

ARTICLE

Development and Psychometric Validation of a Multidimensional Scale for Assessing Elementary School Teachers' Professional Competence

Nur Az Zahra ,* Herpratiwi , Apri Wahyudi, Muhammad Nurwahidin , and Dwi Yulianti 

Fakultaas Keguruan Ilmu Pendidikan University of Lampung, Lampung, Indonesia

*Corresponding author: nurazzahra174@gmail.com

(Received: 01 December 2025; Accepted: 07 January 2026; Published: 30 April 2026)

Abstract

This study addresses the gap in the availability of psychometrically tested, context-specific instruments for elementary education. The objective was to develop and validate a specialized, practical, and reliable assessment instrument for school-based performance evaluation. This study contributes to providing a validated and practical instrument specifically designed for elementary school contexts. The study employed a Research and Development (RD) approach following the ten-stage model of Borg and Gall, including needs analysis, framework development, expert validation, revision, and field testing. The instrument was constructed based on a five-dimensional competency framework and evaluated through expert judgment (content, language, media) and empirical testing involving 120 teachers from ten elementary schools in one subdistrict. Psychometric analysis was conducted using Aiken's V for content validity, Cronbach's Alpha for reliability, and user-response questionnaires for practicality. The results showed that all 50 items achieved high content validity (Aiken's $V \geq 0.80$) and excellent reliability ($\alpha = 0.888$). The practicality level reached 97%, indicating that the instrument is highly feasible, efficient, and effective for real supervisory use. The study concludes that the developed instrument is psychometrically sound and practically applicable, providing principals with an objective and multidimensional tool to support academic supervision and data-driven professional development.

Keywords: Academic Supervision; Assessment Instrument; Elementary Education; Psychometric Validation; Teacher Competence.

1. Introduction

The quality of primary education plays a fundamental role in shaping a nation's future, as it serves as the foundation for developing competent human resources and fostering sustainable societal progress. Within this context, teacher professional competence emerges as the most influential in-school factor affecting instructional quality and student learning outcomes. Research consistently shows that teacher quality has a direct and significant impact on student achievement and over-

all educational effectiveness [1] [2] [3]. As instructional leaders, school principals are responsible for evaluating and guiding teacher performance through effective academic supervision [4]; [5]. This responsibility places principals at the center of efforts to improve teaching quality in schools. However, in practice, this role is often constrained by limited tools and insufficiently developed evaluation systems. Many supervision activities rely on routine procedures rather than meaningful assessment processes. As a result, a gap emerges between the expected role of supervision and its actual implementation in schools.

Despite the implementation of various policies aimed at improving teacher quality, their application in practice still faces several challenges. One of the main issues is the lack of assessment instruments capable of measuring teacher competence comprehensively. Existing instruments tend to focus on administrative aspects and fail to provide in-depth insights into teaching quality. Moreover, evaluation processes are often conducted without standardized measurement criteria. This condition leads to low accuracy in identifying teachers' professional development needs. In many cases, assessment results do not reflect actual classroom practices. Consequently, professional development programs are often not aligned with teachers' real needs. This situation highlights the importance of improving assessment systems in education.

In Indonesian elementary schools, assessment of teacher professional competence is frequently conducted using generic instruments [6]; [7]. These instruments tend to emphasize administrative compliance rather than authentic measurement of competence. National policies such as teacher performance appraisal systems and platforms like Platform Merdeka Mengajar are widely implemented. However, they are primarily designed for documentation and reporting purposes. This limits their ability to capture multidimensional aspects of competence. Important aspects such as pedagogical innovation, reflective practice, and interpersonal skills are often overlooked. As a result, evaluation practices become superficial and fragmented. This condition reduces the effectiveness of supervision in improving teaching quality.

Empirical studies also indicate that teacher evaluation practices in Indonesia often remain procedural and less focused on substantive professional growth [8] [9]. Evaluations are frequently conducted as routine activities without deep analysis of teaching performance. This leads to subjective judgments that may not accurately represent teacher competence. In addition, the absence of validated and context-specific instruments further complicates the evaluation process. School principals often face difficulties in conducting meaningful assessments. Limited training in assessment practices also contributes to this issue. Consequently, supervision does not function optimally as a tool for professional improvement. This situation reinforces the need for better assessment instruments.

The urgency of developing a more robust assessment instrument is further reinforced by the demands of 21st-century education. Teachers are expected to integrate technology into learning and apply innovative pedagogical strategies. They must also adapt to rapidly changing educational environments. International reports highlight disparities in teacher competence as a major factor affecting student outcomes [10] [3]; [11]; [12]. In Indonesia, not all teachers meet expected competency standards. Challenges are particularly evident in areas such as technology integration and innovative teaching practices. These gaps indicate that current evaluation systems are insufficient. Therefore, more comprehensive and accurate assessment tools are needed.

