EDUKASIA: Jurnal Pendidikan dan Pembelajaran
Vol. 7, 1 (January-June, 2026), pp. 139-156
ISSN: 2721-1150; EISSN: 2721-1169; DOI: 10.62775/edukasia.v7i1.1802

Development of Evaluation and Self-Reflection Instruments for Diagnosing Teacher Candidate Competencies in Microteaching Courses

Agus Maramba Meha[1], Daniel Frengky Kamengko[1], Theodora S.N Manu[1], Norcy Beeh1

[1] Universitas Kristen Artha Wacana, Kupang, Indonesia

ARTICLE INFO

Keywords:

Teacher competency evaluation;
Microteaching assessment;
Self-reflection instrument;
ADDIE model;
Teacher candidate diagnosis

Article history:

Received 2025-11-04
Revised 2025-01-09
Accepted 2026-02-15

ABSTRACT

This study addresses the limited availability of validated diagnostic instruments for assessing teacher candidate competencies in microteaching courses, a gap that has resulted in inconsistent evaluation practices and weak competency identification. In many teacher education programs, microteaching assessments still rely on informal observations and subjective judgments, which limits the accuracy of competency diagnosis and weakens feedback for professional development. The absence of standardized and validated diagnostic tools has made it difficult to systematically identify specific strengths and weaknesses of teacher candidates. The research aimed to develop an evaluation and self-reflection instrument covering four key teacher competencies: pedagogical, professional, personality, and social. Using the ADDIE development model, the instrument was designed, validated by expert reviewers, and tested with Biology Education students in their microteaching course. Validation processes confirmed high feasibility across design, content, and language aspects. User testing with students and lecturers also indicated strong acceptability and practical relevance. The resulting instrument provides a structured and multi-dimensional framework for diagnosing competency strengths and weaknesses, offering more systematic support for improving teacher candidate performance in microteaching settings.

Corresponding Author:
Agus Maramba Meha
Universitas Kristen Artha Wacana, Kupang, Indonesia; agusmeha@ukaw.ac.id

1. INTRODUCTION

Teacher competence is a key determinant of education quality, and effective teacher preparation requires systematic competency development and assessment (Talis, 2019). Contemporary teaching demands mastery of subject matter, pedagogical skills, professional ethics, and strong personal and social attributes (Orynbekova et al., 2024). However, global evidence continues to show a gap between theoretical coursework and actual teaching readiness among new teachers (Darling-Hammond & Bransford, 2007). Studies consistently report that many pre-service teachers feel underprepared for classroom management and pedagogical content application.

Microteaching plays a crucial role in bridging this gap by offering controlled environments for practicing essential teaching skills (Bell, 2007; Nofiana, 2017). Research shows that well-designed microteaching significantly strengthens self-efficacy and instructional competence (Qomari, 2008; Ralph, 2014). Yet the effectiveness of microteaching depends heavily on the quality of assessment instruments used to diagnose competency development. In many programs, assessments rely on subjective observations and informal feedback that lack reliability (Dayanindhi & Hegde, 2018; Fernandez, 2010).

Despite its importance, validated diagnostic instruments specifically tailored for microteaching remain scarce (Tschannen-Moran & Hoy, 2001). Existing tools often overlook reflective practice, provide limited coverage of competency domains, or lack rigorous validation processes. This problem is also evident in the Indonesian context, where teacher standards emphasize four required competencies—pedagogical, professional, personality, and social—yet systematic diagnostic assessment is still uncommon (Baek et al., 2025; Kiggundu & Nayimuli, 2009). Many microteaching courses continue to use unstructured observation sheets that fail to identify precise competency gaps.

Recent Indonesian studies reveal persistent weaknesses in basic pedagogical and professional skills among pre-service teachers, highlighting the need for more diagnostic and comprehensive evaluation tools. Although several instruments have been developed for specific constructs such as concept understanding or higher-order thinking skills (Machsunah et al., 2023; Rudiyanto, 2020; Walid et al., 2019), no existing instrument integrates all four teacher competencies with both evaluation and self-reflection components. This gap limits institutions' ability to conduct targeted interventions and monitor competency growth.

Therefore, this research seeks to develop and validate a comprehensive diagnostic instrument that assesses the four mandatory teacher competencies through self-assessment, peer assessment, and lecturer evaluation. The objectives include: (1) analyzing current assessment needs, (2) designing a multidimensional evaluation framework, (3) conducting expert validation and user testing, and (4) evaluating feasibility and reliability. The study hypothesizes that a systematically developed instrument—with demonstrated content validity, reliability, and multi-perspective assessment—will provide more accurate and comprehensive diagnostic information than traditional observation-based approaches. Comprehensive means the instrument's ability to map the competencies of prospective teachers in a complete manner through the assessment of four competency domains (pedagogical, professional, personality, and social) with triangulation of assessors (self, colleagues, and lecturers) supported by behavioral indicators and self-reflection.

2. METHODS

This study employed a Research and Development (R&D) approach using the ADDIE model to develop an evaluation and self-reflection instrument for microteaching courses. The ADDIE model (Analysis, Design, Development, Implementation, Evaluation) was selected due to its systematic structure and suitability for developing educational assessment tools (Branch, 2009).

2.1 Research Design and Participants

The participants consisted of 25 sixth-semester Biology Education students and two microteaching lecturers at Universitas Kristen Artha Wacana. Student selection used purposive criteria: completion of

prerequisite pedagogy courses, a minimum GPA of 3.0, voluntary participation, and willingness to complete all research phases. A sample size of 25 students is considered adequate for the initial stage of validity and reliability testing in research and development (R&D) because this study focuses on the development and feasibility of the instrument (not population generalization), the main emphasis is on the quality of the response, the suitability of the indicators, and the initial stability of reliability, so that the limited sample size is still methodologically acceptable in the context of R&D. Although purposive sampling may introduce self-selection bias, it is appropriate for R&D studies that require informed participants familiar with pedagogical practice.

Three expert validators participated: an instrument development expert, a content expert, and a language expert—each holding a doctoral degree and over 10 years of experience. Experts were mapped to specific validation dimensions: (1) design/structure, (2) content alignment, (3) linguistic clarity. Ethical approval was obtained from the university ethics committee, and all participants provided informed consent. Data confidentiality was maintained through anonymized coding.

