Article

Journal of Interdisciplinary Studies in Education

Volume 15, Issue 8 (2026)
*Journal of Interdisciplinary Studies in Education*
ISSN: 2166-2681Print 2690-0408 Online | https://ojed.org/jise
https://doi.org/10.32674/8hxymt77

# Teacher Assessment Literacy: A Systematic Review and Meta-Analysis Across the ALI, CALI, and TALQ Instruments

Zineb Labbar
*Mohammed V University, Faculty of Education Sciences, CEDoc Man, Society and Education, Morocco*
ORCID: 0009-0006-5960-568X

Abdellatif El Moudni
*Mohammed V University, CEDOC-Man, Society, and Education, Rabat, Morocco*

## ABSTRACT

*Teacher assessment literacy has received increasing attention, yet evidence remains fragmented across studies. This systematic review and meta-analysis synthesized 18 studies comprising 21 independent samples using the ALI, CALI, and TALQ instruments. The pooled mean score was 16.48 (95% CI [14.86, 18.10]) on a 35-point scale, indicating low assessment literacy, with substantial between-study heterogeneity ($I^2$ = 99.21%). Publication year, geographical region, instrument type, teacher status, teaching level, and subject specialization were not significant moderators. This review provides the first quantitative benchmark of teacher assessment literacy across the ALI, CALI, and TALQ instruments and highlights the need to strengthen assessment literacy through pre-service education, continuous professional development, and context-responsive educational policies.*

**Keywords:** assessment literacy, ALI, CALI, TALQ, investigating teachers' assessment literacy level

© Author(s), 2026. Published by Star Scholars Press.
This article is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. https://creativecommons.org/licenses/by/4.0/

## INTRODUCTION

Across educational systems worldwide, student achievement is treated as a primary indicator of institutional effectiveness. National and international frameworks increasingly rely on learning outcomes to evaluate policy success and guide reform initiatives (Gouëdard, 2021). As a result, performance metrics have become central to accountability systems, shaping public discourse, institutional evaluation, and teacher preparation priorities.

However, while student outcomes receive extensive scrutiny, the assessment practices that generate these outcomes are far less critically examined. The competencies required for valid, reliable, and ethically

sound assessment are often assumed rather than systematically evaluated. Popham (2011) has long argued that insufficient preparation in assessment remains one of the most overlooked weaknesses in teacher education, a concern echoed by Stiggins (1993) regarding the neglect of formal assessment training. More recently, Barnes and Qasserras (2024) reported persistent fragmentation in teacher candidates' understanding of assessment principles. These patterns suggest a structural imbalance: educational systems prioritize measurable performance while insufficiently consolidating evidence regarding teachers' assessment competence.

Yet despite decades of scholarship raising concerns about teachers' assessment literacy, the field lacks cumulative evidence capable of determining whether these concerns reflect localized deficiencies or a broader systemic pattern. Without quantitative synthesis and pooled analysis, it remains unclear whether reported weaknesses are context-specific, instrument-dependent, or indicative of a more widespread structural challenge. More recently, assessment literacy research has continued to evolve with the development of new instruments designed to capture broader dimensions of assessment literacy, particularly formative assessment competencies, reflecting the expanding conceptualization of the construct (Fu & Koh, 2026).

## LITERATURE REVIEW

In response to global accountability shifts, the quality of teacher education has become a growing concern, bringing increased attention to assessment literacy. Early conceptualizations reflected two complementary perspectives. One emphasized teacher's understanding of assessment principles and the use of assessment to improve student learning (Stiggins, 1990), whereas the other focused on technical competence in areas such as test construction, score interpretation, and grading practices (Paterno, 2001). Together, these perspectives established the conceptual foundation of teacher assessment literacy and continue to shape contemporary discussions of the construct.

Subsequent scholarship has progressively expanded assessment literacy beyond technical competence to encompass broader professional and contextual dimensions. Rather than viewing assessment solely as a set of measurement skills, contemporary perspectives emphasize its role in supporting educational decision-making (DeLuca et al., 2016; Owusu-Acheampong & Kwapong, 2021; Popham, 2009), recognizing assessment as a socially situated practice influenced by context and culture (Issa & Castulo, 2026; Xu & Brown, 2016). Consequently, assessment literacy is increasingly understood to include ethical awareness, contextual responsiveness (Looney et al., 2018), and professional judgment. Reflecting this evolution, Coombs and DeLuca (2022) proposed the concept of assessment capability, which extends assessment literacy to encompass ethical sensitivity, professional judgment, and responsiveness to learners alongside technical expertise.

Despite the growing recognition of assessment literacy as a cornerstone of effective teaching, evidence consistently indicates that many teachers continue to experience difficulties understanding and applying assessment principles in classroom practice. MacLellan (2004) first highlighted a disconnect between teachers' assessment conceptions and the theoretical and policy frameworks guiding educational practice. More recent research suggests that this gap persists, with teachers continuing to encounter challenges in translating assessment knowledge into effective classroom practice (Kang & Lam, 2024; Pastore, 2023). Consequently, evaluating teachers' assessment literacy remains essential for determining whether teacher education and professional development adequately prepare teachers to meet contemporary assessment expectations.

However, consensus remains limited regarding how teacher assessment literacy should be measured in a way that adequately captures both technical competence and contextual understanding (DeLuca et al., 2016; Gotch & French, 2014). This lack of consensus has resulted in the development of multiple measurement approaches, including self-report inventories, performance-based rubrics, and objective tests. Each offers distinct advantages and limitations: self-report measures capture teachers'

perceptions but are susceptible to response bias, rubric-based assessments evaluate applied competence but often lack standardization, whereas objective tests facilitate standardized comparison despite concerns regarding psychometric quality and contextual authenticity (Gotch & French, 2014). Reflecting this diversity, DeLuca et al. (2016) identified only eight validated objective assessment literacy instruments developed over a 21-year period, highlighting the fragmented nature of assessment literacy measurement. Among the assessment literacy instruments identified in previous reviews (DeLuca et al., 2016), the Teacher Assessment Literacy Questionnaire (TALQ), Assessment Literacy Inventory (ALI), and Classroom Assessment Literacy Inventory (CALI) were selected for the present review because they are objective, knowledge-based instruments that share a common theoretical foundation in the *Standards for Teacher Competence in Educational Assessment of Students* (AFT, NCME, & NEA, 1990), employ a comparable 35-item scoring framework, and permit direct quantitative comparison across studies.

The development of these instruments was guided by the Standards for Teacher Competence in the Educational Assessment of Students (AFT, NCME, & NEA, 1990), which were established to define the assessment competencies expected of classroom teachers. Developed through a review of the assessment literature and an analysis of teachers' professional assessment responsibilities, the Standards organize teacher assessment competence into seven domains covering assessment selection, development, administration, interpretation, grading, communication, and ethical practice. The developers explicitly described the Standards as a "conceptual framework or scaffolding from which specific skills can be derived" (AFT, NCME, & NEA, 1990, p. 3). Since their publication, this competency-based framework has provided the conceptual foundation for the TALQ, ALI, and CALI, making these instruments the principal objective measures of teacher assessment literacy in the literature.

The Teacher Assessment Literacy Questionnaire (TALQ) was the first widely adopted objective instrument developed from the 1990 Standards and demonstrated relatively low internal consistency (KR-20 = .54) (Plake et al., 1993). Campbell et al. (2002) subsequently applied the same instrument to preservice teachers under the name Assessment Literacy Inventory (ALI), reporting improved reliability (KR-20 = .74). Mertler (2003) likewise employed a slightly modified version of the TALQ with both pre-service and in-service teachers, obtaining comparable results. Recognizing limitations of the original instrument, including its length, readability, and decontextualized items, Mertler and Campbell (2005) subsequently redeveloped the Assessment Literacy Inventory into a scenario-based instrument while retaining the same competency framework and 35-item structure aligned with the seven Standards.

Despite their widespread use, the continued reliance on the 1990 Standards for Teacher Competence in the Educational Assessment of Students has attracted increasing criticism as conceptions of assessment literacy have evolved. The Standards primarily conceptualize assessment literacy as a set of teacher competencies related to assessment design, administration, interpretation, grading, communication, and ethical practice. More recent scholarship, however, argues that effective assessment practice also requires professional judgment, formative assessment expertise, sociocultural awareness, digital assessment competence, and responsiveness to diverse learning contexts (Brookhart, 2011; Coombs & DeLuca, 2022; DeLuca et al., 2016; Yan & Brown, 2021). Nevertheless, the Standards continue to serve as the principal competency framework underpinning the most widely used objective assessment literacy instruments. Their continued use reflects both their historical influence and the absence of a universally accepted alternative framework that has been operationalized into validated objective measures. Consequently, although the Standards may not fully represent the expanding scope of contemporary assessment literacy, they provide a common conceptual framework for comparing findings across the TALQ, ALI, and CALI instruments because all three instruments were developed from this shared competency framework.

