



Developing and Validating SPADE–IPP: A Holistic Framework for Student-Generated Assessment Literacy in Teacher Education

Ronald Quileste 

School of Education, Xavier University – Ateneo de Cagayan, Philippines,
rquileste@xu.edu.ph

This study developed and validated the SPADE–IPP Framework, a cyclical model empowering pre-service teachers to craft student-generated assessments aligned with Outcomes-Based Education (OBE), the Philippine Professional Standards for Teachers (PPST), and the Ignatian Pedagogical Paradigm (IPP). Addressing gaps in holistic assessment literacy—where technical skills often eclipse ethical reflection—a mixed-methods developmental design was employed across four phases at Xavier University–Ateneo de Cagayan's School of Education. Phase 1 derived elements from 60 student-generated outputs; Phase 2 synthesized the six components (Standards, Purpose, Assessment Format, Design Principles, Evidence, IPP overlay); Phase 3 validated via expert Delphi (CVI=0.92; Polit, et al., 2007), with high reliability (overall Cronbach's $\alpha = 0.89$; Cronbach, 1951); and Phase 4 implemented with pre-/post-tests ($n=50$) and reflections. Findings revealed pre-framework fragmentation (Course Learning Outcomes (CLO) alignment $M=72\text{--}81\%$; PPST/IPP $M=35\text{--}56\%$), with significant post-gains (Wilcoxon $Z=-4.56$, $p<0.001$; $d=0.82$; Cohen, 1988) and themes of enhanced alignment (68%), IPP confidence (55%), and format balance (62%). Bachelor of Secondary Education (BSED) subgroups (58%) showed amplified IPP integration ($d=0.91$). SPADE–IPP bridges psychometric rigor with *cura personalis*, offering replicable guidance for Jesuit/value-oriented programs and advancing transformative teacher preparation.

Keywords: SPADE-IPP, student-generated assessments, assessment literacy, pre-service teachers, ignatian pedagogical paradigm, outcomes-based education

INTRODUCTION

Assessment is central to effective teaching and learning, serving in teacher education to both measure pre-service teachers' competencies and model how they will assess future students (Xu & Brown, 2016). With the shift toward Outcome-Based Education (OBE), assessments must be rigorous, purposeful, and aligned with intended learning outcomes (Baguio, 2024; Xiong et al., 2024). Yet challenges remain in balancing validity, reliability, and fairness with reflection, authentic learning, and professional readiness (Messick, 1994; Pastore, 2023).

Citation: Quileste, R. (2026). Developing and validating SPADE–IPP: A holistic framework for student-generated assessment literacy in teacher education. *International Journal of Instruction*, 19(3), 439-454.

In the Philippines, national frameworks such as the Philippine Professional Standards for Teachers (PPST) and the Commission on Higher Education (CHED)’s OBE policies require coherence among learning outcomes, assessment tasks, and teaching strategies (CHED Memorandum Order No. 46, 2012; Department of Education, 2017). Teacher education programs must therefore cultivate assessment literacy while modeling learner-centered, ethical, and context-responsive practices (DeLuca et al., 2018).

At Xavier University–Ateneo de Cagayan, this mandate intersects with the Ignatian Pedagogical Paradigm (IPP), which emphasizes context, experience, reflection, action, and evaluation, positioning assessment as a formative act attentive to the whole person (*cura personalis*) (McArthur, 2023; CHED, 2017a, 2017b, 2017c, 2017d; Ajjawi & McArthur, 2024). Existing frameworks remain fragmented: objective test models ensure reliability but risk superficiality, while authentic assessments like GRASPS emphasize real-world application but often lack scalability and reflection integration (Wiggins & McTighe, 2005; Fawns et al., 2024; Black & Wiliam, 2018). Consequently, pre-service teachers may treat objective and performance assessments as disconnected skills (Kurtz et al., 2019).

To address this gap, this study proposes and validates the SPADE–IPP Framework—a cyclical model integrating Standards Alignment, Purpose and Proficiency, Assessment Format, Design Principles, and Evidence of Learning with IPP. SPADE–IPP unites OBE, PPST, and Jesuit values into a coherent process for designing meaningful, context-sensitive assessments, bridging psychometric rigor, authentic performance, and humanistic reflection (Baguio, 2024; McArthur, 2023; Atjonen et al., 2022).

This study aims to produce a validated, transferable framework demonstrating how assessment can be simultaneously rigorous, authentic, and formative. Guided by the following research questions, it positions SPADE–IPP as both a localized innovation and a framework with broader relevance for advancing assessment literacy in OBE contexts:

1. How do pre-service teachers at Xavier University currently design assessments in relation to OBE, PPST, and IPP?
2. How can the SPADE–IPP Framework be developed and validated in terms of content, construct, and reliability across Assessment 1 and Assessment 2 courses?
3. In what ways does the SPADE–IPP Framework integrate objective assessments, authentic tasks, and Ignatian pedagogy into a coherent model of assessment literacy for pre-service teachers?