2.2 ADDIE Model Implementation

To enhance procedural clarity, the ADDIE stages were implemented as follows:

a. Analysis Stage

Needs analysis was conducted using:

- structured classroom observations,
- questionnaires for microteaching students,
- interviews with 10 senior students,
- focus group discussions with faculty,
- curriculum document analysis from five comparable institutions.

Findings revealed that 52% of students reported the absence of structured evaluation instruments in microteaching.

b. Design Stage

Design activities included:

- determining assessment indicators for the four teacher competencies,
- drafting items for self-assessment, peer assessment, and lecturer assessment,
- creating layout prototypes.

Prototypes were tested with a small group (n=6) to refine usability, clarity, and navigation flow.

c. Development Stage

The instrument was developed in digital and print formats. Development involved:

- iterative drafting and refinement,
- expert validation for design, content, and language,
- beta testing with five students and one lecturer.

A total of 50–60 draft items were reviewed, and revisions were made based on expert scoring and comments.

d. Implementation Stage

Instrument testing was conducted over 12 weeks. The implementation included:

- orientation and training sessions for students and lecturers,
- administration of self, peer, and lecturer assessments during microteaching practice,
- scheduled check-ins for support.

Following Donnelly & Fitzmaurice (2005) and González-Fernández et al. (2024), peer assessment was incorporated to strengthen diagnostic accuracy through multi-perspective evaluation.

e. Evaluation Stage

Evaluation used mixed methods:

- quantitative scoring from assessments,
- qualitative feedback from users,
- expert comments from validation forms.

Findings guided refinement of item wording, scoring scales, and instrument structure.

2.3 Data Collection and Analysis

Data sources included:

- expert validation forms,
- student and lecturer response questionnaires,
- observation protocols.

Quantitative analysis used SPSS 28.0 and included:

- descriptive statistics,
- Cronbach's alpha for internal consistency,
- inter-rater reliability using ICC(2,1), a two-way random effects model appropriate for expert agreement,
- Content Validity Index (CVI) following Lynn (in Almanasreh et al., 2019), with a minimum acceptable level of 0.80.

The total number of items evaluated by experts (52 items) was used to compute item-level CVI (I-CVI) and scale-level CVI (S-CVI)

Qualitative data (interviews, open-ended feedback) were analyzed using Braun and Clarke's (2006) six-phase thematic analysis: familiarization, coding, theme generation, review, definition, and reporting. Triangulation was applied by comparing expert input, student responses, and lecturer evaluations.

3. FINDINGS AND DISCUSSION

3.1. Needs Analysis Results

Observation of the microteaching course showed that all students were given opportunities to practice teaching in the microteaching laboratory. However, the assessment practices used by lecturers were mostly informal and oral, without structured tools such as self-reflection instruments. Questionnaire results confirmed this issue, with 52% of students reporting that microteaching evaluations did not use standardized instruments. These results indicate that more than half of the students experienced microteaching evaluation processes that were not supported by standardized and documented assessment instruments. This condition indicates inconsistencies in assessment practices, both between lecturers and between microteaching sessions, resulting in feedback that students received tended to be general and difficult to follow up on. The absence of standardized instruments also limited students' ability to clearly understand assessment criteria and hampered self-reflection and systematic monitoring of competency development.

A more detailed analysis revealed several methodological gaps. Structured observations over eight weeks showed that although all students practiced teaching, only 23% received written feedbackwhile 77% received only verbal and unrecorded feedback. This made it difficult for students to track progress, identify weaknesses, or make targeted improvements. The absence of documentation also created inconsistency between lecturers. These findings are consistent with Saban & Çoklar (2013), Arslan (2021) and Demirci & Akgün (2023) who reported that pre-service teachers frequently call for structured, diagnostic assessment tools in microteaching.

Peer assessment was largely absent, with only 12% of students reporting any structured peer feedback. This contrasts with current best practices recommending multi-perspective evaluation to

strengthen diagnostic accuracy. Student responses also emphasized that microteaching assessments should support self-reflection, aligning with Benton-Kupper (2001), Koukpaki & Adams (2020), Fontaine (2018) who highlights the role of reflective assessment in professional growth.

Interviews with faculty members showed significant variation in assessment practices: 67% relied primarily on subjective impressions, while only 33% used any form of structured criteria. This inconsistency contributed to unclear expectations, unequal feedback quality, and limited identification of specific competency gaps.

Overall, the needs analysis indicates an urgent requirement for standardized, comprehensive, and validated assessment instruments that integrate self-assessment, peer assessment, and lecturer evaluation to support consistent competency development in microteaching.

3.2. Instrument Design

During the design phase, the evaluation and self-reflection instrument was developed based on the four mandatory teacher competencies: pedagogical, professional, personality, and social. These competencies were translated into measurable and observable indicators that formed the basis for item construction. To strengthen methodological rigor, each indicator was operationalized into behaviorally anchored statements to ensure clarity, consistency, and replicability of assessment. The design followed learner-centered microteaching principles (Kilic, 2010), ensuring the instrument not only evaluated teaching skills but also supported reflective practice and professional growth.

Table 1. Four Teacher Competency Framework and Assessment Indicators

Competency Domain	Assessment Indicators	Behavioral Descriptors
Pedagogical	Teacher candidates can develop learning materials according to curriculum development	Develop learning materials aligned with learning objectives, curriculum structure, and student characteristics
	Teacher candidates demonstrate mastery of 8 basic teaching skills	Demonstrate opening and closing lessons, questioning, reinforcement, variation, and classroom management skills
	Teacher candidates apply educational learning principles	Apply student-centered, active, and meaningful learning principles
	Teacher candidates can apply varied approaches, strategies, techniques, and methods appropriate to materials and student development	Combine instructional strategies and methods appropriate to the content and learners' developmental levels
	Teacher candidates can use learning media relevant to materials and learning objectives	Select and use instructional media that support learning objectives
	Teacher candidates can provide motivation to students during the learning process	Provide verbal and non-verbal encouragement to enhance student engagement
	Teacher candidates apply varied assessment techniques	Apply formative and summative assessment techniques appropriately
Professional	Teacher candidates understand learning objectives	Clearly communicate learning objectives to students
	Teacher candidates can master learning materials structurally	Present content in a coherent, systematic, and accurate manner

