IJRES International Journal of Research in Education and Science

www.ijres.net

Assessment Literacy in Teacher Education: Integrating Performance and Perceptions via the ALiPP Framework

Ronald M. Quileste

Xavier University, Philippines, 0000-0002-6388-1445
Corresponding author: Ronald M. Quileste (rquileste@xu.edu.ph)

Article Info

Article History

Received:
3 July 2025

Revised:
2 November 2025

Accepted:
16 December 2025

Published:
1 January 2026

Keywords

Assessment literacy
Pre-service teachers
Mixed - methods
ALiPP framework
Outcomes-based education

Abstract

This study investigated assessment literacy development among 250 pre-service teachers at Xavier University, Philippines, from 2020 to 2024, using the Assessment Literacy Progression and Perception (ALiPP) framework. Employing a convergent parallel mixed-methods design, the study integrated quantitative data, including comprehensive examination scores, GPA, program, gender, and year level, with qualitative data from semi-structured interviews and written reflections to examine performance trends, predictors, and perceptions within the Philippine Outcomes-Based Education context. Statistical analyses included paired t-tests, correlation, ANOVA, and multiple regression, while thematic analysis explored challenges, supports, and perceived relevance. The ALiPP framework modeled assessment literacy as iterative cycles of performance and reflection, addressing gaps in longitudinal research. Findings informed curriculum recommendations emphasizing scaffolded learning, enhanced feedback, and early practicum integration. The study contributes to teacher education by offering a context-specific, mixed-methods approach to assessment literacy that is adaptable to diverse educational settings.

Citation: Quileste, R. M. (2026). Assessment literacy in teacher education: Integrating performance and perceptions via the ALiPP framework. *International Journal of Research in Education and Science (IJRES), 12*(1), 158-176. https://doi.org/10.46328/ijres.5244

ISSN: 2148-9955 / © International Journal of Research in Education and Science (IJRES).

Introduction

Assessment literacy, defined as the ability to design, implement, and interpret assessments to enhance student learning, was a cornerstone of effective teacher education (Popham, 2018; Stiggins, 2021). At Xavier University's School of Education in the Philippines, pre-service teachers developed these competencies through Assessment of Learning 1 and 2. Assessment 1 introduced foundational concepts, such as test construction and validity, while Assessment 2 focused on advanced applications, including rubric design and score analysis (DeLuca et al., 2016). These courses aligned with the Philippines' Outcomes-Based Education (OBE) framework, which emphasized measurable competencies within a resource-constrained, culturally diverse context (Latif & Wasim, 2022). The OBE system required teachers to craft assessments that accurately reflected student outcomes, making assessment literacy critical.

Combining actual performance data (comprehensive exam scores) with student voice (interviews and reflections) provided a comprehensive understanding of assessment literacy development, capturing both measurable progress and subjective experiences (Butler et al., 2021). This mixed-methods approach was grounded in the theoretical framing of assessment literacy as a developmental process, evolving through iterative practice and reflection over time (Popham, 2018; Pastore & Andrade, 2019). Unlike linear models, this perspective highlighted dynamic skill-building, yet few studies explored it longitudinally.

A research gap existed in the scarcity of longitudinal, mixed-methods studies integrating real exam data with student reflections from the same cohort, particularly in non-Western settings like the Philippines (Harding & Kremmel, 2022; Giraldo, 2021). This study addressed this gap by examining assessment literacy among 250 pre-service teachers from 2020 to 2024. The objectives were to investigate performance trends, predictors, and perceptions of assessment literacy development.

Research Questions

The study was guided by six research questions:

1. Was there a significant difference between students' Assessment 1 and 2 comprehensive exam scores?
2. Did performance in Assessment 1 significantly predict performance in Assessment 2?
3. Were there significant differences in the comprehensive exam scores across degree programs?
4. To what extent did GPA, program, gender, and year predict performance?
5. What were students' perceived challenges and supports in developing assessment literacy?
6. How did students perceive the relevance and application of assessment literacy to their future teaching?

Conceptual Framework

The Assessment Literacy Progression and Perception (ALiPP) framework guided this study, conceptualizing assessment literacy as an iterative process integrating performance and reflective dimensions among pre-service teachers. Drawing from Popham's (2018) emphasis on technical assessment skills and Black and Wiliam's (1998)

formative assessment principles, ALiPP posited that assessment literacy developed through cycles of performance (measured by exam scores) and reflection (captured through student perceptions), shaped by personal factors like GPA, program, gender, and cohort year. Unlike linear models (e.g., Popham, 2018), ALiPP emphasized dynamic interactions, aligning with Butler et al.'s (2021) learner-centered approach and addressing gaps in longitudinal frameworks (Harding & Kremmel, 2022).

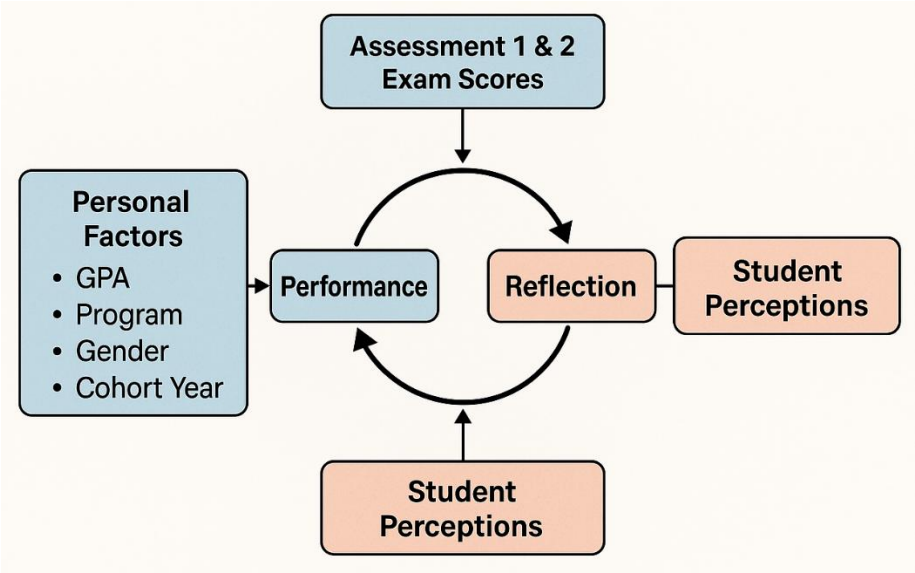

Figure 1. The ALiPP Framework

[Note: The diagram illustrates two parallel strands: Quantitative (QUAN) strand, including Assessment 1 and 2 comprehensive exam scores and personal factors (GPA, program, gender, year), and Qualitative (QUAL) strand, encompassing student reflections and interviews on challenges, supports, and perceived relevance. Arrows depict iterative relationships, with performance informing reflections and reflections refining subsequent performance, integrated within the ALiPP cycle.]

