

THE ASSESSMENT–FEEDBACK LITERACY NEXUS: A LONGITUDINAL MIXED-METHODS STUDY OF RESPONSIBLE ASSESSMENT PRACTICE IN PRESERVICE TEACHER EDUCATION

Ronald Quileste

Xavier University-Ateneo de Cagayan
Philippines

✉ rquileste@xu.edu.ph

ABSTRACT

Assessment literacy and feedback literacy are often treated as separate competencies in teacher education, although both underpin responsible assessment practice. This two-wave, multi-cohort mixed-methods secondary analysis examined their parallel development among 250 preservice teachers at a private university in the southern Philippines. Five successive cohorts from 2020–2024 were matched from Assessment of Learning 1 (A1) to Assessment of Learning 2 (A2), typically completed in consecutive semesters. Quantitative analyses included paired course-phase comparisons, within-domain regression-based prediction, and programme comparisons. Qualitative interviews and written reflections from the ALiPP corpus were reanalysed thematically and integrated through joint displays. Standardized A2 assessment-literacy performance and feedback-literacy scores exceeded A1 scores, while A1 performance strongly predicted A2 outcomes within each domain. Programme-level patterns indicated needs for targeted support in smaller and specialized programmes. Qualitative findings showed shifts from procedural, compliance-oriented experiences toward rubric use, feedback dialogue, fairness, and professional responsibility. Because A1 and A2 assessment-literacy examinations covered sequential but partly different content and were not psychometrically equated, differences indicate standardized course-based performance rather than growth on an invariant scale. The Assessment-Feedback Literacy Nexus integrates structural and dialogic dimensions of responsible assessment practice as an explanatory framework.

KEYWORDS

Assessment literacy, Feedback literacy, Preservice teachers, Professional responsibility, Teacher education

HOW TO CITE

Quileste R. (2026) 'The Assessment–Feedback Literacy Nexus: a Longitudinal Mixed-Methods Study of Responsible Assessment Practice in Preservice Teacher Education', *Journal on Efficiency and Responsibility in Education and Science*, vol. 19, no. 3, pp. 210–221. <http://dx.doi.org/10.7160/eriesj.2026.190303>

Article history

Received

October 24, 2025

Received in revised form

July 15, 2026

Accepted

August 27, 2026

Available on-line

September 30, 2026

Highlights

- Matched two-wave data distinguish within-student course-phase differences from between-cohort context.
- A2 standardized assessment performance and feedback-literacy scores were higher than A1 values.
- AFLN links assessment structure with feedback dialogue and professional responsibility.
- Programme patterns and qualitative evidence point to targeted literacy support.

INTRODUCTION

Assessment is central to teachers' professional work because it shapes how learning is planned, monitored, interpreted, communicated, and supported. For preservice teachers, assessment literacy is therefore not limited to constructing tests, computing scores, or preparing rubrics. It involves the capacity to design valid and fair assessment tasks, interpret evidence responsibly, communicate results clearly, and use assessment

information to improve learning (Popham, 2009; Brookhart, 2011; DePascale, 2018). This broader view is especially important in teacher education, where future teachers must learn to make assessment decisions that are technically sound, ethically defensible, and responsive to learners' needs.

Despite this importance, assessment literacy is often taught and studied as a procedural competence. In many teacher education contexts, students learn tables of specifications,

item analysis, validity, reliability, rubrics, and grading systems as separate course requirements. These competencies are necessary, but they may not be sufficient if preservice teachers are not also prepared to use assessment evidence in ways that support learner agency, dialogue, and improvement. Assessment can become compliance-oriented when it is disconnected from feedback and professional judgment. This concern is consistent with studies showing that assessment literacy remains a continuing challenge in teacher education and that more integrative approaches are needed to connect assessment knowledge with classroom decision-making and responsibility (Khalid, Latif and Yusof, 2021; Girgla et al., 2021; Gan and Lam, 2024).

Feedback literacy extends this discussion by focusing on how learners and teachers understand, interpret, and use feedback. Carless and Boud (2018) define student feedback literacy as the understandings, capacities, and dispositions needed to make sense of feedback and use it for improvement. From the teacher's perspective, feedback literacy involves designing feedback processes that learners can understand, discuss, and act on (Boud and Dawson, 2023; Dawson et al., 2023). Feedback is therefore not simply a comment added after assessment. It is part of how assessment evidence becomes meaningful, usable, and educationally responsible.

However, assessment literacy and feedback literacy are still commonly treated as separate areas of scholarship. Assessment literacy research tends to emphasise evidence, criteria, validity, reliability, and judgment, while feedback literacy research often emphasizes uptake, dialogue, agency, and evaluative judgment. These emphases are complementary, yet the relationship between the two literacies remains underexamined. This separation creates a conceptual gap: preservice teachers may learn how to produce assessment evidence without fully learning how that evidence should be communicated, interpreted, and used to support learning. Conversely, feedback may be discussed as dialogue without sufficient attention to the quality and fairness of the assessment evidence on which it is based.

The need to connect assessment and feedback is supported by emerging work on evaluative judgment and feedback processes. Deneen and Hoo (2021) argued that assessment literacy is connected with self-evaluation and peer feedback, suggesting that assessment competence develops through interaction with evaluative processes. Panadero and Lipnevich (2022) emphasised that feedback should be understood as a system of processes rather than as a single message. Nieminen and Carless (2022) further argued that feedback literacy must be examined critically because feedback is shaped by agency, power, context, and access. These perspectives suggest that assessment and feedback should be viewed as interdependent practices within teacher education.

This interdependence is especially important when assessment is understood as a matter of professional responsibility. Responsible assessment requires attention not only to technical quality but also to fairness, transparency, consequences, and ethical use of evidence. Messick (1995) emphasised that validity includes the meaning and consequences of assessment interpretation and use. The *Standards for Educational and*

Psychological Testing similarly foregrounds validity, fairness, and appropriate interpretation as central to responsible assessment practice (American Educational Research Association, American Psychological Association and National Council on Measurement in Education, 2014). In teacher education, this means that preservice teachers must learn to ask not only whether an assessment tool works, but also whether it supports learners fairly, communicates expectations clearly, and informs better teaching decisions.

The same concern is visible in the Philippine teacher education context. The Commission on Higher Education requires teacher education programmes to prepare graduates who can demonstrate professional and pedagogical competencies across elementary education, secondary education, early childhood education, and special needs education (Commission on Higher Education, 2012, 2017a, 2017b, 2017c, 2017d). The Philippine Professional Standards for Teachers also emphasises assessment, reporting, learner diversity, and the use of evidence to support learning (Department of Education, 2017). These expectations make assessment and feedback literacy essential to preservice teacher preparation. Yet they also create the risk that assessment becomes framed mainly as compliance with outcomes, rubrics, and documentation rather than as a dialogic and ethical practice.