The growing body of assessment literacy research has also prompted evidence syntheses. Most notably, Pastore (2023) conducted a systematic review of empirical studies published between 2013 and 2022, documenting the evolution of teacher assessment literacy research, its dominant conceptualizations, and commonly reported challenges. However, that review provided a qualitative synthesis across diverse assessment literacy measures and did not quantitatively synthesize assessment literacy outcomes or estimate

an overall level of teacher assessment literacy. Moreover, it encompassed a broad range of assessment literacy instruments based on different conceptual frameworks. In contrast, the present study focuses specifically on the three most widely used objective instruments (TALQ, ALI, and CALI), which share a common competency-based foundation in the 1990 Standards. By statistically synthesizing findings across these comparable instruments, this study provides the first pooled empirical benchmark of teacher assessment literacy while examining the extent to which contextual and study characteristics contribute to variation in reported outcomes.

Assessment literacy is increasingly recognized as a multidimensional and context-dependent construct shaped by the interaction of teachers' knowledge, professional experiences, and the sociocultural and institutional contexts in which assessment is enacted (Xu & Brown, 2016). Consequently, variation in reported assessment literacy levels may reasonably be expected across teacher populations and educational settings. Previous assessment literacy research provides both conceptual and empirical support for examining such variation. For example, Muhsin et al. (2025) investigated the influence of teacher characteristics, including gender, educational background, assessment training, and work experience, on teachers' assessment literacy, reporting significant effects for gender and assessment training but not for educational background or work experience. More broadly, Hull and Vígh (2026) systematically reviewed 39 empirical studies and conceptualized preservice teachers' assessment literacy development as a dynamic process shaped by interacting personal factors (e.g., beliefs, self-efficacy, and prior experience), contextual factors (e.g., teacher education, institutional culture, and learning opportunities), and external factors (e.g., educational policy and regulatory agencies). Their review further highlighted variation in how assessment literacy has been conceptualized and operationalized across studies, reflecting the continued evolution of the construct and its measurement. Consistent with these developments, DeLuca et al. (2016) also documented the evolution of teacher assessment literacy and the diversity of approaches used to conceptualize and measure the construct. Collectively, this body of literature provides conceptual and empirical justification for examining whether teacher-, contextual-, and study-level characteristics contribute to between-study variation in reported assessment literacy scores. Accordingly, the present review examined instrument type, publication year, geographical region, teacher status, teaching level, and subject specialization as potential moderators of between-study heterogeneity.

To address these gaps, the present study conducted a systematic review and meta-analysis of studies using the TALQ, ALI, and CALI instruments. Specifically, the study sought to answer the following research questions:

1. What is the pooled level of teacher assessment literacy across studies using the ALI, CALI, and TALQ instruments?
2. To what extent do contextual and study characteristics (e.g., instrument, publication year, geographical region, teacher status, teaching level, and subject specialization) explain variability in reported assessment literacy scores?

## RESEARCH METHOD

To address the research questions, this study employed a systematic review and meta-analysis. The study selection followed the four-stage PRISMA 2020 framework (Page et al., 2021), comprising identification, screening, eligibility, and inclusion. During the identification stage, records were retrieved from the predefined databases. Titles and abstracts were then screened against the eligibility criteria, followed by full-text assessment of potentially relevant studies. Reasons for exclusion at the full-text stage were documented, and only studies meeting all inclusion criteria were retained for the quantitative synthesis. The study selection process is summarized in the PRISMA flow diagram (Figure 1). The meta-analysis quantified overall assessment literacy levels as measured by the Assessment Literacy Inventory (ALI), the Classroom Assessment Literacy Inventory (CALI), and the Teacher Assessment Literacy Questionnaire (TALQ).

Although the ALI, CALI, and TALQ differ in wording, administration, and psychometric characteristics, they were all developed from the same conceptual framework, the Standards for Teacher Competence in the Educational Assessment of Students (American Federation of Teachers [AFT], National Council on Measurement in Education [NCME], & National Education Association [NEA], 1990). Moreover, they retain a common structural blueprint of 35 dichotomously scored items distributed across seven assessment standards. Consistent with guidance from the Cochrane Handbook for Systematic Reviews of Interventions, studies using different instruments may be synthesized when the instruments are judged to measure the same underlying construct based on their conceptual foundations, development, and intended purpose, particularly when formal evidence of psychometric equivalence is unavailable (Higgins et al., 2024). In the present review, the ALI, CALI, and TALQ were considered conceptually comparable because they were all explicitly developed to assess teacher assessment literacy using the same competency framework. Independent reviews likewise classify these instruments as objective measures of teacher assessment literacy while acknowledging differences in their psychometric properties (DeLuca et al., 2016; Gotch & French, 2014; Pastore, 2023). Consequently, these instruments were synthesized as alternative operationalizations of a common construct rather than as measures of different constructs. To evaluate whether instrument-related differences contributed to between-study variability, instrument type was also examined as a potential moderator. Relevant instrument characteristics are summarized in Table 1.

The meta-analysis was conducted using study-level effect sizes (mean assessment literacy scores and their corresponding sampling variances), estimated under a random-effects model. Consequently, the pooled estimate was derived from weighted effect sizes rather than simple averaging of raw scores.

**Table 1**

*Comparative Characteristics of Teacher Assessment Literacy Instruments Included in the Meta-Analysis*

| Characteristic | ALI | CALI | TALQ |
|---|---|---|---|
| Underlying construct | Teacher Assessment Literacy | Teacher Assessment Literacy | Teacher Assessment Literacy |
| Theoretical basis | | AFT/NCME/NEA (1990) | |
| Items | | 35 | |
| Score range | | 0–35 | |
| Standards | | 7 | |
| Instrument development | Campbell et al., 2002 | Mertler, 2003 | Plake et al., 1993 |

*Note.* Instrument characteristics were synthesized from Plake et al. (1993), Campbell et al. (2002), and Mertler (2003). The table summarizes the shared conceptual characteristics that justified pooling the three instruments in the present meta-analysis. As documented by Mertler and Campbell (2005), Campbell et al. (2002) and Mertler (2003) used the same original TALQ instrument (the latter with slight modifications), whereas the scenario-based ALI represented a subsequent redevelopment designed to improve readability, contextualization, and psychometric performance while retaining the same 35-item, seven-standard framework.

## Eligibility Criteria

Eligibility criteria included studies of in-service or pre-service teachers that assessed assessment literacy using the ALI, CALI, or TALQ instruments. Publications from 1990 onward, including peer-reviewed articles, conference papers, and theses, were eligible. Studies were excluded if they focused on student assessment tools, lacked sufficient quantitative data, were not published in English or French, or did not report a clear assessment methodology.

## Search Strategy

This review followed PRISMA 2020 guidelines. Searches were conducted in Scopus, ScienceDirect, and Google Scholar between September 1 and October 1, 2025. The protocol was registered on the Open Science Framework (OSF) prior to screening. Boolean search strings combined terms related to assessment literacy and the three target instruments, with syntax adapted to each database. For Scopus and ScienceDirect, the search combined "assessment literacy," "teachers' assessment literacy," "assessment literacy inventory," "classroom assessment literacy inventory," and "teacher assessment literacy questionnaire." For Google Scholar, the search included "assessment literacy," "teachers' assessment literacy," "assessment literacy inventory," "investigating teachers' assessment literacy level," "classroom assessment literacy inventory," and "teacher assessment literacy questionnaire." Google Scholar results were screened sequentially until additional pages yielded no new eligible studies. The complete search strategies are provided in Table A1.

## Study Selection

The search yielded 1,438 records (Scopus = 852; ScienceDirect = 436; Google Scholar = 150). After removal of 163 duplicates, 1,275 records underwent title and abstract screening. Thirty-eight articles were retained for full-text review, of which 18 met the inclusion criteria and were included in the final meta-analysis. Exclusions were primarily due to irrelevant focus or use of non-eligible instruments. The full selection process is presented in the PRISMA flow diagram (Figure 1).

## Study Selection Reliability

All records were independently screened by two reviewers using predefined inclusion and exclusion criteria. Discrepancies at the title and abstract stage were resolved through discussion until consensus was achieved; full-text screening resulted in complete agreement.