Competency Domain	Assessment Indicators	Behavioral Descriptors
	Teacher candidates are authoritative when implementing classroom learning	Demonstrate confidence, consistency, and classroom authority
	Teacher candidates show responsibility during learning implementation	Manage instructional time and activities responsibly according to the lesson plan
	Teacher candidates apply examples appropriate to taught materials	Provide relevant and contextual examples to support understanding
Personality	Teacher candidates act according to applicable norms	Demonstrate ethical, respectful, and appropriate behavior
	Teacher candidates can demonstrate honest, noble character and become role models for students	Display honesty and integrity and act as a positive role model
	Teacher candidates are authoritative and can be role models during learning	Exhibit personal authority and exemplary conduct during instruction
	Teacher candidates show good work ethics during learning	Demonstrate commitment, diligence, and responsibility in teaching
	Teacher candidates demonstrate disciplined character in learning	Show punctuality and adherence to classroom rules
Social	Teacher candidates have good oral communication abilities	Communicate ideas clearly and effectively in spoken language
	Teacher candidates have good written communication abilities	Produce clear and well-structured written instructional materials
	Teacher candidates use communication and information technology functionally in learning	Use information and communication technology effectively to support learning
	Teacher candidates have abilities to interact effectively with students and peers	Interact openly, empathetically, and cooperatively with students and peers
	Teacher candidates demonstrate two-way interaction skills in learning	Encourage dialogue, discussion, and responsive interaction with students

A summarized version of the competency indicators is presented in Table 1, covering:

- Pedagogical (7 indicators): lesson planning, basic teaching skills, instructional strategies, use of media, motivation, assessment techniques.
- Professional (5 indicators): content mastery, instructional clarity, responsibility, classroom authority.
- Personality (5 indicators): ethics, honesty, discipline, role modeling.
- Social (5 indicators): oral and written communication, ICT use, interaction with peers/students.

These indicators reflect national teacher competency standards and align with best practices in microteaching assessment. Expert consultations and literature reviews were used to refine indicator boundaries and avoid overlap between domains.

The instrument was designed in three complementary components:

1. Self-assessment (22 items)
   Allows teacher candidates to evaluate their own competencies using a 4-point Likert scale.

2. Peer assessment (18 items)
   Focuses on observable behaviors, collaborative skills, and communication effectiveness.
3. Lecturer assessment (25 items)
   Provides expert evaluation with both quantitative scoring and qualitative comments.

To enhance diagnostic precision, the instrument includes clear scoring descriptors and short behavioral anchors (e.g., "uses varied instructional media effectively," "provides learning motivation consistently"). The design process underwent multiple iterations:

- Draft items were reviewed internally by the research team.
- Early prototypes were tested with small student groups to evaluate clarity, navigation, and item comprehension.
- Feedback resulted in refinement of language, simplification of item structure, and restructuring of instructions.

A sample of item prototypes was developed—for example, pedagogical competency items asking students to rate their ability to apply varied strategies, or social competency items evaluating clarity of oral communication (Chan et al., 2017). These sample items were later included in the appendix to support replication. Overall, the design phase ensured that the instrument captured comprehensive competency dimensions while remaining practical and user-friendly for microteaching contexts.

3.3. Validation Results

The developed instrument was evaluated through expert validation involving three specialists: (1) an instrument design expert, (2) a content expert, and (3) a language expert. Validation focused on design quality, content alignment, and linguistic clarity. Each expert used a standardized rubric rating item relevance, clarity, structure, and alignment to competencies, with scores converted to feasibility percentages.

Table 2. Expert Validation Summary

Validation Aspect	Score Achieved	Feasibility (%)	Category
Instrument Design and Structure	47/48	97.0%	Highly Feasible
Content Alignment and Relevance	34/36	94.4%	Highly Feasible
Language Clarity and Accuracy	28/32	87.5%	Highly Feasible

Table 2 shows that all validation components exceeded the minimum 80% feasibility threshold, demonstrating strong foundational quality. Design validation (97%) confirmed that the framework, scoring system, and layout were appropriate and coherent. Content validation (94.4%) indicated strong alignment with the four mandatory teacher competencies and microteaching objectives. Language validation (87.5%) reflected clear and culturally appropriate wording.

The expert validators provided detailed qualitative feedback, including 12 design recommendations, 8 content-related suggestions, and 15 language refinement notes. These contributions informed improvements in item phrasing, indicator clarity, response scale structure, and the overall usability of the instrument. To strengthen reliability, inter-rater agreement was calculated using Cohen's kappa:

- $\kappa = 0.82$ (design)
- $\kappa = 0.78$ (content)
- $\kappa = 0.75$ (language)

Based on the interpretation of Vetter & Schober (2018), these values are included in the substantial agreement category, which indicates a high and consistent level of agreement between validators, and indicates that the experts' assessments are not coincidental, but are based on a relatively uniform understanding of the criteria and indicators used in the instrument.

CVI analysis followed Lynn (in Wang & Sahid, 2024), using item-level CVI (I-CVI) and scale-level CVI (S-CVI):

- All items met the minimum **I-CVI ≥ 0.80** threshold.
- The overall **S-CVI exceeded 0.90**, indicating excellent content representativeness.

Overall, the validation process confirmed that the instrument demonstrates strong technical quality, clear content alignment, and acceptable linguistic clarity, making it ready for user testing in microteaching settings.

3.4. Implementation and User Testing

The validated instrument was implemented with 25 Biology Education students and two microteaching lecturers to examine practical applicability, clarity, and diagnostic accuracy. Implementation procedures followed a structured sequence consisting of orientation, guided use, independent completion, and feedback collection. This structured deployment ensured that both quantitative and qualitative data were obtained under consistent implementation conditions, thereby enhancing the credibility of findings.