To address this gap, this study proposes the Assessment–Feedback Literacy Nexus (AFLN) as an integrative framework for preservice teacher education. AFLN positions assessment literacy as the structural foundation of responsible assessment practice and feedback literacy as the dialogic process through which assessment evidence becomes meaningful for learning. Assessment literacy includes knowledge and practice related to learning outcomes, test construction, validity, reliability, rubrics, scoring, grading, and interpretation. Feedback literacy includes understanding criteria, interpreting comments, engaging in dialogue, using feedback for revision, and supporting learner agency. Their connection appears when preservice teachers learn to link assessment evidence with feedback action. For example, a rubric is not only a scoring tool; it is also a feedback tool that communicates expectations and guides improvement. Item analysis is not only a statistical procedure; it can inform instructional decisions and feedback to learners.

Figure 1 summarizes the proposed relationship among assessment literacy, feedback literacy, contextual factors, and responsible assessment practice.

AFLN does not assume that one literacy simply causes the other. Rather, it proposes that assessment and feedback literacy develop through repeated cycles of design, interpretation, communication, revision, and reflection. In this sense, AFLN is both conceptual and pedagogical. Conceptually, it explains why assessment and feedback should be studied together. Pedagogically, it suggests that teacher education courses should provide repeated opportunities for preservice teachers to design assessment tools, receive and use feedback, revise outputs, discuss criteria, and apply assessment practices in authentic or simulated teaching contexts.

This framework is also informed by prior local work at a private university in the southern Philippines. The ALiPP framework

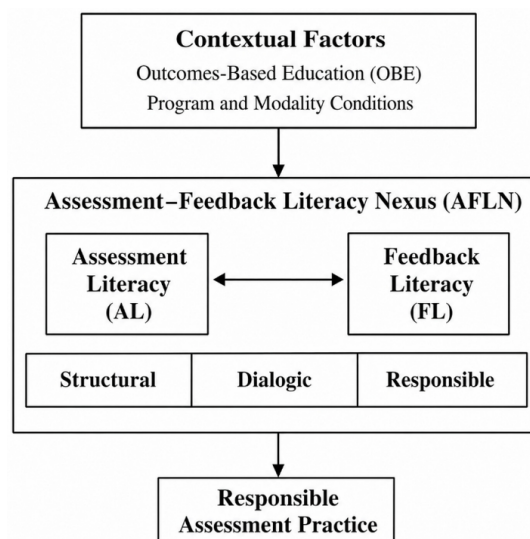


Figure 1: Assessment-Feedback Literacy Nexus framework, 2020-2024 (source: own elaboration)

examined assessment literacy by integrating students' assessment performance and perceptions of assessment learning (Author, 2026). The ASPIRE framework explored how inquiry-based teacher education can support assessment literacy development through active engagement, reflection, and evidence-informed practice (Author, Author and Author, 2025). These frameworks provide useful contextual grounding, but they are not treated as the sole foundation of AFLN. Instead, they are positioned as context-specific precursors. AFLN extends this work by explicitly integrating feedback literacy with assessment literacy and by interpreting both as dimensions of professional responsibility.

The present study reanalyses historical longitudinal mixed-methods data from 250 preservice teachers at a private university in the southern Philippines across five successive cohorts from 2020 to 2024. It examines matched A1-A2 course-phase differences, within-domain longitudinal continuity, programme-level patterns, and qualitative evidence linking assessment, feedback, fairness, and professional responsibility. The study addresses the following questions:

1. How did matched preservice teachers' standardized course-based assessment-literacy performance and overall feedback-literacy scores differ between A1 and A2 across the sequential course sequence?
2. To what extent did A1 scores predict A2 scores within each literacy domain, and what programme-level patterns were observed across assessment literacy and feedback literacy?
3. How do qualitative experiences across the two course phases and successive cohorts explain connections among assessment, feedback, fairness, and professional responsibility?

By addressing these questions, the study contributes to teacher education research by offering a more integrated account of how preservice teachers learn to design, interpret, communicate, and use assessment evidence responsibly. It also responds to the need for research that positions assessment and feedback not as separate technical competencies, but as connected practices of professional responsibility in teacher education.

MATERIALS AND METHODS

Research Design

This study used a two-wave, multi-cohort convergent parallel mixed-methods secondary-analysis design to examine assessment literacy and feedback literacy among preservice teachers at a private university in the southern Philippines. The 2020-2024 window represents five successive cohorts, not one cohort observed continuously for four years. Within each cohort, the same students were matched from Assessment of Learning 1 (A1) to Assessment of Learning 2 (A2), which were normally taken in consecutive semesters; the typical interval between the two measurement phases was therefore approximately one academic semester. Quantitative within-student analyses compared matched A1 and A2 records, whereas cohort year and instructional modality were treated as between-cohort contextual descriptors. The qualitative strand reanalysed archived interview and reflection materials concerning assessment design, feedback use, fairness, rubric development, and professional responsibility. The two strands were analysed separately and integrated during interpretation through joint displays and meta-inferences (Creswell and Creswell, 2018). Qualitative references to A1/A2 or earlier/ later phases are reported as phase- or cohort-level patterns and are not interpreted as individual four-year trajectories unless the same participant could be linked across phases.

Relationship to Prior Studies and Originality

The present article is a secondary reanalysis and extension of the ALiPP source dataset (Quileste, 2026). The 250 matched A1/A2 assessment-literacy records and the original interview/ reflection corpus overlap with ALiPP; accordingly, AL-only descriptive, A1-to-A2 prediction, and programme-level results reported here are identified as previously reported contextual anchors rather than as new findings. The FL_A1 and FL_A2 variables and their longitudinal analyses were not reported in the ALiPP publication and constitute the new quantitative component of the present study. The principal novel contribution is the integration of those feedback-literacy indicators with

the assessment-literacy evidence and the development of AFLN as a responsibility-oriented explanatory framework. ASPIRE (Quileste, Estacio and Pagara, 2025) was a separate mixed-methods study involving 29 preservice teachers under a separate institutional ethics approval; no ASPIRE participant-level data, quantitative results, interviews, or reflections were pooled into or reanalysed in the present dataset. ASPIRE is used only as conceptual and contextual grounding.

Research setting

The study was conducted in a teacher education unit of a private university in the southern Philippines. The research context involved two sequential undergraduate professional education courses: Assessment of Learning 1 (A1) and Assessment of Learning 2 (A2). A1 introduced foundational competencies such as outcomes alignment, table of specifications development, test construction, validity, reliability, and interpretation of assessment results. A2 extended these competencies to applied assessment tasks, including rubric development, performance assessment, grading systems, item analysis, and feedback practices. A1 was normally completed in one semester and A2 in the succeeding semester. The study covered five successive cohorts from 2020 through 2024, including cohorts exposed to fully online delivery during the pandemic and later cohorts returning to face-to-face learning. These cohort differences are treated as contextual conditions rather than as four-year within-person follow-up.

Participants and Sampling

The quantitative sample comprised 250 preservice teachers who completed both A1 and A2 and whose records could be matched across the two course phases. They came from four teacher education programmes: Bachelor of Elementary Education (BEED), Bachelor of Secondary Education (BSED), Bachelor of Special Needs Education (BSNED), and Bachelor of Early Childhood Education (BECED). The programme distribution was BECED $n = 7$, BEED $n = 107$, BSED $n = 88$, and BSNED $n = 48$. Each participant contributed one matched A1-A2 pair within a single cohort year; no participant was treated as contributing observations across multiple cohort years.