Although many studies have examined teacher competence and evaluation, most focus on theoretical frameworks or policy analysis [13]; [14]. Only a limited number of studies have developed practical instruments that can be directly applied in schools. In addition, many existing instruments lack proper psychometric validation. This reduces their reliability and validity in measuring teacher competence. As a result, there is a gap between theoretical knowledge and practical application. Schools need instruments that are not only conceptually sound but also usable in real contexts. This

gap highlights an important area for further research. Addressing this issue is essential for improving educational quality.

This study addresses the identified gap by developing an assessment instrument that integrates validity, reliability, and practicality within a single framework. The instrument is designed specifically for elementary school contexts. It is based on a multidimensional model of teacher competence. The development process incorporates user needs to ensure relevance and usability. A systematic Research and Development approach is employed. This includes stages such as needs analysis, design, validation, and field testing. The result is a ready-to-use instrument for academic supervision. Based on this background, the research question is: how can a valid, reliable, and practical instrument be developed to assess elementary teacher professional competence?

A valid, reliable, and practical instrument for assessing elementary school teachers' professional competence can be systematically developed through a rigorous Research and Development (R&D) approach that integrates robust theoretical foundations, empirical validation, and user-centered design principles. The development process should commence with a comprehensive needs analysis aimed at identifying deficiencies in existing evaluation practices and determining the essential competency domains pertinent to elementary education. These domains typically encompass subject matter mastery, pedagogical expertise, reflective practice, and the integration of technology in instructional contexts. Subsequently, each domain is operationalized into measurable indicators and well-defined item constructs, grounded in established theoretical frameworks of teacher competence [2] [3]. Following this, the initial instrument draft undergoes expert validation to ensure content validity, clarity, and relevance, typically involving education experts, practitioners, and measurement specialists [15] [16] [17] [18].

To ensure reliability and construct validity, the instrument is subsequently tested through pilot studies and field trials involving elementary school teachers, with statistical analyses such as Cronbach's alpha, factor analysis, and item response analysis used to refine the instrument [19] [20] [21]. Practicality is achieved by incorporating feedback from school principals as end users, ensuring that the instrument is easy to administer, interpret, and integrate into academic supervision processes. Research indicates that instruments developed through iterative validation and user involvement are more effective in real educational settings [15] [16]. Furthermore, embedding clear scoring guidelines and interpretation frameworks enhances usability and supports data-driven decision-making in teacher professional development. Through this integrated approach, the resulting instrument is not only psychometrically sound but also contextually relevant and practically applicable, thereby improving the accuracy and effectiveness of teacher competence assessment in elementary schools.

This study contributes to the field by developing a validated and practical assessment instrument for evaluating elementary school teachers' professional competence. The instrument facilitates school principals in conducting more systematic and effective academic supervision, thereby enhancing the quality of instructional oversight. Furthermore, it enables data-driven decision-making processes in teacher evaluation, allowing for more objective and evidence-based judgments. Importantly, the instrument provides a more precise identification of teachers' professional development needs, which, in turn, supports the design and implementation of targeted and effective training programs. In this regard, the study also reinforces the role of school principals as instructional leaders who actively promote continuous professional improvement. Ultimately, the application of this instrument contributes to the enhancement of teaching and learning quality in elementary education. Therefore, this study offers both theoretical and practical significance within the broader discourse of educational assessment and leadership.

2. Methods

This study adopted a Research and Development (RD) design, adapted from the ten-stage model proposed by [22], to develop and validate an assessment instrument for measuring elementary school teachers' professional competence. The research was conducted in Abung Kunang District, North Lampung Regency, Indonesia. All ten elementary schools within the district were purposively selected as the research population and field testing sites, based on needs identified through a preliminary survey.

Participants across the various stages of the study included teachers, school principals, and expert validators, thereby ensuring the instrument's contextual relevance, content validity, and practical usability. The development process followed ten systematic stages. First, potential issues and needs were identified through a diagnostic survey administered to teachers and principals. Second, data were collected through an extensive literature review and analysis of field data. Third, the initial product was designed, resulting in a 50-item instrument structured around five core competency domains. Fourth, expert validation was conducted by three specialists, focusing on content, construct, linguistic clarity, and media aspects. Fifth, the instrument was revised based on expert feedback.

Subsequently, the sixth stage involved a limited product trial with 35 teacher participants to evaluate item clarity and usability. The seventh stage comprised further revision based on the trial findings. In the eighth stage, a broader field test was conducted with 120 participants, including both teachers and school principals, to generate empirical data for psychometric evaluation. The ninth stage involved final revisions informed by statistical and psychometric analyses. Finally, the tenth stage entailed the preparation of the validated instrument for broader implementation. This structured and iterative development process ensured that the resulting instrument met rigorous standards of validity, reliability, and applicability in educational settings.