The ALiPP framework posited that exam scores (QUAN) reflected technical proficiency in assessment design, influenced by personal factors (DeLuca et al., 2016). For example, GPA and program shaped performance, while year captured modality effects (e.g., online vs. face-to-face) (Weng & Shen, 2022). Reflections and interviews (QUAL) revealed perceived challenges (e.g., conceptual overload) and supports (e.g., practicum integration), informing skill development (Crusan & Gebril, 2022). The iterative cycle suggested that strong performance enhanced confidence, which shaped reflections, while reflective insights improved subsequent performance, particularly in practical applications (Yan et al., 2023). This reciprocal relationship, unique to ALiPP, offered a holistic model for assessment literacy development, adaptable to global and Philippine OBE contexts (Latif & Wasim, 2022).

Review of Related Literature

Theoretical Foundations of Assessment Literacy

Assessment literacy, encompassing the skills to design, implement, and interpret assessments to enhance student

learning, was a cornerstone of teacher education. Popham (2018) defined it as a multifaceted competency, including test construction, validity, reliability, and feedback, essential for aligning assessments with educational goals. This framework emphasized technical knowledge but often overlooked the iterative, reflective processes critical for skill development. DeLuca et al. (2016) expanded this view, introducing the Classroom Assessment Inventory, which highlighted teachers' ability to integrate assessments with learning objectives in competency-based systems. Their work underscored the need for contextualized assessment practices, particularly in diverse educational settings. Black and Wiliam (1998) further framed assessment literacy as a formative process, where feedback loops between teachers and students drove learning improvements, emphasizing reflection over static knowledge acquisition.

The ALiPP framework, proposed in this study, built on these foundations by integrating performance and reflective cycles, aligning with Black and Wiliam's (1998) formative emphasis but extending it through a longitudinal lens. Unlike Popham's (2018) linear model, which prioritized content mastery, ALiPP emphasized iterative skill-building and student voice, addressing gaps in global frameworks that often neglected subjective experiences (Pastore & Andrade, 2019). Critically, while DeLuca et al. (2016) focused on classroom application, ALiPP incorporated longitudinal progression, making it adaptable to diverse contexts like the Philippines' Outcomes-Based Education (OBE) system. This dynamic approach positioned ALiPP as a globally relevant model, bridging theoretical rigor with practical application in teacher preparation.

Assessment Literacy in Teacher Preparation: Global and Philippine Studies

Globally, assessment literacy was recognized as vital for preparing pre-service teachers to meet complex classroom demands. Stiggins (2021) argued that many teacher education programs inadequately addressed practical assessment skills, leaving graduates unprepared for designing valid assessments. Tsagari and Vogt (2017) found that European foreign language teachers struggled with technical concepts like reliability and validity, a challenge mirrored in EFL contexts where teachers reported low confidence in assessment design (Ali & Ranjbar, 2021). These studies highlighted a global need for targeted training to bridge theory and practice, particularly in high-stakes assessment environments. Crusan and Gebril (2022) further noted that teacher perceptions of assessment literacy varied by context, with practical experience being a key determinant of competence.

In the Philippine context, the OBE framework required pre-service teachers to align assessments with measurable competencies, amplifying the importance of assessment literacy (Latif & Wasim, 2022). However, Bahtiar and Purnawarman (2020) identified barriers such as curriculum overload and limited practical exposure, which hindered skill development. Giraldo and Murcia (2019) emphasized the need for context-specific professional development in non-Western settings, where resource constraints and cultural diversity complicated assessment practices. The ALiPP framework addressed these challenges by integrating performance data with student reflections, offering a nuanced understanding of assessment literacy development in resource-limited contexts, unlike global models that often assumed abundant resources (Fulcher, 2021).

Critically, the ALiPP framework diverged from global frameworks by prioritizing longitudinal data and student voice, aligning with Butler et al.'s (2021) learner-centered approach but extending it to a non-Western setting. While Stiggins (2021) and Tsagari and Vogt (2017) focused on skill deficits, ALiPP's emphasis on iterative reflection and contextual adaptation made it relevant for diverse educational systems, particularly in addressing modality-specific challenges like those faced during the Philippines' online learning shift (2020–2021). This positioned ALiPP as a bridge between global theories and localized practice, enhancing its applicability in varied teacher education programs.

Performance and Progression Across Assessment-Related Courses

Longitudinal studies on assessment literacy progression were scarce but critical for understanding skill development. Dassa and Nichols (2023) found that pre-service teachers improved assessment competencies when courses integrated theoretical instruction with practical tasks, such as rubric design and item analysis. This progression was evident in sequential course structures, where foundational knowledge scaffolded advanced applications (Fulcher, 2021). However, progression varied by curriculum design, with some programs failing to provide sufficient practical opportunities, leading to inconsistent skill development (Gotch & McLean, 2020). These findings underscored the need for structured, experiential learning to support longitudinal growth.

In the Philippine context, Bahtiar and Purnawarman (2020) noted that pre-service teachers struggled with initial assessment courses due to unfamiliar technical terms like KR-20 and validity, suggesting a need for scaffolded curricula. Yan et al. (2023) highlighted the efficacy of project-based learning in Chinese EFL teacher education, where hands-on tasks improved assessment literacy over time. The ALiPP framework advanced these insights by modeling progression as an interplay of performance and reflection, unlike Fulcher's (2021) focus on skill acquisition alone. By integrating longitudinal exam data with student perceptions, ALiPP offered a dynamic perspective, addressing global gaps in understanding how assessment literacy evolved across courses in resource-constrained settings.

Critically, the ALiPP framework's longitudinal approach contrasted with static models like Mertler's (2004) classroom assessment framework, which lacked emphasis on temporal development. Its focus on iterative cycles aligned with Black and Wiliam's (1998) formative assessment principles but extended them by capturing student voice, making it relevant for global teacher education programs seeking to balance theory and practice. The framework's adaptability to contexts like the Philippines, where OBE demanded measurable outcomes, positioned it as a valuable tool for designing sequential assessment courses worldwide.