Student responses showed strong positive evaluations across multiple aspects. Specifically, 93% agreed that the instrument aligned with mandatory teacher candidate competencies, 96% found it appropriate for microteaching courses, 96% confirmed the suitability of the four competency domains, and 100% agreed that the instrument integrated all competencies effectively. These high acceptance rates indicate that the instrument successfully addressed gaps in previous assessment practices that relied heavily on informal verbal feedback

Implementation also revealed meaningful improvements in student learning behaviors and self-assessment ability. Pre- and post-implementation surveys indicated a 34% increase in students' ability to identify specific areas requiring professional development. Additionally, the quality of written self-reflections improved by 28%, as shown by increased depth, specificity, and alignment with competency indicators. This demonstrates the instrument's capacity to cultivate reflective habits and strengthen diagnostic awareness—key components of effective teacher preparation.

Students reported that the instrument helped them focus their practice more strategically and set clearer goals for subsequent microteaching cycles. Many students highlighted that structured indicators allowed them to understand precisely which pedagogical, professional, personality, and social competencies required improvement. Peer feedback also became more systematic and constructive because users had access to shared assessment criteria.

However, these findings should be interpreted cautiously due to the potential novelty effect, where students' positive responses may be influenced by the novelty of using a more structured instrument compared to previous evaluation practices. Furthermore, student perception data was collected through self-report data, making it less susceptible to social bias and the tendency to respond in a more predictable manner.

Lecturers similarly provided highly positive evaluations, noting that the instrument supported more objective and consistent assessment practices. They emphasized that the structured format reduced ambiguity in evaluating student performance and improved transparency in grading. Lecturers further stated that the instrument facilitated evidence-based mentoring, allowing them to provide targeted guidance rather than general comments.

Overall, the implementation and user testing phase confirmed that the instrument is not only technically valid but also pedagogically impactful. It enhances student engagement, supports reflective practice, and strengthens the consistency of microteaching evaluations across assessors.

3.5. Detailed Validation Results

a. **Instrument Design Validation**

Instrument design validation was conducted by an expert in educational assessment to evaluate the feasibility of the overall structure, layout, scoring system, and usability. The expert assigned a score of 47 out **of** 48, resulting in a 97% feasibility rate, categorized as "highly feasible." The expert emphasized the importance of maintaining a positive scoring orientation, recommending removal of the zero score to avoid negative bias and improve interpretability.

Additional suggestions included improving item grouping, clarifying rating scale descriptions, and enhancing the visual structure of the instrument to support more efficient use by both students and lecturers.

b. **Content Expert Validation**

Content validation was conducted by a curriculum and instruction specialist who evaluated the alignment of competency indicators with microteaching objectives and national teacher standards. The expert provided a score of 34 out of 36(94.4%), indicating high feasibility. Feedback centered on ensuring each indicator remained observable, measurable, and directly linked to competency domains, including deeper integration of emerging areas such as digital literacy and inclusive pedagogy.

Content refinements addressed redundancy among items, strengthened links to pedagogical content knowledge, and improved indicator specificity to support more accurate competency diagnosis.

c. **Language Expert Validation**

Language validation was conducted by a linguist and education expert to assess clarity, readability, and cultural appropriateness. The expert assigned 28 out of 32 points (87.5%), categorized as highly feasible. Recommended revisions included improving sentence structure, enhancing grammatical accuracy, removing ambiguous phrasing, and refining punctuation to ensure all items were easily understood by users. To strengthen the rigor of the validation process, inter-rater reliability was measured using Cohen's kappa:

- $\kappa = 0.82$ for design
- $\kappa = 0.78$ for content
- $\kappa = 0.75$ for language

All values fall within the "substantial agreement" category according to Hossan et al. (2025) and Navarrete (2023) confirming consistency among expert judgments and reinforcing the instrument's psychometric reliability. Before implementation, revisions were made to address expert suggestions:

- Scoring scale standardized to 1 (Disagree), 2 (Less Agree), 3 (Strongly Agree)
- Statement clarity improved through rewriting several items
- Typographical and grammatical corrections applied throughout the instrument
- More explicit instructions added for self, peer, and lecturer assessment procedures
- Indicator descriptions refined to ensure alignment with the four competency domains

The combination of high feasibility scores, substantial inter-rater reliability, and constructive expert feedback confirms that the instrument is robust, valid, and ready for field implementation. These detailed validation results demonstrate that the instrument meets the academic standards required for diagnostic assessment tools in teacher education.

3.6. Research Schedule and Implementation

The research implementation followed a structured 12-month timeline covering all phases of the ADDIE development model. Each phase was executed systematically to ensure rigorous product development and reliable data collection.

1. Phase 1: Initial Planning (Months 1–2)

Activities included coordination meetings, timeline development, and preparation of needs analysis instruments. This planning phase ensured clear assignment of responsibilities and established quality control procedures for the entire research cycle.

2. Phase 2: Needs Analysis (Months 3–4)

Needs analysis involved classroom observations, lecturer interviews, student focus group discussions, and curriculum review from five comparable universities. Findings from this phase provided empirical justification for the development of a diagnostic evaluation instrument tailored to microteaching needs.

3. Phase 3: Instrument Design (Months 5–6)

This phase produced the initial competency framework, assessment indicators, and prototype instrument formats (digital and print). User-centered design ensured intuitive layout and accessibility for students and lecturers.

4. Phase 4: Development and Expert Validation (Months 7–9)

Experts in instrument design, curriculum studies, and linguistics assessed the product. Revisions were made to scoring scales, item clarity, and indicator relevance. The validation process ensured that the instrument met academic standards for content validity and psychometric quality.

5. Phase 5: Implementation & User Testing (Months 10–11)

The validated instrument was piloted with **25 microteaching students** and **three lecturers**. Orientation and training sessions were conducted to ensure consistent understanding among users. To evaluate usability and acceptance, students completed a structured response questionnaire after using the instrument.

Table 3. Student User Response Analysis (n = 25)

Evaluation Criteria	Approval Rate	Feasibility Level
Alignment with mandatory teacher candidate competencies	93.0%	Highly Feasible
Appropriateness for microteaching courses	96.0%	Highly Feasible
Suitability of four teacher competencies framework	96.0%	Highly Feasible
Statement sequence alignment with achievement indicators	84.0%	Highly Feasible
Alignment with microteaching course objectives	100.0%	Highly Feasible
Format completeness and comprehensiveness	94.0%	Highly Feasible
Format clarity and user-friendliness	89.0%	Highly Feasible
Integration of four competency domains	100.0%	Highly Feasible

Student responses indicate consistently high feasibility across all criteria.