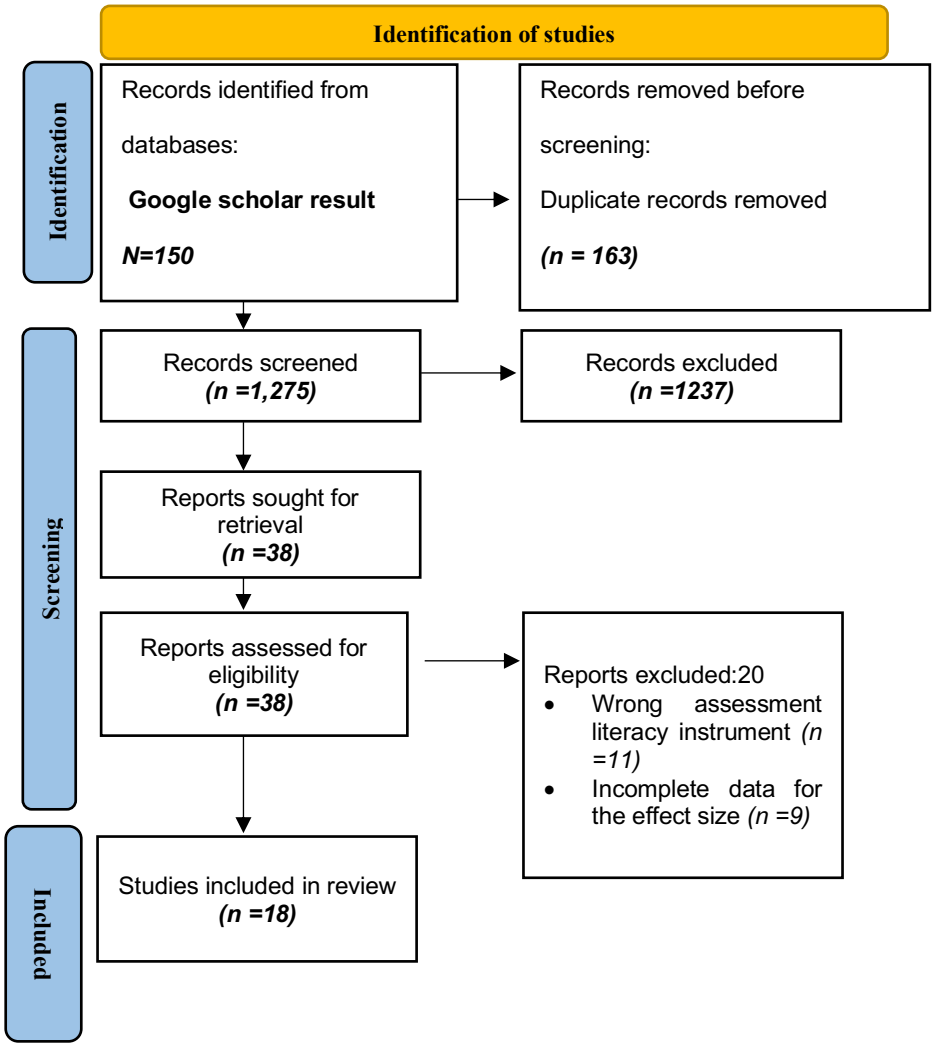
## Quality Assessment

Included studies were evaluated for reporting transparency, sampling clarity, instrument specification, and reliability reporting. All provided sufficient statistical information (means, standard deviations, and internal consistency coefficients) for meta-analysis. Although reporting detail varied, no study was excluded based solely on quality; study characteristics were considered when interpreting heterogeneity.

## Data Extraction

Data extraction followed a structured form adapted from Cochrane guidelines**.** No separate validated study-level quality appraisal instrument was applied. Of the 18 included studies, four reported results for two independent subgroups (e.g., in-service vs. pre-service teachers). Because these subgroups consisted of non-overlapping participants, they were treated as separate samples, yielding 21 independent samples for the meta-analysis. When primary studies reported the standard error of the mean (SEM), or when reported measures of dispersion were verified from the original study to represent SEM rather than SD, standard deviations were calculated using, $SD = SEM \times \sqrt{N}$, following standard meta-analytic procedures.

**Figure 1**
*PRISMA research process diagram (Page MJ, et al. BMJ 2021)*




## Statistical Analysis

Meta-analysis was conducted using the metafor package in R using (Version 4.5.1; R Core Team, 2025) with the *metafor* package (Version 4.8-0; Viechtbauer, 2010). Data preparation was performed using the *dplyr* package (Version 1.2.1). A random-effects model was used to estimate the pooled assessment literacy level across studies. Heterogeneity was assessed using Cochran's $Q$ and the $I^2$ statistic. Moderator analyses were performed by instrument, geographical region, and publication year, while publication bias was evaluated using Egger's regression test and the trim-and-fill method. Although the included instruments differ slightly in format, all are derived from the same seven-domain framework established by AFT/NCME/NEA (1990) and employ comparable scoring structures, supporting the appropriateness of cross-instrument mean comparison.

## RESULTS

**Table 2**

*Characteristics of the Included Studies*

| Study | Country | Instrument | *N* | Status | Teaching level | Subject | *Mean* | *SD* | Percentage | Assessment literacy category |
|---|---|---|---|---|---|---|---|---|---|---|
| Yawmtim V, Wongwanish S ,2014 | Thailand | CALI | 19 | Inservice | Primary | Mixed | 17.11 | 3.62 | 48.89 | Low |
| Sub 1 Afshar, H, Ranjbar, N 2021 | Iran | CALI | 23 | Inservice | Higher education | Technical English | 19.08 | 2.81 | 54.51 | Low |
| Sub2 Afshar, H, Ranjbar, N 2021 | Iran | CALI | 27 | Inservice | Higher education | English | 24.11 | 3.74 | 68.89 | Medium |
| Sub 1 Piosang, 2017 | Philippines | CALI | 40 | Inservice | Secondary education | English | 13.78 | 5.93 | 39.37 | Low |
| Sub 2 Piosang, T, L,2017 | Philippines | CALI | 40 | Inservice | Higher education | English | 10.87 | 6.57 | 31.06 | Low |
| Shams, J.A, al 2018 | Pakistan | CALI | 430 | Inservice | Higher education | None specified | 13.64 | 1.79 | 38.97 | Low |
| Rezai Fard, Tabatabaei 2018 | Iran | CALI | 52 | Inservice | Mixed school levels | English | 18.39 | 3.74 | 52.54 | Low |
| Sub 1 Butuner, S, al 2010 | Turkey | ALI | 34 | Preservice | Secondary education | Mathematics | 15.94 | 2.56 | 53.14 | Low |
| Sub 2 Butuner, S, al 2010 | Turkey | ALI | 41 | Inservice | Secondary education | Mathematics | 15.73 | 3.03 | 52.43 | Low |
| Buyukkarci, K. 2016 | Turkey | ALI | 32 | Inservice | Mixed school levels | English | 12.9 | 3.88 | 36.86 | Low |
| Ergul, A.O, Cetin, S 2021 | Turkey | ALI | 189 | Inservice | secondary education | mixed | 12.94 | 3.66 | 43.14 | Low |

| Study | Country | Instrument | *N* | Status | Teaching level | Subject | *Mean* | *SD* | Percentage | Assessment literacy category |
|---|---|---|---|---|---|---|---|---|---|---|
| Soodmand Afshar al 2018 | Iran | CALI | 106 | Inservice | higher education | English | 12.12 | 2.36 | 34.63 | Low |
| Akayuure, P,2021 | Ghana | CALI | 96 | Inservice | secondary education | mathematics | 17.5 | 4.5 | 50 | Low |
| Sub 1Mertler, C, A 2004 | USA | CALI | 67 | preservice | unclear | none specified | 18.96 | 4.65 | 54.17 | Low |
| Sub 2 Mertler, C, A 2004 | USA | CALI | 101 | Inservice | secondary education | mixed | 21.67 | 3.08 | 61.91 | Medium |
| Kruse, L, al 2020 | USA | ALI | 96 | preservice | secondary education | mixed | 18.05 | 3.76 | 51.57 | Low |
| Mertler, C, A, Campbell, C 2005 | USA | ALI | 249 | preservice | unclear | none specified | 23.83 | 4.64 | 68.09 | Medium |
| Muhammad, F, H, BARDAKCI, M 2019 | Iraq | CALI | 101 | Inservice | secondary education | english | 14.346 | 4.459 | 40.99 | Low |
| Napanoy, J, B, Peckley, M, K 2020 | Philippines | CALI | 101 | Inservice | Primary | mixed | 10.72 | 8.63 | 30.63 | Low |
| Jawhar, S, S, Subahi, A, M, 2020 | Saudi Arabia | CALI | 54 | Inservice | higher education | mixed | 16.8 | 4.5 | 48 | Low |
| Clores, M, A; Reganit, A, A, R, 2020 | Philippines | TALQ | 241 | Inservice | secondary education | mixed | 17.15 | 3.73 | 49 | Low |