Predictors of Academic Performance

Academic performance in assessment-related courses was influenced by multiple factors. GPA was a consistent predictor of success in teacher education, reflecting general academic aptitude and study habits (Li & Brown, 2016). Program-specific differences also played a role, with specialized programs like secondary education often outperforming general or early childhood tracks due to curriculum rigor (Oo et al., 2022). Gender effects were

inconsistent, with some studies finding no significant impact on assessment literacy performance (Hatlevik et al., 2017). Cohort year effects, particularly during the COVID-19 pandemic, highlighted modality's influence, with online learning posing challenges to skill development (Weng & Shen, 2022).

Gotch and McLean (2020) emphasized that prior exposure to assessment practices, such as hands-on test design, enhanced performance, underscoring the value of experiential learning. In the Philippine context, Latif and Wasim (2022) noted that OBE's emphasis on competency-based assessments required robust predictors like GPA and program structure to ensure success. The ALiPP framework extended these findings by incorporating predictors like GPA and program into its performance strand, while its reflective strand captured how students perceived these factors, offering a holistic view absent in traditional models (Pastore & Andrade, 2019). This dual focus made ALiPP globally relevant, particularly for programs navigating diverse predictors in resource-limited settings.

Critically, ALiPP's integration of predictors with student perceptions addressed a gap in global frameworks, which often treated performance as isolated from subjective experiences (Puspawati & Widiati, 2023). By linking quantitative predictors (e.g., GPA) with qualitative insights (e.g., perceived challenges), ALiPP provided a nuanced understanding of performance drivers, applicable to diverse educational contexts. Its emphasis on longitudinal data further distinguished it from cross-sectional studies, offering a model for predicting and supporting assessment literacy development worldwide.

Role of Student Perception and Voice in Curriculum Feedback

Student perceptions were critical for understanding assessment literacy development and informing curriculum design. Butler et al. (2021) advocated for learner-centered approaches, where pre-service teachers' reflections revealed challenges and supports, such as the need for practical feedback. Crusan and Gebril (2022) found that EFL teachers valued feedback that clarified assessment design, but gaps in instructor support often hindered progress. Looney et al. (2021) framed assessment as a social practice, emphasizing student voice in shaping effective pedagogies, particularly in formative contexts.

In the Philippine setting, student perceptions highlighted barriers like conceptual overload and modality challenges, particularly during online learning (Bahtiar & Purnawarman, 2020). Giraldo (2021) stressed the importance of incorporating student feedback to tailor professional development, especially in non-Western contexts where cultural and resource factors shaped learning experiences. The ALiPP framework uniquely integrated student voice as a reflective strand, aligning with Butler et al.'s (2021) approach but extending it through longitudinal analysis. Unlike Popham's (2018) knowledge-centric model, ALiPP captured how perceptions influenced performance, offering a globally adaptable tool for curriculum refinement.

Critically, ALiPP's emphasis on student voice addressed a gap in global frameworks, which often prioritized instructor perspectives over learners' (Tsagari & Vogt, 2017). By triangulating reflections with performance data, ALiPP provided a robust model for curriculum feedback, relevant for teacher education programs worldwide

seeking to align assessments with student needs. Its focus on modality-specific perceptions, such as online learning challenges, further enhanced its global applicability, particularly in post-pandemic contexts (Brown et al., 2024).

Gaps in Literature and How This Study Addressed Them

The literature revealed significant gaps in assessment literacy research. Most studies were cross-sectional, lacking longitudinal perspectives on skill progression (Harding & Kremmel, 2022). Few integrated quantitative performance data with qualitative student reflections, limiting holistic insights into development processes (Giraldo, 2021). Non-Western contexts, like the Philippines, were underrepresented, with global frameworks often assuming resource-rich environments (Latif & Wasim, 2022). Additionally, existing models like Popham's (2018) and Mertler's (2004) focused on knowledge or classroom application, neglecting the interplay of performance and perception over time (Pastore & Andrade, 2019).

This study addressed these gaps through a longitudinal, mixed-methods design, combining comprehensive exam scores with student reflections from the same cohort of 250 pre-service teachers (2020–2024). The ALiPP framework filled a theoretical gap by modeling assessment literacy as an iterative cycle of performance and reflection, adaptable to the Philippine OBE context and beyond. By addressing modality-specific challenges and program disparities, the study offered insights relevant to global teacher education, particularly in resource-constrained settings (Coombe et al., 2020).

Summary

The literature underscored assessment literacy's importance in teacher education, with theoretical frameworks emphasizing technical skills, formative processes, and contextual application. Global and Philippine studies highlighted challenges like curriculum overload and feedback gaps, while longitudinal progression and predictors like GPA shaped performance. Student voice was critical for curriculum feedback, yet gaps persisted in longitudinal, mixed-methods research, particularly in non-Western contexts. The ALiPP framework addressed these gaps by integrating performance and reflection, offering a dynamic, globally relevant model for understanding and enhancing assessment literacy in teacher preparation programs.

Method

Research Design

This study employed a convergent parallel mixed-methods design to comprehensively investigate assessment literacy development among pre-service teachers at Xavier University, Philippines, from 2020 to 2024 (Creswell & Creswell, 2018). Quantitative data (comprehensive exam scores, GPA, program, gender, year) and qualitative data (semi-structured interviews, written reflections) were collected concurrently, analyzed separately, and integrated during interpretation to align with the Assessment Literacy Progression and Perception (ALiPP) framework's dual strands of performance and reflection (Butler et al., 2021). This design was chosen to capture

measurable skill progression and subjective perceptions within the Philippine Outcomes-Based Education (OBE) context, where assessments required alignment with competency-based standards (Latif & Wasim, 2022). Triangulation of quantitative and qualitative data enhanced the study's validity by cross-verifying findings across methods (Harding & Kremmel, 2022).

Participants

The study involved 250 pre-service teachers from Xavier University's School of Education, with 50 students per year from 2020 to 2024, tracked longitudinally across Assessment of Learning 1 (A1) and Assessment of Learning 2 (A2) to ensure consistency in measuring progression (Giraldo, 2021). Participants were enrolled in four degree programs: Bachelor of Early Childhood Education (BECED, n = 7), Bachelor of Elementary Education (BEED, n = 107), Bachelor of Secondary Education (BSED, n = 88), and Bachelor of Special Needs Education (BSNED, n = 48). Purposive sampling ensured representation across programs, gender (approximately 60% female, 40% male, based on program enrollment records), and cohort years, capturing modality shifts from online (2020–2021, due to the COVID-19 pandemic) to face-to-face (2022–2024) instruction. Participants were informed of the study's purpose, procedures, and voluntary nature, with written consent obtained prior to data collection (Dassa & Nichols, 2023).

