Design-Centered Professional Development for Digital, Outcome-Aligned Classroom Assessment in Tomohon Primary Schools

Feibry Feronika Wiwenly Senduk[1]

[1] Universitas Negeri Manado, Indonesia
Correspoding Author: feibry.senduk@unima.ac.id

ABSTRACT

The global shift toward competency-based education underscores the need for authentic, transparent assessment, yet teacher readiness to enact such practices remains uneven in Indonesian primary schools implementing the Merdeka Curriculum. To address this gap, we aimed to strengthen teachers' capacity to design digital, outcome-aligned evaluations—combining LKPD/task sheets with clear analytic rubrics—and to document resulting changes in classroom practice. Twenty-eight Tomohon primary teachers participated in a practice-proximal workshop followed by mentoring, with data drawn from pre–post rubric-based competency scores across five domains, audits of Canva-produced artifacts, brief classroom pilot logs, and surveys of self-efficacy and intention to use. Results showed significant median improvements in all domains, with the largest gains in rubric structure and constructive alignment; by program's end, 92.9% of participants produced complete digital evaluation packs and 71.4% advanced beyond static layouts to interactive or workflow-integrated use. Classroom pilots indicated high on-task behavior and more frequent student reference to success criteria, while teachers reported higher self-efficacy and strong intentions for sustained adoption. These findings suggest that design-centered professional development, paired with low-barrier authoring tools, can accelerate authentic assessment capacity and enhance classroom processes in resource-varied primary settings. Benefits to schools include a shared language of quality, faster planning-to-practice cycles, and clearer, fairer evaluations for learners and families. We recommend scaling bimtek-plus-mentoring with rubric libraries, embedding brief accessibility micro-clinics, formalizing peer moderation, and extending future cycles to feedback quality and inclusive/multimodal rubrics.

Keywords: *assessment; digital learning; primary education; rubrics; teacher professional development.*

INTRODUCTION

Community service in education—often framed as community-engaged learning or service-learning—links the expertise of universities and teacher educators with urgent, real-world needs of schools. Properly designed, it not only strengthens local capacity but also accelerates teachers' adoption of evidence-based practice, particularly where resources are scarce. At the global level, the Education 2030 agenda and SDG 4 call for inclusive, equitable, high-quality learning opportunities and for strengthening teachers' professional learning ecosystems; both emphasize authentic assessment, digital competence, and the effective use of technology to improve learning outcomes. These priorities require context-responsive support, not just policy ambition (UNESCO, 2015; UNESCO, 2024). Indonesia's Kurikulum Merdeka (Freedom to Learn) operationalizes many of these global principles at the system level. It streamlines content, centers learning on students' needs, elevates authentic, formative, and project-based assessment, and gives schools latitude to adapt assessment to local contexts. Yet policy intent alone does not guarantee classroom change; teachers still need concrete tools, models, and mentoring to translate curriculum goals into day-to-day evaluative practice—especially in primary schools (Kemdikbudristek, 2022; Kemdikbudristek, 2023; Vargheese, 2022). In practice, Indonesian primary teachers frequently face multiple, intersecting challenges: uneven access to infrastructure and devices, variable digital assessment literacy, and limited time to design reliable rubrics, performance tasks, portfolios, and project assessments aligned with learning outcomes. The situation can be more acute outside major urban centers, where connectivity and local professional-development opportunities are uneven. These realities create a pressing need for capacity-building models that are (a) tightly anchored to Kurikulum Merdeka

assessment principles; (b) feasible with the digital tools teachers already use; and (c) scaffolded by coaching so that new practices stick (UNICEF, 2021; Habibi et al., 2023; Nowell et al., 2022).

Against this backdrop, many schools report a persistent gap between policy and practice in three areas. First, teachers often struggle to formulate measurable learning outcomes and success criteria, then align them with task-appropriate instruments (e.g., analytic rubrics) and evidence sources (e.g., portfolios). Second, they under-utilize digital platforms to streamline assessment design, standardize documentation, and provide timely feedback to students and parents. Third, they face difficulty ensuring that classroom tasks are truly "authentic" (mirroring real-world application) and fairly scored across classes or grade levels (Double et al., 2020; Kingston & Nash, 2011). A general solution is a school-embedded, community-service program that pairs targeted training with mentored implementation. Concretely, teachers co-design authentic tasks (e.g., inquiry projects), craft analytic rubrics with clear performance descriptors, publish assessment artifacts digitally (templates, rubrics, checklists), and pilot them in their own classes with peer observation and feedback. This model leverages simple, low-barrier tools (e.g., Canva for layout and standardization; an LMS such as Moodle for distribution, submission, and record-keeping) and emphasizes iterative improvement. It fits the professional-learning expectations in SDG 4 and aligns with Kurikulum Merdeka's emphasis on student-centered assessment while respecting local constraints (devices, bandwidth, teacher workload) (UNESCO, 2015).

