These prior experiences are closely linked to PTs' broader epistemological belief systems. Tessema et al. (2024) revealed that assessment conceptions are associated with epistemological beliefs. Specifically, the improvement conception and the school accountability conception moderately correlated with simple knowledge and omniscient authority. These findings suggest that fostering sophisticated epistemological beliefs may support more adaptive assessment strategies, thereby enhancing teaching and learning outcomes. However, the reviewed literature also highlights that PTs' assessment conceptions may differ from those of in-service teachers due to differences in teaching experience, societal and cultural priorities for assessment use (Chen & Cowie, 2016), and flexibility in adapting conception characteristics (Atjonen et al., 2022).

Beyond prior experiences, PTs' AL is related to psychological constructs, including mindset, attitudes, and efficacy. DeLuca et al. (2019) reported a relationship between PTs' mindset and their approach to assessment regarding assessment purposes and fairness. Furthermore, attitudes toward assessment play a crucial role; while Alkharusi (2011a) found a positive relationship between AL and attitudes, Ng et al. (2020) revealed that engaging in peer assessment activities can change PTs' attitudes and their intention to implement such practices in their future teaching. Central to these psychological factors is self-efficacy, which supports PTs' AL development (Schneider et al., 2020), suggesting that success in teacher education programs on assessment helps improve AL. However, Volante and Fazio (2007) reported that self-efficacy levels remained relatively low for teacher candidates across the four years of the program, with most candidates favoring summative purposes, observational techniques, and personal communication over formative purposes. Nevertheless, low self-efficacy levels can be addressed through targeted professional development; as DeLuca et al. (2013) revealed, teacher candidates can expand their assessment conceptions from a narrow focus solely on testing to recognizing multiple forms of assessment serving varying purposes while simultaneously developing greater confidence in assessment practices. Moreover, PTs' assessment efficacy is not only associated with their AL (Alkharusi, 2011a) but also specifically relates to assessment attitudes (Alkharusi, 2010).

The most integrated dimension of personal factors concerns the relationship between assessment and teaching conceptions. The reviewed literature demonstrates that PTs' assessment conceptions are related to their AL (Daniels & Poth, 2017) and their broader conceptions of teaching (Chen & Brown, 2013; Eren, 2010). According to Daniels and Poth (2017), PTs' AL is directly related to their approaches to assessment, particularly whether they view it as a process for student and school accountability or as a means to improve teaching through mastery and performance approaches. Specifically, Chen and Brown (2013) identified relationships between four conceptions of excellent teaching (professional, model, examination, and life-long) and four assessment purposes (diagnostic and formative, irrelevant, control, and life character). They also revealed that the examination factor positively predicted the irrelevant conception and life character development, while negatively predicting diagnostic and formative purposes (Chen & Brown, 2013). Moreover, both consonance and dissonance can exist between prospective teachers' teaching/learning and assessment conceptions. For example, the PTs in Eren's (2010) study valued constructivist principles, such as making learning explicit and promoting

learning autonomy, more than they actually practiced them; instead, their practices were more aligned with traditional teaching and performance orientation. This interplay is further influenced by academic self-regulation; while student and school accountability perceptions relate positively to academic achievement, a tendency to ignore assessment exhibits a negative effect on self-regulation (Al-Haddad et al., 2023). As also highlighted in 4.2.2, ensuring the validity of instruments targeting assessment conceptions often remains a challenge; Reeves et al. (2023) noted that prior measurement models often show poor fit with PTs, suggesting that their conceptions may differ fundamentally from those of in-service teachers. Ultimately, as found by Levy-Vered and Alhija (2018), PTs' dominant view is that assessment improves learning and teaching and improvement and accountability are significantly positively correlated. These assessment conceptions may further evolve during an assessment course (Levy-Vered & Alhija, 2018).

4.3.2. Contextual factors shaping PTs' AL

The reviewed literature consistently highlights that PTs' AL develops primarily through formal assessment courses and professional development modules within teacher education programs. Several studies (Gallego-Noche et al., 2022; Hill et al., 2017; Lomax, 1996; Sondergeld et al., 2020) demonstrate that such training significantly impacts AL across diverse assessment topics. Hill et al. (2017) revealed that teacher candidates' assessment capability development was the outcome of teacher training, while Sondergeld et al. (2020) indicated that objective standard-setting approaches are particularly effective in helping PTs establish performance proficiency levels aligned with student content mastery. The impact of targeted interventions is further evident in diverse settings; as found by Oo et al. (2023), a needs-based professional development program in Myanmar successfully increased the PTs' AfL literacy. Furthermore, training can foster progress in the socio-environmental and ethical dimensions of assessment, especially when principles of education for sustainability and inclusive education are integrated into the design of classroom assessment activities (Gallego-Noche et al., 2022). Regarding the specific content of such training, DeLuca and Klinger (2010) identified that direct instruction is essential in areas such as reporting achievement, ensuring reliability and validity, and articulating a personal philosophy of assessment. To support this learning process, DeLuca, Chavez, Bellara, and Cao (2013) identified four crucial pedagogical constructs: perspective-building conversations, practice activities, modeling, and critical reflection. When these pedagogical aspects are implemented in courses organized around an AfL approach, they encourage prospective teachers to apply alternative assessment tasks more frequently, although Izci and Caliskan (2017) noted that these activities were influenced by the complexity of PTs' assessment conceptions, experiences, and cultural backgrounds. Ultimately, through these professional development opportunities, PTs tend to perceive an increase in their own assessment competence during their teacher education (Schneider & Bodensohn, 2017).