Characteristics of Included Studies. Table 2 summarizes the characteristics of the 18 included studies (21 independent samples), including country, assessment literacy instrument, sample size, teacher status, teaching level, subject specialization, mean assessment literacy score, standard deviation. Overall, the studies represented diverse educational contexts, teacher populations, and geographical regions. With respect to teaching level, nine samples (42.9%) examined secondary school teachers, six (28.6%) focused on higher education instructors, two (9.5%) included primary school teachers, two (9.5%) involved mixed secondary and higher education samples, and two (9.5%) did not report a specific teaching level. This distribution demonstrates that the TALQ, ALI, and CALI instruments have been applied across a broad

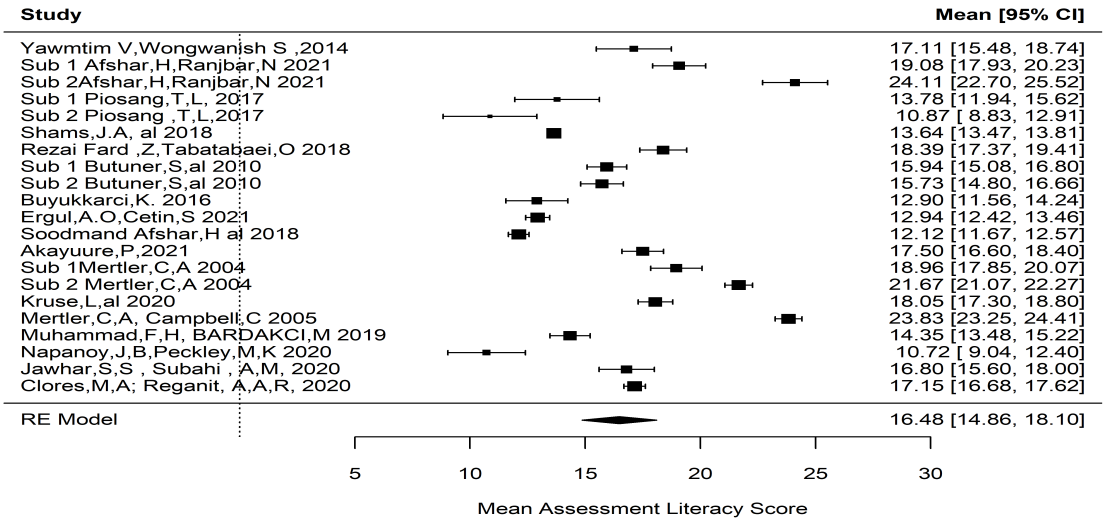
range of educational settings, although secondary and higher education contexts accounted for the majority of the available evidence.

**Meta- Analysis of teacher's assessment literacy scores**

The pooled mean assessment literacy score was 16.48 (95% CI [14.86, 18.10], $p < .001$) on a 35-point assessment literacy scale, corresponding to a low level of teacher assessment literacy according to the classification proposed by Yamtim and Wongwanich (2014). The forest plot (Figure 2) displays individual study means with their corresponding 95% confidence intervals, illustrating both the dispersion of scores and the overall pooled estimate.

**Figure 2**

*Forest plot of pooled assessment literacy score across 21 studies*




Across the 21 independent samples, the pooled mean assessment literacy score was 16.48 *(sd = 0.827*) on a 35-point assessment literacy scale, indicating that teachers generally demonstrated poor (needs improvement) assessment literacy according to the classification proposed by Yamtim and Wongwanich (2014). Consistent with their criteria, scores below 60% of the maximum possible score were classified as poor (needs improvement), scores between 60% and 79% as fair, and scores of 80% or higher as good. On the common 0–35 scale used by the ALI, CALI, and TALQ instruments, these thresholds correspond to scores of <21, 21–27, and ≥28, respectively.

***Heterogeneity***

The Cochran's statistic is $Q= 2387.08, p<0.0001$ $I^2=99.21\%,$ suggesting that the variability in this study is not due to a sample error. The results justified the need for moderator analyses to explain the observed heterogeneity.

***Moderator Analyses***

To explore potential sources of heterogeneity, mixed-effects meta-regression analyses were conducted using instrument, publication year, geographical region, teacher status, teaching level, and subject specialization as moderators.

Instrument type did not explain variability in assessment literacy scores, *QM* (2) = 0.04, *p = .981,* indicating that the TALQ, ALI, and CALI yielded comparable estimates of teacher assessment literacy despite differences in their design. Likewise, subject specialization was not significantly associated with assessment literacy levels, *QM* (4) = 2.14*, p = .710*, accounting for virtually none of the observed between-study heterogeneity *($R^2$ = 0%).*

Publication year and geographical region demonstrated marginal associations with assessment literacy. Publication year accounted for *10.7%* of the between-study heterogeneity but did not reach statistical significance, *QM* (1) = 3.25*, p = .071.* The negative regression coefficient *($\beta = -0.25$, p = .070)* suggests a non-significant negative trend in reported teacher assessment literacy over time. However, because this association did not reach statistical significance, it should be interpreted cautiously and not as evidence of a systematic decline in teacher assessment literacy. Similarly, geographical region explained 24.5% of the observed heterogeneity, although the overall moderator effect remained non-significant, *QM* (8) = 14.48*, p = .070.*

Teacher characteristics also explained only a limited proportion of the between-study variability. Teacher status (pre-service versus in-service) accounted for 8.6% of the heterogeneity but did not significantly predict assessment literacy, *QM* (1) = 2.82*, p = .093.* Although the estimated coefficient suggested higher assessment literacy among pre-service teachers, this difference was not statistically significant. Likewise, teaching level (primary, secondary, higher education, mixed, and unclear) explained only 3.3% of the heterogeneity and was not a significant moderator, *QM* (4) = 4.62*, p = .328.*

Furthermore, In the combined region-year model results shows that regio-year accounted for 20.99% of variability across studies. The combined model is not statistically significant as the test moderator is *QM* =14.34*, p=0.11* indicating that together time and geography didn't predict assessment literacy scores.

Taken together, these findings indicate that none of the examined moderators independently explained the substantial between-study heterogeneity. Although geographical region, publication year, and teacher status showed modest explanatory potential, the persistently high residual heterogeneity suggests that additional study-level factors not consistently reported in the primary studies are likely to contribute to differences in teacher assessment literacy across contexts. (table 3)

**Table 3**

*Moderator Analyses of Teacher Assessment Literacy*

| Moderator | Instrument | Country | Year | Status | Teaching level | Subject |
|---|---|---|---|---|---|---|
| QM | 0.04 | 14.48 | 3.25 | 2.82 | 4.62 | 2.14 |
| P | 0.98 | 0.070 | 0.071 | 0.093 | 0.328 | 0.710 |
| $R^2$ | 0 | 24.5 | 10.7 | 8.6 | 3.3 | 0 |

**Exploratory descriptive analysis by assessment literacy category**

To facilitate interpretation of the substantial between-study heterogeneity, an exploratory subgroup analysis was conducted using the predefined assessment literacy categories (Low and Medium). Because these categories were derived directly from the observed study mean scores, this analysis was not intended to identify an independent source of heterogeneity, but rather to provide a descriptive summary of how studies clustered according to the practical interpretation of teachers' assessment literacy levels.

Meta-regression using assessment literacy category as the grouping variable identified a significant distinction between the two categories, *QM* (1) = 23.90, $p < .001$. Although the model yielded an $R^2$ value of 54.96%, this statistic is reported for descriptive purposes only, as the assessment literacy categories were derived from the same study mean scores used in the analysis. Studies classified as demonstrating low assessment literacy yielded a pooled mean score of 15.37 (95% CI [14.13, 16.61]), whereas studies

classified as demonstrating medium assessment literacy produced a pooled mean score of 23.14 (95% CI [21.60, 24.69]).

Despite this distinction, substantial heterogeneity remained within both *categories (Low: $I^2$ = 98.36%; Medium: $I^2$ = 91.90%),* indicating considerable variability among studies sharing the same assessment literacy classification. Accordingly, these findings should be interpreted as descriptive. Rather than identifying an independent source of between-study heterogeneity, the subgroup analysis demonstrates that, despite variation in reported mean scores, the majority of studies converged on the same practical interpretation of teachers' assessment literacy levels.