Conceptual Framework

This study integrates three assessment traditions: Outcomes-Based Education (OBE), Goal, Role, Audience, Situation, Performance/Product, Standards (GRASPS) performance tasks, and the Ignatian Pedagogical Paradigm (IPP). OBE ensures alignment among learning outcomes, competencies, and assessments (CHED, 2012; Xiong et al., 2024), GRASPS promotes authentic, higher-order tasks (Wiggins &

McTighe, 2005), and IPP embeds reflection, context, action, and evaluation to foster holistic learner formation (Department of Education, 2017; McArthur, 2023). The SPADE–IPP Model applies this integration across Assessment 1 (objective tests) and Assessment 2 (performance-based tasks) (McTighe & Ferrara, 1998; Pastore, 2023). Its five cyclical components—Standards Alignment (S), Purpose and Proficiency (P), Assessment Format (A), Design Principles (D), and Evidence of Learning (E)—overlaid with IPP elements, ensure technical rigor while promoting reflective practice. SPADE–IPP bridges global assessment standards with Filipino pre-service teacher education contexts (DeLuca et al., 2016; Xu & Brown, 2016).

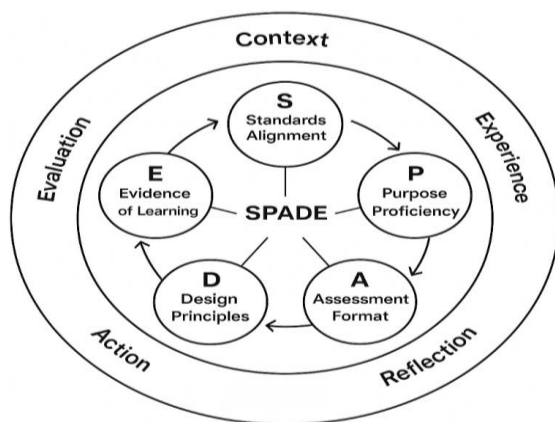


Figure 1
The Cyclical SPADE–IPP Model for Student-Generated Assessment Design.

The SPADE–IPP Framework integrates five interconnected components with the Ignatian Pedagogical Paradigm (IPP) to guide assessment design. Standards Alignment (S) anchors tasks in course learning outcomes (CLOs), OBE frameworks, and national benchmarks like the Philippine Professional Standards for Teachers (PPST), exemplified by objective tests in Assessment 1 and GRASPS tasks in Assessment 2 (Department of Education, 2017; Baguio, 2024; CHED, 2012, 2017a, 2017b, 2017c, 2017d). Purpose and Proficiency (P) clarifies whether assessments are diagnostic, formative, or summative and specifies expected proficiency levels, guided by Bloom’s Taxonomy and the Structure of the Observed Learning Outcome (SOLO) framework, such as quizzes for comprehension and GRASPS projects for synthesis or evaluation (Wiggins, 1998; Black & Wiliam, 2018; Biggs & Tang, 2011). Assessment Format (A) aligns formats to purpose and outcomes, emphasizing objective tests in Assessment 1 and authentic tasks in Assessment 2, ensuring flexible and complementary evaluation (McTighe & Ferrara, 1998; Wiggins & McTighe, 2005; Fawns et al., 2024). Design Principles (D) embed criteria for validity, reliability, fairness, clarity, and practicality, with rubrics and items designed to be observable, unbiased, and inclusive (Messick, 1994; Kennedy & Shiel, 2022; Ling, 2025). Evidence of Learning (E) ensures actionable feedback through answer keys, rubrics, and scoring guides, tracking reliability and proficiency across tasks (Heritage, 2010; Black & Wiliam, 2018).

The IPP overlay—context, experience, reflection, action, and evaluation—grounds assessments in Jesuit values of *cura personalis*, emphasizing learner backgrounds, authentic experiences, metacognition, real-world application, and feedback-informed growth (McArthur, 2023; Ajjawi & McArthur, 2024). Cyclically, SPADE–IPP guides pre-service teachers to align standards (S), define purpose and proficiency (P), select formats (A), apply design principles (D), and gather evidence (E) within an IPP-informed reflective cycle, producing assessments that are technically robust, philosophically grounded, and pedagogically transformative (DeLuca et al., 2018; Pastore, 2023).

METHOD

Research Design

This study employed a mixed-methods developmental research design to design, validate, and evaluate the SPADE–IPP framework. Developmental research was appropriate as it aimed to produce a model that is both theoretically grounded and practically applicable in teacher education. Following design-based research traditions (DeLuca et al., 2016) and Creswell and Creswell’s (2018) mixed-methods approach, the study proceeded through iterative cycles of derivation, synthesis, validation, and implementation, ensuring the framework was empirically tested and contextually responsive.

Research Setting

The study was conducted at the School of Education, Xavier University–Ateneo de Cagayan, a Jesuit institution integrating OBE and the Ignatian Pedagogical Paradigm (IPP). The focal courses were Assessment of Student Learning 1 (Assessment 1) and Assessment 2, enrolling 50 third-year pre-service teachers (Section EDA: 29; Section EDB: 21). Assessment 1 emphasized objective test construction, while Assessment 2 focused on authentic assessments, including GRASPS tasks, allowing longitudinal tracking of student progression in assessment literacy.