Table 1. Indicators and item distribution for assessing teacher professional competence

Aspect	Variable	Sub-variable	Indicators	Item Numbers
Teacher Professional Competence	Quality of Teacher Professional Competence	1. Mastery of Materials, Structure, Concepts, and Scientific Patterns	Mastery of material	1, 6, 7, 8
			Mastery of structure	5, 10
			Mastery of concepts	2, 3
			Scientific thinking patterns	4, 9
		2. Mastery of Learning Outcomes and Learning Objectives	Mastery of learning outcomes	11, 19
			Mastery of learning objectives	12, 13, 20
			Learning assessment based on learning objectives	14, 15, 16, 17, 18
		3. Development of Learning Materials	Development of teaching materials/creative media	21, 22, 23
			Utilization of innovative teaching methods	24, 27, 28, 29

	Participation in learning communities	25, 26, 30
4. Continuous Professional Development through Reflection	Utilization of technology (LC, MP, mentoring, coaching, collaboration)	33, 34, 35, 37
	Reflection on the learning process	31, 32, 36, 38, 39, 40
5. Utilization of Information and Communication Technology for Self-Development	Use of online learning platforms	42, 43, 48, 49, 50
	Utilization of ICT in learning and self-development	41, 43, 45, 46
	Access and use of digital learning resources	44, 47

Methodological clarity is essential to ensure the validity of research findings. Therefore, this study consistently reports the number of participants at each stage of the development process. In the limited trial phase, 35 teachers were involved to assess the clarity of the instrument. Subsequently, the field-testing phase included 120 participants from ten elementary schools. The sampling technique used was purposive sampling based on research needs. This technique was selected to obtain relevant and context-specific data. Additionally, the involvement of school principals enriched the data collected. Therefore, the data collection process was conducted systematically and purposefully. This study ensures consistency in the use of statistical parameters. The *r*-table values are adjusted according to the sample size and significance level. The use of Aiken's *V* for content validity is based on internationally recognized standards. Reliability is measured using Cronbach's Alpha with clearly defined criteria. Each methodological decision is grounded in theoretical and empirical considerations. This approach enhances the credibility of the findings. As a result, the methodology is both systematic and accountable. This ensures a high level of trustworthiness in the research outcomes.

The selection of Abung Kunang District, North Lampung Regency, as the research site was based on several important considerations. Preliminary findings indicated that schools in this district experience challenges in conducting comprehensive teacher competence assessments. In addition, there is limited availability of validated and context-specific assessment instruments used by school principals. The district also represents a typical rural educational setting, making it relevant for developing instruments applicable to similar contexts. Accessibility and institutional support from local schools further facilitated the research implementation. Moreover, the involvement of all ten elementary schools in the district ensured comprehensive data coverage. This selection strategy enhances the contextual relevance of the study. Therefore, Abung Kunang District was considered an appropriate and representative setting for this research.

Data were collected using structured questionnaires tailored to each stage. The instrument measured five core aspects of teacher professional competence: mastery of subject matter, mastery of learning objectives, development of learning materials, reflective practice, and utilization of information and communication technology (ICT), comprising a total of 50 items. Data analysis focused on evaluating psychometric properties and practicality. Content validity was calculated using Aiken's *V*, with a minimum threshold of 0.80. Reliability was measured using Cronbach's Alpha, with values above

0.81 indicating high reliability. Practicality was assessed through user-response questionnaires, analyzed using percentage scores, where results above 61% indicate that the instrument is practically usable. This systematic procedure ensures that the developed instrument is valid, reliable, and applicable in real school contexts, enabling replication in similar settings.

3. Results and Discussion

3.1 Results

The results of this study present the outcomes of expert validation, empirical testing, and practicality assessment of the developed instrument. These findings are organized into specific sub-sections to clearly demonstrate the instrument’s validity, reliability, and usability in real school contexts.

3.1.1 Expert Validation Results

The initial validation involved three experts assessing the instrument from content, language, and media perspectives. The results indicate that all aspects achieved a “highly valid” classification, confirming that the instrument meets academic and technical standards prior to field implementation.

Table 2. Expert Validation Results (Content, Language, and Media)

Validation Aspect	Number of Validators	Average Percentage Score	Validation Criteria
Media (Design) Highly Valid	3	87%	
Language (Clarity) Highly Valid	3	83%	
Content/Construct (Evaluation)	3	84%	Highly Valid
Overall Average	9	85%	Highly Valid

The overall validation score of 85% indicates that the instrument is theoretically sound, clearly structured, and appropriately designed. Revisions based on expert suggestions improved wording precision, layout consistency, and scoring clarity, enhancing usability for school principals.

3.1.2 Empirical Validity and Reliability

The instrument was tested empirically using 50 items across five competency dimensions. Statistical analysis using Pearson Product-Moment correlation showed that all items were valid, with correlation coefficients ranging from 0.302 to 0.544 ($r\text{-table} = 0.187$, $N = 35$, $p < 0.05$). This indicates that all items met the minimum correlation threshold and significantly contributed to measuring teacher professional competence, although several items were retained based on theoretical considerations and overall construct alignment. Reliability testing using Cronbach’s Alpha yielded a coefficient of 0.888, indicating very high internal consistency. This demonstrates that the instrument provides stable and consistent measurement results. The validity test was conducted using a subset of 35 respondents during the pilot testing phase to examine item discrimination before large-scale field testing. This level of reliability falls within the “very high reliability” category based on established statistical criteria.