Purposive sampling was used because the analysis required complete paired longitudinal records. Inclusion required enrolment in both sequential assessment courses, a matched student identifier across A1 and A2, complete assessment-literacy scores, corresponding FL_A1 and FL_A2 overall scores, and membership in one of the 2020-2024 cohorts. Incomplete or unmatched records were excluded from the quantitative analysis.

The secondary qualitative reanalysis drew on the ALiPP source corpus of semi-structured interviews and written reflections. The interview sample comprised 20 participants, four from each cohort, purposively selected to represent variation in programme and gender. Interviews lasted approximately 30-45 minutes and were conducted via Zoom for the 2020-2021 cohorts and face-to-face for the 2022-2024 cohorts. Written reflections were collected after A1 and A2 as course requirements and were used to triangulate the interview themes.

In the de-identified secondary archive, the reflection materials had been consolidated during the source-study analysis rather than retained as a reliably countable set of individual documents; therefore, this revision does not infer an exact reflection-document count. No standalone course artifacts are treated as a separate qualitative dataset in the present analysis.

Data Sources and Instruments

The study used four data sources: standardized course-based assessment-literacy scores, overall feedback-literacy scores, qualitative interviews and written reflections, and programme/cohort contextual variables.

Assessment literacy was represented by comprehensive examination scores from A1 and A2. Each course used a 60-item multiple-choice comprehensive examination administered as a final assessment. Raw scores ranged from 0 to 60 and were standardized to a 0-100 scale for analysis. The A1 examination sampled foundational assessment competencies, including test construction, validity principles, and table of specifications development; the A2 examination sampled more advanced and applied competencies, including rubric design, item analysis, and KR-20 calculations (Quileste, 2026). The two examinations were separately developed for their respective courses and did not use a psychometric equating or common-item linking procedure. Accordingly, the shared 0-100 transformation facilitates descriptive comparison but does not establish measurement invariance or psychometric equivalence. A1-A2 differences are therefore interpreted as differences in standardized course-based assessment performance across sequential phases, not as gain on an invariant assessment-literacy scale.

Feedback literacy was represented by FL_A1 and FL_A2 overall scores from an archived, locally adapted 20-item Feedback Literacy Scale used at the two course phases. Items were rated on a five-point agreement scale from 1 (strongly disagree) to 5 (strongly agree), and the secondary dataset retained an overall phase score expressed on the same 1-5 metric, with higher values indicating stronger feedback literacy. Archived examples included items addressing evaluation of the usefulness of feedback and recognition of emotional barriers to responding to feedback. The local instrument was conceptually informed by the feedback-literacy literature (Carless and Boud, 2018; Dawson et al., 2023) but was not a direct administration of the later 24-item Feedback Literacy Behaviour Scale reported by Dawson et al. (2023). FL_A1 corresponded to the first-course phase and FL_A2 to the second-course phase. Because only overall phase scores, rather than complete item- or subscale-level responses, were retained in the secondary dataset, the present analysis uses total feedback-literacy scores and makes no claims about specific subdimensions such as emotion management or appraisal.

Qualitative data consisted of semi-structured interview transcripts and post-A1/post-A2 written reflections from the ALiPP source corpus. These materials addressed students' experiences with assessment design, rubric development, feedback, peer review, instructor guidance, online and face-to-face learning conditions, fairness, and application to future teaching.

Programme, cohort year, instructional modality, and gender were retained as descriptive contextual variables. Programme comparisons were examined directly; cohort year and modality were used to contextualise qualitative and descriptive patterns rather than as repeated observations from the same individuals across years.

Validity and Reliability

Several procedures were used to document measurement quality and to delimit the claims that can be made from the secondary dataset. For assessment literacy, both 60-item course examinations were faculty-developed and aligned with the intended outcomes and tables of specifications of A1 and A2. The source study reported KR-20 coefficients above .75 for both examinations (Quileste, 2026). This supports acceptable internal consistency within each course examination, but it does not establish equivalence between the two examinations. Because the forms sampled partly different content and were not psychometrically equated, the study does not interpret the A1-A2 assessment-literacy difference as invariant-scale growth.

For feedback literacy, archived source-study documentation described a locally adapted 20-item, five-point scale reviewed by five faculty experts, with a reported content-validity ratio above .80 and an overall Cronbach's alpha of .85. These values are reported as legacy validation information from the source documentation. Phase-specific alpha coefficients for FL_A1 and FL_A2 could not be recomputed because item-level responses were not retained in the de-identified secondary dataset. This constraint is reported explicitly as a limitation, and the analysis is restricted to the overall FL_A1 and FL_A2 scores.

The quantitative dataset was screened for duplicate records, missing values, inconsistent identifiers, and out-of-range values. Only complete matched cases were retained. Descriptive statistics and assumption checks were reviewed before inferential analysis. Shapiro-Wilk tests were used as an initial normality check, but interpretation also considered distributional shape, sample size, and the robustness of the planned analyses.

The qualitative approach is described in this revision as codebook-oriented thematic analysis rather than reflexive thematic analysis. The six-phase sequence of familiarisation, initial coding, theme construction, theme review, theme definition, and reporting provided the procedural structure (Braun and Clarke, 2006). In the source analysis, the primary researcher coded the corpus and a second researcher independently reviewed 20% of the material; the archived analysis reported 85% agreement (Cohen's kappa = .82), with discrepancies discussed to refine the shared coding framework. Member-checking with five interview participants was also used in the source study. In the present secondary reanalysis, interview and reflection codes were compared within a common thematic matrix so that source type, course phase, cohort context, and divergent cases could be considered together. The agreement check is reported as a quality-control feature of the codebook-oriented approach, not as a criterion for reflexive thematic analysis.

Mixed-methods validity was addressed through integration after the separate analyses. Quantitative patterns were compared with qualitative themes through joint displays, with attention to convergence, complementarity, and divergence. Because the revised quantitative analysis does not report an ambiguously specified cross-domain correlation or a pooled repeated-observation regression, AFLN is supported through the convergence of parallel two-wave score patterns, within-domain continuity, programme comparisons, and qualitative evidence rather than through a claim of direct causal coupling between assessment literacy and feedback literacy.

Data Preparation and Screening

Before analysis, student records were matched across A1 and A2 using student identifiers. Duplicate entries were reviewed and removed when they represented repeated identifiers rather than distinct participants. The final analytical dataset contained 250 matched students, each with one AL_A1, AL_A2, FL_A1, and FL_A2 record. This structure yields 250 paired cases, not 500 independent participants or observations for ordinary least-squares modelling.

AL_A1 and AL_A2 represented standardized course-based assessment performance at the first and second course phases; FL_A1 and FL_A2 represented overall feedback-literacy scores at the corresponding phases. For feedback literacy, the within-student phase difference can be interpreted on the same 1-5 score metric. For assessment literacy, no construct-level gain score is interpreted because A1 and A2 were separately developed, non-equated course examinations. Because feedback-literacy scores were aggregated on a five-point scale, they were treated as approximately continuous for parametric analysis. This decision was based on the use of overall scale scores rather than single Likert-type items and the sample size; the limitation is acknowledged in the Discussion.