**Influence Diagnostics**

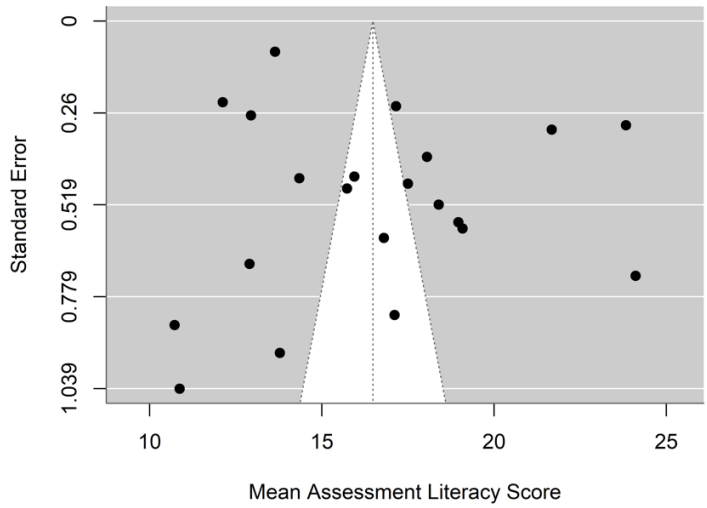
To assess whether individual studies affected pooled results we conducted influence diagnostic. The pooled mean remained stable across leave-one-out iterations, ranging from 16.11 to 16.76, compared to the overall estimate of 16.48 on a 35-point assessment literacy scale. These results suggest that the findings are robust and not driven by any single study. Nevertheless, no study exceeded 1 for cook's distance indicating that no individual study exerted undue influence on the pooled estimate. (Figure A1).

**Publication Bias Assessment**

Publication bias was examined using three complementary approaches: visual inspection of a funnel plot, Egger's regression test, and the Duval and Tweedie trim-and-fill procedure. Visual inspection of the funnel plot (Figure 3) indicated a reasonably symmetrical distribution of effect sizes around the pooled mean, with no clear evidence of missing small-study effects. Egger's regression test was non-significant *($z = –0.92$, $p = .36$)*, suggesting no statistical evidence of funnel plot asymmetry.

**Figure 3**

*The Funnel plot assessing publication bias*




Additionally, the trim-and-fill analysis imputed one potentially missing study on the right side of the distribution. However, the adjusted pooled mean (16.76, 95% CI [15.12, 18.40]) was nearly identical to the original estimate (16.48, 95% CI [14.86, 18.10*]),* indicating that any potential publication bias had

minimal impact on the overall results (Figure A2). Collectively, these analyses suggest that publication bias is unlikely to have materially influenced the pooled estimate of teacher assessment literacy.

**Cumulative Meta-Analysis**

A cumulative meta -analysis was performed to identify how the pooled mean assessment literacy evolves as studies added in chronological order (Figure A3). The mean stabilized around *M = 16.4* (95% CI [14.86, 18.1]) after roughly the tenth study, although early cumulative estimates varied more (from *M = 18.96* to 21.52). This suggests that rather than significantly changing the overall estimate mean, additional study has mainly supported earlier findings. considering the high level of between-study heterogeneity *($I^2$ > 99%),* the cumulative pattern indicates that research on teacher assessment literacy, as determined by the TALQ, CALI, and ALI instruments, has empirically stabilized although subsequently decreased and stabilized near 16.5 as additional studies were incorporated. The cumulative mean assessment literacy scores showed considerable fluctuation in the early stages as studies were added sequentially by year, reflecting the limited amount of available data. However, after approximately the tenth study, the mean stabilized around *M* = 16.5, closely matching the overall random-effects estimate (Figure A4). This stabilization indicates that the pooled estimate became increasingly stable as additional studies were incorporated, suggesting that subsequent evidence reinforced the overall estimate rather than substantially changing it.

## DISCUSSION AND CONCLUSIONS

The meta-analysis revealed a pooled mean score of 16.48 on a 35-point assessment literacy scale, indicating poor teacher assessment literacy across the included studies according to the classification proposed by Yamtim and Wongwanich (2014). This finding suggests that, despite more than three decades of research and continued emphasis on assessment in teacher education, many teachers continue to demonstrate insufficient assessment competence. Given the central role of assessment in informing instructional decisions, monitoring student learning, and supporting formative assessment practices, this finding raises concerns regarding the extent to which current teacher preparation and professional development adequately equip teachers with the knowledge and skills required for effective classroom assessment. This interpretation is consistent with previous research reporting similarly low levels of teacher assessment literacy across diverse educational contexts (DeLuca et al., 2016; Hull & Vígh, 2026; Muhson et al., 2025; Pastore, 2023)

Although the present meta-analysis did not directly examine the mechanisms underlying teachers' assessment literacy, previous research has suggested that one possible explanation for the consistently low levels observed across studies is the persistent gap between theoretical assessment knowledge and its implementation in classroom practice. Furthermore, teachers may possess basic knowledge of assessment concepts, previous research suggests that they frequently experience difficulties translating this knowledge into authentic instructional settings (Çibukçiu & Koliqi, 2026; Kang & Lam, 2024; Pastore, 2023). These challenges extend beyond the design of assessment tasks to the interpretation and use of assessment evidence for instructional decision-making (Clark et al., 2022; Mandinach & Gummer, 2016). Consequently, the low pooled assessment literacy identified in the present meta-analysis may reflect not only limitations in teachers' technical assessment knowledge but also broader difficulties applying assessment information to support teaching and learning.

The present meta-analysis identified substantial between-study heterogeneity *($I^2$ = 99.21%),* indicating considerable variation in reported teacher assessment literacy scores across studies. Although such heterogeneity may initially appear to limit the interpretability of the pooled estimate, it is not unexpected in educational research synthesizing evidence across diverse educational systems, teacher populations, and institutional contexts (Borenstein et al., 2009). The moderator analyses conducted in the

present study further suggest that this variability cannot be attributed to a single study characteristic. Instrument type, teaching level, teacher status, and subject specialization did not significantly explain between-study differences, while geographical region and publication year demonstrated only modest explanatory potential. Collectively, these findings indicate that teacher assessment literacy is shaped by a complex interaction of contextual factors rather than by any individual study-level characteristic.

To better understand the practical implications of this heterogeneity, an exploratory subgroup analysis was conducted using predefined assessment literacy categories. Because these categories were derived from the observed study mean scores, this analysis was intended to facilitate interpretation rather than identify an independent source of heterogeneity. Although the included studies differed considerably in their reported mean scores, the majority converged on the same substantive conclusion regarding teachers' assessment literacy. Specifically, 18 of the 21 independent samples (85.7%) were classified as demonstrating low assessment literacy, whereas only three samples (14.3%) reached the medium category, and none were classified as high. Moreover, the subgroup analysis yielded an $R^2$ value of approximately 55%; however, because the assessment literacy categories were derived from the same study mean scores, this statistic should be interpreted descriptively rather than as evidence that assessment literacy category independently explained between-study heterogeneity.

Taken together, these findings illustrate that the substantial statistical heterogeneity did not preclude a consistent substantive interpretation of teachers' assessment literacy across the included studies. In other words, although reported assessment literacy scores varied across educational contexts, countries, and teacher populations, most studies consistently indicated inadequate levels of teacher assessment literacy. Consequently, the pooled estimate should be interpreted as an empirical benchmark of teacher assessment literacy rather than as a universal value applicable to every educational system. The remaining unexplained heterogeneity likely reflects contextual factors that were not consistently reported in the primary studies, including differences in national assessment policies, teacher education program, professional development opportunities, institutional support, and assessment culture.

The robustness of these findings was further supported by the absence of substantial publication bias. Egger's regression test was non-significant, and trim-and-fill analysis identified only one potentially missing study, resulting in a negligible change in the pooled estimate. Together with the sensitivity analyses, these findings suggest that publication bias is unlikely to have materially influenced the overall estimate of teacher assessment literacy, supporting the stability of the pooled results despite the substantial between-study heterogeneity.