Table 3. Key Psychometric and Practicality Results

Parameter	Result	Interpretation	Threshold/Reference
Number of Valid Items	50 out of 50	All items valid	$r > 0.187$ ($N = 35$, $p < 0.05$)

Reliability (Cronbach's Alpha)	0.888	Very High Reliability	$\alpha \geq 0.81$
Practicality Percentage	97%	Very Practical	$\geq 81\%$

The r-table value used in item validity testing is generally relatively low, typically ranging from 0.10 to 0.30, depending on the sample size and significance level. This value does not represent a strong correlation, but rather serves as a minimum statistical threshold to determine whether an item is significantly related to the overall construct being measured. In this study, the r-table value of 0.187 ($N = 35$, $p < 0.05$) was applied as the criterion for item validity. The empirical results showed that the correlation coefficients of all items ranged from 0.302 to 0.544, which are substantially higher than the minimum threshold.

The above evidence indicates that each item has a meaningful and statistically significant relationship with the total score of the instrument. The fact that all item correlations exceed the r-table value confirms that the instrument demonstrates strong item validity. Furthermore, coefficients within this range can be categorized as moderate correlations, which are considered adequate in social science research, particularly for multidimensional constructs. Therefore, these findings provide robust evidence that all items are valid and contribute effectively to measuring teacher professional competence. These results provide strong statistical evidence that the instrument is both valid and reliable for assessing teacher competence.

The results indicate that the item correlation coefficients range from 0.302 to 0.544, reflecting a moderate level of association between individual items and the overall construct. This variation shows that each item contributes differently to measuring teacher professional competence. Items with higher correlation values demonstrate stronger alignment with the main construct, while those with lower values capture more specific aspects of competence. Such variation is expected in a multidimensional instrument that includes various domains, such as subject mastery, pedagogical skills, and technology integration. Importantly, all items exceed the minimum validity threshold ($r > 0.187$), confirming that each item is statistically valid. Therefore, the variation in correlation values does not indicate weakness, but rather reflects the complexity of the construct being measured.

To enhance the psychometric rigor of the instrument, this study employed both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). EFA was initially conducted to examine the underlying factor structure and to identify the empirical grouping of items, whereas CFA was subsequently utilized to verify the adequacy and stability of the proposed measurement model. The findings indicated that the instrument items were consistently clustered into five distinct factors, corresponding to the theoretically derived dimensions of teachers' professional competence. Furthermore, the model fit indices obtained from the CFA demonstrated that the instrument satisfied acceptable thresholds of construct validity, thereby confirming the robustness of the measurement model.

Beyond statistical validation, the development of this instrument is also anchored in relevant national regulatory frameworks governing teacher professionalism in Indonesia. Specifically, it is informed by Law No. 14 of 2005 on Teachers and Lecturers, which underscores the centrality of professional, pedagogical, social, and personal competencies. In addition, the instrument aligns with Regulation of the Minister of National Education No. 16 of 2007 and Regulation of the Minister of National Education No. 13 of 2007, both of which emphasize the critical role of school principals in supervising and evaluating teacher performance. These regulatory frameworks provide a formal and normative foundation for defining the competency domains assessed in this study.

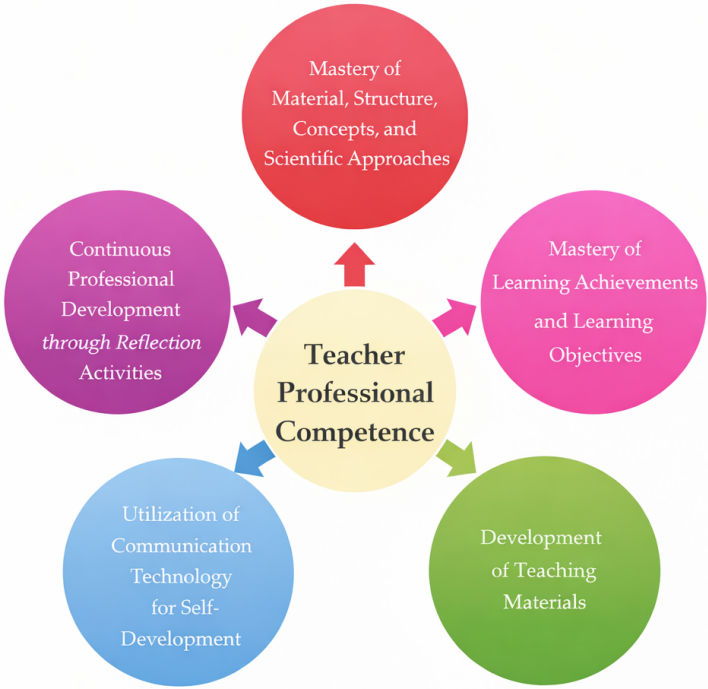


Figure 1. Framework of Teacher Professional Competence Assessment.