Quantitative Data Analysis

Quantitative analyses were conducted in four steps. First, descriptive statistics were computed for AL_A1, AL_A2, FL_A1, and FL_A2 to summarize the score patterns across the two course phases.

Second, paired-samples t-tests were used to compare matched A1 and A2 scores. For feedback literacy, the same five-point total-score metric was used at both phases, so the test evaluates a within-student phase difference on a common score metric. For assessment literacy, the paired test is retained only as a comparison of standardized course-phase performance; because the A1 and A2 examinations were not psychometrically equated, the resulting difference and Cohen's *d* are not interpreted as evidence of growth on an invariant assessment-literacy measure.

Third, separate within-domain regression models examined longitudinal continuity across the two course phases. AL_A2 was regressed on AL_A1, and FL_A2 was regressed on FL_A1. These models test whether first-course performance predicts second-course performance within the same literacy domain; they do not test cross-domain causality.

Fourth, programme differences were examined separately for AL_A1, AL_A2, FL_A1, and FL_A2 using Welch's ANOVA because of the markedly unequal subgroup sizes. Post hoc comparisons were conducted where significant omnibus effects were found. No pooled 500-observation ordinary least-squares regression is reported in the revised analysis because each student contributes repeated A1/A2 observations and those observations are not independent.

Qualitative Data Analysis

Qualitative data were analysed using a codebook-oriented thematic approach structured around Braun and Clarke's six phases (Braun and Clarke, 2006). The primary researcher conducted the full reanalysis of the de-identified interview/reflection corpus using a codebook focused on students' experiences of assessment and feedback across the two course phases. Initial codes included conceptual overload, difficulty applying assessment concepts, feedback gaps, statistical anxiety, modality challenges, instructor modelling, LMS resources, peer collaboration, digital tool use, practicum integration, fairness and validity awareness, rubrics as learning tools, and feedback as dialogue. A second researcher independently reviewed 20% of the source material in the original ALiPP analysis; discrepancies were discussed to refine coding decisions, and the archived analysis reported Cohen's kappa = .82. Interview and reflection evidence was then compared across source type, course phase, and cohort context before themes were finalized. Because the design comprises successive cohorts, statements about A1/A2 or pandemic/post-pandemic patterns are treated as corpus-level or cohort-level patterns rather than as evidence that the same interview participants were followed continuously for four years.

Mixed-methods Integration

Integration occurred after the separate quantitative and qualitative analyses. Joint displays were used to compare statistical patterns with qualitative themes. Integration was organised around convergence, complementarity, and divergence.

Convergence was identified when quantitative and qualitative findings supported the same interpretation. Complementarity was identified when one strand expanded or explained the other. Divergence was identified when one strand complicated the other. Meta-inferences were then developed by weighing both strands together.

This procedure allowed the study to examine whether quantitative course-phase patterns converged with students' accounts of fairness, feedback dialogue, assessment responsibility, and professional meaning-making. Integration was interpretive: it did not assume that parallel quantitative patterns demonstrate a direct causal relation between the two literacies.

Researcher Positionality and Ethical Considerations

The study involved an insider research context because the researcher was familiar with the assessment courses and teacher education programme. This positionality provided contextual knowledge of course design and assessment practices, but it also created potential power, social-desirability, and interpretive biases. These risks are therefore treated explicitly rather than assumed to be eliminated.

Participant-generated responses in the source materials were collected through course-based Google Forms that included electronic consent language authorizing the use of responses for research purposes and assuring the confidentiality of personal identifiers. Historical assessment records originated from routine instructional processes. For the present secondary analysis, identifiers used for record matching were removed before the analytical dataset was prepared, quantitative findings are reported in aggregate, and qualitative excerpts use participant codes. No new participant recruitment or data collection was undertaken for the present study, and the analysis did not affect students' grades or academic standing. The researcher's insider/instructor role is acknowledged as a limitation.

The present manuscript reports the consent and data-governance procedures directly documented for the records used in this secondary analysis. The separate ASPIRE study (Quileste, Estacio and Pagara, 2025), cited only as conceptual and contextual grounding, received institutional ethics approval under XU REC Package No. SOE-202400222; that approval is not represented as covering the ALiPP/AFLN dataset, and no ASPIRE participant data are included in the present analysis. Archival course records are reported only in aggregate, qualitative materials are pseudonymised, and subgroup reporting is intentionally cautious where small programmes could increase re-identification risk.

RESULTS

The results present the quantitative and qualitative findings for the matched two-wave dataset. The quantitative strand is organised around A1-A2 course-phase score differences, within-domain longitudinal prediction, and programme comparisons. The qualitative strand presents the secondary thematic reanalysis of interviews and written reflections, followed by mixed-methods integration. Cohort year and modality are used only as contextual descriptors in the qualitative interpretation.

Quantitative Results

A1-A2 Score Differences within Matched Students

Table 1 presents the matched A1 and A2 scores for standardized assessment-literacy course performance and overall feedback-literacy scores.

Measure	<i>N</i>	A1 Mean (SD)	A2 Mean (SD)	<i>t</i>	<i>df</i>	<i>p</i>	Cohen's <i>d</i>
Assessment literacy score	250	49.2 (7.23)	52.0 (6.92)	-21.3	249	< .001	1.35
Feedback literacy score	250	3.4 (0.70)	4.1 (0.60)	-15.2	249	< .001	0.96

Table 1: A1 and A2 scores for matched students, 2020-2024 (source: secondary analysis; AL statistics previously reported in Author, 2026)

Standardized assessment-literacy course performance was higher at A2 ($M = 52.0$, $SD = 6.92$) than at A1 ($M = 49.2$, $SD = 7.23$), $t(249) = -21.3$, $p < .001$, $d = 1.35$. Because the two examinations sampled partly different content and were not equated, this result is interpreted as a statistically detectable course-phase score difference, not as psychometrically invariant assessment-literacy growth. Overall feedback-literacy scores,

which were recorded on the same five-point metric at both phases, were also higher at A2 ($M = 4.1$, $SD = 0.60$) than at A1 ($M = 3.4$, $SD = 0.70$), $t(249) = -15.2$, $p < .001$, $d = 0.96$.

Within-domain Longitudinal Prediction

Separate regressions examined whether A1 scores predicted A2 scores within each literacy domain. Table 2 presents the results.

Model	Predictor	Estimate	SE	t	p	R ²
AL model	Intercept	6.78	0.85	7.97	<.001	.921
AL model	A1 assessment literacy	0.92	0.02	53.68	<.001	.921
FL model	Intercept	0.92	0.12	7.67	<.001	.884
FL model	A1 feedback literacy	0.88	0.03	29.33	<.001	.884

Table 2: Within-domain A1 scores predicting A2 scores, 2020-2024 (source: secondary analysis; AL model previously reported in Author, 2026)

A1 standardized assessment-literacy performance strongly predicted A2 standardized assessment-literacy performance, explaining 92.1% of variance in the later course score ($R^2 = .921$). A1 overall feedback-literacy score likewise strongly predicted A2 overall feedback-literacy score ($R^2 = .884$). These results indicate strong within-domain continuity across the two sequential course phases. For assessment literacy, the result should be

understood as prediction across related but non-equated course assessments rather than as stability on an identical test form.