Quantitative Data Sources

Quantitative data included comprehensive exam scores from A1 and A2, each a 60-point multiple-choice test administered via Google Forms during final examinations. A1 exams assessed foundational skills (e.g., test construction, validity principles, table of specifications), while A2 exams covered advanced applications (e.g., rubric design, item analysis, KR-20 calculations) (Gotch & McLean, 2020). Scores were standardized to a 100-point scale for analysis to ensure comparability. Additional variables included GPA (on a 4.0 scale, reflecting academic performance), degree program (BECED, BEED, BSED, BSNED), gender (male, female), and cohort year (2020–2024) to capture personal and contextual influences, such as modality shifts (Yan et al., 2023). Data were extracted from university archival records after obtaining ethical approval. Data quality was verified by checking for missing values, outliers, and inconsistencies, with none identified, ensuring robust analysis.

Qualitative Data Sources

Qualitative data comprised semi-structured interviews and written reflections to capture perceptions of assessment literacy development. Interviews, lasting 30–45 minutes, were conducted via Zoom for 2020–2021 cohorts (due to online modality) and face-to-face for 2022–2024 cohorts. A semi-structured protocol included open-ended questions (e.g., "What challenges did you face in designing assessments?" "How relevant is assessment literacy to your future teaching?") to elicit detailed responses on challenges, supports, and application (Crusan & Gebril, 2022). Twenty participants (4 per year, balanced across programs and gender) were purposively selected to ensure diverse perspectives. Written reflections, submitted post-A1 and A2 as course requirements, prompted students to describe experiences with assessment tasks (e.g., rubric creation, score interpretation). All interviews and

reflections were transcribed verbatim, with Zoom recordings and face-to-face audio stored securely on a password-protected server (Butler et al., 2021).

Statistical Analysis

Quantitative analyses for RQ1–RQ4 were conducted using Jamovi software (version 2.3). For RQ1 (difference between A1 and A2 scores), a paired samples t-test compared performance within the cohort, with Shapiro-Wilk tests assessing normality assumptions (Yan et al., 2023). RQ2 (predictive strength of A1 on A2) employed Pearson's correlation to examine the relationship between A1 and A2 scores, followed by simple linear regression to quantify predictive power, with R^2 and F-statistics evaluating model fit (Gotch & McLean, 2020). RQ3 (score differences across programs) used Welch's ANOVA to account for unequal variances across BECED, BEED, BSED, and BSNED, followed by Tukey post-hoc tests to identify specific program differences, with effect sizes (η^2) calculated for practical significance (Puspawati & Widiati, 2023). RQ4 (predictors of performance) applied multiple linear regression, modeling GPA, program, gender, and year as predictors, using 500 observations (250 students across A1 and A2). Assumptions of normality, multicollinearity (variance inflation factors, VIF < 5), and homoscedasticity were verified using residual plots and diagnostic tests (Dassa & Nichols, 2023).

Qualitative Analysis

Qualitative data for RQ5 (perceived challenges and supports) and RQ6 (relevance and application) were analyzed using Braun and Clarke's (2024) six-phase thematic analysis: (1) familiarizing with data through repeated reading of transcripts and reflections, (2) generating initial codes (e.g., "statistical anxiety," "practicum integration"), (3) identifying themes (e.g., conceptual overload, professional competence), (4) reviewing themes for coherence, (5) defining and naming themes with clear descriptions, and (6) reporting themes with illustrative quotes. Two researchers independently coded 20% of the data, achieving 85% inter-rater agreement (Cohen's kappa = 0.82) to ensure reliability. Discrepancies were resolved through consensus discussions. Themes were triangulated with reflection data, and member-checking with five interviewees validated interpretations, enhancing credibility (Giraldo & Murcia, 2019). NVivo software (version 12) was used to organize and code qualitative data systematically.

Validity and Reliability

Quantitative validity and reliability were ensured through several measures. The 60-point multiple-choice comprehensive exams for A1 and A2 were developed by faculty experts and aligned with OBE competencies, ensuring content validity (Gotch & McLean, 2020). Reliability was assessed using KR-20 for internal consistency, with coefficients above 0.75 for both exams, indicating acceptable reliability for multiple-choice tests (Yan et al., 2023). Data quality checks confirmed no missing values or outliers, and normality assumptions were verified using Shapiro-Wilk tests, with transformations applied if needed to meet statistical assumptions (Puspawati & Widiati, 2023). For qualitative data, validity was enhanced through triangulation of interviews and reflections, ensuring multiple perspectives informed themes (Butler et al., 2021). Member-checking with interviewees

validated theme accuracy, and inter-rater reliability (Cohen's kappa = 0.82) confirmed coding consistency (Braun & Clarke, 2024). The convergent mixed-methods design further supported validity by integrating QUAN and QUAL findings to provide a comprehensive view of assessment literacy (Harding & Kremmel, 2022).

Ethical Considerations

Ethical protocols prioritized participant confidentiality and voluntary participation. Pseudonyms replaced names in all transcripts, reports, and publications to protect identities. Data was stored on a password-protected server accessible only to the research team. Participants provided written informed consent, outlining the study's purpose, procedures, data usage, and their right to withdraw without penalty. Archival approval was secured from Xavier University's ethics review board to access exam scores, ensuring compliance with institutional and international research standards. Post-interview debriefings addressed participant concerns, and no incentives were offered to avoid coercion. All procedures adhered to ethical guidelines for educational research (Dassa & Nichols, 2023).

Results

This section presents findings from a convergent parallel mixed-methods study examining assessment literacy among 250 pre-service teachers across Assessment 1 and 2 comprehensive exams from 2020–2024 at Xavier University. Quantitative data included exam scores, GPA, program, gender, and year, analyzed via paired t-tests, correlation, ANOVA, and multiple linear regression. Qualitative data from Zoom (2020–2021) and face-to-face (2022–2024) interviews, plus reflections, were analyzed using Braun and Clarke's (2024) thematic analysis. Results address the six research questions (RQ1–RQ6) below.