Research has shown that the specific cultural and contextual background also impacts PTs' AL. DeLuca et al. (2020) revealed significant differences in assessment purposes and approaches to fairness between PTs in different national contexts, specifically Canada and Germany. These contextual influences may contribute to the emergence of distinct profiles; Coombs et al. (2020) identified three subgroups: eager, contemporary, and hesitant assessors. Crucially, these approaches to assessment are not viewed as stable characteristics but are shaped by the interplay of learning experiences, institutional context, and personal dispositions (Coombs et al., 2020).

The final contextual factor involves the translation of values into classroom practices. Lomax (1996) identified the usefulness of assessment courses by highlighting the need for hands-on experience, particularly in grading, conducting parent conferences, handling the pressures of mandated standardized tests, and working with cooperating teachers whose assessment philosophy differed from those of PTs. Furthermore,

as Eren (2013) reported, the relationships between prospective teachers' assessment values and practices are mediated by their efficacy beliefs regarding making learning explicit and promoting learning autonomy. Similarly, Rasooli et al. (2023) found significant relationships between approaches to grading and assessment conceptions focused on supporting students' learning.

4.3.3. External factors shaping PTs' AL

The reviewed literature suggests that PTs' assessment conceptions are shaped by the broader cultural context (Daniels et al., 2014; Pastore, 2020), assessment policy (Vandeyar & Killen, 2006), and controlling agencies such as supervising teachers (Oo et al., 2021). Regarding cultural dimensions, Daniels et al. (2014) indicated differences between Canadian and New Zealand PTs' assessment conceptions; specifically, while both groups viewed assessment similarly in most aspects, their conceptions differed in terms of the inaccuracy of assessment. Similarly, Liu et al. (2016) found that cultural dimensions of assessment shape the different perceptions of ethical issues in assessment, as evidenced by differences between Chinese and American PTs. Beyond cultural factors, the influence of the institutional and policy environment is also identified in the reviewed literature. Pastore (2020) reported that Italian teacher trainees, while generally agreeing that assessment improves teaching and learning, conceived of assessment as being misaligned with the accountability system. The tension between PTs' personal beliefs and assessment policy is further illustrated by Vandeyar and Killen (2006), who found that while most student-teachers' beliefs aligned with assessment principles, some PTs held beliefs inconsistent with these established standards. Finally, controlling agencies often directly influence PTs' AL development; Oo et al. (2021) revealed that although PTs' decision-making in the classroom was largely influenced by their own assessment beliefs and values, their actual use of assessment strategies was importantly constrained by their supervising teachers.

5. Discussion and conclusion

5.1. Approaches and instruments for operationalizing PTs' AL

Addressing the first research question, this section discusses the shifts in methodological approaches and instruments used in the reviewed studies. Researchers have mainly used AL instruments originally developed for in-service teachers when examining PTs' AL. Instruments applied for PTs primarily operationalize teachers' competence based on the assessment standards and guidelines introduced by the American Federation of Teachers (AFT et al., 1990), Stiggins (1991), Brookhart (2011), and assessment policies in different contexts (e.g., Vandeyar & Killen, 2006). Early studies operationalized PTs' AL based on the AFT et al. (1990) standards, primarily focusing on competencies related to assessment processes and use (Mertler and Campbell, 2005; Plake & Impara, 1997). Subsequently, methodological approaches in AL shifted from focusing on teachers' understanding of how to construct and use assessment toward emphasizing teachers' knowledge of AfL purposes and their further professional development in assessment (DeLuca et al., 2013; DeLuca et al., 2016a; Volante & Fazio, 2007). This expanding focus reflects a shift in how PTs' AL has been operationalized, from a traditional assessment perspective (i.e., focusing on competence in the context of AoL) toward a broader perspective that captures how PTs conduct AfL and AaL (Brookhart, 2011; DeLuca et al., 2016b). This shift in operationalization also reflects evolving understandings of how teachers assess students' learning in line with up-to-date assessment standards and guidelines (DeLuca et al., 2016b). However, this trend is largely confined to European and U.S. contexts, with relatively limited research on PTs' AL in Asian settings.