Programme Differences

Programme differences were examined separately at A1 and A2 for the four teacher education programmes. Table 3 presents the descriptive statistics.

Programme	N	A1 AL Mean (SD)	A2 AL Mean (SD)	A1 FL Mean (SD)	A2 FL Mean (SD)
BECED	7	39.6 (6.52)	43.5 (6.94)	3.1 (0.60)	3.6 (0.50)
BEED	107	51.5 (5.68)	54.1 (5.54)	3.6 (0.70)	4.2 (0.60)
BSED	88	50.7 (6.44)	53.1 (6.05)	3.5 (0.70)	4.3 (0.60)
BSNED	48	42.9 (7.23)	46.4 (7.48)	3.2 (0.80)	3.8 (0.70)

Table 3: Assessment- and feedback-literacy scores by programme, 2020-2024 (source: secondary analysis; AL statistics previously reported in Author, 2026)

Welch's ANOVA showed significant programme differences for A1 assessment-literacy performance, $F(3, 27.5) = 22.8$, $p < .001$; A2 assessment-literacy performance, $F(3, 27.3) = 16.9$, $p < .001$; A1 feedback literacy, $F(3, 28.1) = 19.4$, $p < .001$; and A2 feedback literacy, $F(3, 27.9) = 15.2$, $p < .001$. Post hoc comparisons indicated that BEED and BSED students generally had higher observed scores than BECED and BSNED students. These are unadjusted programme comparisons and should not be interpreted as evidence of programme quality or causal programme effects. In particular, the BECED subgroup was small ($n = 7$), so the results are used to identify possible support needs rather than to rank programmes.

Qualitative Results

The qualitative findings derive from a secondary thematic reanalysis of the ALiPP interview/reflection corpus. They explain how students described assessment and feedback across course phases and cohort contexts. Because the interview sample was not a single panel followed continuously from 2020 to 2024, temporal statements below refer to phase- or cohort-level patterns in the corpus rather than to four-year within-person change.

Challenges and Supports in Literacy Co-development

A detailed thematic summary of students' difficulties and sources of support is presented in Appendix A.

A1-phase and online-cohort materials frequently described assessment as technically demanding. Concepts such as validity, reliability, tables of specifications, item analysis, and KR-20 were often described as difficult, and students reported uncertainty about whether assessment tools and rubrics were valid or usable in practice. Feedback gaps, statistical anxiety, and limited opportunities for immediate clarification were recurring challenges, particularly in the 2020-2021 online context.

Across the corpus, students identified LMS resources, templates, instructor modelling, peer collaboration, digital tools, and practicum experiences as supports for interpreting assessment tasks and using feedback. A2-phase and later face-to-face cohort materials more often emphasized rubric use, dialogue, revision, and professional application. These patterns are interpreted as contextual and phase-level tendencies rather than as proof that every individual participant followed the same developmental sequence.

Perceived Relevance and Responsible Application

Table 4 presents themes related to students' understanding of assessment and feedback as future teaching responsibilities.

The qualitative corpus showed a phase- and cohort-level contrast between procedural/compliance-oriented accounts and accounts emphasizing professional responsibility. A1-

Theme	Description	Illustrative quote	Pattern	AFLN interpretation
Developing real skills	Shift from completing requirements to taking responsibility for assessment work	"Assessment is a skill I'll need every day." (S3, 2022, F2F)	Evident from 2022	AL and FL became connected to professional competence
Understanding fairness in testing	Recognition that unclear or poorly designed assessment can affect learners	"One wrong test item confused my kids." (S9, 2024, F2F)	Evident from 2021	Feedback experiences deepened fairness awareness
Using rubrics as helpful guides	Rubrics were understood as learning and communication tools, not only grading tools	"Students liked knowing how they were graded." (S2, 2023, F2F)	Strong after 2022	Rubrics joined structural clarity with feedback dialogue
Feedback through conversation	Feedback was increasingly viewed as a two-way process	"I learned to ask good questions." (S7, 2024, F2F)	Strong in F2F cohorts	Dialogic feedback supported reflective assessment practice
Limits of online exams	Online assessments were seen as limited in capturing thinking and applied competence	"Rubrics showed more about how I think." (S4, 2021, Zoom)	Strong in 2020-2021	Modality affected how assessment and feedback were experienced

Table 4: Responsible application themes across phases and cohorts, 2020-2024 (source: secondary thematic reanalysis)

phase materials more often emphasized difficulty, uncertainty, and completion of course requirements, whereas A2-phase and later-cohort materials more often framed assessment as a skill needed for teaching. This pattern is not treated as a four-year within-person trajectory; rather, it is evidence that different course phases and instructional contexts were associated with different emphases in participants' accounts.

Fairness emerged as an important ethical theme. Students recognised that assessment quality was not only a matter of technical correctness but also of learner protection and professional accountability. Rubrics were increasingly

understood as tools for communicating expectations and supporting learning. Feedback was also described as a conversation that helped students revise, clarify, and make better pedagogical decisions.

Mixed-methods Integration

The quantitative and qualitative results were integrated to examine how course-phase patterns in assessment literacy and feedback literacy corresponded with students' accounts of assessment practice. Table 5 presents the joint display of quantitative patterns, qualitative themes, and AFLN interpretations.

Quantitative pattern	Qualitative theme	Integrated AFLN interpretation
A2 standardized AL performance exceeded A1 (paired standardized course-score difference, $d = 1.35$); FL total scores were higher at A2 than A1 ($d = 0.96$).	Practicum integration, rubric use, and development of real skills	Both domains showed parallel A1-A2 phase patterns, while the non-equated AL examinations require cautious interpretation of the AL difference.
A1 AL strongly predicted A2 AL ($R^2 = .921$); A1 FL strongly predicted A2 FL ($R^2 = .884$).	LMS resources, instructor modelling, and peer collaboration	Strong within-domain continuity is consistent with cumulative course learning, but it does not demonstrate cross-domain causality.
Programme differences were significant across AL and FL phase scores.	Feedback gaps and uneven learning support	Programme patterns identify possible differences in learning conditions and support needs; they are not interpreted as causal programme effects.
The quantitative revision does not report the previously ambiguous cross-domain correlation or the 500-observation pooled OLS model.	Students nevertheless described assessment and feedback as connected through rubrics, dialogue, revision, and fairness.	AFLN is supported as an explanatory mixed-methods framework through convergence of evidence, not as a validated causal model.

Table 5: Integrated AFLN findings, 2020-2024 (source: secondary analysis and thematic reanalysis)

The integrated findings are consistent with AFLN as an explanatory framework, but they do not establish direct cross-domain causality. Quantitatively, A2 scores were higher than A1 scores in both domains, first-phase scores strongly predicted second-phase scores within each domain, and programme differences were evident. Qualitatively, participants described how feedback, rubrics, peer collaboration, instructor modelling, and practicum experiences shaped their understanding of assessment as a fair, dialogic, and professionally responsible practice. The convergence of these strands supports an integrated interpretation of the two literacies without requiring an ambiguously specified cross-domain correlation.