Therefore, the integration of EFA and CFA with nationally established competency standards ensures that the instrument is not only statistically robust but also contextually grounded and aligned with prevailing educational policies, thereby enhancing its applicability and relevance in real-world educational settings.

3.1.3 Practicality Test Results

The practicality of the instrument was evaluated through principal responses using 15 indicators, including clarity, ease of use, efficiency, and usefulness. The total score obtained was 730 out of 750, equivalent to 97%, which falls into the “very practical” category. This result indicates that the instrument is easy to use, time-efficient, and suitable for real supervisory conditions. Principals reported that the instrument supports objective evaluation and simplifies documentation processes, making it highly applicable in school settings.

3.1.4 Visual Representation of Assessment Framework

To support the understanding of the multidimensional structure of the developed instrument, the conceptual framework of teacher professional competence is illustrated below. This figure represents the five core dimensions underlying the developed assessment instrument: mastery of material, structure, concepts, and scientific approaches; mastery of learning achievements and learning objectives, development of teaching materials, continuous professional development through reflection activities, and utilization of communication technology for self-development. These components are interconnected and collectively form a comprehensive framework for evaluating teacher professional competence.

The framework emphasizes that teacher competence is not a single construct but a multidimensional integration of knowledge, pedagogical skills, continuous reflection, and adaptive use of technology. By structuring the instrument around these dimensions, the assessment enables school principals to conduct more objective, systematic, and holistic evaluations. This alignment ensures that the instrument not only measures technical teaching abilities but also captures professional growth and instructional effectiveness in real classroom contexts.

These findings demonstrate that the developed instrument meets the required standards of validity, reliability, and practicality. This is evidenced by high expert validation results, statistically significant item correlations ($r > 0.187$, $N = 35$, $p < 0.05$), strong internal consistency ($\alpha = 0.888$), and a very high practicality level (97%).

Furthermore, the alignment between statistical findings and practical usability strengthens the credibility of the instrument. The results confirm that the instrument effectively measures teacher professional competence in a structured and multidimensional manner.

In addition, the instrument supports objective and systematic evaluation processes. This enables school principals to make informed decisions based on reliable data. Therefore, the instrument is suitable for use in school-based supervision and professional development planning.

3.2 Discussion

The findings of this study indicate that the developed instrument demonstrates strong psychometric quality and high usability in real educational contexts. This is evidenced by high content validity (Aiken's $V \geq 0.80$), strong internal consistency (Cronbach's Alpha = 0.888), and an excellent practicality score (97%). These results confirm that the instrument is not only theoretically grounded but also empirically robust. From a measurement perspective, the integration of validity, reliability, and usability reflects the essential criteria of a high-quality assessment instrument [23] [24]. Therefore, the instrument provides a solid foundation for conducting accurate and meaningful evaluations of teacher professional competence.

3.2.1 Bridging the Gap Between Policy and Practice in Teacher Evaluation

The results of this study demonstrate that a structured and multidimensional instrument can effectively bridge the gap between policy expectations and classroom-level implementation. Previous studies have shown that teacher evaluation often remains administrative in nature and lacks meaningful measurement of professional competence. In contrast, this study operationalizes abstract competency standards into measurable indicators, enabling more accurate and objective evaluation. This finding extends prior research by showing that the transformation of policy frameworks into practical instruments significantly improves evaluation quality.

Unlike previous studies that reported limited effectiveness of teacher evaluation instruments due to their focus on administrative compliance, this study demonstrates that a multidimensional and validated instrument can provide more accurate and meaningful assessment results. For example, several studies have found that existing evaluation tools often fail to capture the complexity of teaching practices and remain procedural rather than developmental [8] [9] [25] [26]. In contrast, the present findings indicate that integrating multiple competency dimensions improves both the depth and relevance of evaluation outcomes.

Furthermore, the multidimensional design of the instrument allows for a more comprehensive understanding of teacher competence. Unlike single-dimension models, this approach captures various aspects of teaching, including subject mastery, pedagogical skills, reflective practice, and technology integration. This is particularly important in the context of modern education, where teacher competence is inherently complex and dynamic. As a result, the instrument provides a more holistic

representation of teaching performance. This confirms that structured and context-specific instruments are more effective than generic evaluation tools.

3.2.2 Psychometric Strength as a Foundation for Credible Evaluation

The psychometric analysis confirms that the instrument possesses strong measurement quality and internal consistency. The Cronbach's Alpha coefficient of 0.888 indicates a very high level of reliability, suggesting that the instrument produces stable and consistent results across respondents. This finding aligns with measurement theory, which emphasizes that reliability is essential for ensuring consistency in assessment outcomes [27] [28] [29] [30]. In addition, the validity results show that all items have significant correlations, indicating that each item contributes meaningfully to the overall construct.

