

Assessment Instrument Development Competence among Pre-Service Language Teachers: Advancing Reflective Practice through Contextualized and Progressive Learning

Ariyana^{1✉}, Soleh Ibrahim², Nakamura Kaito³, Kim Ji-woo Kirana⁴

^{1,2}Faculty of Education, Universitas Muhammadiyah Tangerang, Indonesia

³Faculty of Education, Osaka Kyoiku University, Japan

⁴College of Education, Kyungpook National University, South Korea

DOI: 10.23917/ijolae.v8i3.20344

Received: August 3rd, 2026. Revised: September 20th, 2026. Accepted: September 23rd, 2026

Available Online: September 27th, 2026. Published Regularly: September, 2026

Abstract

Assessment competence is critical for pre-service language teachers to construct valid, contextualized, and meaningful assessments that support advanced and responsive learning. This research assessed pre-service language teachers' perception of assessment competence, the dimensions that comprise the construct and the role of reflection in bridging assessment theory and practice. Sequential explanatory mixed-method approach was used in this research in which 25 sixth-semester Language Education major students participated. Data on pre-service language teachers' assessment competence were collected using a 10-item likert-scale questionnaire and analyzed using exploratory factor analysis and qualitative data on reflection in relation to the three dimensions were obtained from 10 interviews and analysis of 25 assessment rubrics. Although high levels of perceived competence were reported, substantial challenges remained in contextualizing assessment decisions and translating assessment knowledge into authentic learning situations. Three preliminary dimensions of assessment competence were identified: Evaluation Knowledge, Instrument Design Skills and Contextualization of Language Skills Assessment. It was also found out that there was a disparity between assessment knowledge and its implementation. Reflective practice emerged as a crucial mechanism for integrating these dimensions through contextualized, progressive, and practice-oriented learning experiences.

Keywords: assessment competence, pre-service teachers, contextualized learning, progressive learning, meaningful learning experience, advanced learning.



This is an open access article under the CC-BY license

✉Corresponding Author:

Ariyana, Faculty of Education, Universitas Muhammadiyah Tangerang, Indonesia

Email: ariyana.mpd@gmail.com

1. Introduction

Language proficiency can be referred to as the main competence in educational settings especially when talking about the language teachers who not only have to master language competencies but also learn how to teach, assess and evaluate learners' language proficiency. Competence in assessment is one of the crucial elements of

the professionalism of the language teachers as the results of the assessment provide evidence on diagnosing students' proficiency, revealing their needs and designing proper interventions (Haerazi, H. et al., 2023). Within the framework of the Merdeka Curriculum in Indonesia, the importance of assessment has increased because learning is expected to respond to diverse learner char-

acteristics, needs, and competency levels, requiring more advanced, adaptive, and context-sensitive approaches to learning and assessment. The use of diagnostic assessment allows teachers to identify students' initial conditions and use assessment evidence to design differentiated, responsive, and context-sensitive learning experiences while simultaneously developing their assessment competence (Ochoa et al., 2024). Therefore, the ability to develop assessment instruments is one of the important dimensions of the pre-service language teachers' professional preparation.

The recent literature reveals growing attention to the development of systematic, context-appropriate, and learning-oriented language assessment instruments that can support advanced assessment practices and more meaningful learning experiences. For instance, portfolio assessment instruments are claimed to increase students' motivation and contribute to speaking competency development as the learners are able to demonstrate their progress through multiple pieces of evidence (Kurniawan et al., 2025). Moreover, the use of the scoring rubrics is found to facilitate the assessment of writing competence and provide systematic and constructive feedback to the students helping them to improve their performance further on (Lustyantie et al., 2025). The ongoing tendency towards assessment which is based on the principles of validity, reliability and transparency is reflected in the development of Higher Order Thinking Skills (HOTS) assessment items in Indonesian language courses at senior high school and vocational school levels (Sutami et al., 2020). Besides, the research on the evaluation of thematic learning at elementary schools shows that the development of the appropriate assessment rubrics helps to increase teachers' understanding of the assessment process and

leads to better results (Nurhayati et al., 2019).