The integration suggests that assessment literacy can be interpreted as the structural dimension of responsible assessment practice, while feedback literacy provides a dialogic and reflective dimension through which assessment evidence is discussed, revised, and used. This is an interpretive mixed-methods inference: the quantitative analyses establish parallel phase patterns and within-domain continuity, while the qualitative findings explain how students connected assessment and feedback in practice.

DISCUSSION

The present study examined assessment literacy and feedback literacy within a two-wave, multi-cohort teacher

education context. Quantitative findings showed higher A2 than A1 scores in both domains and strong A1-to-A2 prediction within each domain, while qualitative findings explained how different course phases and cohort contexts were associated with changing emphases on fairness, rubric use, peer collaboration, instructor feedback, and professional application. Because the A1 and A2 assessment-literacy examinations were not psychometrically equated, the AL phase difference is interpreted as a standardized course-performance difference rather than as invariant-scale growth. Taken together, the findings support treating assessment literacy and feedback literacy as complementary dimensions of responsible assessment practice, without claiming direct causal co-development.

Connected Development of Assessment and Feedback Literacy

The quantitative findings show parallel course-phase patterns rather than a demonstrated cross-domain statistical effect. A2 assessment-literacy course performance and overall feedback-literacy scores were higher than A1 values, and first-phase scores strongly predicted second-phase scores within each domain. Assessment literacy has traditionally been conceptualized as knowledge of assessment design, interpretation, and decision-making (Popham, 2009; Brookhart, 2011; DePascale, 2018), while feedback literacy emphasizes the capacity to interpret, seek, and use feedback (Carless and Boud, 2018; Boud and Dawson, 2023). The present analysis does not report a cross-domain correlation whose score pairing cannot be unambiguously reconstructed from the de-identified secondary archive. The claim of connection therefore rests on parallel quantitative patterns plus qualitative and theoretical integration, not on an asserted direct AL-FL coefficient.

This interpretation is strengthened by the qualitative findings. Students consistently described how rubric clarification, instructor modelling, peer review, and opportunities to revise assessment outputs transformed assessment from a technically demanding requirement into a more understandable and professionally meaningful practice. These experiences illustrate that assessment competence is reinforced when learners are actively engaged in feedback processes rather than merely receiving assessment outcomes. Such findings are consistent with the growing literature describing feedback as an active process of meaning-making rather than a one-way transmission of comments (Carless and Boud, 2018; Winstone, Mathlin and Nash, 2019; Boud and Dawson, 2023).

The strong within-domain longitudinal continuity suggests that first-course performance was closely associated with second-course performance. For assessment literacy, this continuity spans related but non-equated examinations; for feedback literacy, it spans repeated total scores on the same local score metric. These findings are compatible with cumulative learning across sequential courses, but they should not be read as evidence that one literacy causes change in the other. This interpretation aligns with Deneen and Hoo (2021), who emphasize participation in evaluative activities, and with Panadero and Lipnevich (2022), who view assessment and feedback as interconnected educational processes.

AFLN and Responsible Assessment Practice

The findings provide mixed-methods evidence consistent with the Assessment-Feedback Literacy Nexus (AFLN) proposed in this study. AFLN extends assessment-literacy and feedback-literacy scholarship by positioning the two literacies within a single responsibility-oriented framework. Assessment literacy provides a structural foundation for generating and interpreting defensible evidence, while feedback literacy provides a dialogic process through which that evidence can be discussed, revised, and used for learning. The framework is explanatory: it does not assert a statistically demonstrated bidirectional causal mechanism.

This perspective responds to a gap in the literature concerning how assessment and feedback can be understood together in preservice teacher education. Rather than claiming that the present data prove that the two literacies develop as a single latent process, AFLN proposes that responsible assessment practice requires both sound assessment design and meaningful feedback engagement. The present quantitative patterns and qualitative accounts are treated as convergent evidence for that integrative proposition.

The qualitative findings illustrate this interaction. Students increasingly viewed rubrics as communication tools rather than grading instruments, feedback as dialogue rather than correction, and assessment as a professional responsibility rather than a procedural requirement. These observations support broader arguments that responsible assessment extends beyond measurement to include fairness, transparency, learner participation, and ethical use of evidence (Messick, 1995; Shepard, 2000; McArthur, 2016). Consequently, AFLN contributes not only a conceptual integration of two literacy traditions but also a professional perspective that aligns assessment practice with educational responsibility.

Program Differences and Targeted Support

Programme differences observed in the quantitative analyses indicate that assessment and feedback literacy did not develop uniformly across teacher education programmes. While differences should not be interpreted as indicators of programme quality, they suggest that preservice teachers encounter different opportunities, curricular emphases, and learning conditions throughout their preparation.

The qualitative findings provide possible explanations for these differences. Students repeatedly identified instructor modelling, peer collaboration, access to exemplars, structured feedback, and practicum experiences as important supports for literacy development. Programmes providing more frequent opportunities for collaborative assessment design and iterative feedback may therefore create stronger conditions for literacy development than programs with fewer such opportunities.

These findings have practical implications for teacher education. Rather than adopting a uniform approach to assessment instruction, institutions may consider differentiated support based on programme characteristics. Additional scaffolded practice, structured feedback cycles, rubric calibration activities, and authentic assessment experiences may be particularly beneficial in programmes where students have fewer opportunities to engage in assessment design

and interpretation. Such recommendations are consistent with recent work advocating sustained assessment literacy development throughout teacher preparation rather than within isolated assessment courses (Gan and Lam, 2024; Soicher, Baker and Thomas, 2024).

Implications for Teacher Education

The study has several implications for teacher education curriculum and practice. First, assessment courses should integrate assessment literacy and feedback literacy rather than treating them as separate instructional topics. Preservice teachers should repeatedly design assessment tools, receive structured feedback, revise their work, and reflect on how assessment evidence informs future instructional decisions.

Second, teacher educators should position feedback as a central component of assessment instruction rather than as an activity occurring after assessment has been completed. Structured peer review, instructor modelling, rubric discussion, and opportunities for revision appear to support both technical competence and professional judgment. Such approaches encourage students to understand assessment as an iterative process rather than a sequence of isolated tasks.

Third, the findings reinforce the importance of professional responsibility within assessment education. Students increasingly recognized that assessment decisions influence learner confidence, participation, and educational opportunity. Consequently, teacher education should continue to emphasise validity, fairness, transparency, and learner agency as integral components of assessment competence rather than as supplementary ethical considerations.

Finally, AFLN offers a useful conceptual lens for future curriculum development. Rather than organizing assessment courses around disconnected technical skills, teacher education programmes may benefit from structuring learning experiences that explicitly connect assessment design, feedback processes, reflection, and responsible professional practice.

The Contribution of the AFLN Framework

The findings are consistent with AFLN as an integrative framework for understanding assessment and feedback literacy development in teacher education. Its contribution is not to claim that assessment literacy and feedback literacy follow a fixed causal sequence. Rather, AFLN explains how assessment and feedback become connected in the formation of responsible teaching practice.