Participants and Sampling

Participants included 50 pre-service teachers (32% Bachelor in Elementary Education (BEED), 58% Bachelor in Secondary Education (BSED), 6% Bachelor in Special Needs Education (BSNED), 4% Bachelor in Early Childhood Education (BECED) and others) purposively selected for enrollment in both courses. Four content experts in assessment design and Jesuit education validated SPADE–IPP components. Additionally, 30 pre-service teachers from another institution in Northern Mindanao piloted pre- and post-tests to establish instrument reliability.

Instruments

Data were collected using multiple instruments aligned with SPADE–IPP components. A document corpus of 30 multiple-choice tests and 30 authentic GRASPS tasks was analyzed for standards mapping, purpose, format, principles, evidence, and IPP integration. Pre- and post-tests, each with 60 multiple-choice items measuring assessment literacy, were validated, pilot-tested, and showed high reliability.

(Cronbach's $\alpha = 0.92\text{--}0.94$; Cronbach, 1951). Sample items included knowledge-level questions (e.g., "Which assessment format is most appropriate for measuring recall of key concepts?") and application-level scenarios (e.g., "How would you design a task that integrates PPST indicators with real-world application?"). A rubric evaluated student outputs across five domains—CLO/PPST alignment, format/principles integration, evidence quality, IPP reflection/action, and overall reflective thinking—with strong inter-rater reliability ($\kappa = 0.86$). For illustration, rubric criteria used a 4-point scale with descriptors such as: 4 = Clear, complete, and explicit alignment/integration; 3 = Adequate but partial; 2 = Limited; 1 = Absent or unclear. An expert checklist, based on Content Validity Index (CVI) criteria, assessed SPADE-IPP components for clarity, relevance, and IPP alignment. Qualitative data were gathered through structured reflections, semi-structured interviews, and focus groups (~150 pages of transcripts), capturing participants' experiences and perceptions of SPADE-IPP.

Data Collection Procedures

1. Derivation (Sep–Oct 2024): Student outputs analyzed using thematic coding to identify SPADE elements.
2. Model Development (Nov 2024): Elements synthesized into a provisional framework; refined through expert consultations.
3. Validation (Dec 2024–Jan 2025): Two-round Delphi method with four experts; consensus achieved at 80%, S-CVI = 0.94.
4. Implementation and Evaluation (Feb–May 2025): SPADE-IPP applied in workshops and guided assessment design; pre-/post-tests, rubric-scored outputs, reflections, and focus groups collected.

Data Analysis

Quantitative data were analyzed using SPSS v.28. Pre- and post-test comparisons primarily employed the Wilcoxon signed-rank test due to non-normal distributions in several variables, supplemented by paired t-tests where normality assumptions were met; effect sizes were reported as Cohen's *d*. Rubric scores were similarly compared using the Wilcoxon signed-rank test or paired t-tests as appropriate. Subgroup analyses explored program-specific impacts.

Qualitative data were thematically analyzed using NVivo 14, generating 1,200 codes organized into SPADE-IPP nodes. Trustworthiness was ensured via peer debriefing, member checks (85% agreement), and triangulation, with joint displays integrating qualitative and quantitative findings.

Ethical Considerations

The study received Institutional Research Ethics Board (IREB) approval and adhered to the ethical principles outlined in the Belmont Report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979)—respect for persons, beneficence, and justice. Participation was voluntary, with informed

consent obtained from all participants. Pre-/post-tests were ungraded, and participants could withdraw at any time without consequence. Data were anonymized, securely stored, and retained for five years. Cultural and Jesuit sensitivity were prioritized in reflections, interviews, and focus groups, ensuring ethical compliance. No adverse events occurred, and debriefing sessions were conducted post-study.

FINDINGS

The results are organized according to the study's three research questions. Data from pre-test/post-test surveys, rubric-scored outputs ($n=100$), and focus group discussions ($n=4$; 32 participants) were triangulated to capture both quantitative and qualitative dimensions of student-generated assessment literacy and SPADE–IPP validation. Quantitative analyses included descriptive statistics (means, SDs, percentages) and inferential tests (Wilcoxon signed-rank for non-normal data), while qualitative findings were analyzed thematically following Braun and Clarke's (2006, 2021) six-phase approach. Tables and charts are provided to illustrate key trends.

RQ1. How do pre-service teachers in Xavier University currently design assessments in relation to OBE, PPST, and IPP?

Analysis of 50 student-generated artifacts per course showed uneven assessment design before SPADE–IPP. In Assessment 1, most students (92%) used objective formats, achieving moderate CLO alignment ($M=72\%$, $SD=12\%$) but weak PPST ($M=41\%$, $SD=15\%$) and IPP integration ($M=35\%$, $SD=18\%$), often limited to superficial reflection. BSED students ($n=29$) outperformed BEED ($n=16$) in CLO alignment (78% vs. 68%), and section EDA ($n=29$) exceeded EDB ($n=21$) in PPST links (48% vs. 38%). By Assessment 2, authentic tasks (78%) improved CLO alignment ($M=81\%$, $SD=9\%$), but PPST ($M=56\%$, $SD=13\%$) and IPP ($M=44\%$, $SD=16\%$) integration remained inconsistent, often appended rather than embedded. Overall, assessments were technically competent but fragmented, underscoring the need for an integrative framework.