The scope of AL instruments has expanded from focusing on core assessment competencies to covering other aspects of AL sub-components, such as assessment conceptions and efficacy. This trend is in line with the model of AL in practice proposed by Xu and Brown

(2016), as teachers' AL is conceptualized in relation to other aspects, such as assessment efficacy and conceptions. More recently, the integration of ICT into instruments targeting PTs' assessment competence has emerged, underscoring the growing importance of digital knowledge in evaluating student achievement (Hoang et al., 2025). Consistent with this shift, new standards of assessment competence increasingly include ICT components, reflecting teachers' widespread use of digital tools to assess student learning and foster twenty-first-century skills (Redecker & Johannessen, 2013).

Furthermore, two types of AL instruments—tests and questionnaires—were used in the empirical studies. The results indicated that researchers shifted from tests of PTs' AL toward self-report instruments capturing their perceptions and confidence in assessment. This trend may stem from a growing preference for a holistic perspective over an analytic one, which covers only assessment knowledge, excluding other aspects that reflect the context of practice and teachers' assessment conceptions, as has been discussed by DeLuca et al. (2020), since testing teachers' competence alone is considered analytic in teachers' AL research. While a holistic perspective provides broader assessment dimensions, the analytic perspective focusing on assessment competence typically demonstrates stronger psychometric properties and is easier to administer (DeLuca et al., 2020).

The results of this review indicated that relatively few AL instruments have been newly developed specifically for PTs (DeLuca et al., 2016a). Additionally, existing AL tools have not been sufficiently validated with PTs across different contexts. Even where validation has occurred, few studies on PTs' AL have been conducted in Asian settings, and little evidence exists regarding the suitability of these instruments in these cultural contexts (e.g., Oo et al., 2021, 2023). Evidence suggests that certain instruments designed for in-service teachers lack validity when applied among PTs, likely due to differences in background, training, and experience (e.g., Daniels & Poth, 2017; Eren, 2013). Other instruments demonstrate validity only in their original study contexts, requiring adaptation for use elsewhere (e.g., Daniels & Poth, 2017; Eren, 2013), or resulting in data that do not fit existing models for samples from different contexts (e.g., Brown & Remesal, 2012).

Based on the synthesized findings, we propose two guiding principles for the development of future PTs' AL instruments. First, AL tools should be context-driven to capture PTs' assessment knowledge in close alignment with curriculum content. Second, building on patterns identified across existing methodological approaches, future AL instruments should extend beyond assessment knowledge to cover assessment-related constructs, such as assessment beliefs, attitudes, efficacy, learning, technology knowledge, and situated assessment practices.

5.2. Factors influencing PTs' AL conceptualization and development

Building on the findings related to the second research question, this section provides a refined conceptual framework for PTs' AL that integrates assessment knowledge/competence with cognitive, affective, and contextual dimensions. Our synthesis demonstrates that PTs' AL development is a dynamic process shaped by the complex interplay of personal, contextual, and external influences. While grounded in established models of teachers' AL (Sultana, 2019; Tsagari, 2021; Xu & Brown, 2016), the proposed framework highlights features specific to PTs' transitional professional status and teacher education contexts.

At the personal level, PTs' AL is driven by their assessment conceptions, efficacy, attitudes, mindsets, and experiences. PTs' AL develops on the basis of their values in assessment and their assessment learning preferences. These findings are consistent with the literature on the relationships between AL and other constructs, such as assessment conceptions (e.g., Levy-Vered & Alhija, 2018; Quilter & Gallini, 2000) and efficacy (DeLuca & Klinger, 2010; Gullickson, 1993). Their understanding of the purposes of assessment might also lead them to develop AL in practice; thus, they tend to use the assessment knowledge they gain in teaching practices, as they know the importance of assessment in

the teaching and learning process (DeLuca & Lam, 2014). Furthermore, PTs' AL might depend on how important they find assessment in the teaching and learning processes; they might be aware that good teaching quality depends on the quality of assessment. PTs' AL might also be shaped by their assessment experiences as students (Quilter & Gallini, 2000); these experiences—specifically their assessment perceptions of what was effective or ineffective as students—are applied to the current position.

At the contextual level, PTs' AL is influenced by assessment training, practices, and cultural dimensions of assessment. Assessment training contributes significantly to developing PTs' AL since training is the source of their assessment knowledge (Weng & Shen, 2022; Xu & Brown, 2016). Assessment practices impact their training needs and efficacy, as PTs want to learn what they think is practical for their daily use (Gullickson, 1993). Additionally, the education culture might shape PTs' AL since the education system is developed according to different cultural assessment dimensions in other countries (Sultana, 2019; Tsagari, 2021).

At the external level, PTs' AL is driven by assessment policy and the supervising agency. Thus, PTs are expected to follow the practices of supervising teachers, who may have different perspectives on assessment during their teaching practices (Oo et al., 2021; Lomax, 1996). This characteristic is unique to PTs since they are required to work with supervising teachers during their teaching practices. Like in-service teachers, they have to conform to the assessment policy regarding their practices during their practicum (Forsberg & Wermke, 2012; Lomax, 1996).