However, the availability of assessment frameworks and instruments alone is insufficient to ensure successful implementation, because translating assessment principles into context-sensitive educational practice remains a challenging aspect of contemporary teacher learning. The available empirical data suggests that many educators tend to use relatively traditional assessment practices, namely, the written tests while using the standardized rubrics and other systematic procedures rather rarely (Hasim et al., 2025). This condition reveals the existence of the gap between the knowledge of the principles of assessment possessed by teachers and their capability to translate this knowledge in the context-appropriate assessment practices. The similar situation can be observed in relation to the evaluation instruments in electronic textbooks. Although these instruments demonstrate relatively good quality, the effectiveness of their implementation is limited by the lack of other learning and assessment resources (Listeani et al., 2021). All these findings reveal the fact that assessment competence cannot be considered as teachers' knowledge of the principles of assessment and assessment instruments. Assessment competence is therefore better understood as the ability to interpret assessment principles critically, apply them to authentic instructional settings, make context-sensitive assessment decisions, and reflect systematically on the consequences of these decisions for students' learning. Such competence represents an advanced form of professional learning that integrates knowledge, application, judgment, and reflection.

In this regard, the reflective practice is one of the important yet understudied dimensions of the assessment competence of

pre-service language teachers. The previous research paid attention to assessment instrument development and teacher reflection separately; however, there was relatively little research conducted in relation to the reflective practice of pre-service language teachers during the process of development of assessment instruments (Jamil, 2025). The existing literature revealed the study of general evaluation practices of pre-service teachers or explored the ways of how they could apply reflective learning in assessment instrument development (Mursyida et al., 2020). At the same time, the insufficient preparation and limited knowledge of reflective practice were claimed to be the factors which may prevent pre-service teachers from using the appropriate assessment practices in teaching contexts (Wahyuni & Arieffiani, 2022). Thus, a research gap remains regarding the dimensions of assessment instrument development competence and the role of reflective practice in bridging theoretical assessment knowledge with context-sensitive decision-making, authentic practice, and progressive learning experiences in teacher education. Filling this gap is important in the context of teacher education since the assessment practices of pre-service teachers are not only dependent on their theoretical knowledge but also on their interpretation of this knowledge and ability to apply it in real-life situations.

The present study is based on two perspectives: the Sociocultural Theory and the Teacher Cognition framework by Borg. According to the Sociocultural Theory, the learning and development are achieved through social interaction and participation in the activities which are situated in culture (Vygotsky, 1980). In this regard, the concept of the Zone of Proximal Development (ZPD) is particularly important for understanding the development of professional competence

of pre-service teachers through the interaction with other people, such as their peers, lecturers and mentors. From this perspective, the reflective practice can act as a mediational mechanism which converts the externally mediated experiences into professional knowledge and decision-making of the pre-service teachers. The Teacher Cognition framework by Borg, which complements this perspective, highlights the importance of teachers' knowledge, beliefs and thinking in relation to their instructional decisions and classroom practices (Borg, 2023). This perspective can be used in order to understand why the pre-service teachers are able to demonstrate adequate theoretical knowledge of the assessment while experiencing difficulties in implementing this knowledge in practice. Thus, integrating these two perspectives allows considering assessment instrument development as a socially-mediated and cognitively-reflective process.

Having considered the abovementioned, the present study aims to fill the research gap through the examination of the assessment instrument development competence of pre-service language teachers and the underlying dimensions. Thus, the study aims to: (1) analyze pre-service language teachers' perceptions of their competence in developing assessment instruments; (2) reveal the dimensional constructs which underlie this competence through exploratory factor analysis; and (3) explore the possible role of reflective practice in bridging the gap between the assessment theory and its implementation in practice. Through the identification of the latent dimensions of the assessment instrument development competence, the study provides empirical evidence of how pre-service teachers perceive and implement the assessment-related knowledge. Besides, exploring the

reflective practice as the bridge between theoretical knowledge and its application in practice allows providing the more comprehensive view on the development of assessment competence. The findings are expected to be used for the further refinement of teacher education programs through the design of the assessment-related learning experiences which would integrate theoretical knowledge, context-based application, social interaction and reflection.