Table 1

Alignment of Pre-SPADE–IPP Student-Generated Assessments to CLOs, PPST, and IPP ($n=50$ per Course; Cohort Breakdown: 58% BSED [$n=29$], 32% BEED [$n=16$], 6% BSNED [$n=3$], 2% BECED [$n=1$])

Course	CLO Alignment (M %, SD)	PPST Alignment (M %, SD)	IPP Integration (M %, SD)	Dominant Assessment Types
Assessment 1	72% (12%)	41% (15%)	35% (18%)	Objective tests (MCQs, T/F, ID; 92% usage)
Assessment 2	81% (9%)	56% (13%)	44% (16%)	Authentic tasks, performance tasks (78% usage)

The chart below illustrates alignment trends across courses, highlighting progressive but uneven pre-framework baselines.

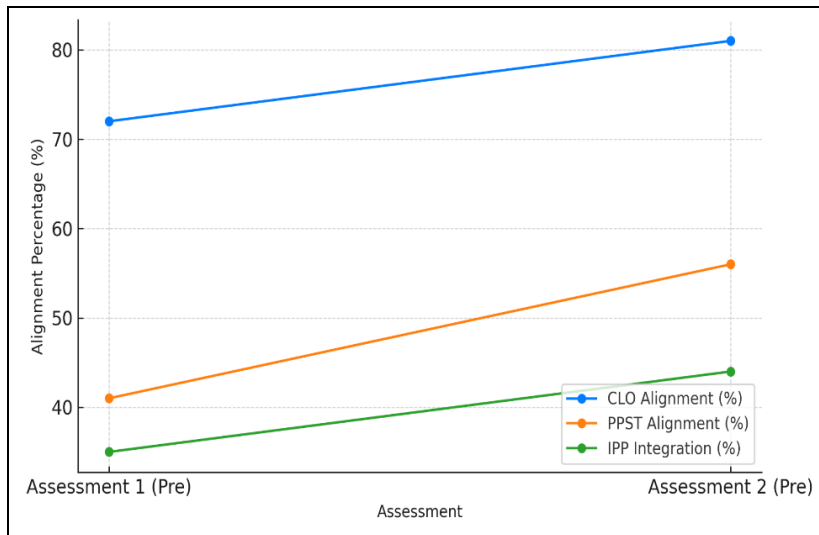


Figure 2
Pre-SPADE-IPP Alignment Trends in Student-Generated Assessments Across Assessment 1 and 2 ($n=50$ per course; Cohort: 58% BSED [$n=29$], 32% BEED [$n=16$], 6% BSNEED [$n=3$], 2% BECED [$n=1$]). Lines represent mean percentages for CLO Alignment, PPST Alignment, and IPP Integration; similar patterns observed across subgroups.

RQ2. How can the SPADE-IPP Framework be developed and validated in terms of content, construct, and reliability across Assessment 1 and Assessment 2 courses?

Content Validation

Four assessment and pedagogy experts rated SPADE-IPP for clarity, comprehensiveness, relevance, and alignment, averaging 4.6/5. The overall Content Validity Index (CVI) was 0.92 (Item-CVI 0.83–1.00), surpassing the 0.80 threshold (Polit, et al., 2007) and confirming strong expert agreement that SPADE-IPP captured essential domains of student-generated assessment literacy.

Construct Validation

Construct validity was supported through multi-source triangulation. Thematic analysis of 60 student-generated artifacts consistently reflected SPADE-IPP components (85% pre-synthesis). Delphi consensus among experts (S-CVI=0.94) affirmed cyclical alignment with OBE and IPP. Student reflections ($n=38$) and focus groups reported SPADE-IPP as a “unifying tool” for balancing assessment formats, reinforcing framework coherence.

Reliability Analysis

Internal consistency was high across constructs (overall Cronbach’s $\alpha = 0.89$; Cronbach, 1951). Pre-/post-test scores showed significant gains (Wilcoxon $Z=-4.56$,

$p < 0.001$; $d = 0.82$), and rubric evaluations of student outputs improved from $M = 2.7/4$ ($SD = 0.6$) to $M = 3.5/4$ ($SD = 0.5$; $Z = -5.12$, $p < 0.001$). Largest effects occurred in Evidence of Learning ($d = 1.05$) and IPP Integration ($d = 0.91$; Cohen, 1988; small = 0.2, medium = 0.5, large = 0.8). BSED students ($n = 29$) showed higher gains in Standards Alignment ($d = 0.92$) than BEED/others ($n = 20$; $d = 0.78$), and EDA ($n = 29$) and EDB ($n = 21$) sections had comparable gains in Design Principles ($d = 0.89$ vs. 0.76).