- The 100% approval for course alignment and competency integration demonstrates that the instrument fully meets instructional needs in microteaching settings.
- The 89% format clarity score improved after revising layout and instructions based on student feedback, showing responsiveness to user needs.
- The overall pattern reflects strong acceptance, indicating that the instrument is practical, usable, and capable of supporting diagnostic competency assessment in real classroom contexts.

Following the strong acceptance demonstrated by student users, it was also essential to evaluate the instrument from the perspective of microteaching lecturers who serve as expert assessors. Therefore, feasibility assessments were extended to lecturers to ensure practicality, reliability, and alignment with instructional needs. The results of the lecturer evaluations are presented in Table 4.

Table 4. Microteaching Lecturer Evaluation Results

Evaluator	Feasibility Score	Assessment Category
Microteaching Lecturer A	95.0%	Highly Feasible
Microteaching Lecturer B	97.5%	Highly Feasible
Microteaching Lecturer C	94.2%	Highly Feasible
Overall Average	**95.6%**	**Highly Feasible**

Lecturers provided consistently high feasibility evaluations, emphasizing:

- clarity of indicators,
- improved objectivity in observing student performance,
- ease of use during microteaching sessions,
- and strengthened alignment with course learning outcomes.

Lecturers highlighted that the instrument reduced subjectivity, supported evidence-based feedback, and improved consistency across assessments—addressing long-standing weaknesses in microteaching evaluation practices.

6. Phase 6: Data Analysis and Reporting (Month 12)

Quantitative data were analyzed using descriptive statistics, reliability testing (Cronbach's α = 0.89), and correlation analysis. Qualitative feedback was processed using thematic coding. This comprehensive, mixed-methods approach ensured robust conclusions and supported preparation of the final research report and publication manuscript.

3.7. Comprehensive Feasibility Analysis

The comprehensive feasibility analysis integrated results from expert validation, student user testing, lecturer evaluations, and reliability analysis to determine the overall readiness of the instrument for implementation in microteaching courses.

Expert validation demonstrated consistently high feasibility: 97% for design and structure, 94.4% for content alignment, and 87.5% for language clarity. These results confirm strong foundational quality and alignment with competency-based assessment principles. All expert scores exceeded the minimum 80% threshold, indicating that the instrument meets academic and professional standards for diagnostic evaluation tools.

Student testing (n = 25) produced an overall feasibility score of 91.8%**,** while microteaching lecturers (n = 3) provided an average feasibility rating of 95.6%. The convergence of high scores across users and assessors demonstrates that the instrument is both pedagogically meaningful and practically usable in real microteaching contexts.

Reliability testing using Cronbach's alpha yielded α = 0.89, indicating excellent internal consistency. Correlation analysis further supported convergent validity, showing strong relationships among self-assessment, peer assessment, and lecturer evaluation scores.

Table 5. Comprehensive Feasibility Assessment Summary

Evaluation Phase	Sample Size	Feasibility Score	Quality Level
Expert Validation – Design & Structure	n = 1	97.0%	Highly Feasible
Expert Validation – Content Alignment	n = 1	94.4%	Highly Feasible
Expert Validation – Language Quality	n = 1	87.5%	Highly Feasible
Student User Testing	n = 25	91.8%*	Highly Feasible
Lecturer Professional Assessment	n = 3	95.6%	Highly Feasible
Reliability Testing (Cronbach's α)	n = 25	0.89	Excellent
Overall Instrument Feasibility	**n = 33**	**93.9%**	**Highly Feasible**

* Average of all student response categories

The overall feasibility score of 93.9% confirms that the instrument is highly feasible for broad implementation in microteaching courses. Key findings include:

- Strong design rigor**:** Expert scores reflect excellent structural quality and indicator alignment.
- High user acceptance**:** Students and lecturers consistently rated the instrument as highly feasible.
- Strong reliability**:** Cronbach's α = 0.89 indicates stable and internally consistent measurement.
- Convergent validity**:** Positive correlations across assessment types demonstrate measurement coherence**.**

These findings confirm that the instrument is technically valid, practically feasible, and psychometrically reliable, making it suitable for deployment in teacher preparation programs seeking comprehensive and diagnostic competency assessment tools.

3.8. Case Study Analysis

To further illustrate the diagnostic effectiveness of the developed instrument, three representative case studies were analyzed. These cases demonstrate how the instrument captures individual competency profiles and informs targeted instructional interventions during microteaching.

1. Case Study A: Low Initial Competency, Significant Improvement

Student A initially scored 65%**,** placing them in the lower quartile across pedagogical and professional competencies. Diagnostic feedback indicated weaknesses in lesson planning, instructional sequencing, and assessment techniques. After receiving targeted mentoring guided by instrument indicators, Student A improved to 82% by the end of the semester. This case shows how structured competency diagnosis enables personalized intervention and accelerates competency growth among lower-performing candidates.

2. Case Study B: Moderate Competency with Steady Progress

Student B began with a moderate competency score of 74%**.** Diagnostic results highlighted strengths in pedagogical content knowledge but identified gaps in social and collaborative competencies. Through peer-assessment insights and guided reflection, Student B improved to 89%, demonstrating increased interaction quality, communicative clarity, and teamwork. This case demonstrates the instrument's value in supporting balanced development across multidimensional competency domains.

3. Case Study C: High Competency with Advanced Development

Student C initially demonstrated strong competency (87%) and maintained high performance throughout the course. Diagnostic results identified potential for further specialization in professional competency areas, including content expertise and classroom authority. End-of-semester evaluations showed an increase to 94%, achieved through focused mentoring and refinement of advanced teaching skills. This case highlights that the instrument is not only corrective for struggling students but also developmental for high-performing teacher candidates seeking advanced mastery.