International literature substantiates the elements of this approach. First, authentic assessment and formative assessment have consistent, positive—if sometimes modest—effects on learning, engagement, and transfer, especially when teachers are supported to align outcomes, tasks, and criteria. Meta-analytic and review evidence highlights the benefits of performance tasks, portfolios, and well-designed rubrics for achievement and self-regulation (Kingston & Nash, 2011; Panadero et al., 2023). Second, rubric quality matters. Reviews point out that analytic, task-specific rubrics—paired with exemplars and rater calibration—improve scoring reliability and instructional usefulness. Recent syntheses also identify design features that strengthen formative use (e.g., student-friendly criteria and feedback cycles). These findings justify hands-on workshops where teachers build rubrics from exemplars, pilot them, then refine based on student work (Jonsson, 2025; Panadero et al., 2023). Third, digital platforms can amplify assessment quality and efficiency. Systematic reviews show that Moodle is widely used to manage content, run quizzes and assignments, and support online assessment with generally positive effects on engagement and performance; beyond higher education, lessons translate to K–12 when training is scaffolded. Digital peer assessment and e-portfolios similarly expand feedback opportunities and transparency (Gamage et al., 2022; Double et al., 2020; Janssens et al., 2022). Fourth, ICT-mediated learning and evaluation are associated with improved academic outcomes in primary and secondary education when implemented with pedagogical intent. Recent meta-analyses in K–12 report positive effects of educational technology on reading and mathematics and highlight the need to support teachers' classroom integration strategies—precisely the aim of a mentored, school-embedded program (Cheung & Slavin, 2013a, 2013b). Fifth, simple creation tools such as Canva, when used as part of a structured pedagogy, can help teachers standardize and visualize assessment instruments (e.g., rubrics, project briefs, portfolios). Emerging studies from Indonesia and elsewhere report improved student engagement and clarity of expectations, suggesting a practical on-ramp for teachers with limited design backgrounds. While this evidence base is still growing, it aligns with broader findings on the benefits of clear goals and criteria (Jamaludin & Sedek, 2024; Churiyah et al., 2022; Zainal Abidin, 2025). Finally, augmented reality (AR) and interactive media provide optional layers of authenticity—e.g., contextualizing tasks in local phenomena—and have shown medium-to-large effects on learning in meta-analyses. When bandwidth or devices are limited, teachers can still adopt the design logic (situated tasks + concrete criteria) without the hardware-heavy components, ensuring the model is adaptable across contexts (Garzón & Acevedo, 2019).

Despite this maturing evidence, two gaps persist in Indonesia. First, there is a documented digital divide across regions and between urban and rural schools—differences in connectivity, device access, and teachers' digital self-efficacy—that complicates "lift-and-shift" adoption of assessment innovations. UNICEF situation analyses, UNESCO case studies, and recent Indonesian research all note that infrastructure limitations and uneven digital skills impede effective, equitable implementation. This is especially salient in eastern and outer-island regions, where professional

development is less accessible (UNICEF, 2021; UNESCO, 2021; Habibi et al., 2023; Juharyanto et al., 2023). Second, while Indonesia has many descriptive accounts of technology-enhanced instruction, there is comparatively less published work on community-service models that co-develop context-fit assessment tools with primary teachers, measure gains in teachers' assessment literacy, and document fidelity of classroom implementation in remote or small-city settings. Internationally, service-learning in teacher education is well-studied and shows promise for bridging theory–practice gaps; however, localized, school-embedded capacity-building that couples authentic assessment design with lightweight digital tooling remains underreported in Indonesian primary schools—particularly in eastern Indonesia. This gap motivates a community-service intervention that is both evidence-aligned and context-aware (Nowell et al., 2022). The case of Tomohon (North Sulawesi) typifies these needs: primary teachers request practical help to (a) translate competency targets into measurable learning outcomes; (b) craft valid, fair rubrics; (c) design appealing, usable digital assessment packages (project briefs, checklists, portfolios); and (d) use simple apps (e.g., Canva for templating; an LMS for submission and feedback) to standardize documents and streamline workflows. Addressing such needs through a mentored, school-based program would also generate actionable, locally validated exemplars that neighboring schools can adopt with minimal adaptation (Gamage et al., 2022).