These features justify treating PTs' AL as a related but distinct construct from in-service teachers' AL in terms of assessment beliefs, efficacy, attitude, practices, and learning, thereby reinforcing the rationale for examining PTs' AL as a separate construct in empirical research. Unlike in-service teachers' AL, PTs' assessment conceptions are partly influenced by their experience as students; thus, their beliefs are in the transitional stage of AL development, which requires facilitation from teaching agents such as co-teachers and trainers. In addition, PTs' AL may encounter tensions from supervising teachers that could hinder the application of assessment theories acquired during training if supervising teachers hold different philosophies or experiences regarding assessment. PTs' AL is also less likely to be fully developed, as their opportunities to engage in assessment practice are limited. With experiences in assessment practices and learning as students, their identity as an assessor may differ from that of in-service teachers, particularly regarding how they utilize assessments to support students' learning improvement (Atjonen et al., 2022).

Based on these insights, and as a conceptual contribution of this review, we propose that the conceptualization of PTs' AL should encompass three main dimensions of assessment. First, regarding the knowledge base, PTs' AL should include not only assessment knowledge/competence but assessment-related technology literacy; that is, the ability to construct and conduct digital assessments. Second, the cognitive and affective domain should cover assessment conceptions, efficacy, attitudes, and their identity as an assessor. Third, for externally influential dimensions on assessment, PTs' AL should consider contextual and cultural issues and significant individuals, such as cooperating and supervising teachers, with reference to their assessment practices. Together, these dimensions extend existing models of teachers' AL by explicitly incorporating PTs' transitional professional status and constrained assessment practice opportunities.

6. Limitations

In this study, we limited our review to empirical studies focusing on PTs' AL in a general, non-discipline-specific context. Consequently, discipline-specific studies and studies addressing related constructs outside an explicit AL framework were excluded. Although such studies may contain insights relevant to the research questions, including them

would have reduced the conceptual comparability of findings. These decisions may limit the generalizability of the findings. Methodologically, although a consensus-based protocol was used during eligibility assessment, primary data extraction was conducted by the first author, which may have resulted in interpretive bias in thematic synthesis. Furthermore, restricting the inclusion criteria to English-language publications may have resulted in the omission of relevant studies published in other languages. Accordingly, the findings should be interpreted as reflecting trends in general, cross-disciplinary research on PTs' AL, rather than providing comprehensive coverage of subject-specific or locally grounded AL conceptualizations.

7. Implications and recommendations

Based on the synthesized findings and the proposed conceptualization of PTs' AL, this study offers implications for both teacher education programs and researchers examining PTs' AL.

For teacher education programs, assessment education should explicitly address the interplay of personal, contextual, and external influences on PTs' AL development. At the personal level, programs must consider PTs' assessment conceptions, assessment efficacy, conceptions of teaching, and attitudes toward assessment. Addressing how PTs value assessment can help them improve their AL by amplifying the importance of assessment in the teaching and learning processes. Identifying PTs' assessment learning needs can inform the refinement of program content and ensure alignment with necessary assessment knowledge. Understanding PTs' attitudes toward assessment can facilitate the development of reflective and up-to-date assessment practices. Contextually, in the current digital era, integrating ICT into assessment education programs is essential, enabling PTs to use digital tools effectively in their classroom assessment practices. To reduce the control effect of external factors, autonomy in assessment practices should also be permitted for PTs to integrate their assessment competence from training into their assessment practices and use the current assessment theories. Emphasizing the importance of assessment competence in teaching quality may further encourage PTs to focus on sharpening their assessment ability.

For researchers, AL tools should be validated in different contexts, and new instruments should be developed in alignment with both international and local assessment standards and guidelines. Further investigations of PTs' AL should be conducted to determine the causal relationships among AL, assessment conceptions, and other related constructs. Such research can explain how AL is related to assessment conceptions and how other factors, such as attitudes, efficacy, and practices, influence these relationships. Understanding the direction of these relationships would help researchers understand better how these constructs are meaningfully related. In addition, the limited empirical evidence on PTs' AL in Asian contexts highlights a significant gap. Studies should utilize AL instruments that are developed following the research context to match their guidelines and principles of assessment and validated to improve the quality of the empirical evidence. Another possible research area could focus on relationships between AL, assessment conceptions, and the conceptions of feedback, which are absent from the current literature and might play a significant role in these relationships in the assessment research paradigm. The relationships between these constructs are relevant because feedback is an emerging construct closely associated with teachers' assessment practices. Teacher education programs should also consider focusing on developing teachers' feedback competence, especially for AaL and AfL, to develop PTs' AL.

CRedit authorship contribution statement

Piseth Hull: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Tibor Vigh: Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Ethics approval

This review paper synthesizes data from previously published studies. As the data analyzed in this review are already publicly available, no additional ethics approval is required for this review.