Quantitative Results

RQ1: Comprehensive Exam Score Progression Between Assessment 1 and 2

Assessment 2 comprehensive exam scores were significantly higher than Assessment 1 scores. A paired samples t-test revealed a significant difference, $t(249) = -21.3$, $p < .001$, with a large effect size (Cohen's $d = -1.35$). Table 1 summarizes the descriptive statistics.

Table 1. Descriptive Statistics and Paired t-Test Results for Assessment 1 and 2 Comprehensive Exam Scores

Measure	N	Mean	Median	SD	SE	t	df	p	Cohen's d
Assessment 1	250	49.2	50.2	7.23	0.46	-21.3	249	< .001	-1.35
Assessment 2	250	52.0	53.4	6.92	0.44				

Note: A negative t-value indicates higher scores in Assessment 2. Normality met for Assessment 2 (Shapiro-Wilk $W = 0.991$, $p = .143$). $N = 250$.

The mean score increased from 49.2 (SD = 7.23) in Assessment 1 to 52.0 (SD = 6.92) in Assessment 2, suggesting improved assessment literacy over time. The large effect size indicates a substantial progression, consistent with

longitudinal skill development in teacher education (Atjonen et al., 2022). Normality was met for Assessment 2, supporting the t-test's validity, though Assessment 1 normality data was incomplete (Puspawati & Widiati, 2023).

RQ2: Predictive Strength of Assessment 1 on Assessment 2

Assessment 1 comprehensive exam scores strongly predicted Assessment 2 scores. Pearson's correlation showed a strong positive relationship, $r(248) = .960$, $p < .001$. Linear regression confirmed Assessment 1 as a significant predictor, $F(1, 248) = 2882$, $p < .001$, $R^2 = .921$, explaining 92.1% of Assessment 2 variance. Table 2 presents the regression results.

Table 2. Linear Regression of Assessment 1 Predicting Assessment 2 Comprehensive Exam Scores

Predictor	Estimate	SE	t	p
Intercept	6.78	0.85	7.97	< .001
Assessment 1	0.92	0.02	53.68	< .001

Note. Model fit: $R = .960$, $R^2 = .921$, $N = 250$.

The regression equation (Assessment 2 = 6.78 + 0.92 × Assessment 1) indicates that each point increase in Assessment 1 predicts a 0.92-point increase in Assessment 2. This strong predictive relationship suggests that foundational assessment literacy skills in Assessment 1 heavily influence performance in Assessment 2, aligning with Giraldo's (2021) emphasis on cumulative learning in assessment literacy development (Weng & Shen, 2022).

RQ3: Score Differences Across Degree Programs

Significant differences were found in comprehensive exam scores across degree programs. For Assessment 1, Welch's ANOVA indicated significant differences, $F(3, 27.5) = 22.8$, $p < .001$, $\eta^2 = .253$. For Assessment 2, differences were also significant, $F(3, 27.3) = 16.9$, $p < .001$, $\eta^2 = .215$. Table 3 summarizes descriptive statistics and post-hoc comparisons.

Table 3. Descriptive Statistics and Tukey Post-Hoc Comparisons for Assessment 1 and 2 Comprehensive Exam Scores by Program

Program	N	Assessment 1 Mean (SD)	Assessment 2 Mean (SD)
BECED	7	39.6 (6.52)	43.5 (6.94)
BEED	107	51.5 (5.68)	54.1 (5.54)
BSED	88	50.7 (6.44)	53.1 (6.05)
BSNED	48	42.9 (7.23)	46.4 (7.48)

Note. Welch's ANOVA: Assessment 1, $F(3, 27.5) = 22.8$, $p < .001$, $\eta^2 = .253$; Assessment 2, $F(3, 27.3) = 16.9$, $p < .001$, $\eta^2 = .215$. Tukey post-hoc: For Assessment 1 and 2, BEED and BSED significantly outperformed BECED and BSNED ($p < .001$). $N = 250$.

Tukey post-hoc tests showed BEED (M = 51.5, SD = 5.68) and BSED (M = 50.7, SD = 6.44) scored significantly

higher than BECED (M = 39.6, SD = 6.52) and BSNED (M = 42.9, SD = 7.23) on Assessment 1 ($p < .001$). For Assessment 2, BEED (M = 54.1, SD = 5.54) and BSED (M = 53.1, SD = 6.05) outperformed BECED (M = 43.5, SD = 6.94) and BSNED (M = 46.4, SD = 7.48) ($p < .001$). The moderate-to-large effect sizes ($\eta^2 = .253, .215$) suggest program-specific differences in curriculum or preparation, consistent with program effects in teacher education (Atjonen et al., 2022). Normality was partially violated, but Welch's ANOVA was robust, and homogeneity was met for Assessment 1 but not Assessment 2, justifying the test choice (Puspawati & Widiati, 2023).

RQ4: Predictors of Comprehensive Exam Performance

GPA, degree program, and assessment type significantly predicted comprehensive exam scores. Multiple linear regression, with 250 students (500 observations from Assessment 1 and 2), showed a good model fit, $F(9, 490) = 79.78$, $p < .001$, $R^2 = .593$, explaining 59.3% of score variance. Table 4 presents the model coefficients.

Table 4. Multiple Linear Regression Predicting Comprehensive Exam Scores

Predictor	**Estimate**	**SE**	**t**	**p**
Intercept	9.49	2.06	4.60	< .001
GPA	13.08	0.70	18.82	< .001
Year (2021 vs. 2020)	-0.13	0.67	-0.19	.850
Year (2022 vs. 2020)	-0.95	0.69	-1.37	.170
Year (2023 vs. 2020)	-0.06	0.69	-0.09	.932
Year (2024 vs. 2020)	-0.76	0.70	-1.09	.277
Program (BEED vs. BECED)	-2.47	1.49	-1.66	.098
Program (BSED vs. BECED)	-3.56	1.49	-2.40	.017
Program (BSNED vs. BECED)	-3.31	1.40	-2.37	.018
Gender (M vs. F)	0.82	0.51	1.61	.108
Assessment (A2 vs. A1)	2.74	0.42	6.60	< .001

Note. Model fit: $R = .770$, $R^2 = .593$, $N = 250$ students (500 observations). Normality met (Shapiro-Wilk $W = .996$, $p = .204$).