Because neither geographical region nor publication year emerged as statistically significant moderators, the following interpretation should be regarded as exploratory and intended to provide contextual insights rather than confirm definitive differences between educational contexts. Although geographical region accounted for 24.5% of the between-study heterogeneity, the overall moderator effect did not reach statistical significance ($QM$ (8) = 14.48, $p = .070$). Consequently, regional differences should be interpreted cautiously and regarded as exploratory rather than confirmatory. Nevertheless, previous research suggests that variation in teacher assessment literacy may reflect differences in assessment policy, teacher education, professional development opportunities, and institutional support rather than geographic location itself. For example, studies conducted in the United States have reported relatively stronger conceptual assessment knowledge despite persistent challenges in interpreting assessment data and applying assessment information in classroom practice (Clark et al., 2022; Gareis & Grant, 2015; Kim et al., 2020). Likewise, some studies from Iran have reported comparatively higher assessment literacy scores alongside continuing concerns regarding teachers' preparation and reliance on traditional assessment practices (Salimi & Farsi, 2020). In contrast, studies from Saudi Arabia, India, Indonesia, Thailand, and Ghana have identified recurring barriers, including limited assessment training, inadequate institutional support, examination-oriented educational cultures, and restricted opportunities for continuing professional development (Alsowat, 2022; Imsa-ard & Tangkiengsirisina, 2023; Prastikawati et al., 2024; Quainoo et al., 2025; Suherman et al., 2025; Varshini & Venkatalakshmi, 2024). However, because the geographical

region moderator did not reach statistical significance, these country-level comparisons should be interpreted as illustrative examples from the broader literature rather than evidence of systematic regional differences in teacher assessment literacy. Future research should therefore investigate the specific policy and professional mechanisms underlying these contextual differences rather than geographic location itself.

Publication year demonstrated a modest but non-significant association with assessment literacy, accounting for approximately 10.7% of the between-study heterogeneity ($QM$ (1) = 3.25, $p = .071$). The negative regression coefficient ($\beta = -0.25$) suggests a slight decline in reported assessment literacy over time; however, this trend should be interpreted cautiously because it did not reach statistical significance. Rather than indicating a genuine deterioration in teachers' assessment competence, this pattern may reflect the evolving conceptualization of assessment literacy. As noted by DeLuca et al. (2016), conceptions of assessment literacy have expanded considerably over recent decades, moving beyond technical measurement skills to encompass formative assessment, professional judgment, ethical decision-making, and contextual responsiveness. Consequently, more recent studies may be evaluating a broader and more demanding construct than earlier investigations, making direct comparisons across time increasingly complex. Regardless of whether this trend reflects changing expectations or actual differences in teacher competence, the findings reinforce the need for teacher education and professional development programs address assessment literacy as a comprehensive professional capability rather than a narrowly defined technical skill.

## IMPLICATIONS

The consistently low levels of teacher assessment literacy identified across studies, with 85.7% of independent samples classified as demonstrating poor assessment literacy, highlight the need for sustained investment in both pre-service teacher education and continuous professional development. Given that this pattern was observed across diverse educational contexts and that no individual moderator independently explained the substantial between-study heterogeneity, strengthening assessment literacy should extend beyond isolated training initiatives. Rather than treating assessment as a peripheral pedagogical skill, teacher preparation program should integrate assessment literacy as a core professional competence, emphasizing the design, implementation, interpretation, and ethical use of assessment to support student learning (Kainulainen & Tarnanen, 2026). At the same time, the substantial unexplained heterogeneity observed across studies suggests that professional development should be context-responsive, allowing program to address local educational needs while maintaining shared standards for assessment competence.

## LIMITATIONS

This review was limited to studies using three established instruments, which, while foundational, may not capture newer or domain-specific measures of teacher assessment literacy. Although these instruments assess the same underlying construct, differences in content coverage, psychometric properties, and scoring procedures may have contributed to variability across studies. Nevertheless, instrument type was not a significant moderator, suggesting that these differences did not substantially influence the pooled estimates. Although the search was comprehensive, studies published outside English or French or in non-indexed sources may have been missed. Substantial heterogeneity ($I^2 = 99.21\%$) also remained largely unexplained despite examining several moderators, indicating that important contextual factors were either inconsistently reported or unavailable in the primary studies. Consequently, interpretations of regional and temporal patterns should be regarded as exploratory and hypothesis-generating rather than confirmatory. Finally, the predominance of cross-sectional designs limits conclusions regarding longitudinal changes in teacher assessment literacy. Future research should report participant and contextual characteristics more consistently to facilitate more comprehensive moderator analyses and meta-analytic syntheses.

**Contribution of this Review**

This review provides the first quantitative synthesis of teacher assessment literacy measured using the three most widely adopted objective instruments (TALQ, ALI, and CALI). By establishing an empirical benchmark and examining between-study heterogeneity across comparable measures, it offers a methodological and empirical foundation for future assessment literacy research, instrument development, and evidence-informed teacher education and professional development.

## CONCLUSION

This systematic review and meta-analysis synthesized evidence from 18 studies using the ALI, CALI, and TALQ instruments to examine teachers' assessment literacy. The findings indicate consistently low levels of teachers' reported assessment literacy across studies, with a pooled mean score of 16.48 and substantial heterogeneity across educational contexts. Although assessment literacy varied by instrument, region, and publication year, no single factor independently explained the observed variability. These findings underscore the need to strengthen assessment literacy through pre-service teacher education, continuous professional development, and context-responsive educational policies.

**Author Contributions:** Conceptualization, Z.L.; methodology, Z.L.; investigation, Z.L.; writing-original draft preparation, Z.L.; writing-review and editing, Z.L.; supervision, A.E.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors disclose that they have not received any funding for this manuscript.

**Acknowledgments:** The first author would like to thank the second author for his valuable supervision. During the preparation of this manuscript, the authors used ChatGPT for the purposes of grammar, language, and formatting assistance. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

**Conflicts of Interest:** In accordance with my ethical obligation as a researcher, I am reporting that the authors disclose that they have no actual or perceived conflicts of interest.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the publisher and/or the editor(s). The publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## REFERENCES

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

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, *9*(4), 30–32. https://doi.org/10.1111/j.1745-3992.1990.tb00391.x

Akayuure, P. (2021). Classroom assessment literacy levels of mathematics teachers in Ghanaian senior high schools. *Contemporary Mathematics and Science Education*, *2*(2), Article ep21013. https://doi.org/10.30935/conmaths/11286

Alsowat, H. H. (2022). An investigation of Saudi EFL teachers' perceptions of learning-oriented language assessment. *European Journal of English Language and Literature Studies*, *10*(3), 16–32. https://doi.org/10.37745/ejells.2013/vol10no3pp.16-32

Barnes, M. E., & Qasserras, L. (2024). Low assessment literacy in an assessment-obsessed field: What teacher candidates learn about assessment during teacher education. *Journal of Curriculum and Pedagogy*. Advance online publication. https://doi.org/10.1080/15505170.2024.2354187

Beziat, T. L. R., & Coleman, B. K. (2015). Classroom assessment literacy: Evaluating preservice teachers. *The Researcher*, *27*(1), 25–30.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Wiley. https://doi.org/10.1002/9780470743386

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

Bütüner, S. Ö., Eldem, E., & Arslan, S. (2010). Öğretmen adaylarının ölçme ve değerlendirme okuryazarlıklarının değerlendirilmesi: Assessment Literacy Inventory'nin Türkçe uyarlaması [Assessment Literacy Inventory–Turkish adaptation]. *Kuram ve Uygulamada Eğitim Bilimleri*, *10*(3), 1635–1664.

Büyükkarcı, K. (2016). Identifying the areas for English language teacher development: A study of assessment literacy. *Pegem Eğitim ve Öğretim Dergisi*, *6*(3), 333–346. https://doi.org/10.14527/pegegog.2016.017

Campbell, C., Murphy, J. A., & Holt, J. K. (2002, October). *Psychometric analysis of an assessment literacy instrument: Applicability to pre-service teachers* [Paper presentation]. Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH, United States.

Clark, A., Kobrin, J. L., Karvonen, M., & Hirt, A. (2023). Teacher use of diagnostic score reports for instructional decision-making in the subsequent academic year. *Practical Assessment, Research, and Evaluation*, *28*(1), Article 6. https://doi.org/10.7275/pare.1255

Clark, A. K., Nash, B., Burnes, J., & Karvonen, M. (2022). Teacher assessment literacy: Implications for diagnostic assessment systems. *Applied Measurement in Education*, *35*(1), 17–32. https://doi.org/10.1080/08957347.2022.2034823

Clores, M. A., & Reganit, A. A. R. (2020). Investigating the assessment literacy of teachers in private junior high schools in the Philippines. *Humanities, Arts and Social Sciences Studies*, *20*(2), 461–476.