This community-service program aims to strengthen primary-teacher capacity in Tomohon to design and implement Kurikulum Merdeka-aligned authentic assessment—centered on clear outcomes, analytic rubrics, project/portfolio evidence, and formative feedback—using accessible digital tools and mentored cycles of design, pilot, reflection, and refinement (Kemdikbudristek, 2022). The program integrates four strands that are rarely combined in Indonesian primary-school contexts: (1) explicit assessment-literacy coaching (outcomes ↔ tasks ↔ criteria) grounded in current research; (2) low-barrier digital authoring for standardization and clarity (e.g., Canva templates for rubrics, task sheets, and portfolio covers); (3) lightweight LMS use (e.g., Moodle) for distributing tasks, collecting evidence, and managing feedback; and (4) a service-learning partnership model in which university facilitators co-design with teachers in-situ, generating locally valid exemplars for reuse across schools. This braided approach responds to known barriers (time, design expertise, bandwidth) while capitalizing on proven affordances of authentic assessment and digital platforms (Gamage et al., 2022; Jamaludin & Sedek, 2024; Churiyah et al., 2022). Meta-analyses and reviews support each strand: authentic and formative assessment improve learning; rubrics and portfolios enhance performance and self-regulation when well-designed; Moodle-supported assessment and digital peer assessment expand feedback and engagement; and school-embedded support accelerates transfer from workshop to classroom. Moreover, SDG 4 and Kurikulum Merdeka both prioritize teacher empowerment and assessment for learning—making a community-service implementation not only desirable but consistent with national and global policy. Finally, addressing regional inequities in digital readiness is an equity imperative; a mentored, context-adapted model helps ensure that teachers outside metropolitan centers can implement high-quality assessment practices (Kemdikbudristek, 2023; Double et al., 2020; Panadero et al., 2023; Gamage et al., 2022; UNESCO, 2015). The service targets public and private primary schools in Tomohon over a short, intensive arc (e.g., preparation, training, mentored classroom pilots, and reflection), with deliverables including: (a) a bank of standards-aligned task sheets and analytic rubrics; (b) class-level portfolio structures and exemplars; (c) LMS course shells for distributing tasks and collecting evidence; and (d) teacher-friendly Canva templates to standardize documentation across grade levels. Evaluation will track growth in teachers' assessment literacy and classroom implementation fidelity, alongside teacher and student perceptions of clarity and fairness. The resulting assets and implementation notes will be shared with local education offices to support scale-out (Janssens et al., 2022; Gamage et al., 2022).

METHOD

Service Design

The program was designed as a three-stage bimbingan teknis (Bimtek) integrating seminar, hands-on workshop, guided practice, and school-embedded mentoring. This blended design aimed to ensure not only conceptual understanding but also practical classroom application. The framework was informed by Kirkpatrick's four-level evaluation model—reaction, learning, behavior, and results

(Kirkpatrick & Kirkpatrick, 2006)—and educational research and development (R&D) principles emphasizing iterative prototyping and feedback cycles (Borg & Gall, 2003). 1) Seminar/mini-lecture: to provide conceptual grounding in authentic assessment and alignment of Competency Achievement (CP), Learning Objectives (TP), and evaluation. 2) Live demonstration: to illustrate Canva-based digital authoring workflows (layout, accessibility, consistency), with optional demonstration of integrating tasks into a Learning Management System (LMS). 3) Guided workshop: to support participants in producing a complete assessment package (learning goals, indicators, LKPD/task sheet, analytic rubric, parent guide), reviewed through structured peer feedback. 4) Remote mentoring: via WhatsApp, enabling follow-up support, artifact submission, and iterative improvement during classroom pilot testing.

Participants

The program targeted primary school teachers in Tomohon, North Sulawesi, Indonesia, including both public and private schools, coordinated through the City Education Office and the SD principals' working group. Each school nominated at least one teacher in charge of classroom assessment (Grades 1–6). Inclusion criteria: active teaching status, willingness to participate in all stages, completion of pre- and post-tests, and consent to submit teaching artifacts. Non-teaching staff were excluded. A target of 25–40 teachers was set, balancing group size for meaningful peer exchange with the feasibility of personalized feedback. Information was collected on sex, age band, teaching experience, grade taught, and prior exposure to Kurikulum Merdeka assessment practices. Teachers also self-rated their digital readiness for Canva and LMS use. School-level descriptors (urban vs. peri-urban, device availability, internet connectivity) were included to contextualize outcomes. Participation was voluntary with informed consent. All data were anonymized for analysis. Teachers retained ownership of their classroom artifacts but granted permission for anonymized samples to be included in research reporting.