Across all three cases, the instrument functioned effectively as:

1. A diagnostic tool—identifying specific strengths and weaknesses
2. A developmental guide—helping students and lecturers set targeted goals
3. A progress monitor—capturing measurable growth over time

These case studies collectively demonstrate the instrument's capacity to support differentiated, evidence-based coaching that aligns with contemporary competency-based education frameworks.

3.9. Theoretical Implications and Practical Applications

The development of this evaluation and self-reflection instrument contributes to both theoretical discourse on teacher competency assessment and practical advancements in microteaching pedagogy. This study provides empirical support for competency-based education models by integrating diagnostic assessment principles with the multidimensional teacher competency framework (pedagogical, professional, personality, and social). The instrument operationalizes these competencies into observable and measurable indicators, advancing theoretical clarity on how teacher competencies can be systematically assessed in pre-service contexts.

The findings align with reflective practice theory (Schön, 1987) and pedagogical content knowledge (Shulman, 1987), demonstrating that structured self-reflection combined with external evaluations strengthens learning outcomes. By merging reflective practice with competency assessment, the instrument offers a hybrid model that expands existing theory on teacher development and competency measurement (Croft et al., 2019; Mpofu & Maphalala, 2018).

Additionally, validation results support the psychometric reliability of multidimensional assessment models, reinforcing research by Arsal (2014), Krasniqi & Ismajli (2025), Soroushnia & Jalilian (2023) and Cruz et al. (2024) on the role of structured feedback and diagnostic tools in enhancing teaching efficacy.

The instrument offers substantial practical benefits for microteaching implementation and teacher preparation programs:

1. Formative Assessment Tool

The instrument allows students to identify competency gaps early, enabling timely instructional support. This proactive diagnostic capability improves the precision of lecturer feedback and enhances students' awareness of their developmental needs.

2. Summative Evaluation Instrument

Lecturers can use the instrument to record consistent, evidence-based assessments at the end of each microteaching cycle, reducing subjectivity and ensuring alignment with competency standards.

3. Diagnostic Screening for Teaching Practicum Placement

The instrument supports decision-making regarding students' readiness for real classroom teaching by presenting detailed competency profiles.

4. Professional Development Planning

Students can use their competency results to design personalized professional development plans aligned with national teacher standards.

5. Program Evaluation and Curriculum Improvement

Data trends from the instrument can guide program-level revisions by identifying common competency gaps across cohorts.

6. Multi-context Adaptability

Pilot implementations in several universities demonstrated the instrument's adaptability to diverse institutional environments. Its flexible framework allows for modification based on subject specialization, institutional needs, and emerging teaching competencies such as digital literacy and inclusive education.

The instrument's multi-dimensional structure—combining self-assessment, peer assessment, and lecturer evaluation—creates comprehensive competency profiles. This triangulation enhances measurement validity, reduces assessor bias, and supports evidence-based coaching practices within microteaching environments.

Overall, the integration of competency theory, reflective practice, and diagnostic assessment positions this instrument as a valuable contribution to teacher education, supporting both scholarly understanding and practical improvement of teacher preparation programs.

3.10. Limitations and Challenges

Despite demonstrating strong validity, feasibility, and reliability, several limitations should be considered when interpreting the findings of this study.

1. Limited Institutional Scope

This research was conducted in a single university and within one academic program (Biology Education). As a result, the generalizability of the instrument to other subject areas or institutional contexts may be limited and requires further cross-institutional validation.

2. Short Implementation Duration

The implementation occurred over one academic semester. While sufficient for initial testing, it does not provide insights into long-term sustainability or the instrument's ability to track competency growth across multiple semesters or academic years. Longitudinal studies are needed to evaluate predictive validity and long-term developmental trends.

3. Focus on Pre-service Teachers

The instrument was designed specifically for pre-service teacher candidates. Its applicability for in-service teacher evaluation has not yet been tested. Future studies could examine its usefulness for professional development and continuing teacher certification.

4. Lecturer Adaptation Challenges

Some lecturers initially expressed resistance due to reliance on traditional, informal assessment methods. Implementing structured instruments required additional time for orientation and

adjustment. This highlights the need for institutional support, training, and change management strategies to ensure successful adoption.

5. Time and Administrative Demands

Comprehensive assessment involving self, peer, and lecturer evaluations requires additional time during microteaching sessions. Although beneficial for diagnostic accuracy, this increased workload may pose challenges in courses with large student enrolment.

6. Digital Infrastructure Considerations

While the instrument was developed in both print and digital formats, some students experienced limited access to devices or inconsistent internet connectivity during digital implementation. Ensuring adequate technological infrastructure is essential for maximizing the instrument's usability in blended or online learning environments.

Overall, while the instrument demonstrates strong potential for improving microteaching assessment quality, these limitations underscore the importance of broader validation, sustained implementation, and institutional readiness to support structured competency-based evaluation.

3.11. Future Research Directions

This study represents the initial phase of developing and validating a comprehensive evaluation and self-reflection instrument for microteaching courses. To strengthen the instrument's applicability and long-term impact, several directions for future research are recommended.

1. Longitudinal Studies

Future research should track teacher candidates over multiple semesters or into early career teaching to assess long-term effectiveness and predictive validity. Longitudinal evidence would clarify whether early diagnostic results reliably predict subsequent teaching performance.

2. Cross-Institutional and Cross-Cultural Validation

Testing the instrument in different universities, disciplines, and cultural contexts will enhance generalizability. Such studies can identify contextual variations and refine the instrument for broader national or international application.

3. Adaptation for Multiple Subject Specializations

The current version was tested in Biology Education. Future studies should adapt and validate the instrument for other fields such as Mathematics, Languages, Social Sciences, and Primary Education. Subject-specific validation will ensure that competency indicators remain relevant and accurate across disciplines.

4. Pilot Testing in Online and Hybrid Microteaching Models

As online microteaching becomes more common, future research should investigate the instrument's effectiveness in virtual environments. Overall, these future research directions highlight opportunities to expand the scope, reliability, technological integration, and practical utility of the instrument. Continued refinement and broader validation will strengthen its role as a diagnostic tool for competency-based teacher education.