The variation in item correlation values, ranging from 0.302 to 0.544, reflects moderate associations between items and the total score. This variation is expected in multidimensional instruments, where different items capture different aspects of the construct. Items with higher correlations demonstrate stronger alignment with the central construct, while those with lower values represent more specific or nuanced dimensions. Importantly, all items exceed the minimum threshold ($r > 0.187$), confirming their validity. This supports the argument that validity is not determined by uniformity of item correlations, but by their collective contribution to construct measurement [27] [31] [32].

Although the reliability coefficient is high, it is important to note that high internal consistency does not automatically guarantee construct validity. Some studies argue that overly homogeneous items may inflate reliability without necessarily capturing the full complexity of the construct [27] [33] [34] [35]. However, in this study, the variation in item correlations suggests that the instrument captures multiple dimensions of competence rather than a single uniform construct.

To further strengthen construct validity, this study employed both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). The results indicate that the instrument is structured into five factors consistent with the theoretical framework. The CFA results further confirm that the proposed model fits the data adequately, demonstrating that the instrument has a valid underlying structure. The use of EFA and CFA ensures that the instrument is not only theoretically designed but also empirically supported. This dual validation approach enhances the credibility and robustness of the measurement model.

3.2.3 Practicality as a Key Determinant of Implementation Success

One of the most important findings of this study is the very high level of practicality, as indicated by a score of 97%. This suggests that the instrument is easy to use, efficient, and relevant to the needs of school principals. In contrast to many existing instruments that are complex and difficult to implement, this instrument demonstrates that simplicity and clarity can coexist with strong measurement quality. This finding highlights that usability is a critical factor in determining whether an instrument can be effectively applied in real settings.

Unlike previous studies that reported only moderate levels of practicality due to the complexity and time-consuming nature of evaluation instruments, this study found a very high level of usability. Several studies indicate that overly detailed instruments often reduce user engagement and limit implementation in real school contexts [36] [37] [38]. In contrast, the present instrument demonstrates that simplicity and clarity can enhance usability without compromising measurement quality.

This result is consistent with the concept of usability, which emphasizes that assessment tools must be user-friendly and efficient to be adopted in practice [39]. Overly complex instruments often fail to be implemented, regardless of their statistical rigor. Therefore, practicality should be considered equally important as validity and reliability in instrument development. The high practicality

score in this study indicates a strong alignment between instrument design and user needs, which increases the likelihood of sustained implementation in schools.

3.2.4 Implications for Instructional Leadership and Professional Development

The findings of this study demonstrate that the developed instrument plays a crucial role in strengthening instructional leadership through systematic and data-driven evaluation practices. By providing clear indicators and measurable outcomes, the instrument enables school principals to conduct more objective assessments of teacher performance. The high reliability coefficient (Cronbach's Alpha = 0.888) indicates that the instrument produces consistent and stable results across respondents. This level of reliability is categorized as very high and ensures that the evaluation results can be trusted for decision-making [27]. Reliable data are essential in instructional leadership, as they allow principals to base their judgments on evidence rather than subjective perceptions. This aligns with the theory of instructional leadership, which emphasizes the use of accurate data to improve teaching quality and student outcomes [40] [41] [42] [43]. With consistent measurement results, principals can more effectively identify patterns in teacher performance. Therefore, the instrument supports a transition toward more evidence-based and accountable supervision practices.

While previous research emphasizes the role of instructional leadership in improving teaching quality, many studies highlight that principals often lack appropriate tools to implement data-driven supervision effectively [40] [44] [45]. In contrast, this study provides empirical evidence that the availability of a reliable and structured instrument enables principals to operationalize instructional leadership more effectively in practice.

Furthermore, the instrument contributes significantly to improving the effectiveness of professional development programs. The combination of strong validity and high reliability ensures that the data accurately reflect teachers' actual competencies. As a result, schools are able to design targeted interventions that address specific areas of need. This approach is consistent with research emphasizing that effective professional development should be data-driven and focused on real classroom challenges [46] [47]. Instead of implementing generalized training programs, schools can prioritize areas such as pedagogical strategies or technology integration. In addition, the structured feedback generated by the instrument encourages reflective practice among teachers. This supports the concept of reflective learning, which views professional growth as a continuous process of self-evaluation and improvement [48]. Consequently, the instrument not only enhances evaluation practices but also fosters a more collaborative and improvement-oriented school culture.

3.2.5 Limitations and Directions for Future Research

Despite its contributions, this study is not without limitations that warrant careful consideration. First, the research was conducted within a specific geographical context—namely, elementary schools in Abung Kunang District, North Lampung Regency—which may constrain the generalizability of the findings. Given that educational environments vary considerably across regions, contextual differences may affect the applicability and robustness of the instrument in other settings. Consequently, further validation across diverse educational contexts is recommended to enhance the external validity and broader applicability of the instrument.