Procedures

Stage 1 – Preparation (Weeks −2 to 0), formal communication with the Education Office and school principals' working group to confirm participants, logistics, and focal contacts. Needs assessment: Rapid online/phone survey and short interviews with selected teachers to map training needs and prioritize product themes (e.g., LKPD, rubrics, quizzes), as recommended in needs analysis frameworks (Watkins, Meiers, & Visser, 2012); Creation of slide decks on authentic evaluation in Kurikulum Merdeka, templates for rubrics and LKPDs, pretest instruments, and administrative tools (attendance, consent forms); Venue preparation, Wi-Fi/hotspot arrangement, digital folder access, WhatsApp group setup, and distribution of printed quick guides with QR codes linking to Canva templates. Stage 2 – Implementation (Bimtek, Days 1–2); Principles of authentic assessment, CP–TP–Assessment alignment, and examples of analytic rubrics. Canva-based design workflow (layout grids, fonts, accessibility considerations), plus optional LMS integration. Teachers produced a complete assessment package (goals, indicators, task sheet, rubric, parent guide) and engaged in structured peer review. Facilitators provided targeted feedback for revision. Collection of pretest responses (if not completed earlier) and initial artifacts. Stage 3 – Monitoring and Mentoring (Weeks +1 to +3); Remote mentoring: Ongoing consultation via WhatsApp during classroom pilots. Artifact collection: Teachers submitted Canva files, anonymized student work, and reflective notes. Posttest and reflection: Administration of posttest and short interviews or audio reflections on classroom implementation. Observation: When feasible, structured classroom observations of assessment implementation. Close-out session: Portfolio sharing, cross-school feedback, and planning for sustained practice.

Instruments

To evaluate teachers' knowledge and skills, a pretest and posttest design was applied. The knowledge test consisted of 20–25 items in multiple-choice and short-answer formats, focusing on authentic assessment, rubric design, feedback principles, and the alignment of Competency Achievement (CP) with Learning Objectives (TP). In addition, a performance rubric was used to score artifacts produced by teachers during the training. The rubric covered five key dimensions: clarity and measurability of outcomes and indicators (S1), authenticity and cognitive demand of tasks

(S2), rubric structure including criteria, levels, and descriptors (S3), alignment between outcomes, tasks, and rubric (S4), and accessibility and usability of documents in terms of layout, language clarity, and parent guidance (S5). Each dimension was rated on a four-point analytic scale, with content validity established through expert review and internal consistency examined post-hoc (DeVellis, 2017). During classroom implementation, an observation checklist was employed to capture the clarity of tasks, student engagement, teachers' use of rubrics, and the provision of feedback. Each item was rated on a Likert scale from 1 to 4, supplemented with open-ended notes. When possible, two observers were deployed to enhance reliability through inter-rater agreement checks. Teachers completed surveys designed to capture their reactions and perceptions of the training. Items included Level 1 (Reaction) variables from Kirkpatrick's model—satisfaction, perceived relevance, and usability of the training—as well as measures of self-efficacy regarding digital assessment design and intentions to continue using the approaches. Barriers to implementation were also self-reported to inform subsequent program improvements. Semi-structured interviews and written reflections provided qualitative insights. Teachers were asked to share their experiences with alignment challenges, difficulties in drafting rubrics, perceived changes in student engagement, and needs for scaling up implementation. To reduce participant burden, reflections were also accepted as voice recordings. This flexibility supported richer data collection and encouraged teachers to share authentic experiences (Braun & Clarke, 2006). Each participant submitted an artifact portfolio consisting of finalized Canva files, including rubrics, LKPD or task sheets, and parent guides. Peer-review forms and anonymized samples of student work were also included. These artifacts served as direct evidence of skill application and were used to triangulate with other data sources.

Data Analysis

Quantitative data were analyzed using descriptive and inferential statistics. Descriptive measures such as median (Mdn), interquartile range (IQR), mean, and standard deviation (SD) were reported for all tests and survey results. To measure changes between pretest and posttest, the Wilcoxon Signed-Rank Test was employed due to the ordinal nature of rubric scores and the potential for non-normality in the data (Wilcoxon, 1945). Reported statistics included the Wilcoxon W value, exact *p*-value, effect size (*r*), and the Hodges–Lehmann median difference with 95% confidence intervals. Sensitivity analyses using paired *t*-tests were conducted where data assumptions were met. Reliability checks included Cronbach's α for internal consistency and inter-rater reliability measures (weighted κ or ICC) based on a 20% subsample of artifacts. Qualitative data from interviews and reflections were analyzed using reflexive thematic analysis. This process included familiarization with the data, coding, development of themes, reviewing and refining, and naming themes. To ensure rigor, triangulation of analysts was applied, and member checks were performed by sharing summaries with participants for validation (Braun & Clarke, 2006).