Table 2
Reliability Indices of SPADE–IPP Constructs ($n = 50$)

Construct	Cronbach's α	KR-20 (Dichotomous Items)	Split-Half Reliability
Objective Assessments	0.84	0.81	0.79
Authentic Performance Tasks	0.87	–	0.82
Ignatian Integration	0.82	–	0.77
Overall	0.89	–	0.85

Table 3
Pre- and Post-SPADE-IPP Rubric Scores for Student-Generated Outputs by Dimension ($n = 50$; Cohort: 58% BSED, 32% BEED, 6% BSNEED, 2% BECED)

Dimension	Pre M (SD)	Post M (SD)	Z (p)	d	BSED Gain (d)	BEED/Other Gain (d)
Standards Alignment	2.8 (0.6)	3.7 (0.4)	-4.12 (<0.001)	0.85	0.92	0.78
Purpose & Proficiency	2.6 (0.7)	3.5 (0.5)	-3.98 (<0.001)	0.76	0.81	0.72
Assessment Format	2.9 (0.5)	3.6 (0.4)	-4.23 (<0.001)	0.89	0.87	0.83
Design Principles	2.5 (0.8)	3.4 (0.6)	-4.67 (<0.001)	0.95	0.94	0.88
Evidence of Learning	2.4 (0.7)	3.6 (0.4)	-5.01 (<0.001)	1.05	1.02	0.99
IPP Integration	2.3 (0.9)	3.5 (0.5)	-4.89 (<0.001)	0.91	0.91	0.72

Note: 4-point scale; $d =$ Cohen's effect size; "Other" includes BSNEED/BECED/Unspecified.

RQ3. In what ways does the SPADE–IPP Framework integrate objective assessments, authentic tasks, and Ignatian pedagogy into a coherent model of assessment literacy for pre-service teachers?

Analysis of 38 reflections and four focus groups (~120 minutes) revealed three overarching themes with 87% code saturation (Braun & Clarke, 2021). Table 4 summarizes how SPADE–IPP functioned as a transformative scaffold for student-generated assessment literacy. Section EDA ($n = 29$) emphasized alignment and clarity more than EDB ($n = 21$; 65% vs. 52%), reflecting program differences.

Theme 1: Enhanced Clarity of Alignment (68%)

SPADE–IPP helped participants visualize intersections among CLOs, PPST, and IPP, shifting assessment design from reactive to deliberate.

- Mapping to CLOs: Participants began aligning tasks to learning outcomes and standards systematically.
- Awareness of PPST: Framework clarified inclusion of professional standards, particularly in elementary contexts.

Theme 2: Confidence in Integrating Ignatian Pedagogy (55%)

Participants reported greater agency in embedding IPP elements across formats.

- Reflection in tests: Even objective tests prompted reflection.
- Contextualized scenarios: GRASPS tasks and rubrics incorporated diverse learner contexts.

Theme 3: Balanced View of Assessment Formats (62%)

SPADE–IPP encouraged participants to see objective and authentic tasks as complementary.

- Complementarity of formats: Objective tests measured knowledge; performance tasks demonstrated applied learning.
- Equal valuing of methods: Design principles ensured fairness and actionable assessment across formats.

Table 4
Themes and Subthemes from Thematic Analysis of SPADE–IPP Experiences (n=38 Reflections; 4 FGDs; Cohort Ties Included)

Theme	Subthemes	Representative Extracts
Enhanced Clarity of Alignment	– Mapping to CLOs – Awareness of PPST standards	“...I always start by checking CLOs and PPST indicators first.” (FGD12, BSED, EDA) “...linking PPST to elementary contexts felt overwhelming; S and P components made it click.” (Ref 8, BEED, EDB)
Confidence in Integrating IPP	– Reflection in tests – Contextualized scenarios	“...even a multiple-choice exam can ask students to reflect after answering.” (Ref 22, BSED, EDA) “...adapt a GRASPS task for multicultural classrooms, weaving in IPP’s context.” (Ref 15, BSED-English, EDA)
Balanced View of Assessment Formats	– Complementarity of formats – Equal valuing of methods	“...objective tests and performance tasks as partners.” (FGD2, BEED, EDA) “No more format bias—D principles ensure even MCQs are fair.” (Ref 31, BECED, EDB)

DISCUSSION

This study developed and validated the SPADE–IPP Framework, an integrative model supporting student-generated assessment literacy among pre-service teachers at Xavier University, a Jesuit institution guided by Outcomes-Based Education (OBE), the Philippine Professional Standards for Teachers (PPST), and the Ignatian Pedagogical

Paradigm (IPP) (Biggs & Tang, 2011; Department of Education, 2017; Katindoy, 2019). The study addressed three research questions: (1) the current state of pre-service teachers' assessment design relative to OBE, PPST, and IPP; (2) the development and validation of SPADE–IPP; and (3) the integration of objective assessments, authentic tasks, and Ignatian pedagogy into a coherent framework. Findings are interpreted within local and international scholarship on assessment literacy and teacher preparation, highlighting SPADE–IPP's role in fostering pre-service teachers as reflective, competent assessment designers (DeLuca et al., 2018; Xu & Brown, 2017; Darling-Hammond, 2017).