Çibukçiu, B., & Koliqi, D. (2026). Measuring teacher formative assessment literacy in primary education: A multidimensional study using the T-FALS scale. *Frontiers in Education*, *11*, 1787953. https://doi.org/10.3389/feduc.2026.1787953

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*, 279–301. https://doi.org/10.1007/s11092-022-09389-9

Dang, J., & Li, S. (2025). Research on the current status of assessment literacy of Chinese primary and secondary school teachers. *Education Reform and Development*, *7(*5), 214–225. https://doi.org/10.26689/erd.v7i5.10472

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). 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

Efendi, R., Rustaman, N. Y., & Kaniawati, I. (2017). Profile of prospective physics teachers on assessment literacy. *Journal of Physics: Conference Series*, *812*, Article 012061. https://doi.org/10.1088/1742-6596/812/1/012061

Ergül, A. Ö., & Çetin, S. (2021). Measurement and assessment literacy levels of teachers in terms of some variables. *Adiyaman University Journal of Educational Sciences*, *11*(1), 26–35.

Fu, R., & Koh, K. (2026). Developing and validating teacher formative assessment literacy questionnaire in the high-stakes examination culture: A case of China. *Frontiers in Psychology*, *17*, 1771941. https://doi.org/10.3389/fpsyg.2026.1771941

Gareis, C. R., & Grant, L. W. (2015). Assessment literacy for teacher candidates: A focused approach. *Teacher Educators' Journal*, *2015*, 4–21.

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

Gouëdard, P. (2021). *Developing indicators to support the implementation of education policies* (OECD Education Working Papers No. 255). OECD Publishing. https://doi.org/10.1787/b9f04dd0-en

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2024). *Cochrane handbook for systematic reviews of interventions* (Version 6.5). Cochrane. https://training.cochrane.org/handbook

Hull, P., & Vígh, T. (2026). Preservice teachers' assessment literacy: A systematic review. *Studies in Educational Evaluation*, *89*, 101600. https://doi.org/10.1016/j.stueduc.2026.101600

Imsa-ard, P., & Tangkiengsirisina, S. (2023). Language assessment literacy: Thai EFL university lecturers' knowledge, beliefs, and practices. *PASAA: Journal of Language Teaching and Learning in Thailand*, *66*, 1–34. https://doi.org/10.58837/chula.pasaa.66.1.2

Issa, H. M., & Castulo, N. J. (2026). Investigating teaching practices in large classes: A case study of public primary schools in Tanzania. *Journal of Interdisciplinary Studies in Education*, *15*(1), 1–28. https://doi.org/10.32674/zr9wew97

Jawhar, S. S., & Subahi, A. M. (2020). The impact of specialty, sex, qualification, and experience on teachers' assessment literacy at Saudi higher education. *International Journal of Learning, Teaching and Educational Research*, *19*(6), 1–20. https://doi.org/10.26803/ijlter.19.6.1

Kainulainen, J., & Tarnanen, M. (2026). *Becoming an assessment-literate teacher in teacher education*. *Teaching and Teacher Education, 171*, Article 105341. https://doi.org/10.1016/j.tate.2025.105341

Kang, M., & Lam, R. (2024). Understanding university English instructors' assessment literacy: A formative assessment perspective. *Language Testing in Asia*, *14*, Article 52. https://doi.org/10.1186/s40468-024-00323-y

Kim, A. A., Chapman, M., Kondo, A., & Wilmes, C. (2020). Examining the assessment literacy required for interpreting score reports: A focus on educators of K–12 English learners. *Language Testing*, *37*(1), 54–75. https://doi.org/10.1177/0265532219859881

Kruse, L., Impellizeri, W., Witherel, C. E., & Sondergeld, T. A. (2020). Evaluating the impact of an assessment course on preservice teachers' classroom assessment literacy and self-efficacy. *Mid-Western Educational Researcher*, *32*(2), 107–125. https://scholarworks.bgsu.edu/mwer/vol32/iss2/3

Looney, A., Cumming, J., van der Kleij, F., & Harris, K. (2018). Reconceptualising assessment capacity: Building on teacher assessment literacy. *Assessment in Education: Principles, Policy & Practice*, *25*(3), 291–305. https://doi.org/10.1080/0969594X.2016.1268090

MacLellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualizations. *Teaching and Teacher Education*, *20*(5), 523–535. https://doi.org/10.1016/j.tate.2004.04.008

Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate? Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, *60*, 366–376. https://doi.org/10.1016/j.tate.2016.07.011

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. (2003). *Preservice versus in-service teachers' assessment literacy: Does classroom experience make a difference?* [Paper presentation]. Annual meeting of the Mid-Western Educational Research Association, Columbus, OH. https://files.eric.ed.gov/fulltext/ED482914.pdf

Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, *33*(1), 49–64.

Mertler, C. A., & Campbell, C. (2005, April). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the Assessment Literacy Inventory* [Paper presentation]. Annual Meeting of the American Educational Research Association, Montréal, Quebec, Canada. https://eric.ed.gov/?id=ED490355

Muhammad, F. H. N., & Bardakçı, M. (2019). Iraqi EFL teachers' assessment literacy: Perceptions and practices. *Arab World English Journal*, *10*(2), 420–430. https://doi.org/10.24093/awej/vol10no2.33

Muhson, A., Kusuma, M., Kusumawardani, D. A., & Putri, R. A. (2025). Teachers' assessment literacy: The impact of gender, education, training and work experience. *International Journal of Evaluation and Research in Education*, *14*(3), 2058–2066. https://doi.org/10.11591/ijere.v14i3.22040

Napanoy, J. B., & Peckley, M. K. (2020). Assessment literacy of public elementary school teachers in the Indigenous communities in Northern Philippines. *Universal Journal of Educational Research*, *8*(11B), 5693–5703. https://doi.org/10.13189/ujer.2020.082203

Owusu-Acheampong, E., & Kwapong, O. A. T. F. (2021). Promoting assessment for learning to ensure access, equity, and improvement in educational outcomes in Ghana. *Journal of Interdisciplinary Studies in Education*, *10*(SI), 39–56.

Pastore, S. (2023). Teacher assessment literacy: A systematic review. *Frontiers in Education*, *8*, Article 1217167. https://doi.org/10.3389/feduc.2023.1217167

Paterno, J. (2001). Measuring success: A glossary of assessment terms. In *Building cathedrals: Compassion for the 21st century*. http://www.angelfire.com/wa2/buildingcathedrals/measuringsuccess.html

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, 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*, Article n71. https://doi.org/10.1136/bmj.n71

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, *12*(4), 10–39. https://doi.org/10.1111/j.1745-3992.1993.tb00548.x

Piosang, T. L. (2017). A cross-sectional analysis of classroom assessment literacy of English teachers in secondary and tertiary levels. *Educational Measurement and Evaluation Review*, *8*(1), 30–48.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, *48*(1), 4–11. https://doi.org/10.1080/00405840802577536

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, *46*(4), 265–273. https://doi.org/10.1080/08878730.2011.605048

Prastikawati, E. F., Mujiyanto, Y., Saleh, M., & Fitriati, S. W. (2024). Teaching practicum: Does it impact pre-service EFL teachers' language assessment literacy? *World Journal of English Language*, *14*(3), 242–252. https://doi.org/10.5430/wjel.v14n3p242

Quainoo, E. Y., Ekuban, A., & Yidana, W. (2025). Classroom assessment literacy of basic school teachers in Cape Coast: A latent profile analysis. *Education Research International*, *2025*,

Article 8862411. https://doi.org/10.1155/2025/8862411
Rezaei Fard, Z., & Tabatabaei, O. (2018). Investigating assessment literacy of EFL teachers in Iran. *Journal of Applied Linguistics and Language Research*, *5*(3), 91–100.
Salimi, E. A., & Farsi, M. (2020). Probing into EFL teachers' assessment literacy and teaching experience: The case of native ESL and non-native EFL teachers. *Journal of English Language Teaching and Learning*, *12*(25), 269–288. https://doi.org/10.22034/elt.2020.10684
Shams, J. A., Iqbal, M. Z., & Iqbal, M. Z. (2018). Investigation of classroom assessment literacy of university teachers of Punjab. *Pakistan Journal of Distance and Online Learning*, *4*(2), 103–118.
Soodmand Afshar, H., Tofighi, S., Asoudeh, M., & Ranjbar, N. (2018). The impact of alternative assessment knowledge, teaching experience, gender, and academic degree on EAP teachers' assessment literacy. *Iranian Journal of English for Academic Purposes*, *7*(2), 1–15.
Stiggins, R. J. (1990). Assessment literacy. *Phi Delta Kappan*, *72*(7), 534–539.
Stiggins, R. J. (1993). Teacher training in assessment: Overcoming the neglect. In S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 1–12). Buros Institute of Mental Measurements. https://digitalcommons.unl.edu/burosteachertraining/4/
Suherman, A., Surayya, S. A., Dwimaulani, A., & Setiawan, B. (2025). A study of formative assessment literacy in Indonesia: Insights from pre-service EFL teachers. *Scope: Journal of English Language Teaching*, *9*(2), 856–866. https://doi.org/10.30998/scope.v9i2.26043
Varshini, V., & Venkatalakshmi, R. (2024). Assessment literacy of tertiary-level educators: Insights from Tamil Nadu. *Shanlax International Journal of Education*, *12*(2), 45–53. https://doi.org/10.58256/btp1d832
Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03
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
Yan, Z., & Brown, G. T. L. (2021). Assessment for learning in the Hong Kong assessment reform: A case of policy borrowing. *Studies in Educational Evaluation*, *68*, Article 100985. https://doi.org/10.1016/j.stueduc.2021.100985
Yamtim, V., & Wongwanich, S. (2014). A study of classroom assessment literacy of primary school teachers. *Procedia - Social and Behavioral Sciences*, *116*, 2998–3004. https://doi.org/10.1016/j.sbspro.2014.01.696