RESULTS AND DISCUSSION

Cohort and Data Sources

The community-service initiative involved 28 primary school teachers from ten public and private elementary schools (kelas I–VI) in Tomohon, representing a range of subjects and experience levels. Evidence for this study was triangulated across four sources: (a) pre–post competency assessments using a five-domain analytic rubric, (b) audit of teacher-produced artifacts—namely LKPD/task sheets, analytic rubrics, and digital evaluation templates designed in Canva, (c) brief implementation logs maintained during limited classroom try-outs, and (d) participant surveys capturing satisfaction, self-efficacy, and intention to use the approaches. The five rubric domains mirrored the program's proximal learning outcomes—S1 (clarity and measurability of outcomes/indicators), S2 (authenticity and cognitive demand), S3 (rubric structure: criteria, levels, descriptors), S4 (constructive alignment across outcomes, task, and rubric), and S5 (accessibility and usability, including layout, readability, and inclusivity)—and were scored on a 0–4 analytic scale. Given the ordinal nature of the ratings and the modest cohort size (n = 28), analyses prioritized medians, interquartile ranges, and nonparametric tests, with effect sizes reported as rank-biserial *r* where applicable.

Teacher Competencies: Pre–Post Gains

Across the five domains, teachers demonstrated statistically and practically meaningful improvement (Table 1). The largest gains appeared in S3 (Rubric Structure) and S4 (Alignment), which increased from median scores around 1.5 at baseline to 3.0–3.5 at post, underscoring growth in writing performance descriptors and aligning tasks with intended competencies. Improvements in S1 and S2 indicate a shift away from recall-heavy checks toward tasks that embody real-world contexts and higher-order thinking. Although S5 (Accessibility/Usability) improved, wider IQRs at post suggest heterogeneity in design comfort and information-design skill; some participants quickly adopted principles of visual hierarchy and plain language, whereas others continued to produce dense text layouts. These patterns are consistent with the program's design-centered emphasis on unpacking CP/TP, modelling exemplar rubrics, and calibrating rubric levels through facilitated review (e.g., Panadero & Jonsson, 2013).

Table 1. Pre–Post Performance by Domain

Domain	Pre Median (IQR)	Post Median (IQR)	Δ (Post–Pre)	Wilcoxon p	Effect size r
S1 Outcomes/Indicators	2.0 (1.5–2.5)	3.0 (2.5–3.5)	+1.0	< .001	0.58
S2 Authenticity/Cognitive Demand	2.0 (1.5–2.0)	3.0 (2.5–3.0)	+1.0	< .001	0.60
S3 Rubric Structure	1.5 (1.0–2.0)	3.0 (3.0–3.5)	+1.5	< .001	0.72
S4 Alignment (CP–Task–Rubric)	1.5 (1.0–2.0)	3.0 (2.5–3.5)	+1.5	< .001	0.69
S5 Accessibility/Usability	2.0 (1.5–2.5)	3.0 (2.5–3.5)	+1.0	< .001	0.55

Digital Adoption and Product Quality

By the end of the workshop cycles, 26 of 28 teachers (92.9%) submitted a complete digital evaluation package comprising aligned learning outcomes, an LKPD/task brief, and an analytic rubric; two participants submitted partially refined drafts. To capture breadth and depth of technology use, an adoption index categorized artifacts into Level 0 (no digital product), Level 1 (static template/layout), Level 2 (interactive linking to resources/forms), and Level 3 (integrated workflow enabling task prompt, digital submission, and rubric-informed feedback). Post-distribution showed L0 = 0, L1 = 8, L2 = 13, and L3 = 7, indicating that 71.4% moved beyond static documents and 25.0% implemented closed-loop submission–feedback flows. Teachers most frequently cited ease of reusability, visual clarity, and time savings as reasons to persist with Canva-based authoring—factors known to drive diffusion of "good-enough" digital tools in school settings (Means et al., 2013; OECD, 2020).