Current Assessment Practices and Gaps

Prior to SPADE–IPP, pre-service teachers demonstrated moderate alignment with CLOs ($M = 72\text{--}81\%$) but weaker integration of PPST indicators ($M = 41\text{--}56\%$) and IPP principles ($M = 35\text{--}44\%$). Assessment 1 relied mainly on objective formats, while Assessment 2 incorporated authentic tasks, though reflection and contextualization were often treated as add-ons (Brookhart, 2011; DeLuca et al., 2016). BSED students (58%; $n = 29$) outperformed BEED/others (38%; $n = 20$) in CLO alignment, and section differences suggested curriculum delivery effects (Gepila, 2020).

These initial fragmentation patterns align with broader findings in pre-service teacher assessment literacy. For instance, Diaz et al. (2023), in a study of pre-service teachers' self-reported views on English language assessment published in the *International Journal of Instruction*, similarly identified procedural gaps and inconsistent integration of assessment formats, where technical skills dominated over reflective or holistic practices. Likewise, Adnan et al. (2019), examining teacher competence in authentic and integrative assessment for Indonesian language learning (also in the *International Journal of Instruction*), reported challenges in balancing objective and performance-based tasks, mirroring the pre-framework silos observed here.

These trends reflect global patterns: pre-service teachers acquire technical skills but struggle with holistic, standards-aligned assessments (DeLuca et al., 2018; Xu & Brown, 2017). In the Philippine context, translating PPST competencies into practice remains challenging (Department of Education, 2017), and IPP elements are often compartmentalized rather than fully integrated (Katindoy, 2019). Consequently, assessments were technically competent but ethically shallow, underscoring the need for frameworks that foster reflective, agentic design (DeLuca et al., 2016; Brookhart, 2011). SPADE–IPP addresses this by uniting technical, professional, and philosophical dimensions into cohesive, transferable artifacts (Biggs & Tang, 2011).

Validation of the SPADE–IPP Framework

The SPADE–IPP Framework showed strong content validity ($CVI = 0.92$; $S\text{-}CVI = 0.94$) and construct validity, with 85% of themes from 60 student artifacts matching its components, expert consensus achieved, and 92% of students endorsing its utility (Brookhart, 2011; DeLuca et al., 2018; Xu & Brown, 2017). Reliability was high ($\alpha = 0.89$), and pre/post gains were significant ($Z = -4.56$, $p < 0.001$; $d = 0.82$), with rubric scores improving from 2.7/4 to 3.5/4. BSED students showed larger gains in Standards

Alignment ($d = 0.92$) than BEED/others ($d = 0.78$), reflecting SPADE–IPP’s adaptability. By embedding IPP within a cyclical, constructively aligned framework, SPADE–IPP integrates cognitive, professional, and philosophical dimensions, creating a versatile instrument–framework hybrid applicable to other OBE contexts, such as Ubuntu-inspired curricula (Biggs & Tang, 2011; Katindoy, 2019; McArthur, 2023).

The validation outcomes and significant gains parallel interventions in similar contexts. Oo et al. (2023), in their *International Journal of Instruction* study on a needs-based professional development program, reported comparable improvements in pre-service teachers’ assessment for learning literacy through targeted scaffolding, with enhanced utility and adaptability across subgroups—echoing SPADE–IPP’s large effect sizes and program-specific differences.

Integration of Objective, Authentic, and Ignatian Elements

SPADE–IPP fosters balanced integration, with thematic analysis showing enhanced alignment clarity (68%), IPP confidence (55%), and format complementarity (62%). Pre-service teachers began viewing objective and authentic tasks as complementary, e.g., BSED-English majors linking MCQs to IPP reflections and adapting GRASPS for multicultural equity (McTighe & Wiggins, 2005; Katindoy, 2019). Three key contributions emerge: (1) synthesizing CLOs, PPST, and IPP as mutually reinforcing, with section-level scalability (EDA 65% vs. EDB 52%); (2) building confidence in embedding IPP across formats, expanding Jesuit pedagogy beyond performance tasks; and (3) promoting judicious format selection, enhancing adaptability and holistic literacy (Brookhart, 2011; Xu & Brown, 2017). While IPP emphasis may challenge non-value-oriented contexts, comparable subgroup gains indicate flexibility, bridging psychometrics and pedagogy and advancing “expanded” assessment literacy (Stiggins, 1991; Heritage, 2010; DeLuca et al., 2018).

Insights and Contributions

Several insights emerge:

1. **Beyond Mechanics to Agency:** SPADE–IPP reframes student-generated assessment literacy as macro-level creation, elevating micro-skills (e.g., rubric design) to pedagogical agency—evident in post-framework outputs’ IPP depth (DeLuca et al., 2018; Brookhart, 2011).
2. **Contextualized Global Reach:** Rooted in Philippine/Jesuit realities, it embeds local values (IPP) into OBE/PPST without dilution, modeling “glocal” innovation for OBE worldwide (Katindoy, 2019; Biggs & Tang, 2011).
3. **Technical–Pedagogical Synthesis:** By co-locating validity/reliability with *cura personalis*, it resolves literacy’s silos, with cohort gains (e.g., BSED $d = 0.91$ in IPP) proving ethical-technical harmony (Heritage, 2010; Stiggins, 1991).