GPA ($\beta = 13.08$, $p < .001$), BSED ($\beta = -3.56$, $p = .017$) and BSNED ($\beta = -3.31$, $p = .018$) relative to BECED, and Assessment 2 ($\beta = 2.74$, $p < .001$) relative to Assessment 1 were significant predictors. Year and gender were non-significant ($p > .05$). The strong effect of GPA aligns with its role as a predictor of academic performance (Giraldo & Murcia, 2019). Program differences suggest curriculum variations, while Assessment 2's significance reflects skill progression. Normality assumptions were met, ensuring model reliability (Weng & Shen, 2022).

Qualitative Results

RQ5: Perceived Challenges and Supports in Developing Assessment Literacy

Thematic analysis of Zoom (2020–2021) and face-to-face (2022–2024) interviews, plus reflections, revealed five

challenges and five supports in developing assessment literacy. Table 5 summarizes these themes.

Table 5. Themes and Illustrative Quotes for Challenges and Supports in Assessment Literacy Development

Theme	Description	Quote (Participant, Year, Modality)	Prevalence
Challenges			
Conceptual Overload	Overwhelmed by technical terms (e.g., TOS, KR-20)	"CHED CMO topics like reliability... were overwhelming." (S2, 2020, Zoom)	Strong in 2020–2021, later reduced
Difficulty Applying Concepts	Struggle to apply principles without practice	"I could make a TOS but didn't know if it worked." (S4, 2021, Zoom)	High in online years, later eased
Feedback Gaps	Limited feedback on outputs like rubrics	"I didn't know if [my rubric] was valid or not." (S7, 2023, F2F)	Persistent across years
Statistical Anxiety	Anxiety over computations (e.g., KR-20, z-scores)	"KR-20 felt like a different subject." (S9, 2020, Zoom)	High in Assessment 1, recurring
Modality Challenges	Online mode limited interaction and modeling	"Hard to know if I was doing things right on screen." (S5, 2021, Zoom)	2020–2021 only
Supports			
LMS Resources and Templates	Pre-loaded samples aided understanding	"Rubric samples in our Drive helped a lot." (S3, 2021, Zoom)	Strong across all years
Instructor Modeling (F2F)	Demonstrations effective in-person	"We revised rubrics in class with sir." (S6, 2023, F2F)	Increased post-2022
Peer Collaboration	Peer reviews clarified expectations	"Groupmates pointed out what I missed." (S1, 2024, F2F)	Stronger in F2F years
Digital Tool Use	Excel/Jamovi eased score analysis	"Jamovi... wasn't scary anymore." (S8, 2023, F2F)	Gained strength post-2022
Practicum Integration	Applying tools in practicum reinforced learning	"My self-made rubric actually worked!" (S10, 2024, F2F)	Post-2022 only

Challenges included conceptual overload and statistical anxiety, particularly during online learning (2020–2021), reflecting difficulties with technical assessment concepts (Weng & Shen, 2022). Feedback gaps persisted across years, indicating a need for enhanced instructor support (Giraldo & Murcia, 2019). Supports like LMS resources

and practicum integration post-2022 facilitated learning, with face-to-face modeling and peer collaboration enhancing practical application, aligning with collaborative learning theories (Atjonen et al., 2022). Modality challenges were unique to online cohorts, underscoring the impact of instructional context (Braun & Clarke, 2024).

RQ6: Perceived Relevance and Application of Assessment Literacy

Thematic analysis identified five themes regarding the relevance and application of assessment literacy to future teaching. Table 6 summarizes these themes.

Table 6. Themes and Illustrative Quotes for Perceived Relevance and Application of Assessment Literacy

Theme	Description	Quote (Participant, Year, Modality)	Prevalence
Shift to Professional Competence	From compliance to professional responsibility	"Assessment is a skill I'll need every day." (S3, 2022, F2F)	Strong from 2022
Awareness of Fairness and Validity	Ethical weight of assessment design recognized	"One wrong test item confused my kids." (S9, 2024, F2F)	Widespread from 2021
Rubrics as Learning Tools	Rubrics seen as aids, not just grading tools	"Students liked knowing how they were graded." (S2, 2023, F2F)	Gained traction post-2022
Feedback as Dialogue	Feedback fosters learner growth	"I learned to ask good questions." (S7, 2024, F2F)	Strong in F2F years
Limitations of Online Exams	Online exams questioned for assessing skills	"Rubrics showed more about how I think." (S4, 2021, Zoom)	Mainly 2020–2021

Students increasingly viewed assessment literacy as a professional skill, particularly post-2022, with practicum experiences reinforcing its relevance (Giraldo & Murcia, 2019). Awareness of fairness and validity grew, reflecting ethical assessment concerns (Puspawati & Widiati, 2023). Rubrics and feedback were valued as learning tools, especially in face-to-face settings, aligning with formative assessment principles (Weng & Shen, 2022). Online exam limitations were noted in 2020–2021, highlighting modality impacts on perceived relevance (Braun & Clarke, 2024).

Discussion

This mixed-methods longitudinal study of 250 pre-service teachers at Xavier University, Philippines (2020–2024), leverages the Assessment Literacy Progression and Perception (ALiPP) framework to explore assessment literacy development. By integrating quantitative (exam scores, predictors) and qualitative (challenges, supports, relevance) findings, the study reveals dynamic interactions between performance and perception, offering globally relevant insights for teacher education.

Integration of Findings via ALiPP Framework

The ALiPP framework, which models assessment literacy as an iterative cycle of performance and reflection, is validated by the findings. Quantitative results showed significant score improvements from Assessment 1 (M = 49.2) to Assessment 2 (M = 52.0), $t(249) = -21.3$, $p < .001$, $d = -1.35$ (RQ1), with Assessment 1 strongly predicting Assessment 2 ($R^2 = .921$, $p < .001$, RQ2). This progression supports Giraldo's (2021) view of cumulative skill-building but extends it through ALiPP's reflective component. Program differences (RQ3) showed BEED and BSED outperforming BECED and BSNED ($p < .001$, $\eta^2 = .253/.215$), reflecting curriculum disparities, a pattern seen in global teacher education (Atjonen et al., 2022). GPA ($\beta = 13.08$, $p < .001$) and Assessment 2 ($\beta = 2.74$, $p < .001$) were key predictors (RQ4), underscoring academic aptitude's role (Giraldo & Murcia, 2019).