Table 2. Digital Adoption Index

Level	Description	n	%
0	No artifact	0	0.0
1	Static document/template	8	28.6
2	Interactive links (resources/forms)	13	46.4
3	Integrated workflow (submit & feedback)	7	25.0

Classroom Implementation and Student Engagement

Limited classroom try-outs (approximately 21 classes, predominantly in grades 3–6) yielded positive engagement indicators. Observers recorded on-task behavior, frequency of student comprehension questions per 30-minute block, evidence of rubric use by students, and clarity of instructions using a 1–4 scale. Results summarized in Table 3 indicate strong on-task rates (median ≈ 82%) and high observer ratings for clarity and rubric use. Notably, classes in which teachers explicitly discussed success criteria prior to work time showed smoother task flow and fewer requests for clarification, suggesting that visible goals and criteria strengthened self-regulation. Upper-primary students adapted more readily to criteria-referenced checking, while lower-primary classes

benefited from pictorial rubrics and teacher modelling, echoing universal design recommendations for early readers (CAST, 2018; Mayer, 2009).

Table 3. Implementation Indicators (Pilot Try-Outs, n Classes ≈ 21)

Indicator	Mean (SD)	Median (IQR)
On-task behavior (%)	81.3 (9.8)	82 (75–88)
Student questions showing task comprehension (per 30' block)	6.1 (2.0)	6 (5–7)
Evidence of rubric use by students (1–4)	3.1 (0.6)	3 (3–4)
Clarity of instructions (observer rating, 1–4)	3.3 (0.5)	3 (3–4)

Self-Efficacy, Intention to Use, and Satisfaction

Survey responses on five-point scales reflected high satisfaction (M = 4.6, SD = 0.5), increased self-efficacy for rubric design (≈ +1.1 points from baseline), and strong intention to integrate digital evaluation in the forthcoming term (86% "likely/very likely"). Open comments clustered around three themes: improved ability to translate CP into observable success criteria, reduced drafting time and greater consistency via templates, and enhanced transparency in explaining grades to students and parents. These outcomes align with professional learning studies showing that practice-proximal PD with modelling and rehearsal robustly improves teacher confidence and uptake (Desimone, 2009; Darling-Hammond, Hyler, & Gardner, 2017).

Alignment with Expectations and Unexpected Findings

The intervention hypothesized that targeted, design-centered training would elevate authentic assessment competencies and digital authoring. Observed patterns met expectations, with largest gains in the domains initially weakest (rubric literacy and constructive alignment) and adoption exceeding conservative projections for interactive use. Two unanticipated findings emerged. First, teachers in several schools organically formed "rubric-swap" peer circles, exchanging drafts for critique; this professional learning community (PLC) behavior was associated with higher S3/S4 scores, echoing evidence on coaching and PLCs as multipliers of PD impact (Kraft, Blazar, & Hogan, 2018; Vescio, Ross, & Adams, 2008). Second, despite strong templates, some artifacts revealed accessibility issues—overly decorative layouts and heavy color use that increased extraneous cognitive load—highlighting the need for micro-clinics on plain language and universal design (Mayer, 2009; CAST, 2018).

Authentic Assessment, Alignment, and Rubric Quality

The substantial gains in constructive alignment and rubric structure parallel foundational assertions that intended outcomes, learning activities, and assessment tasks must be coherently aligned to drive learning (Biggs, 1996; Biggs & Tang, 2011). The move toward authentic, performance-based tasks with explicit criteria resonates with the formative assessment tradition, which emphasizes clarity of goals and success criteria to improve feedback quality and learner self-regulation (Black & Wiliam, 1998, 2009; Hattie & Timperley, 2007). Moreover, research consistently finds that well-designed rubrics enhance transparency and enable students to monitor progress—outcomes mirrored in our observation of students referencing criteria during work time (Andrade, 2005; Panadero & Jonsson, 2013). A minority of teachers initially produced criteria loosely connected to task demands, a novice pattern widely reported in rubric research; targeted peer review and facilitator feedback helped rectify this misalignment (Jonsson & Svingby, 2007; Brookhart, 2013).

Feedback, Self-Efficacy, and Student Engagement

The improved classroom flow and student self-checks observed during try-outs align with evidence that criteria-referenced, timely feedback facilitates learning gains across age groups (Hattie, 2009; Shute, 2008; Wiliam, 2011). Teachers' reported self-efficacy gains are consistent with PD syntheses showing that active learning, coherence, and sustained support outperform decontextualized workshops (Desimone, 2009; Darling-Hammond et al., 2017). Developmental nuance was evident: upper-primary pupils leveraged textual rubrics more readily, whereas younger

learners benefited from pictorial supports and modelling, echoing universal design and multimedia learning recommendations (CAST, 2018; Mayer, 2009).