Funding

The work was supported by the University of Szeged Open Access Fund (grant number: 8036). This research was supported by the Research Programme for Public Education Development, Hungarian Academy of Sciences.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

While working on this paper, Piseth Hull was a recipient of the Stipendium Hungaricum Scholarship. The authors would like to thank Enago (www.enago.com) for the English language review.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.stueduc.2026.101600](https://doi.org/10.1016/j.stueduc.2026.101600).

Data availability

The complete coding dataset for all 39 studies included in this review, covering all categories defined in the coding framework (Table 3), is provided as [Supplementary Material](#).

References

- The papers in our review are marked with an asterisk (*).
- Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In In. D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching* (pp. 205–221). Springer. https://doi.org/10.1007/978-90-481-3927-9_12.
- (*) Al-Haddad, S. S., Afari, E., Khine, M. S., & Eksail, F. A. A. (2023). Self-regulation, self-confidence, and academic achievement on assessment conceptions: An investigation study of pre-service teachers. *Journal of Applied Research in Higher Education*, 15(3), 813–826. <https://doi.org/10.1108/JARHE-09-2021-0343>.
- (*) Alkharusi, H. (2010). Factors associated with preservice teachers' attitude towards educational measurement. *The International Journal of Learning: Annual Review*, 17(1), 95–108. <https://doi.org/10.18848/1447-9494/CGP/v17i01/46803>.
- (*) Alkharusi, H. (2011a). An analysis of the internal and external structure of the Teacher Assessment Literacy Questionnaire. *The International Journal of Learning: Annual Review*, 18(1), 515–528. <https://doi.org/10.18848/1447-9494/CGP/v18i01/47461>.
- (*) Alkharusi, H. (2011b). Psychometric properties of the Teacher Assessment Literacy Questionnaire for preservice teachers in Oman. *Procedia - Social and Behavioral Sciences*, 29, 1614–1624. <https://doi.org/10.1016/j.sbspro.2011.11.404>.
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association (AFT, NCME, & NEA) (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30–32. doi:10.1111/j.1745-3992.1990.tb00391.x.
- (*) Atjonen, P., Pöntinen, S., Kontkanen, S., & Ruotsalainen, P. (2022). enhancing preservice teachers' assessment literacy: Focus on knowledge base, conceptions of assessment, and teacher learning. *Frontiers in Education*, 7, Article 891391. <https://doi.org/10.3389/educ.2022.891391>.
- Baker, B. A., & Riches, C. (2018). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing*, 35(4), 557–581. <https://doi.org/10.1177/0265532217716732>.
- Bijsterbosch, E., Béneker, T., Kuiper, W., & van der Schee, J. (2019). Teacher professional growth on assessment literacy: A Case study of prevocational geography education in the Netherlands. *The Teacher Educator*, 54(4), 420–445. <https://doi.org/10.1080/08878730.2019.1606373>