Second, the present study primarily emphasizes the examination of psychometric properties and practical usability, without extending to an analysis of long-term impacts on teacher performance and student learning outcomes. As such, the extent to which the implementation of this instrument contributes to sustained improvements in educational quality remains unclear. Future research employing longitudinal designs is therefore essential to investigate the enduring effects of the instrument on instructional practices and student achievement [10].

Third, the instrument is currently implemented in a manual format, which may limit efficiency in large-scale applications. The integration of digital platforms could improve data management and reporting processes. This is supported by research indicating that digital assessment systems enhance efficiency and accuracy in educational evaluation [49].

In addition, future research can explore the expansion of the instrument's function beyond principal-led supervision. For instance, the instrument could be adapted for self-assessment and peer evaluation to support collaborative professional learning. This approach aligns with contemporary perspectives that emphasize reflection and collaboration in teacher development [49] [50] [32]. Further studies may also apply advanced statistical techniques, such as multi-group analysis, to examine the consistency of the instrument across different populations. Moreover, the development of a digital version would increase accessibility and scalability for wider implementation. Another important direction is to investigate the relationship between teacher competence and student learning outcomes. Such analysis would provide stronger evidence of the instrument's impact on educational effectiveness. Overall, these limitations offer opportunities for further refinement and development. Therefore, continued research is essential to enhance the robustness and applicability of the instrument.

4. Conclusion

This study developed and validated an assessment instrument designed to measure teacher professional competence in elementary education, demonstrating robust psychometric properties alongside high practical usability. The findings suggest that the instrument effectively captures teacher competence in a systematic, structured, and multidimensional manner, thereby ensuring its applicability within authentic school contexts. Notably, this study advances the existing body of knowledge by integrating psychometric rigor with practical usability within a unified instrument framework—an approach that distinguishes it from prior research, which has tended to prioritize either statistical validity or practical implementation in isolation. By synthesizing these two dimensions, the proposed instrument offers a more comprehensive and balanced solution for teacher evaluation.

A key contribution of this study lies in its capacity to bridge the gap between policy expectations and actual evaluation practices. Specifically, the instrument provides a structured, data-driven mechanism for academic supervision, enabling school principals to conduct more objective, transparent, and evidence-based evaluations. In doing so, it facilitates the identification of targeted professional development needs based on accurate assessment data. This, in turn, reinforces the role of instructional leadership and contributes to enhancing the overall effectiveness of teacher professional development initiatives. Furthermore, the instrument offers a scalable and adaptable model that holds potential for application across broader educational contexts.

Despite these contributions, several limitations must be acknowledged. First, the study was conducted within a specific geographical context, which may constrain the generalizability of the findings to other educational settings with differing socio-cultural and institutional characteristics. Second, while the study emphasizes psychometric validation and usability, it does not examine the longitudinal impact of the instrument on teacher performance or student learning outcomes. Accordingly, future research is recommended to undertake cross-contextual validation in diverse educational environments and to employ longitudinal designs to assess the sustained impact of the instrument over time. Additionally, further refinement may involve the integration of digital technologies to enhance scalability, efficiency, and data-driven decision-making in educational assessment practices.