Digital Adoption and Design for Learning

Adoption patterns favored simple, well-scaffolded tools: most teachers moved beyond static documents toward interactive linking and a quarter implemented integrated submission–feedback cycles. This trajectory is consistent with diffusion studies suggesting that "good-enough" technologies with clear pedagogical payoff and manageable workload are more likely to persist (Means et al., 2013; OECD, 2020). For a subset, the redesign of assessment workflow approximated the "modification" stage of the SAMR model, where technology reshapes task and feedback flows rather than merely digitizing paper processes (Puentedura, 2010). At the same time, instances of aesthetic over-design underline the importance of information design and cognitive load principles during teacher-authoring (Sweller, 2011).

Professional Communities, Coaching, and Sustainability

The emergent rubric-swap circles illustrate how peer structures and coaching can amplify PD effects, providing ongoing cycles of modelling, practice, and feedback that support behavior change beyond initial workshops (Kraft et al., 2018; Vescio et al., 2008). International syntheses highlight duration, collective participation, and active learning as hallmarks of effective PD; the Tomohon format—Bimtek with WhatsApp-based mentoring—fits these descriptors while remaining lightweight and context-appropriate (Desimone, 2009; Darling-Hammond et al., 2017).

Position within National and Global Reforms

The shift toward competency-oriented, authentic assessment aligns closely with Indonesia's *Kurikulum Merdeka* and global competency-based education trends (OECD, 2018; UNESCO, 2021). Findings suggest that light standardization through co-designed templates and rubrics can support, rather than constrain, teacher professional judgment—addressing the standardization–autonomy tension documented in international debates (Baird et al., 2017).

Convergences and Tensions in the Evidence

Taken together, the results converge with broader literature on the value of aligned tasks and visible criteria, the practicality of low-barrier tools for scaling adoption, and the role of PLCs/coaching in sustainability. Persistent tensions remain between visual appeal and accessibility, prescriptive criteria and creative latitude, and speed of diffusion versus depth of formative feedback culture. Addressing these tensions will require deliberate cycles of co-design, testing, and reflection, supported by concise guidance on universal design and feedback literacy (Boud & Molloy, 2013; Wiliam, 2011).

The Importance

For Tomohon schools, the primary contribution is a shared language of quality in assessment design. Gains in S3 and S4 indicate that teachers increasingly articulate and negotiate criteria and levels, enabling moderation across schools and fairer grading. Canva-based templating compressed the time from planning to practice, a crucial efficiency in contexts where teachers carry heavy workloads. Student-facing clarity—through improved layout and explicit criteria—appears to elevate engagement and self-checking, supporting smoother task flow and reducing time lost to clarification.

Because *Kurikulum Merdeka* depends on assessment that authentically captures competencies, these results illustrate a practical approach for districts: start with authentic tasks, scaffold with co-designed rubrics, and package them as accessible digital artifacts. Two policy-relevant points emerge. First, short, regionally coordinated Bimtek coupled with mentoring can seed a critical mass of exemplars for adaptation, consistent with open educational practices. Second, providing modular rubric libraries for recurring competencies can lift baseline quality while preserving teacher contextualization and autonomy.

Internationally, the policy–practice gap persists where ambitious frameworks are implemented via thin PD. This case adds to evidence for practice-proximal, artifact-centered, and peer-supported models that yield measurable change in teacher competencies and classroom processes. It also

reinforces the notion that "good-enough technologies" can catalyze instructional redesign without heavy infrastructure—an important consideration for low- and middle-income settings (OECD, 2020; Means et al., 2013).

The study contributes localized, primary-grade evidence on digital LKPD plus rubric workflows, a relatively under-documented area in Indonesian contexts. It also highlights the design–accessibility interface, revealing common pitfalls (over-decoration) that warrant targeted micro-competencies. Finally, it documents emergent PLC dynamics in a community-service setting, suggesting that lightweight, messaging-app–mediated PLCs can counter the typical post-PD drop-off.

Practically, the findings argue for embedding a one-page alignment map (CP → task → rubric) as a standard gate before classroom use; curating a grade-banded rubric library for recurrent competencies; and adding short micro-clinics on plain language, contrast, and iconography to strengthen accessibility. Institutionalizing periodic rubric-swap cycles and moderation meetings can stabilize expectations and reduce grading variability. Expanding closed-loop digital workflows will help teachers close the feedback loop efficiently and transparently.

At the system level, districts can scale the Bimtek + mentoring architecture by training regional coaches in constructive alignment and feedback literacy. Establishing lightweight quality assurance rubrics for school-level review of assessment artifacts, alongside an open, versioned repository—such as a Tomohon Open Assessment Bank—can accelerate diffusion while sustaining quality. Monitoring should prioritize process indicators (alignment checks passed, rubric use) over mere counts of certificates.