Implications for Teacher Education

SPADE–IPP scaffolds student-generated assessments as capstone practices, guiding instructors through iterative templates and reducing fragmentation (Brookhart, 2011; Xu

& Brown, 2017). Institutions can map curricula longitudinally, tailoring modules to program-specific needs (Darling-Hammond, 2017), while policymakers gain a measurable PPST-operational tool for CHED accreditation (Department of Education, 2017; Gepila, 2020). Quasi-experimental studies could track its transfer to in-service practice (DeLuca et al., 2018).

The framework's emphasis on integrative competence resonates with Adnan et al. (2019), who highlighted the need for structured support to develop authentic assessment skills in pre-service teachers, suggesting SPADE–IPP's cyclical model could extend to language or other disciplines.

Beyond classrooms, SPADE–IPP informs curriculum design and faculty development by embedding checkpoints in outcomes-based syllabi, aligning assessments with CLOs, Ignatian values, and PPST competencies. It standardizes assessment discourse, enhances inter-program coherence, and promotes evidence-based teaching (Biggs & Tang, 2011; DeLuca et al., 2016). At the policy level, SPADE–IPP operationalizes PPST Domains 5 and 7, providing measurable indicators for teacher readiness and ethical competence, thus advancing holistic, values-driven assessment in Philippine teacher education (Darling-Hammond, 2017; Katindoy, 2019; McArthur, 2023).

Limitations and Future Directions

Limitations of this study include its single-institution sample ($n = 50$; 58% BSED), which may reduce generalizability, particularly to rural contexts (Gepila, 2020). The developmental design emphasized framework validation over long-term tracking, suggesting a need for longitudinal studies to examine sustained assessment literacy in graduates (DeLuca et al., 2018; Xu & Brown, 2017). Researcher-instructor duality was mitigated through blinded scoring, but external replication is advisable (BERA, 2018). Expanding testing across disciplines and digital modalities, including AI-assisted designs, would further strengthen SPADE–IPP's applicability (McArthur, 2023). Despite these limitations, SPADE–IPP advances student-generated assessment literacy as a holistic practice, integrating objective and authentic tasks within OBE, PPST, and IPP frameworks to equip pre-service—and potentially global—teachers with transformative assessment design tools (Brookhart, 2011; Heritage, 2010; Biggs & Tang, 2011; DeLuca et al., 2016).

CONCLUSION

This study developed and validated the SPADE–IPP Framework, a cyclical model empowering pre-service teachers to craft student-generated assessments aligned with Outcomes-Based Education (OBE), the Philippine Professional Standards for Teachers (PPST), and the Ignatian Pedagogical Paradigm (IPP). Through a mixed-methods developmental design across two assessment courses, SPADE–IPP emerged as a rigorous, contextually grounded tool, bridging technical design with ethical reflection. Key contributions include its structured components—Standards, Purpose, Assessment Format, Design Principles, Evidence of Learning, and IPP overlay—which scaffold holistic assessment creation, fostering validity alongside *cura personalis*. Validation confirmed strong coherence and impact, while thematic analysis revealed enhanced

alignment clarity, IPP confidence, and balanced views of assessment formats, with cohort variations underscoring the framework's adaptability across programs such as BSED and BEED. Limitations include the single-institution sample (n=50) and urban focus, limiting generalizability, alongside the absence of long-term transfer tracking. Future research could explore longitudinal effects, cross-institutional replication, or adaptations for digital and non-Jesuit contexts. Ultimately, SPADE–IPP reimagines student-generated assessment literacy as transformative practice, harmonizing rigor with compassion. It equips pre-service teachers to design assessments that ignite learner agency, meeting OBE and PPST imperatives while advancing teacher preparation toward more equitable, reflective, and values-driven classrooms.

REFERENCES

- Ajjawi, R., & McArthur, J. (2024). From authentic assessment to authenticity in assessment. *Assessment & Evaluation in Higher Education*, 49(4), 511–522. <https://doi.org/10.1080/02602938.2023.2271193>
- Atjonen, P., Pöntinen, S., Kontkanen, S., & Ruotsalainen, P. (2022). In 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/educ.2022.891391>
- Baguio, J. B. (2024). Outcomes-Based Education: Teachers' attitude and implementation. *University of Bohol Multidisciplinary Research Journal*, 7(1). <https://doi.org/10.15631/ubmrj.v7i1.113>
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press/McGraw-Hill Education.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & 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*, 13(2), 201–216. <https://doi.org/10.1080/2159676X.2019.1704846>
- British Educational Research Association. (2018). *Ethical guidelines for educational research* (4th ed.). <https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018>
- Brookhart, S. M. (2011). Educational Assessment Knowledge and Skills for Teachers. *Educational Measurement: Issues and Practice*, 30, 3–12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>

Commission on Higher Education. (2012). *CHED Memorandum Order No. 46, s. 2012: Policy-standard to enhance quality assurance (outcomes-based and typology-based QA)*. Commission on Higher Education. <https://ched.gov.ph/>