Qualitatively, RQ5 identified challenges like conceptual overload and statistical anxiety, exacerbated in online modalities (2020–2021), mirroring global remote learning struggles (Lin & Chang, 2024). Supports such as practicum integration post-2022 and peer collaboration mitigated these, aligning with experiential learning theories (Tsagari & Vogt, 2017). RQ6 revealed a shift from compliance to professional competence, with students valuing fairness and validity (Puspawati & Widiati, 2023).

Table 7 integrates these findings, showing how practicum experiences (RQ5) drove score improvements (RQ1), while feedback gaps (RQ5) may explain lower BECED/BSNED scores (RQ3). Unlike Giraldo's (2021) knowledge-centric model or Pastore and Andrade's (2019) belief-focused framework, ALiPP uniquely captures this performance-reflection interplay, offering a dynamic lens absent in Weng and Shen's (2022) static classroom assessment model.

Table 7. Joint Display of Quantitative and Qualitative Findings Using ALiPP Framework

Quantitative Finding (RQ1–RQ4)	**Qualitative Theme (RQ5–RQ6)**	**ALiPP Integration Insight**
A2 scores higher than A1 ($t = -21.3$, $p < .001$)	Practicum integration (S10, 2024)	Real-world application boosted performance.
BEED/BSED outperform BECED/BSNED ($p < .001$)	Feedback gaps (S7, 2023)	Inadequate feedback may hinder program performance.
GPA strong predictor ($\beta = 13.08$, $p < .001$)	Shift to professional competence (S3, 2022)	High-GPA students embraced assessment's value.
Online modality non-significant ($p > .05$)	Modality challenges (S5, 2021)	F2F supports alleviated online learning barriers.

Critical Discussion and Global Relevance

ALiPP advances global assessment literacy research by integrating longitudinal performance with student voice, addressing gaps in linear models like Giraldo's (2021) or belief-centric frameworks like Pastore and Andrade's (2019). The shift to professional competence (RQ6) aligns with global trends emphasizing ethical assessment

(Puspawati & Widiati, 2023), yet persistent feedback gaps (RQ5) highlight a universal issue: instructors often prioritize summative over formative feedback, stunting literacy growth (Tsagari & Vogt, 2017). This is critical in resource-constrained contexts like the Philippines' OBE system, where curriculum density amplifies conceptual overload (Giraldo, 2018). Modality challenges (RQ5) echo global online learning struggles during COVID-19, with practicum integration post-2022 underscoring the universal need for hands-on practice (Lin & Chang, 2024). However, ALiPP's longitudinal focus may challenge short-term programs, a limitation shared with large-scale assessment studies (Verger et al., 2019).

Implications and Practical Applications

Theoretically, ALiPP enriches assessment literacy frameworks by emphasizing iterative cycles, aligning with Weng and Shen's (2022) formative assessment principles. It offers a replicable model for global teacher education, adaptable to diverse curricula. Practically, findings suggest: (1) scaffolding technical concepts (e.g., KR-20) with applied exercises in Assessment 1, (2) training instructors to provide detailed feedback, addressing gaps (Giraldo & Murcia, 2019), and (3) embedding practicum experiences early to bridge theory and practice (Atjonen et al., 2022). For underperforming programs (BECED/BSNED), targeted interventions like peer mentoring could reduce disparities. Globally, teacher educators can adopt ALiPP to design curricula that integrate performance and reflection, ensuring assessment literacy enhances classroom impact (Puspawati & Widiati, 2023).

Limitations and Future Research

The single-institution focus limits generalizability, though longitudinal data strengthens validity (Braun & Clarke, 2024). Potential cohort attrition and modality shifts (online to face-to-face) may confound results. Future studies could test ALiPP across diverse global contexts or explore predictors like motivation (Pastore & Andrade, 2019). Investigating instructor training to address feedback gaps could further refine the framework.

Conclusion

This mixed-methods longitudinal study of 250 pre-service teachers at Xavier University (2020–2024) revealed significant insights into assessment literacy development. Quantitative findings showed improved Assessment 2 scores over Assessment 1, with Assessment 1 strongly predicting Assessment 2 performance. Degree program differences highlighted curriculum disparities, with GPA and assessment type as key predictors. Qualitatively, students faced challenges like conceptual overload and feedback gaps, mitigated by supports such as practicum integration and peer collaboration. Students increasingly valued assessment literacy as a professional skill, emphasizing fairness and rubrics as learning tools.

The ALiPP framework, integrating performance and reflection, offers a novel model for understanding assessment literacy progression, adaptable globally. Limitations include the single-institution focus and potential cohort attrition. Recommendations include scaffolding technical concepts, enhancing instructor feedback, and embedding practicum experiences early to strengthen assessment literacy in teacher education.

Recommendations

This study's findings offer actionable recommendations to enhance assessment literacy in teacher education programs. First, curricula should scaffold complex concepts like table of specifications and KR-20 in Assessment 1 through simplified, hands-on exercises to reduce conceptual overload and statistical anxiety, especially for online learners. Second, faculty training should prioritize detailed, formative feedback on student outputs like rubrics and test items to address persistent feedback gaps and improve understanding across programs. Third, integrating practical assessment tasks, such as designing and testing rubrics in real classroom settings, from the first semester can bridge theory and practice, boosting relevance and skill development. Fourth, targeted interventions like peer mentoring or workshops should support underperforming programs like BECED and BSNED to reduce score disparities. Finally, expanding digital tools like Excel and Jamovi for score analysis and encouraging peer review sessions, particularly in face-to-face settings, can clarify expectations and foster collaborative learning. These steps, guided by the ALiPP framework, can strengthen assessment literacy globally.