2. Method

This study used the design of a sequential explanatory mixed methods where there were two consecutive phases, namely a quantitative phase followed by a qualitative phase. In the first phase, quantitative data were collected and analyzed in order to identify patterns and the dimensions of the pre-service language teachers' perceived competence in developing assessment instruments. Then, the subsequent qualitative phase was performed to further analyze participants' reflective experience and quality of assessment instruments that they had developed. A sequential explanatory design was chosen as it allows integration of patterns in quantitative data with qualitative evidence so that both the dimensional structure of the competence and reflective processes can be explained better (Likita & et al., 2024).

a. Participants and Sampling

The population included all students in the sixth semester at a university in Tangerang City who had joined the Indonesian Language Education Study Program and had already taken the course of Language Learning Evaluation. The participants were selected by purposive sampling method based on some predetermined inclusion criteria. Namely, the

participants (1) had taken the Language Learning Evaluation course; (2) had completed or were in the process of completing an assessment instrument development task including four language skills, that is, listening, speaking, reading, and writing; and (3) agreed to take part actively in all parts of the study, including the completion of the questionnaire, interviews and submission of documents.

Twenty-five participants met the inclusion criteria and took part in the quantitative phase. Among the 25 participants, 10 of them were chosen for the subsequent qualitative phase. In fact, the qualitative participants were purposively selected based on the response pattern of participants in the quantitative phase in order to represent the diversity of their perceptions of assessment competence, especially the high, moderate, and low perceptions of competence.

b. Research Instruments and Data Sources

Data were collected from three sources, namely a self-reflection questionnaire, semi-structured interviews and documentary evidence of assessment instruments development. The first instrument was the four-point Likert-scale self-reflection questionnaire which consisted of 10 statements aimed to measure the pre-service language teachers' perceptions of their competence in developing assessment instruments. Four-point scale was used to force the participants to indicate their definite opinion about each statement.

The validity of the questionnaire was determined via its evaluation by two specialists in educational assessment. The experts were asked to evaluate the relevance and clarity of the questionnaire items in terms of the competence in the development

of assessment instrument. After the expert evaluation was done, it was concluded that the questionnaire could be used in the quantitative phase.

Reliability of the instrument was measured by calculating the Cronbach's alpha which resulted in $\alpha = 0.87$. This figure shows that there is high internal consistency among the items.

In the second phase of data collection, the participants were interviewed semi-structuredly to receive the information regarding their experiences in developing assessment instrument, their comprehension of assessment principles, difficulties, decision-making and reflections in their development of the assessment instrument.

Each of these interviews lasted 30-45 minutes, was recorded audiovisually after obtaining the consent of participants and was transcribed.

Thirdly, the participants' assessment products were used as documentary evidence. In particular, the rubrics of assessment, test items and observation protocols were developed by the participants for the four language skills. This approach is similar to the study conducted previously and examining the quality of items and their alignment with the cognitive taxonomies in the Minimum Competency Assessment context (Aini & Mukhlis, 2022).

c. Data Collection Procedure

The procedure of data collection included three stages one after another. At the first stage, participants completed the 10-item self-reflection questionnaire. The quantitative data collected at this stage were analyzed in order to find out patterns in participants' perceptions and to find out the dimensions that require qualitative investigation.

At the second stage, 10 participants were chosen based on their response pattern in the questionnaire. The semi-structured interviews were conducted to get detailed information about the experiences, reasoning, difficulties and reflective processes of participants during the development of the assessment instruments. The interview questions were designed to elaborate the issues which emerged at the first stage of data collection.