- Bolivar, E. M. R. (2020). Monitoring preservice teachers' language assessment literacy development through journal writing. *Malaysian Journal of ELT Research*, 17(1), 38–52.
- 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. (2002). *Teachers' conceptions of Assessment (Unpublished doctoral dissertation) University of Auckland*.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301–318. <https://doi.org/10.1080/0969594042000304609>
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged version. *Psychological Reports*, 99(1), 166–170. <https://doi.org/10.2466/pr0.99.1.166-170>
- Brown, G. T. L., & Hirschfeld, G. H. F. (2007). Students' conceptions of assessment and mathematics achievement: Evidence for the power of self-regulation. *Australian Journal of Educational and Developmental Psychology*, 7(1), 63–74.
- (*) Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish Journal of Psychology*, 15(1), 75–89. https://doi.org/10.5209/rev_SJOP.2012.v15.n1.37286.
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002). *Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers* (October). Columbus, OH: Paper session presentation at the meeting of the Mid-Western Educational Research Association.
- (*) Chen, J., & Brown, G. T. L. (2013). High-stakes examination preparation that controls teaching: Chinese prospective teachers' conceptions of excellent teaching and assessment. *Journal of Education for Teaching*, 39(5), 541–556. <https://doi.org/10.1080/02607476.2013.836338>.
- (*) Chen, J., & Cowie, B. (2016). Chinese preservice teachers' beliefs about assessment. *Educational Practice and Theory*, 38(2), 77–93. <https://doi.org/10.7459/ept/38.2.05>.
- 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>
- (*) Coombs, A., DeLuca, C., & MacGregor, S. (2020). A person-centered analysis of teacher candidates' approaches to assessment. *Teaching and Teacher Education*, 87, Article 102952. <https://doi.org/10.1016/j.tate.2019.102952>.
- (*) Daniels, L. M., & Poth, C. A. (2017). Relationships between pre-service teachers' conceptions of assessment, approaches to instruction, and assessment: An achievement goal theory perspective. *Educational Psychology*, 37(7), 835–853. <https://doi.org/10.1080/01443410.2017.1293800>.
- (*) Daniels, L. M., Poth, C., Papile, C., & Hutchison, M. (2014). Validating the Conceptions of Assessment-III Scale in Canadian preservice teachers. *Educational Assessment*, 19(2), 139–158. <https://doi.org/10.1080/10627197.2014.903654>.
- 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., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for preservice assessment education: Supporting teacher candidates' assessment literacy development. *The Teacher Educator*, 48(2), 128–142. <https://doi.org/10.1080/08878730.2012.760024>.
- (*) DeLuca, C., Chavez, T., & Cao, C. (2013). Establishing a foundation for valid teacher judgement on student learning: The role of pre-service assessment education. *Assessment in Education: Principles, Policy & Practice*, 20(1), 107–126. <https://doi.org/10.1080/0969594X.2012.668870>.
- (*) DeLuca, C., Coombs, A., & LaPointe-McEwan, D. (2019). Assessment mindset: Exploring the relationship between teacher mindset and approaches to classroom assessment. *Studies in Educational Evaluation*, 61, 159–169. <https://doi.org/10.1016/j.stueduc.2019.03.012>.
- (*) 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*, 41(3), 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., Schneider, C., Coombs, A., Pozas, M., & Rasooli, A. (2020). A cross-cultural comparison of German and Canadian student teachers' assessment competence. *Assessment in Education: Principles, Policy & Practice*, 27(1), 26–45. <https://doi.org/10.1080/0969594X.2019.1703171>.
- Earl, L. (2003). *Assessment as learning*. Thousand Oaks.
- Earl, L., & Katz, S. (2005). Rethinking classroom assessment with purpose in mind. *Western and Northern Canadian Protocol for Collaboration in Education*.
- Engelsen, K. S., & Smith, K. (2014). Assessment literacy. In In. C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning*, 1 pp. 91–107. Springer. https://doi.org/10.1007/978-94-007-5902-2_6.
- (*) Eren, A. (2010). Consonance and dissonance between Turkish prospective teachers' values and practices: Conceptions about teaching, learning, and assessment. *Australian Journal of Teacher Education*, 35(3), 27–48. <https://doi.org/10.14221/ajte.2010v35n3.3>.