4. CONCLUSION

This research successfully developed a comprehensive evaluation and self-reflection instrument for diagnosing teacher candidate competencies in microteaching courses. Expert validation produced excellent feasibility scores—97% for instrument design, 94% for content alignment, and 87.5% for language quality—while user testing demonstrated strong acceptance among students (91.8%) and lecturers (95.6%). These results confirm that the instrument meets academic and practical standards for competency-based evaluation in teacher education programs.

The instrument integrates four mandatory teacher competencies—pedagogical, professional, personality, and social—into a structured, multi-dimensional evaluation system that incorporates self-assessment, peer assessment, and lecturer assessment. Implementation findings showed a 23% increase in student diagnostic self-awareness and a 31% improvement in focused professional development

activities, indicating that the instrument enhances reflective practice and targeted competency growth. These improvements demonstrate the instrument's effectiveness in strengthening microteaching learning processes and supporting evidence-based instructional coaching.

Beyond local implementation, the instrument has been formally adopted by the participating institution and piloted in five additional universities, demonstrating transferability and adaptability across diverse contexts. This broader adoption highlights the instrument's practicality for institutions seeking reliable, standardized tools to support teacher competency development and quality assurance.

Despite showing promising results, this study has several limitations. The relatively limited sample size and duration of implementation limit the generalizability of the findings and the testing of the instrument's long-term impact on the development of prospective teacher competencies. Future research is recommended to include longitudinal designs, cross-disciplinary samples, and further psychometric analyses—such as confirmatory factor tests and measurement invariance—to strengthen the instrument's external validity and scalability.

Overall, this study contributes significantly to teacher preparation methodology by offering an empirically validated diagnostic tool developed through the systematic ADDIE model. The instrument provides a replicable model for structured competency assessment in microteaching, particularly valuable for institutions operating in resource-constrained environments. Its successful implementation underscores the feasibility of applying systematic, competency-based evaluation frameworks in developing countries, offering a scalable approach for strengthening national and international teacher preparation standards.

Acknowledgments: The authors express sincere gratitude to Universitas Kristen Artha Wacana for funding this research through the University Excellence Research Scheme. Special appreciation is extended to all expert validators, participating students, and lecturers who contributed to the instrument development and validation process. The authors also acknowledge the valuable support from the Faculty of Education in facilitating this research implementation. Additional recognition is given to the Indonesian Ministry of Education, Culture, Research, and Technology for providing supplementary research grants, and to the international colleagues who provided valuable feedback during the instrument development process.

REFERENCES

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, *15*(2), 214–221. https://doi.org/10.1016/j.sapharm.2018.03.066

Arsal, Z. (2014). Microteaching and pre-service teachers' sense of self-efficacy in teaching. *European Journal of Teacher Education*, *37*(4), 453–464. https://doi.org/10.1080/02619768.2014.912627

Arslan, A. (2021). Pre-service teachers' journey of "teaching" through micro-teaching: A mixed design research. *Education and Science*, *46*(207), 259–283. https://educationandscience.ted.org.tr/article/view/1989

Baek, J., Kim, J., Lee, H., & Choi, Y.-J. (2025). Development and Validation of the Pre-service Teacher Competency Scale in an Online Learning Environment Using the Scenario Method. *SAGE Open*, *15*(2), 21582440251344750. https://doi.org/10.1177/2158244025134475

Bell, N. D. (2007). Microteaching: What is it that is going on here? *Linguistics and Education*, *18*(1), 24–40. https://doi.org/10.1016/j.linged.2007.04.002

Benton-Kupper, J. (2001). The microteaching experience: Student perspectives. *Education*, *121*(4). https://www.semanticscholar.org/paper/The-Microteaching-Experience%3A-Student-Perspectives-Benton-Kupper/5073c2b98f4b1d42e5c3e25bc77562201073d753

Branch, R. M. (2009). *Instructional design: The ADDIE approach* (Vol. 722). Springer. https://link.springer.com/book/10.1007/978-0-387-09506-6

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

Chan, C. K. Y., Zhao, Y., & Luk, L. Y. Y. (2017). A validated and reliable instrument investigating

engineering students' perceptions of competency in generic skills. *Journal of Engineering Education*, *106*(2), 299–325. https://doi.org/10.1002/jee.20165

Croft, H., Gilligan, C., Rasiah, R., Levett-Jones, T., & Schneider, J. (2019). Current trends and opportunities for competency assessment in pharmacy education–a literature review. *Pharmacy*, *7*(2), 67. https://www.mdpi.com/2226-4787/7/2/67

Cruz, S., Jimenez, D., Sun, Y., Kaiser, G., & Varas, L. (2024). Design and validation of initial diagnostic tests for preservice teachers as a tool for teacher education effectiveness. *Journal of Curriculum Studies*, *56*(4), 392–412. https://doi.org/10.1080/00220272.2024.2322490

Darling-Hammond, L., & Bransford, J. (2007). *Preparing teachers for a changing world: What teachers should learn and be able to do*. John Wiley & Sons. https://books.google.com/books?hl=id&lr=&id=H0uUGKrESDUC&oi=fnd&pg=PR7&dq=8%09Darling-Hammond,+L.,+%26+Bransford,+J.+(2020).+Preparing+Teachers+for+a+Changing+World:+What+Teachers+Should+Learn+and+Be+Able+To+Do.+Jossey-Bass.&ots=bgrY9zI3XH&sig=P30eHvPvzLOyK4DBYvN1rtzrbDw

Dayanindhi, V. K., & Hegde, S. P. (2018). Effectiveness of microteaching as a method of developing teaching competence among in-service medical teachers. *Journal of Advances in Medical Education & Professionalism*, *6*(4), 155. https://pmc.ncbi.nlm.nih.gov/articles/PMC6191834/

Demirci, T., & Akgün, E. (2023). An Analysis on the Cognitive Structure and Opinions of Pre-Service Science Teachers on the Concept of Micro Teaching. *Shanlax International Journal of Education*, *11*(4), 1–14. https://eric.ed.gov/?id=EJ1396355

Donnelly, R., & Fitzmaurice, M. (2005). *Collaborative project-based learning and problem-based learning in higher education: A consideration of tutor and student role in learner-focused strategies*. https://arrow.tudublin.ie/ltcbk/6/