At the third stage, the participants' assessment products were collected and analyzed as documentary evidence. The assessment instruments were analyzed in order to see how well the participants' perceptions were reflected in their assessment products.

d. Quantitative Data Analysis

The quantitative data were analyzed with the help of exploratory factor analysis with principal component analysis and varimax rotation to explore the dimensional structure of the 10-item questionnaire (Marlina et al., 2023). PCA was used to explore the patterns of variance in the observed variables, and the varimax rotation was applied in order to make the structure of factors more interpretable.

First, the suitability of the dataset for factor analysis was tested with the help of the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. The factor retention and interpretation were based on the component structure, factor loadings and conceptual coherence of the factors.

However, the sample size is relatively small, namely 25 participants, and this sample is below commonly recommended thresholds for factor extraction. As a result, the EFA findings should be treated as preliminary evidence about the dimensional

structure of perceived assessment instrument development competence rather than as the confirmation of factor model. Consequently, the findings of the quantitative phase were not treated in isolation. They were triangulated with the qualitative data collected during the interviews and documentary analysis.

e. Qualitative Data Analysis

The qualitative data obtained from the interviews and the documentary evidence were analyzed thematically using the framework proposed by Braun and Clarke (2022) and including six phases. In particular, the analysis was performed using six interconnected phases: familiarization with the data, generation of initial codes, searching for the potential themes, reviewing themes, defining and naming themes, and production of the final analytical account.

The interviews were repeatedly read to become familiar with the participants' experience and reflective account. The relevant parts of the transcript were coded according to issues related to the assessment knowledge, practice of instrument development, contextual decision-making, perceptions of difficulties, feedback, revision and reflective processes. After that, the codes were grouped in broader themes which reflected the recurring patterns in participants.

Simultaneously, the documentary evidence was analyzed in order to see how the themes obtained during the interview analysis were reflected in the assessment instruments of participants.

The qualitative analysis served two purposes. First, it helped in explaining the themes and dimensions which were found out in the process of quantitative analysis. Second, it helped in comparing the

participants' perceptions and assessment practice.

f. Integration of Quantitative and Qualitative Findings

The integration took place at the interpretation stage through the joint display matrix. This method allowed to bring together the quantitative factor structure, representative response pattern of the participants, qualitative themes and documentary evidence. The researchers could examine the patterns of convergence, complementarity and divergence between the perceptions and assessment practices.

The special attention was paid to cases when the patterns of high and low perceived competence were shown in quantitative findings but required additional explanation in qualitative findings and documentary evidence. Namely, the high level of perceived competence was explained in light of participants' accounts and assessment products. In addition, the low-level dimensions were explored through the difficulties, misconceptions and contextual constraints experienced by participants during the instrument development.

The process of integration made it possible to move beyond the presentation of isolated findings to the interpretation of assessment instrument development competence. Thus, qualitative findings were not used as the supplement to the quantitative but as the interpretative tool for the preliminary factor structure.

3. Result and Discussion

Results and discussion of the present study are described based on the analysis of quantitative and qualitative data obtained with the help of the sequential explanatory mixed-methods design. The quantitative results first describe the perceptions of pre-

service language teachers about their competence in developing assessment instruments and then describe the dimensional structure of this competence through exploratory factor analysis. The findings of the study are further supplemented with qualitative results that show how perceptions of competence were manifested in terms of the practices used during the process of assessment. Integration of the above-mentioned data sources gives an opportunity to consider the relationships between the areas of assessment competence such as knowledge of assessment, skills in designing assessment instruments, understanding of the context and reflective practice (Rahmawati, 2023). In order to analyze the results and discuss them in depth, the following interrelated sections are considered: (1) perceptions of assessment instrument development competence among pre-service language teachers; (2) dimensional structure of assessment instrument

development competence; (3) reflective practice as the tool for integration of assessment competence and application of theory to practice; and (4) pedagogical and theoretical implications for language teacher education.

a. Pre-Service Teachers' Perceived Assessment Instrument Development Competence

The quantitative findings revealed that pre-service language teachers generally considered themselves to be competent in the development of language assessment instruments. As illustrated in Table 1, the answers to all ten statements tended to cluster predominantly within the Strongly Agree and Agree categories. The three indicators of assessment competence Statements 1, 5, and 6 were associated with full agreement, as none of the respondents disagreed with them.