Subsequent cycles could deepen feedback quality (task-, process-, and self-regulation–level feedback), develop inclusive/pictorial rubrics for early grades and multilingual learners, and extend to multimodal domains (arts/PE) and inquiry-heavy subjects (science/IPS). Adding comparison groups or delayed-treatment designs would strengthen causal inferences and support broader generalizability.

The one-group pre–post design limits causal claims, and maturation or Hawthorne effects cannot be fully excluded. The self-selected, relatively motivated sample (n = 28) may overestimate district readiness, and inter-rater reliability—though checked on a subsample—can be enhanced through double coding. Follow-up was short and focused on near-term behavior change rather than longer-term student outcomes. Even so, convergent evidence across teacher scores, adoption levels, classroom observations, and self-efficacy suggests that practice-proximal, digitally enabled PD can quickly improve assessment design quality and classroom processes in Tomohon's primary schools.

CONCLUSION

This community-service program aimed to strengthen Tomohon primary teachers' capacity to design digital, outcome-aligned classroom evaluations consistent with the Merdeka Curriculum. The training produced substantial gains in rubric quality and constructive alignment (largest median increases in S3–S4), high adoption of Canva-based artifacts (71.4% interactive or workflow-integrated), and improved classroom processes (clearer criteria, higher on-task behavior). These results demonstrate that practice-proximal, design-centered PD using low-barrier tools can rapidly translate policy aspirations into classroom assessment routines. The work contributes practical templates and empirical indicators (pre–post domain scores, adoption tiers) to the Indonesian primary context and enriches international PD literature by highlighting lightweight PLC dynamics and the design–accessibility interface in teacher-created assessments.

REFERENCES

Andrade, H. (2005). Teaching with rubrics: The good, the bad, and the ugly. College Teaching, 53(1), 27–31. https://doi.org/10.3200/CTCH.53.1.27-31

Biggs, J. (1996). Enhancing teaching through constructive alignment. Higher Education, 32, 347–364. https://doi.org/10.1007/BF00138871

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability, 21, 5–31. https://doi.org/10.1007/s11092-008-9068-5

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cheung, A. C. K., & Slavin, R. E. (2013a). The effectiveness of educational technology applications for enhancing mathematics achievement in K–12 classrooms: A meta-analysis. Review of Educational Research, 83(2). https://doi.org/10.3102/0034654313483906

Cheung, A. C. K., & Slavin, R. E. (2013b). The effectiveness of educational technology applications on reading outcomes for struggling readers in elementary schools: A best‐evidence synthesis. Review of Educational Research, 83(1). https://doi.org/10.3102/0034654313475814

Double, K. S., McGrane, J., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. Educational Research Review, 31, 100339. https://doi.org/10.1016/j.edurev.2020.100339

Garzón, J., & Acevedo, J. (2019). Meta-analysis of the impact of augmented reality on students' learning effectiveness. IEEE Access, 7, 101480–101492. https://doi.org/10.1109/ACCESS.2019.2931991

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. Computers & Education, 57(4), 2333–2351. https://doi.org/10.1016/j.compedu.2011.06.004

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. Educational Research Review, 2(2), 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. Educational Measurement: Issues and Practice, 30(4), 28–37. https://doi.org/10.1111/j.1745-3992.2011.00220.x

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. Review of Educational Research, 88(4), 547–588. https://doi.org/10.3102/0034654318759268

Martin-Blas, T., & Serrano-Fernández, A. (2009). The role of new technologies in the learning process: Moodle as a teaching tool in Physics. Computers & Education, 52(1), 35–44. https://doi.org/10.1016/j.compedu.2008.06.005

Nicol, D. J., & Macfarlane‐Dick, D. (2006). Formative assessment and self‐regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2), 199‐218. https://doi.org/10.1080/03075070600572090

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. Educational Research Review, 9, 129–144. https://doi.org/10.1016/j.edurev.2013.01.002

Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: A roadmap of where to go next. Educational Research Review, 24, 1–13. https://doi.org/10.1016/j.edurev.2018.02.004

Shute, V. J. (2008). Focus on formative feedback. Review of Educational Research, 78(1), 153–189. https://doi.org/10.3102/0034654307313795

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. Review of Educational Research, 81(1), 4–28. https://doi.org/10.3102/0034654310393361

Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. Computers & Education, 81, 154–176. https://doi.org/10.1016/j.compedu.2014.10.004

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. Teaching and Teacher Education, 24(1), 80–91. https://doi.org/10.1016/j.tate.2007.01.004

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2