---

**ZINEB LABBAR,** is a PhD student at the Faculty of Education Sciences, Mohammed V University in Rabat, Morocco. Her research interests include teacher assessment literacy, classroom assessment, educational assessment, and educational policy. Email: zineblabbar710@gmail.com

**ABDELLATIF EL MOUDNI**, is an Associate Professor at the Faculty of Education Sciences, Mohammed V University in Rabat, Morocco. His research interests include education, educational assessment, and educational policy. Email: *moudnifse@gmail.com*

---

## APPENDIX A

**Table A1**

*Summary of Search Strategy Across Databases*

| Search base | Science direct | Scopus | Google scholar |
|---|---|---|---|
| Research strings | ("assessment literacy" OR "teachers' assessment literacy" OR "assessment literacy inventory" OR "classroom assessment literacy inventory" OR "teacher assessment literacy questionnaire") | ("assessment literacy" OR "teachers' assessment literacy" OR "assessment literacy inventory" OR "classroom assessment literacy inventory" OR "teacher assessment literacy questionnaire") | "Assessment literacy" OR "teachers' assessment literacy" OR "assessment literacy inventory" OR "investigating teachers' assessment literacy level" OR "classroom assessment literacy inventory" OR "teacher assessment literacy questionnaire" |

**Figure A1**

*Influence diagnostic plot showing leverage, residuals and influence measures for each included study*

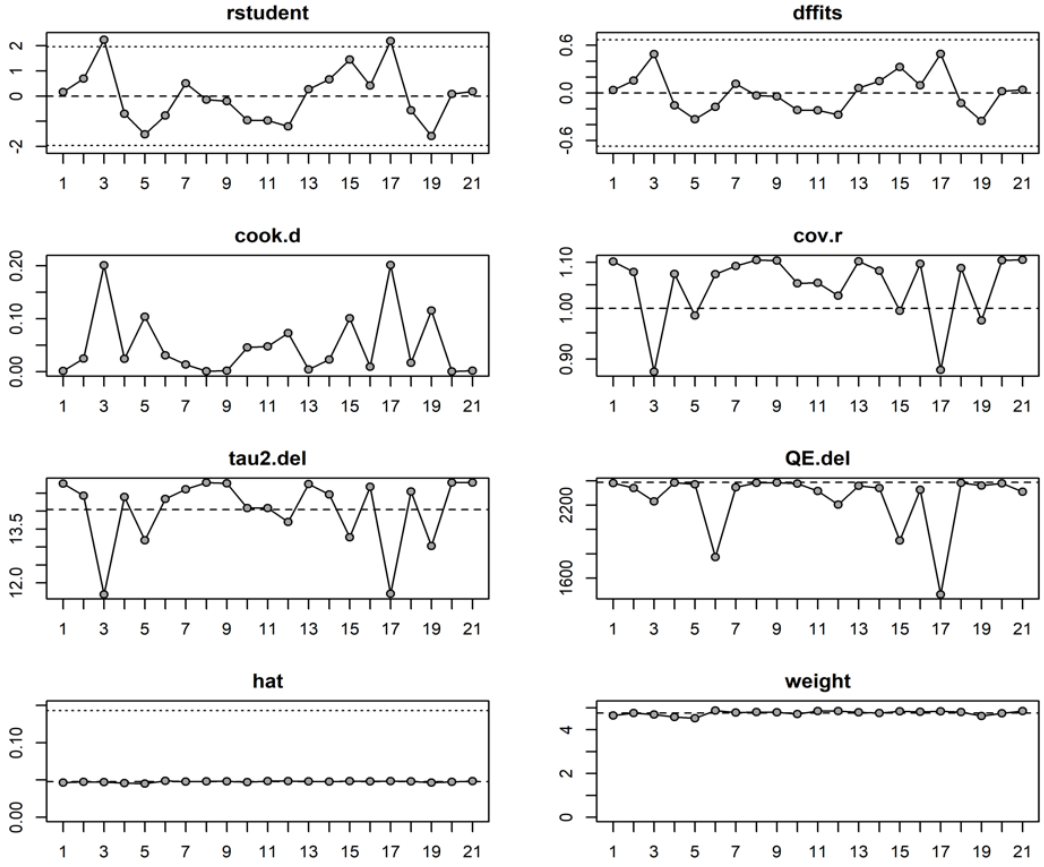


**Figure A2**
*Trim and fill adjusted funnel Plot assessing publication Bias*

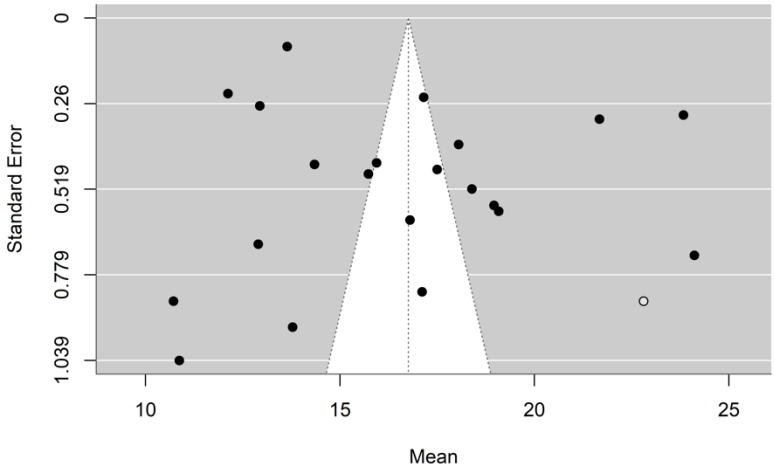



**Figure A3**
*Cumulative Forest plot*

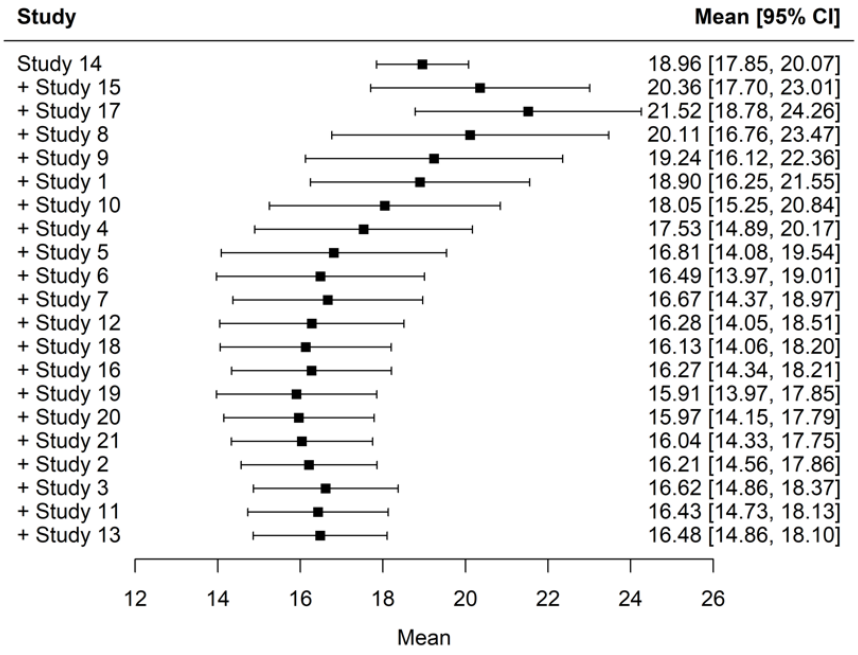


**Figure A4**
*The Trend of Cumulative Mean over Studies*