Table 1. Pre-Service Teachers' Self-Perceptions of Assessment Instrument Development Competence

Statement	Strongly Agree	Agree	Disagree	Strongly Disagree
1. Understanding the purpose of evaluating listening and speaking skills	10 (40%)	15 (60%)	0 (0%)	0 (0%)
2. Determining appropriate assessment instrument types	5 (20%)	19 (76%)	1 (4%)	0 (0%)
3. Developing instruments based on clear learning indicators	4 (16%)	20 (80%)	1 (4%)	0 (0%)
4. Matching instrument forms with language-skill characteristics	11 (44%)	13 (52%)	1 (4%)	0 (0%)
5. Formulating items or criteria in clear language	8 (32%)	17 (68%)	0 (0%)	0 (0%)
6. Understanding differences between reading and writing skills	4 (16%)	21 (84%)	0 (0%)	0 (0%)
7. Selecting instruments for reading and writing objectives	3 (12%)	20 (80%)	2 (8%)	0 (0%)
8. Accurately measuring reading comprehension	4 (16%)	16 (64%)	5 (20%)	0 (0%)
9. Ensuring relevance to reading and writing objectives	2 (8%)	19 (76%)	4 (16%)	0 (0%)
10. Developing scoring rubrics reflecting writing-skill aspects	6 (24%)	17 (68%)	2 (8%)	0 (0%)

In turn, the distribution depicted in Table 1 shows that the participants felt very strong competence regarding assessment basics. Indeed, Statement 1, relating to the purposes of assessing listening and speaking skills, and Statement 6, regarding

distinguishing between reading and writing skills, both received full agreement. Similarly, Statement 5, pertaining to formulating assessment criteria in a clear and student-friendly manner, received full agreement. The above pattern suggests that

the participants have established a relatively high level of conceptual competence in assessment principles, gained during their formal training. Importantly, the findings are significant due to the necessity to ensure

alignment between desired learning outcomes, language skills, and assessment procedures (Lustyantie et al., 2025; Rahmawati, 2023).