- (*) Eren, A. (2013). Prospective teachers' conceptions of assessment and efficacy beliefs: Values and practices. *Education and Science*, 38(170), 19–31.
- 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. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- (*) Gallego-Noche, B., García-González, E., Jiménez-Fontana, R., & Azcárate, P. (2022). Inclusive assessment and sustainability dimensions: Pre-service teachers' ideas and knowledge (Article) *Sustainability*, 14(15), 9651. <https://doi.org/10.3390/su14159651>.
- 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*, 21(2), 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>
- Gullickson, A. R. (1993). Matching measurement instruction to classroom-based evaluation: Perceived discrepancies, needs, and challenges. In S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 1–25). Buros Institute of Mental Measurements.
- (*) Hill, M. F., Ell, F. R., & Eyers, G. (2017). Assessment capability and student self-regulation: The challenge of preparing teachers (Article) *Frontiers in Education*, 2, 21. <https://doi.org/10.3389/educ.2017.00021>.
- Hill, M. F., Smith, L. F., Cowie, B., Gilmore, A., & Gunn, A. (2014). Preparing initial primary and early childhood teacher education students to use assessment. *Assessment Matters*, 7, 4–23. <https://doi.org/10.18296/am.0123>
- (*) Hoang, L. P., Le, P. A., Le, H. T., Nguyen, D. T., Phan, T. C., Le, H. T., Van Tran, H., & Tran-Duong, Q. H. (2025). Evaluating educational assessment competence of pre-service teachers: Extended standards in the context of digital classroom assessment transformation. *Education and Information Technologies*, 30(12), 16347–16374. <https://doi.org/10.1007/s10639-025-13467-y>.
- (*) Izi, K., & Caliskan, G. (2017). Development of prospective teachers' conceptions of assessment and choices of assessment tasks. –464 *International Journal of Research in Education and Science*, 3(2), 464. <https://doi.org/10.21890/ijres.327906>.
- James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *The Curriculum Journal*, 17(2), 109–138. <https://doi.org/10.1080/09585170600792712>
- Jaramillo Delgado, J., & Gil Bedoya, A. M. (2019). Pre-service English language teachers' use of reflective journals in an assessment and testing course. *JSR Fulnam Journal of Students' Research*, (4), 210–218. <https://doi.org/10.21501/25007858.3010>
- Juanjuan, G., & Mohd Yusoff, N. (2022). The shared features of effective improvement programmes for teachers' assessment literacy, 18 *Pedagogies: An International Journal*, 18(4), 710–727. <https://doi.org/10.1080/1554480X.2022.2084398>.
- Khadijeh, B., & Amir, R. (2015). Importance of teachers' assessment literacy. *International Journal of English Language Education*, 3(1), 139–146. <https://doi.org/10.5296/ijelev.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>
- Kyttälä, M., Björn, P. M., Rantamäki, M., Lehesvuori, S., Aro, M., & Lerkkanen, M.-K. (2022). Assessment conceptions of Finnish pre-service teachers. *European Journal of Teacher Education*, 47(3), 529–547. <https://doi.org/10.1080/02619768.2022.2058927>
- Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System*, 81, 78–89. <https://doi.org/10.1016/j.system.2019.01.006>
- Lan, C., & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation*, 61, 112–122. <https://doi.org/10.1016/j.stueduc.2019.03.003>
- (*) 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>.
- Levy-Vered, A., Reinders-Kafri, S., & Inbar-Lourie, O. (2022). From murky salty water to crystal-clear fresh river flow: Pre-service teachers' assessment perceptions. *Journal of Education and Development*, 6(5), 8–23. <https://doi.org/10.20849/jed.v6i5.1293>
- (*) Liu, J., Johnson, R., & Fan, X. (2016). A comparative study of Chinese and United States pre-service teachers' perceptions about ethical issues in classroom assessment. *Studies in Educational Evaluation*, 48, 57–66. <https://doi.org/10.1016/j.stueduc.2016.01.002>.
- Lo, Y. Y., & Leung, C. (2022). Conceptualising assessment literacy of teachers in Content and Language Integrated Learning programmes. *International Journal of Bilingual Education and Bilingualism*, 25(10), 3816–3834. <https://doi.org/10.1080/13670050.2022.2085028>
- (*) Lomax, R. G. (1996). On becoming assessment literate: An initial look at preservice teachers' beliefs and practices. *The Teacher Educator*, 31(4), 292–303. <https://doi.org/10.1080/08878739609555122>.
- (*) 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>.
- Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 1(33), 49–64.

- Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory* (April). Montreal, Quebec, Canada: Paper session presentation at the annual meeting of the American Educational Research Association.
- (*) Ng, W. S., Xie, H., Wang, F. L., & Li, T. (2020). Peer assessment of peer assessment plan: A deep learning approach of teacher assessment literacy. *International Journal of Innovation and Learning*, 27(4), 450–466. <https://doi.org/10.1504/ijil.2020.107617>.
- Nguyễn, T. H. L. (2016). Development of Vietnamese pre-service EFL teachers' assessment literacy. *Open Access Te Herenga Waka-Victoria University of Wellington*. <https://doi.org/10.26686/wgtn.17143502.v1>
- Nier, C. V., Donovan, A. E., & Malone, M. E. (2009). Increasing assessment literacy among LCTL instructors through blended learning. *Journal of the National Council of Less Commonly Taught Languages*, 7(1), 95–118.
- 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), 352–373.
- Oo, C. Z., Alonzo, D., & Davison, C. (2021). Pre-service teachers' decision-making and classroom assessment practices (Article) *Frontiers in Education*, 6, Article 628100. <https://doi.org/10.3389/educ.2021.628100>.
- (*) Oo, C. Z., Alonzo, D., & Davison, C. (2023). Using a needs-based professional development program to enhance pre-service teacher assessment for learning literacy. *International Journal of Instruction*, 16(1), 781–800. <https://doi.org/10.29333/iji.2023.16144a>.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamsler, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372(71). <https://doi.org/10.1136/bmj.n71>
- (*) Pastore, S. (2020). How do Italian teacher trainees conceive assessment?. *International Journal of Instruction*, 13(4), 215–230. <https://doi.org/10.29333/iji.2020.13414a>.
- Pastore, S. (2022). Assessment literacy in the higher education context: A critical review. *Intersection: A Journal at the Intersection of Assessment and Learning*, 4(1). <https://doi.org/10.61669/001c.39702>
- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84, 128–138. <https://doi.org/10.1016/j.tate.2019.05.003>
- Plake, B. S., & Impara, J. C. (1992). University of Nebraska. *Teacher competencies questionnaire description*.
- Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning adjustment, and achievement* (pp. 53–58). Academic Press.
- 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. (2008). *Classroom assessment: What teachers need to know* (5th ed.). Pearson/Allyn & Bacon.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>
- Puspawati, I. (2022). Teachers' language assessment literacy: A systematic review. *Indonesian EFL Journal*, 8(1), 113–126. <https://doi.org/10.25134/iefj.v8i1.5613>
- Qotboddin, J. M., Hossein, K., & Khalil, M. (2020). Assessment literacy of Iranian EFL teachers: A review of recent studies. *The Journal of Asia TEFL*, 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>
- (*) Rasooli, A., DeLuca, C., & Cheng, L. (2023). Beginning teacher candidates' approaches to grading and assessment conceptions: Implications for teacher education in assessment. *Educational Research for Policy and Practice*, 22(1), 63–90. <https://doi.org/10.1007/s10671-022-09320-5>.
- Redecker, C., & Johannessen, Ø. (2013). Changing assessment—Towards a new assessment paradigm using ICT. *European Journal of Education*, 48(1), 79–96. <https://doi.org/10.1111/ejed.12018>
- (*) Reeves, T. D., Onder, Y., & Kraner, C. (2023). Validation and invariance of the Conceptions of Assessment-III Abridged (COA-IIIa) among pre-service and in-service teachers in the United States. *Educational Assessment, Evaluation and Accountability*, 35(3), 419–447. <https://doi.org/10.1007/s11092-023-09407-4>.
- (*) Schneider, C., & Bodensohn, R. (2017). Student teachers' appraisal of the importance of assessment in teacher education and self-reports on the development of assessment competence. *Assessment in Education: Principles, Policy & Practice*, 24(2), 127–146. <https://doi.org/10.1080/0969594X.2017.1293002>.
- (*) Schneider, C., DeLuca, C., Pozas, M., & Coombs, A. (2020). Linking personality to teachers' literacy in classroom assessment: A cross-cultural study. *Educational Research and Evaluation*, 26(1–2), 53–74. <https://doi.org/10.1080/13803611.2021.1902354>.
- SeviMel-ŞahiN, A., & Subaşı, 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/akukeyg.501817>
- Singh, C. K. S., Singh, H. K. J., Singh, T. S. M., Moneyam, S., Abdullah, N. Y., & Zaini, M. F. (2022). ESL teachers' assessment literacy in classroom: A review of past studies. *Journal of Language and Linguistic Studies*, 18(1), 1–17.
- (*) Sondergeld, T. A., Stone, G. E., & Kruse, L. M. (2020). Objective standard setting in educational assessment and decision making. *Educational Policy*, 34(5), 735–759. <https://doi.org/10.1177/0895904818802115>.
- Starck, J. R., Richards, K. A. R., & O'Neil, K. (2018). A conceptual framework for assessment literacy: Opportunities for physical education teacher education. *Quest*, 70(4), 519–535. <https://doi.org/10.1080/00336297.2018.1465830>
- Stiggins, R. J. (1991). *Assessment literacy*. Phi Delta Kappan.
- 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/10.11186/s40468-019-0077-8>
- Sun, H., & Zhang, J. (2022). Assessment literacy of college EFL teachers in China: Status quo and mediating factors (Article) *Studies in Educational Evaluation*, 74, Article 101157. <https://doi.org/10.1016/j.stueduc.2022.101157>.
- Tajeddin, Z., Saeedi, Z., & Panahzadeh, V. (2022). English language teachers' perceived classroom assessment knowledge and practice: Developing and validating a scale. *Profile: Issues in Teachers' Professional Development*, 24(2), 247–264. <https://doi.org/10.15446/profile.v24n2.90518>
- (*) Tessema, G., Michael, K., & Areaya, S. (2024). The relationship between epistemological beliefs and assessment conceptions among pre-service teachers. *Open Education Studies*, 6(1), Article 20240056. <https://doi.org/10.1515/edu-2024-0056>.
- Tsagari, D. (2021). Gauging the assessment literacy levels of English teachers in Norway. *The European Journal of Applied Linguistics and TEFL*, 10(1), 161–191.
- (*) Vandeyar, S., & Killen, R. (2006). Beliefs and attitudes about assessment of a sample of student teachers in South Africa. *Africa Education Review*, 3(1–2), 30–47. <https://doi.org/10.1080/18146620608540441>.
- Vogel, L. R., Rau, W. C., Baker, P. J., & Ashby, D. E. (2006). Bringing assessment literacy to the local school: A decade of reform initiatives in Illinois. *Journal of Education for Students Placed at Risk*, 11(1), 39–55. https://doi.org/10.1207/s15327671espr1101_3
- (*) 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>.
- 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 (Article) *Education Sciences*, 13(2), 190. <https://doi.org/10.3390/educsci13020190>.
- Weng, F., & Shen, B. (2022). Language assessment literacy of teachers (Article) *Frontiers in Psychology*, 13, Article 864582. <https://doi.org/10.3389/fpsyg.2022.864582>.
- 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>
- 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., & He, L. (2019). How pre-service teachers' conceptions of assessment change over practicum: Implications for teacher assessment literacy. *Frontiers in Education*, 4, 145. <https://doi.org/10.3389/educ.2019.00145>
- Xu, Y., & Liu, Y. (2025). Language assessment literacy for teachers. In In. Z. Tajeddin, & T. S. Farrell (Eds.), *Handbook of Language Teacher Education* (pp. 659–692). Springer. https://doi.org/10.1007/978-3-031-47310-4_16.
- Yan, Z., & Pastore, S. (2022). Are teachers literate in formative assessment? The development and validation of the Teacher Formative Assessment Literacy Scale (Article) *Studies in Educational Evaluation*, 74, Article 101183. <https://doi.org/10.1016/j.stueduc.2022.101183>.