Commission on Higher Education. (2017a). *CHED Memorandum Order No. 74, s. 2017: Policies, Standards and Guidelines for Bachelor of Elementary Education (BEEEd)*. Commission on Higher Education. <https://ched.gov.ph/2017-ched-memorandum-orders/>

Commission on Higher Education. (2017b). *CHED Memorandum Order No. 75, s. 2017: Policies, Standards and Guidelines for Bachelor of Secondary Education (BSEd)*. Commission on Higher Education. <https://ched.gov.ph/wp-content/uploads/2017/11/CMO-No.-75-s.-2017.pdf>

Commission on Higher Education. (2017c). *CHED Memorandum Order No. 76, s. 2017: Policies, Standards and Guidelines for Bachelor of Early Childhood Education (BECEd)*. Commission on Higher Education. <https://ched.gov.ph/2017-ched-memorandum-orders/>

Commission on Higher Education. (2017d). *CHED Memorandum Order No. 78, s. 2017: Policies, Standards and Guidelines for Bachelor of Technology and Livelihood Education (BTLEd)*. Commission on Higher Education. <https://ched.gov.ph/2017-ched-memorandum-orders/>

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

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334. <https://doi.org/10.1007/BF02310555>

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Darling-Hammond, L. (2017). *Empowered educators: How high-performing systems shape teaching quality around the world*. Jossey-Bass

DeLuca, C., & Klinger, D. A. (2019). Toward a differential and situated view of assessment literacy. *Frontiers in Education*, *4*, 94. <https://doi.org/10.3389/educ.2019.00094>

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Approaches to Classroom Assessment Inventory: A new instrument to support teacher assessment literacy. *Educational Assessment, Evaluation and Accountability*, *28*(4), 251–272. <https://doi.org/10.1080/10627197.2016.1236677>

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

- Department of Education. (2017). *DepEd Order No. 42, s. 2017: National adoption and implementation of the Philippine Professional Standards for Teachers (PPST)*. Department of Education. <https://www.deped.gov.ph/2017/08/11/do-42-s-2017-national-adoption-and-implementation-of-the-philippine-professional-standards-for-teachers/>
- Díaz, C., Ortiz, M., Gómez, J., & Sanhueza, S. (2023). *English Language Assessment: Pre-service Teachers' Self-Reported Views*. *International Journal of Instruction*, 16(4), 405–422. <https://doi.org/10.29333/iji.2023.16424a>
- Fawns, T., Bearman, M., Dawson, P., Nieminen, J. H., Ashford-Rowe, K., Willey, K., ... Press, N. (2024). Authentic assessment: from panacea to criticality. *Assessment & Evaluation in Higher Education*, 50(3), 396–408. <https://doi.org/10.1080/02602938.2024.2404634>
- Gepila, E. C., Jr. (2020). Assessing teachers using Philippine standards for teachers. *Universal Journal of Educational Research*, 8(3), 739–746. <https://doi.org/10.13189/ujer.2020.080302>
- Heritage, M. (2010). *Formative assessment: Making it work*. Corwin
- Kennedy, E., & Shiel, G. (2022). Writing assessment for communities of writers: Rubric validation to support formative assessment of writing in Pre-K to grade 2. *Assessment in Education: Principles, Policy & Practice*, 29(2), 127–149. <https://doi.org/10.1080/0969594X.2022.2047608>
- Kurtz, J. B., Lourie, M. A., Holman, E. E., Grob, K. L., & Monrad, S. U. (2019). Creating assessments as an active learning strategy: What are students' perceptions? *Medical Education Online*, 24(1), 1630239. <https://doi.org/10.1080/10872981.2019.1630239>
- Ling, J. H. (2025). Assessing the validity and reliability of a rubric-based assessment. *Journal of Multidisciplinary Research and Development*, 2(2), 215–228. <https://doi.org/10.56916/jmrd.v2i2.1253>
- McArthur, J. (2023). Rethinking authentic assessment: Work, well-being, and society. *Higher Education*, 85, 85–101. <https://doi.org/10.1007/s10734-022-00822-y>
- McTighe, J., & Ferrara, S. (1998). Assessing learning in the classroom. *Journal of Quality Learning*, 1(1), 5–12.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23. <https://doi.org/10.3102/0013189X023002013>
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. U.S. Government Printing Office. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html>

- Oo, C. Z., Alonzo, D., & Davison, C. (2023). *Using a Needs-Based Professional Development Program to Enhance Pre-Service Teacher Assessment for Learning Literacy*. *International Journal of Instruction*, 16(1), 781–800. <https://doi.org/10.29333/iji.2023.16144a>
- Pastore, S. (2023). Teacher assessment literacy: A systematic review. *Frontiers in Education*, 8, 1217167. <https://doi.org/10.3389/educ.2023.1217167>
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in nursing & health*, 30(4), 459–467. <https://doi.org/10.1002/nur.20199>
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534–539.
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Alexandria, VA: ASCD.
- Xiong, X., Yu, X., & Xu, Y. (2024). Study based on blended learning design and effect assessment under the concept of OBE. *International Journal of New Developments in Education*, 6(9), 44–49. <https://doi.org/10.25236/IJNDE.2024.060906>
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>