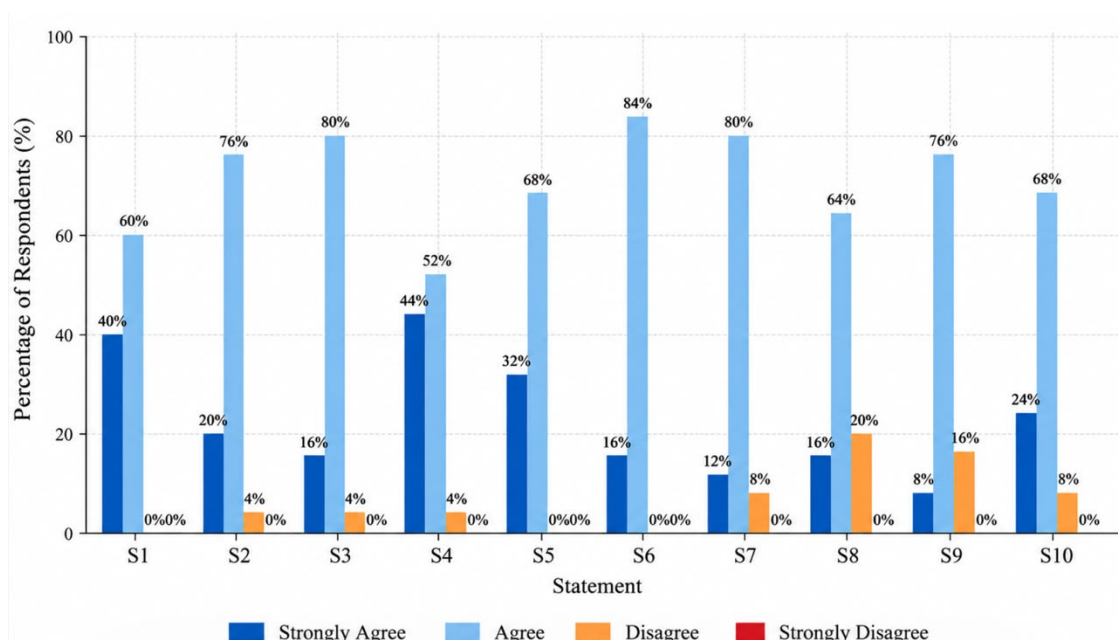


Figure 1. Distribution of Pre-Service Teachers' Responses Across the Ten Assessment Competence Indicators

The above figure provides a more visual representation of the distribution of responses and highlights the general positive pattern across the ten indicators. The absence of Strongly Disagree responses across all items suggests that the participants did not feel any lack of competence whatsoever (Hani, 2024). However, the figure also

reveals the presence of significant variation in the Disagree category proportions. Statement 8 and Statement 9 are notable due to receiving the highest disagreement proportions, thus showing that the participants felt higher uncertainty regarding the issue of instrument construction than about basics of the process.

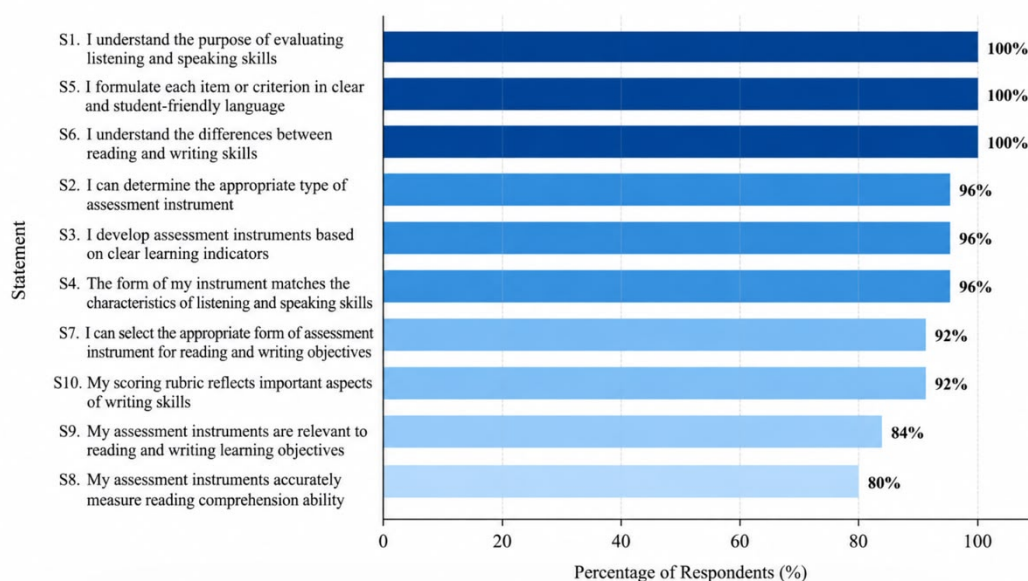


Figure 2. Relative Perceived Competence Across Assessment Indicators

As seen in Figure 2, the highest level of perceived competence was achieved in Statements 1, 5, and 6, all of which received 100% agreement. The next set of high competence consisted of Statements 2, 3, and 4, all of which got 96% agreement, while Statements 7 and 10 reached 92% agreement. In turn, Statement 8 received 80% agreement and Statement 9 received 84% agreement. Thus, participants reported greater confidence in basic assessment principles and instrument-construction procedures than in ensuring that assessment instruments accurately measure specific language abilities and remain sufficiently relevant to particular learning objectives and contexts. This difference highlights a key challenge in developing context-sensitive and progressive assessment practices. The above finding is relevant in terms of the evidence that educators can rely on relatively conventional assessment methods without applying systematic assessment criteria or rubrics (Hasim et al., 2025).