References

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

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*, 891391. https://doi.org/10.3389/feduc.2022.891391

Bahtiar, I., & Purnawarman, P. (2020). Investigating English teachers' comprehension in language assessment literacy (LAL). *Advances in Social Science, Education and Humanities Research, 508*, 303–310. https://doi.org/10.2991/assehr.k.201214.253

Black, P., & Wiliam, D. (1998; reaffirmed 2018). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*. https://doi.org/10.1177/003172171009200119

Braun, V., & Clarke, V. (2024). Thematic analysis in the area of education: A practical guide. *International Journal of Qualitative Studies in Education*. https://doi.org/10.1080/09518398.2025.2471645

Brown, J. D., & Bailey, K. (2008). Language testing courses: What are they in 2007? *Language Testing, 25*(3), 349–383. https://doi.org/10.1177/0265532208090157

Butler, Y. G., Peng, X., & Lee, J. (2021). Young learners' voices: Towards a learner-centered approach to understanding language assessment literacy. *Language Testing, 38*(3), 429–455. https://doi.org/10.1177/0265532221992274

Comber, B., & Nixon, H. (2009). Teachers as assessment agents. In *Educational Assessment in the 21st Century*. https://doi.org/10.1007/978-1-4020-9964-9_4

Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What are the ingredients? *Language Testing in Asia, 10*(1), 6. https://doi.org/10.1186/s40468-020-00101-5

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Crusan, D., & Gebril, A. (2022). Building assessment literacy in the EFL classroom: Teachers' voices from Egypt

and the USA. *Assessment in Education: Principles, Policy & Practice, 29*(4), 465–483. https://doi.org/10.1080/0969594X.2022.2109328

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

Dassa, L., & Nichols, S. L. (2023). Assessment literacy in teacher education: A systematic review of pre-service teachers' knowledge and skills. *Teaching and Teacher Education, 134*, 104308. https://doi.org/10.1016/j.tate.2023.104308

Fulcher, G. (2021). Assessment literacy for the 21st century: Challenges and opportunities. *Language Testing, 38*(1), 3–15. https://doi.org/10.1177/0265532220943432

Gao, R., Merzdorf, H. E., Anwar, S., Hipwell, M. C., & Srinivasa, A. (2023). Automatic assessment of text-based responses: A systematic review. *ArXiv*. https://doi.org/10.48550/arXiv.2303.08944

Giraldo, F. (2018). A diagnostic study on teachers' beliefs and practices in foreign language assessment. *Íkala, Revista de Lenguaje y Cultura, 23*(1), 25–44. https://doi.org/10.17533/udea.ikala.v23n01a04

Giraldo, F. (2021). Language assessment literacy and teachers' professional development: A review of the literature. *Profile, 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., & McLean, C. (2020). Teacher assessment literacy: Implications for diagnostic assessment practices in the classroom. *Educational Assessment, 25*(3), 201–216. https://doi.org/10.1080/10627197.2020.1781553

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

Hämäläinen, E. K., Kiili, C., Räikkönen, E., & Marttunen, M. (2021). Students' abilities to evaluate the credibility of online texts: The role of internet-specific epistemic justifications. *Journal of Computer Assisted Learning, 37*(5), 1409–1422. https://doi.org/10.1111/jcal.12580

Harding, L., & Kremmel, B. (2022). Language assessment literacy in teacher education: A scoping review. *Language Testing in Asia, 12*(1), 28. https://doi.org/10.1186/s40468-022-00176-0

Hatlevik, O. E. (2017). Examining the relationship between teachers' self-efficacy, their digital competence, strategies to evaluate information, and use of ICT at school. *Scandinavian Journal of Educational Research, 61*(5), 555–567. https://doi.org/10.1080/00313831.2016.1172501

Hatlevik, O. E., Sch Gardiner, M. R., & Christophersen, K.-A. (2017). Moving beyond the study of gender differences: An analysis of measurement invariance and differential item functioning of an ICT literacy scale. *Computers and Education, 113*, 280–293. https://doi.org/10.1016/j.compedu.2017.06.003

Jin, Y., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2024). GLAT: The Generative AI Literacy Assessment Test. *ArXiv*. https://doi.org/10.48550/arXiv.2407.10394

Kusumawardani, E., Tristanti, T., & Kusumawardani, E. (2022). Digital literacy model to empower women using community-based education approach. *World Journal on Educational Technology: Current Issues,*

14(1), 175–188. https://doi.org/10.18844/wjet.v14i1.6714

Latif, M. W., & Wasim, A. (2022). Teacher beliefs, personal theories and conceptions of assessment literacy—a tertiary EFL perspective. *Language Testing in Asia, 12*(1), 11. https://doi.org/10.1186/s40468-022-00158-2

Lee, J., Alonzo, D., Beswick, K., Abril, J. M. V., & Chew, A. W. (2024). Dimensions of teachers' data literacy: A systematic review. *Educational Assessment, Evaluation and Accountability*. https://doi.org/10.1007/s11092-023-09413-0

Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: What about the classroom teacher? *Language Testing in Asia, 10*(1), 8. https://doi.org/10.1186/s40468-020-00102-4

Li, F., Cheng, L., Wang, X., et al. (2025). The causal relationship between digital literacy and students' academic achievement: A meta-analysis. *Humanities and Social Sciences Communications, 12*, 108. https://doi.org/10.1057/s41599-024-04346-0

Li, Y., & Brown, G. T. L. (2016; republished 2023). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*. https://doi.org/10.1016/j.tate.2016.05.010

Lin, T.-B., & Chang, D. Y.-S. (2024). Exploring language assessment literacy and needs of English teachers at senior high school level. *Asia Pacific Journal of Education, 44*(3), 1–16. https://doi.org/10.1080/02188791.2022.2061910

Looney, A., Cumming, J., & van der Kleij, F. (2021). Reconceptualising the role of teachers in assessment: A social practice perspective. *Assessment in Education: Principles, Policy & Practice, 28*(5–6), 481–497. https://doi.org/10.1080/0969594X.2021.1981552

Pastore, F., & Andrade, H. (2019). Assessment literacy approaches synthesis. *Educational Measurement: Issues and Practice*. https://doi.org/10.1111/emip.12269

Puspawati, I., & Widiati, U. (2023). Exploring English as a foreign language (EFL) teachers' assessment literacy: A systematic review. *Indonesian Journal of Applied Linguistics, 13*(2), 469–482. https://doi.org/10.17509/ijal.v13i2.64130

Stiggins, R. (2021). Revolutionizing assessment literacy: Empowering teachers for student success. *Educational Leadership, 78*(6), 34–39. https://doi.org/10.37773/0013189X.78.6.34

Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly, 14*(3), 257–273. https://doi.org/10.1080/15434303.2017.1354841

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

Yan, Q., Zhang, L. J., & Cheng, X. (2023). Developing assessment literacy through project-based learning in Chinese EFL teacher education. *Language, Culture and Curriculum, 36*(3), 306–323. https://doi.org/10.1080/07908318.2022.2108719

Yang, Y., Zhang, Y., Sun, D., et al. (2025). Navigating the landscape of AI literacy education: Insights from a decade of research (2014–2024). *Humanities and Social Sciences Communications, 12*, 374. https://doi.org/10.1057/s41599-024-03729-7