Assessment literacy provides the structural dimension of the framework. It helps preservice teachers design valid tasks, align assessment with outcomes, construct rubrics, interpret scores, and make defensible judgments. Feedback literacy provides the dialogic dimension. It helps preservice teachers understand how assessment information is communicated, interpreted, questioned, revised, and used for learning. AFLN brings these dimensions together by positioning assessment and feedback as interdependent practices of professional responsibility.

This contribution extends prior local work while making the overlap explicit. ALiPP (Author, 2026) is the source of the 250 matched assessment-literacy records and the interview/reflection corpus; AL-only results reproduced in this article

are therefore contextual anchors rather than novel findings. ASPIRE (Author, Author and Author, 2025) was a separate 29-participant study and contributes conceptual context only; none of its participant-level data are included here. The new contribution of the present article is the FL_A1/FL_A2 component and its integration with the ALiPP evidence to formulate AFLN as a responsibility-oriented explanatory framework grounded additionally in wider scholarship on feedback literacy, fairness, validity, dialogue, and professional responsibility.

The present findings also clarify the appropriate scope of AFLN. At this stage, AFLN should not be presented as a fully validated causal model. It is better understood as a responsibility-oriented explanatory framework supported by parallel two-wave score patterns, strong within-domain continuity, programme-level differences, and qualitative evidence of dialogic and ethical assessment practice. This positioning preserves the theoretical usefulness of AFLN while remaining consistent with the limits of the secondary data.

Limitations and Future Research

Several limitations should be considered when interpreting the findings. First, the study was conducted within a single teacher education institution, limiting generalisability. Second, programme sizes were markedly unequal, especially BECED ($n = 7$), so programme comparisons are descriptive of this context and should not be used to rank programme quality. Third, A1 and A2 assessment-literacy examinations were separately developed 60-item course assessments with acceptable internal consistency ($KR-20 > .75$ in the source study) but were not psychometrically equated; consequently, their standardized score difference cannot be interpreted as growth on an invariant scale. Fourth, feedback literacy was available only as overall FL_A1 and FL_A2 scores. Archived documentation reported content review and overall $\alpha = .85$, but item-level data were unavailable for recomputing phase-specific reliability or testing measurement invariance. Fifth, the de-identified secondary archive did not permit unambiguous reconstruction of the previously reported cross-domain correlation, so that coefficient is not reported in this revision. Sixth, the qualitative archive contained 20 semi-structured interviews and post-A1/post-A2 reflection materials, but the consolidated archive did not preserve a defensible count of separate reflection documents; phase- and cohort-level patterns are therefore reported cautiously and are not treated as four-year within-person trajectories. Finally, the researcher occupied an insider/instructor position, creating potential power and social-desirability effects despite documented research-use consent for participant-generated responses, de-identification, pseudonymisation, and aggregate reporting.

Future research should examine AFLN across multiple institutions using prospectively designed measurement that places assessment literacy and feedback literacy on well-documented, repeated metrics. At least three measurement occasions would permit stronger modelling of developmental trajectories, and item-level data would allow phase-specific reliability and measurement-invariance testing. Future studies

should also pre-specify mixed-effects or other repeated-measures models where observations are nested within students, document qualitative sample sizes and source counts prospectively, and test cross-domain relationships using clearly defined score pairings. Such work could establish whether the conceptual connections proposed by AFLN hold beyond the present secondary-analysis context.

CONCLUSION

This study examined assessment literacy and feedback literacy among 250 matched preservice teachers across two sequential course phases embedded within five successive cohorts from 2020 to 2024. A2 standardized assessment-literacy course performance and A2 overall feedback-literacy scores were higher than A1 values, and first-course scores strongly predicted second-course scores within each domain. The assessment-literacy difference is interpreted cautiously because the A1 and A2 examinations were not psychometrically equated, and the revised analysis does not claim a direct cross-domain statistical association or causal co-development.

The qualitative reanalysis showed phase- and cohort-level contrasts between compliance-oriented accounts and accounts emphasizing professional responsibility. Rubric use, peer

collaboration, instructor modelling, practicum application, and feedback dialogue helped explain how students understood assessment as involving fairness, clarity, transparency, revision, and learner agency. These qualitative patterns contextualise the quantitative findings without being treated as a four-year within-person trajectory.

The study contributes the Assessment-Feedback Literacy Nexus as an integrative explanatory framework for teacher education. AFLN positions assessment literacy as a structural dimension of responsible assessment practice and feedback literacy as a dialogic dimension through which assessment evidence becomes meaningful for learning. Its contribution is conceptual and integrative rather than causal: the present secondary data support convergence across quantitative and qualitative evidence but do not constitute full validation of a causal model.

For teacher education programmes, the findings point to the need for assessment courses that include repeated feedback cycles, rubric calibration, guided revision, peer review, and practicum-based application. Future studies may extend the AFLN framework across institutions, programmes, and practicum contexts to examine how assessment and feedback literacy transfer into actual teaching practice.

REFERENCES

- American Educational Research Association, American Psychological Association and National Council on Measurement in Education (2014) *Standards for educational and psychological testing*, Washington, DC: American Educational Research Association.
- Boud, D. and Dawson, P. (2023) 'What feedback literate teachers do: An empirically-derived competency framework', *Assessment & Evaluation in Higher Education*, Vol. 48, No. 2, pp. 158–171. <https://doi.org/10.1080/02602938.2021.1910928>
- Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*, Vol. 3, No. 2, pp. 77–101. <https://doi.org/10.1080/2159676x.2019.1704846>
- Braun, V. and Clarke, V. (2021) 'To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales', *Qualitative Research in Sport, Exercise and Health*, Vol. 13, No. 2, pp. 201–216. <https://doi.org/10.1080/2159676x.2019.1704846>
- Brookhart, S. M. (2011) 'Educational assessment knowledge and skills for teachers', *Educational Measurement: Issues and Practice*, Vol. 30, No. 1, pp. 3–12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Carless, D. and Boud, D. (2018) 'The development of student feedback literacy: Enabling uptake of feedback', *Assessment & Evaluation in Higher Education*, Vol. 43, No. 8, pp. 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Commission on Higher Education (2012) *CHED Memorandum Order No. 46, s. 2012: Policy-standard to enhance quality assurance (QA) in Philippine higher education through an outcomes-based and typology-based QA*, Quezon City: Commission on Higher Education.
- Commission on Higher Education (2017a) *CHED Memorandum Order No. 74, s. 2017: Policies, standards and guidelines for Bachelor of Elementary Education*, Quezon City: Commission on Higher Education.
- Commission on Higher Education (2017b) *CHED Memorandum Order No. 75, s. 2017: Policies, standards and guidelines for Bachelor of Secondary Education*, Quezon City: Commission on Higher Education.
- Commission on Higher Education (2017c) *CHED Memorandum Order No. 76, s. 2017: Policies, standards and guidelines for Bachelor of Early Childhood Education*, Quezon City: Commission on Higher Education.
- Commission on Higher Education (2017d) *CHED Memorandum Order No. 77, s. 2017: Policies, standards and guidelines for Bachelor of Special Needs Education*, Quezon City: Commission on Higher Education.
- Creswell, J. W. and Creswell, J. D. (2018) *Research design: Qualitative, quantitative, and mixed methods approaches*, 5th edition, Thousand Oaks, CA: SAGE Publications.
- Dawson, P., Yan, Z., Lipnevich, A., Tai, J., Boud, D. and Mahoney, P. (2023) 'Measuring what learners do in feedback: The feedback literacy behaviour scale', *Assessment & Evaluation in Higher Education*, Vol. 49, No. 3, pp. 348–362. <https://doi.org/10.1080/02602938.2023.2240983>
- Deneen, C. C. and Hoo, H.-T. (2021) 'Connecting teacher and student assessment literacy with self-evaluation and peer feedback', *Assessment & Evaluation in Higher Education*, Vol. 48, No. 2, pp. 214–226. <https://doi.org/10.1080/02602938.2021.1967284>
- DePascale, C., Sharp, A., Ryan, K. and Betebenner, D. (2018) *Building a conceptual framework for assessment literacy*, Dover, NH: National Center for the Improvement of Educational Assessment.