However, the distinction between self-perceptions of competence and real competence becomes even clearer after comparing the quantitative responses with the qualitative and documentary findings. In terms of the latter, it can be stated that most of the participants understood the purposes and principles of assessment and understood the need for alignment between assessment and learning objectives (Harizon et al., 2023). However, some of the participants admitted that their understanding remained largely declarative, namely that they knew what should be done but could not apply the principles adequately to the development of an assessment instrument. This finding indicates that high self-perceived competence cannot be regarded as sufficient evidence of developed professional competence, particularly when advanced learning requires teachers to translate conceptual knowledge into context-sensitive decisions and authentic assessment practices.

Table 3. Comparison of Perceived Competence and Assessment Product Quality

Evidence Source	Indicator	Empirical Result	Interpretation
Questionnaire	Statements 1, 5, and 6	100% Agree/Strongly Agree	Very high perceived competence in foundational assessment knowledge
Questionnaire	Statement 8	80% Agree/Strongly Agree; 20% Disagree	Relatively lower confidence in measuring reading comprehension accurately
Questionnaire	Statement 9	84% Agree/Strongly Agree; 16% Disagree	Relatively lower confidence in contextual relevance
Assessment documents	Rubrics containing essential assessment aspects	18/25 (72%)	Most rubrics contained basic assessment components
Assessment documents	Rubrics with clear and differentiated descriptors	10/25 (40%)	Performance-level differentiation remained limited
Interviews	Understanding of assessment principles	Generally strong	Declarative knowledge was relatively well established
Interviews	Application of assessment principles	Varied	Evidence of a theory–practice discrepancy

As seen from Table 3, the above issue provides an important explanatory dimension to the quantitative findings. While the questionnaire shows high levels of perceived competence, the analysis of 25 assessment rubrics reveals that only 72% contained the aspects required of assessment, namely content, organization, spelling, and language use, while only 40% provided clear and differentiated descriptors for each performance level. However, the discrepancy between the self-perception and product quality is not necessarily indicative of incorrect self-reports (Giraldo & Murcia, 2019). Rather, it suggests that self-perception of competence and demonstrated performance are two related yet different issues.

Importantly, the above discrepancy becomes even more significant due to the fact that documents represent a more direct manifestation of the competence of the participants than self-perception alone. An assessment rubric may contain correct categories but may still fail to differentiate between levels of performance properly (Kazykhankyzy & Alagözlü, 2019). Thus, the presence of content, organization,

spelling, and language-use criteria does not automatically imply that the rubric provides proper guidance for scoring. The above finding is supported by the previous studies emphasizing the importance of scoring rubrics for measuring the writing ability and providing constructive feedback (Lustyantie et al., 2025). At the same time, it is consistent with the findings that the development and application of proper rubrics increase teachers' understanding of assessment processes and outcomes (Nurhayati et al., 2019).

The above information helps to explain why the discrepancy occurred. The participants generally showed knowledge of the terminology and principles related to assessment but some of them faced problems with the justification of their choice of an assessment format, with measuring the intended construct with an instrument, and with differentiating between the assessment criteria for various performance levels. In other words, the competence of the participants was developed mainly at the level of declarative knowledge. The above issue is important because professional competence involves not only what teachers

know but also how they translate knowledge, beliefs, and reasoning into pedagogical decisions. This translation is central to progressive education, in which learning is oriented toward active application, contextual judgment, reflection, and continuous improvement (Borg, 2023).