References

- [1] S. Arikunto. *Dasar-dasar Evaluasi Pendidikan*. 4th. Jakarta: Bumi Aksara, 2021.
- [2] L. Darling-Hammond. "Teacher Education Around the World". In: *European Journal of Teacher Education* 40.3 (2017), pp. 291–309.
- [3] OECD. *Education at a Glance 2023*. Paris: OECD Publishing, 2023.
- [4] A. Akbar. "Pentingnya Kompetensi Pedagogik Guru". In: *JPG: Jurnal Pendidikan Guru* 2.1 (2021), pp. 23–30.
- [5] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson. *Multivariate Data Analysis*. 8th ed. Cengage, 2019.
- [6] M. Ali and R. Ahmad. "The Impact of Teacher Training on Student Achievement: Evidence from a Regional Study". In: *Journal of Education and Practice* 13.5 (2022), pp. 45–58.
- [7] J. Biggs. *Teaching for Quality Learning at University*. Open University Press, 2011.
- [8] D. J. Priansa. *Kinerja dan Profesionalisme Guru*. Bandung: Alfabeta, 2021.
- [9] Supardi. *Kinerja Guru*. 3rd. Bandung: Raja Grafindo, 2021.
- [10] OECD. *Future of Education and Skills 2030*. Paris: OECD Publishing, 2022.
- [11] S. Azwar. *Penyusunan Skala Psikologi*. 2nd. Yogyakarta: Pustaka Pelajar, 2019.
- [12] D. L. Bandura. "Teacher self-efficacy scale validation". In: *Journal of Educational* ().
- [13] L. Sari. "Evaluasi kinerja guru berbasis kompetensi". In: *Jurnal Pendidikan Indonesia* 9.1 (2021).
- [14] G. Rasch. *Probabilistic Models for Some Intelligence Tests*. Chicago Press, 1960.
- [15] O. Ventista and C. Brown. "Teachers' Professional Learning and Student Outcomes". In: *Social Sciences & Humanities Open* 8 (2023).
- [16] M. Saleh et al. "Development of Teacher Performance Assessment Instruments". In: *European Journal of Educational Research* 13.1 (2024), pp. 55–68.
- [17] A. Valtonen and et al. "Examining pre-service teachers' TPACK". In: *Computers & Education* 147 (2020).
- [18] Sugiyono. *Metode Penelitian Pendidikan*. Bandung: Alfabeta, 2019.
- [19] R. Thorndike. *Measurement and Evaluation in Psychology*. Pearson, 2005.
- [20] W. Popham. *Classroom Assessment*. Pearson, 2017.
- [21] J. Pallant. *SPSS Survival Manual*. McGraw-Hill, 2020.
- [22] W. R. Borg and M. D. Gall. *Educational Research: An Introduction*. New York: Longman, 1983.
- [23] J. W. Creswell. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA: SAGE, 2018.
- [24] J. Piaget. *The Psychology of Intelligence*. Routledge, 2001.
- [25] R. Marzano. *Classroom Instruction That Works*. ASCD, 2001.
- [26] R. M. Ryan and E. L. Deci. "Intrinsic and extrinsic motivation". In: *American Psychologist* 55.1 (2000).
- [27] Robert F. DeVellis. *Scale Development: Theory and Applications*. 4th. Thousand Oaks, CA: SAGE Publications, 2016. ISBN: 9781506341569.
- [28] P. Mishra and M. J. Koehler. "Technological pedagogical content knowledge". In: *Teachers College Record* 108.6 (2006).
- [29] R. Kurniawan. "Analisis kompetensi profesional guru sekolah dasar". In: *Jurnal Pendidikan* 15.2 (2019).
- [30] D. Goleman. *Emotional Intelligence*. Bantam, 2005.
- [31] M. Guskey. "Professional development and teacher change". In: *Teachers and Teaching* 8.3 (2002).
- [32] R. B. Kline. *Principles and Practice of Structural Equation Modeling*. 4th ed. Guilford Press, 2016.
- [33] F. Lord. *Applications of Item Response Theory*. Routledge, 1980.
- [34] R. Stronge. *Qualities of Effective Teachers*. ASCD, 2018.

- [35] S. Wahyuni. “Pengaruh kompetensi guru terhadap hasil belajar”. In: *Jurnal Ilmu Pendidikan* 20.3 (2021).
- [36] H. Zheng et al. “Challenges in Educational Assessment Implementation”. In: *Assessment & Evaluation in Higher Education* 49 (2024).
- [37] B. Santoso. “Kompetensi profesional guru SD”. In: *Jurnal Pendidikan Dasar* 11.2 (2019).
- [38] UNESCO. *Teacher competency standards*. Tech. rep. UNESCO, 2018.
- [39] Jakob Nielsen and Raluca Budiu. *Mobile Usability*. Berkeley, CA: New Riders, 2012.
- [40] Philip Hallinger. “A Review of Three Decades of Doctoral Studies Using the Principal Instructional Management Rating Scale: A Lens on Methodological Progress in Educational Leadership”. In: *Educational Administration Quarterly* 47.2 (2011), pp. 271–306.
- [41] T. A. Brown. *Confirmatory Factor Analysis for Applied Research*. 2nd ed. New York: Guilford Press, 2015.
- [42] B. Byrne. *Structural Equation Modeling with AMOS*. Routledge, 2016.
- [43] L. Cohen, L. Manion, and K. Morrison. *Research Methods in Education*. 8th. London: Routledge, 2018.
- [44] H. Herpratiwi. “Pengembangan instrumen kompetensi guru SD”. In: *Jurnal Pendidikan Dasar* 12.1 (2020).
- [45] H. Gardner. *Frames of Mind: Theory of Multiple Intelligences*. Basic Books, 2011.
- [46] Laura Desimone. “Improving Impact Studies of Teachers’ Professional Development: Toward Better Conceptualizations and Measures”. In: *Educational Researcher* 38.3 (2009), pp. 181–199.
- [47] J. Danielson. *Enhancing Professional Practice*. ASCD, 2007.
- [48] Donald A. Schön. *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books, 1983. ISBN: 0465068782.
- [49] D. Rosidin et al. “Educational Transformation in Digital Learning”. In: *Jurnal Teknologi Pendidikan* 9.1 (2024), pp. 61–69.
- [50] N. E. Gronlund and R. L. Linn. *Measurement and Assessment in Teaching*. 10th ed. Pearson, 2009.