- Department of Education (2017) *National adoption and implementation of the Philippine Professional Standards for Teachers*, Department Order No. 42, s. 2017, Pasig City: Department of Education.
- Gan, L. and Lam, R. (2024) ‘Unveiling classroom assessment literacy: Does teachers’ self-directed development play out?’, *Education Sciences*, Vol. 14, No. 9, p. 961. <https://doi.org/10.3390/educsci14090961>
- Girgla, A., Good, L., Krstic, S., McGinley, B., Richardson, S., Sneidze-Gregory, S. and Star, J. (2021) *Developing a teachers’ assessment literacy and design competence framework*, Camberwell: Australian Council for Educational Research.
- Khalid, N. H. M., Latif, A. A. and Yusof, I. J. (2021) ‘Research focusing on assessment literacy: A systematic literature review’, *Estudios de Economía Aplicada*, Vol. 39, No. 10, pp. 1–25. <https://doi.org/10.25115/eea.v39i10.5816>
- McArthur, J. (2016) ‘Assessment for social justice: The role of assessment in achieving social justice’, *Assessment & Evaluation in Higher Education*, Vol. 41, No. 7, pp. 967–981. <https://doi.org/10.1080/02602938.2015.1053429>
- Messick, S. (1995) ‘Validity of psychological assessment: Validation of inferences from persons’ responses and performances as scientific inquiry into score meaning’, *American Psychologist*, Vol. 50, No. 9, pp. 741–749. <https://doi.org/10.1037/0003-066x.50.9.741>
- Nieminen, J. H. and Carless, D. (2022) ‘Feedback literacy: A critical review of an emerging concept’, *Higher Education*, Vol. 85, No. 6, pp. 1381–1400. <https://doi.org/10.1007/s10734-022-00895-9>
- Panadero, E. and Lipnevich, A. A. (2022) ‘A review of feedback models and typologies: Towards an integrative model of feedback elements’, *Educational Research Review*, Vol. 35, p. 100416. <https://doi.org/10.1016/j.edurev.2021.100416>
- Popham, W. J. (2009) ‘Assessment literacy for teachers: Faddish or fundamental?’, *Theory Into Practice*, Vol. 48, No. 1, pp. 4–11. <https://doi.org/10.1080/00405840802577536>
- 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*, Vol. 12, No. 1, pp. 158–176. <https://doi.org/10.46328/ijres.5244>
- Quileste, R., Estacio, M. L. P. and Pagara, C. R. (2025) ‘Exploring the ASPIRE framework: Enhancing assessment literacy through inquiry-based teacher education’, *International Journal of Education and Teaching Zone*, Vol. 4, No. 3, pp. 243–258. <https://doi.org/10.57092/ijetz.v4i3.447>
- Shepard, L. A. (2000) ‘The role of assessment in a learning culture’, *Educational Researcher*, Vol. 29, No. 7, pp. 4–14. <https://doi.org/10.3102/0013189x029007004>
- Soicher, R. N., Baker, A. R. and Thomas, R. C. (2024) ‘A Mixed-Methods Research Design to Advance inclusive and equitable teaching’, *Innovative Higher Education*, Vol. 49, No. 6, pp. 1105–1125. <https://doi.org/10.1007/s10755-024-09741-5>
- Winstone, N. E., Mathlin, G. and Nash, R. A. (2019) ‘Building feedback literacy: Students’ perceptions of the Developing Engagement with Feedback Toolkit (DEFT)’, *Frontiers in Education*, Vol. 4, p. 39. <https://doi.org/10.3389/feduc.2019.00039>

APPENDIX

APPENDIX A

Theme	Description	Illustrative quote	Pattern	AFLN interpretation
Conceptual overload	Difficulty managing technical assessment concepts such as TOS, validity, reliability, and KR-20	“CHED CMO topics like reliability... were overwhelming.” (S2, 2020, Zoom)	Strong in 2020-2021	Technical AL demands created uncertainty and feedback-related anxiety
Difficulty applying concepts	Students could prepare assessment tools but were unsure whether these worked in practice	“I could make a TOS but didn’t know if it worked.” (S4, 2021, Zoom)	High in online cohorts	Limited practice weakened the assessment-feedback connection
Feedback gaps	Students needed more specific feedback on outputs such as rubrics and assessment tools	“I didn’t know if [my rubric] was valid or not.” (S7, 2023, F2F)	Persistent	Feedback gaps limited revision and professional judgment
Statistical anxiety	Computation and interpretation tasks caused anxiety	“KR-20 felt like a different subject.” (S9, 2020, Zoom)	High in A1	Feedback and guided modeling helped reduce assessment anxiety
Modality challenges	Online learning limited immediate clarification and modeling	“Hard to know if I was doing things right on screen.” (S5, 2021, Zoom)	Strong in 2020-2021	Modality shaped access to feedback dialogue
LMS resources	Templates and samples helped students understand assessment expectations	“Rubric samples in our Drive helped a lot.” (S3, 2021, Zoom)	Across years	Structured resources supported the AL-FL bridge
Instructor modeling	Demonstration and guided revision helped students understand criteria	“We revised rubrics in class with sir.” (S6, 2023, F2F)	Strong after 2022	Modeling supported dialogic and responsible assessment practice
Peer collaboration	Peer review helped students notice gaps in their work	“Groupmates pointed out what I missed.” (S1, 2024, F2F)	Strong in F2F cohorts	Peer feedback strengthened interpretation and revision
Digital tools	Excel and Jamovi reduced fear of assessment computations	“Jamovi... wasn’t scary anymore.” (S8, 2023, F2F)	Strong after 2022	Technology supported assessment interpretation
Practicum integration	Real or simulated teaching helped students see the use of assessment tools	“My self-made rubric actually worked!” (S10, 2024, F2F)	Strong after 2022	Practice connected assessment structure with feedback use

Table A1: Challenges and supports across course phases and cohort contexts, 2020-2024 (source: secondary thematic reanalysis)