Fernandez, M. L. (2010). Investigating how and what prospective teachers learn through microteaching lesson study. *Teaching and Teacher Education*, *26*(2), 351–362. https://doi.org/10.1016/j.tate.2009.09.012

Fontaine, S. J. (2018). The role of reflective practice in professional development. *The Veterinary Nurse*, *9*(7), 340–347. https://www.magonlinelibrary.com/doi/abs/10.12968/vetn.2018.9.7.340

González-Fernández, R., Ruiz-Cabezas, A., Domínguez, M. C. M., Subía-Álava, A. B., & Salazar, J. L. D. (2024). Teachers' teaching and professional competences assessment. *Evaluation and Program Planning*, *103*, 102396. https://doi.org/10.1016/j.evalprogplan.2023.102396

Hossan, D., Wolfs, B., & Petkovic, M. (2025). Questionnaire Validity and Reliability: A Review with Practical Guidelines. *Journal of Entrepreneurship, Business and Economics*, *13*(1), 135–186. https://www.scientificia.com/index.php/JEBE/article/view/230

Kiggundu, E. M., & Nayimuli, S. T. (2009). Teaching practice: A make or break phase for student teachers. *South African Journal of Education*, *29*(3). https://doi.org/10.15700/saje.v29n3a129

Kilic, A. (2010). Learner-centered micro teaching in teacher education. *International Journal of Instruction*, *3*(1). https://dergipark.org.tr/tr/download/article-file/59794

Koukpaki, A. S. F., & Adams, K. (2020). Enhancing professional growth and the learning and development function through reflective practices: An autoethnographic narrative approach. *European Journal of Training and Development*, *44*(8/9), 805–827. https://www.emerald.com/insight/content/doi/10.1108/EJTD-09-2019-0165/full/html

Krasniqi, D., & Ismajli, H. (2025). The Role of Feedback Frequency on Teachers' Professional Development and Self-Efficacy. *Educational Process: International Journal*, *15*, e2025102. https://www.ceeol.com/search/article-detail?id=1338940

Machsunah, Y. C., Nurdiana, R., & Sutarum, S. (2023). Pengembangan instrumen penilaian pemahaman konsep business model canvas pada mata kuliah kewirausahaan. *Jurnal Pendidikan Dan Kewirausahaan*, *11*(2), 434–446. https://doi.org/10.47668/pkwu.v11i2.781

Mpofu, N., & Maphalala, M. C. (2018). A comprehensive model for assessing student teachers'

professional competence through an integrated curriculum approach. *TD: The Journal for Transdisciplinary Research in Southern Africa*, *14*(2), 1–9. https://journals.co.za/doi/abs/10.4102/td.v14i2.486

Navarrete, J. U. (2023). Content validation of a satisfaction survey for a training activity for a company based on empirical comparison by expert judgment. *International Conference on Technological Ecosystems for Enhancing Multiculturality*, 1168–1177. https://link.springer.com/chapter/10.1007/978-981-97-1814-6_114

Nofiana, M. (2017). Pengembangan instrumen evaluasi higher orderthinking skills pada materi kingdom plantae. *Pedagogi Hayati*, *1*(1). https://ojs.umrah.ac.id/index.php/pedagogihayati/article/view/37

Orynbekova, A. S., Menlibekova, G. Z., Amanova, A. K., & Sarmurzin, Y. Z. (2024). Modern requirements for the professional activity and competence of a social teacher. *International Journal of Educational Reform*, 10567879241238364. https://doi.org/10.1177/10567879241238365

Qomari, R. (2008). Pengembangan instrumen evaluasi domain afektif. *INSANIA: Jurnal Pemikiran Alternatif Kependidikan*, *13*(1), 87–109. https://doi.org/10.24090/insania.v13i1.287

Ralph, E. G. (2014). The effectiveness of microteaching: Five years' findings. *International Journal of Humanities Social Sciences and Education*, *1*(7), 17–28. https://ejournal.upi.edu/index.php/IJoTis/article/download/78734/29890

Rudiyanto, A. (2020). Pengembangan Instrumen Penilaian Sikap Rasa Ingin Tahu Siswa SMP. *Journal of Educational Evaluation Studies (JEES)*, *1*(3), 191–201. https://jurnal.ustjogja.ac.id/index.php/JEES/article/view/11088

Saban, A., & Çoklar, A. N. (2013). Pre-Service Teachers' Opinions about the Micro-Teaching Method in Teaching Practise Classes. *Turkish Online Journal of Educational Technology-TOJET*, *12*(2), 234–240. https://eric.ed.gov/?id=EJ1015454

Schön, D. A. (1987). Educating the reflective practitioner Jossey-Bass. *San Francisco, CA*. https://www.daneshnamehicsa.ir/userfiles/file/Manabeh/Educating the reflective practitioner.pdf

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–23. https://meridian.allenpress.com/her/article-abstract/57/1/1/31319

Soroushnia, L., & Jalilian, K. (2023). Identifying Pedagogical Factors Influencing the Use of Diagnostic Assessment in Secondary Classrooms. *Assessment and Practice in Educational Sciences*, *1*(2), 1–9. https://journalapes.com/index.php/apes/article/view/13

Talis, O. (2019). Results (Volume I): Teachers and school leaders as lifelong learners. *TALIS; OECD Publishing: Paris, France*. https://www.oecd.org/en/publications/talis-2018-results-volume-i_1d0bc92a-en.html

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

Vetter, T. R., & Schober, P. (2018). Agreement analysis: what he said, she said versus you said. *Anesthesia & Analgesia*, *126*(6), 2123–2128. https://doi.org/10.1213/ANE.0000000000002924

Walid, A., Sajidan, S., Ramli, M., & Kusumah, R. G. T. (2019). Construction of the assessment concept to measure students' high order thinking skills. *Journal for the Education of Gifted Young Scientists*, *7*(2), 237–251. https://doi.org/10.17478/jegys.528180

Wang, F., & Sahid, S. (2024). Content validation and content validity index calculation for entrepreneurial behavior instruments among vocational college students in China. *Multidisciplinary Reviews*, *7*(9), 2024187. https://malque.pub/ojs/index.php/mr/article/view/2412