Additionally, the findings show that assessment instrument development competence should be considered as an evolving professional competence rather than purely technical skill. The strong responses to the indicators regarding the basics of assessment show that formal training can form the conceptual basis for assessment literacy (Rahmawati, 2023). At the same time, the low responses to the reading-comprehension measurement and contextual relevance indicators, as well as the low proportion of rubrics containing differentiated descriptors, reveal the need for further opportunities of guided practice, feedback, and reflection. The above interpretation is consistent with the existing evidence on the effectiveness of assessment training and rubric development in terms of increasing the understanding of assessment processes and improving their application (Nurhayati et al., 2019).

Thus, the first-stage findings establish an important distinction between perceived assessment competence and demonstrated assessment competence. Generally, the participants considered themselves to be capable of developing assessment instruments, especially in terms of the purposes of assessment, the language-skill differences, and the basics of design (Borromeo Ferri, R., Greefrath, 2026). However, the qualitative and documentary findings show that this competence was not developed equally in the context of application and performance-level specification. Consequently, the results of

the first stage of research provide an empirical base for exploring the dimensions of assessment instrument development competence during the exploratory factor analysis. According to the logic of the mixed-methods design, the qualitative findings do not replace the quantitative findings but explain the variation within them.

b. Dimensional Structure of Assessment Instrument Development Competence

After the completion of descriptive analysis of the pre-service teachers' perceptions, the quantitative analysis of the data was expanded through exploratory factor analysis (EFA). In this regard, PCA was applied with Varimax rotation to examine 10 questionnaire items associated with assessment competence. The goal of the exploratory factor analysis was to detect any patterns in the correlations between the variables and see whether the construct of interest can be represented by a limited number of dimensions (Marlina et al., 2023). In this respect, the obtained statistical evidence was interpreted with consideration of qualitative evidence in order not to separate the structure of the construct from pre-service teachers' actual practices.

Preliminary analysis of the suitability of items for factorization has yielded a KMO value of 0.827, and Bartlett's test of sphericity was statistically significant, $\chi^2 = 168.45$, $df = 45$, $p < 0.001$. The results demonstrate the presence of sufficient patterning in the correlation structure of questionnaire items in order to perform exploratory analysis. The KMO value is considerably higher than the commonly used threshold of 0.60 and hence indicates acceptable sampling adequacy for the exploratory analysis of the present data set. Meanwhile, significant Bartlett's test

additionally indicates that the correlation matrix is statistically different from the identity matrix (Kazykhankyzy & Alagözlü, 2019). Therefore, the questionnaire items exhibit interrelatedness that allows their structure exploration.

Table 2. Results of Exploratory Factor Analysis with Varimax Rotation

Item	Item Description	Factor 1	Factor 2	Factor 3
1	Understanding evaluation purposes for listening and speaking	0.82	0.12	0.09
2	Determining appropriate assessment instrument types	0.11	0.79	0.18
3	Designing instruments based on learning indicators	0.20	0.85	0.10
4	Aligning instrument forms with language-skill characteristics	0.14	0.76	0.22
5	Ensuring language clarity in assessment instruments	0.09	0.81	0.13
6	Understanding reading and writing skill types	0.88	0.07	0.05
7	Selecting instrument forms for reading and writing	0.76	0.21	0.19
8	Developing instruments that measure reading comprehension	0.13	0.24	0.84
9	Ensuring instrument relevance to learning objectives	0.18	0.15	0.81
10	Developing rubrics reflecting writing-skill aspects	0.72	0.17	0.20

Table 2 reveals the clear three-factor structure. Each item is strongly correlated with one factor, while weaker correlations are observed with other factors. The loadings of items on the factors range from 0.72 to 0.88 for the dominant factor associations, thus providing a relatively clear empirical ground for interpreting three dimensions.

Factor 1 includes Items 1, 6, 7, and 10; Factor 2 – Items 2, 3, 4, and 5; and Factor 3 – Items 8 and 9. The pattern shows that the assessment instrument development competence does not include a single homogeneous dimension but consists of several competence dimensions.

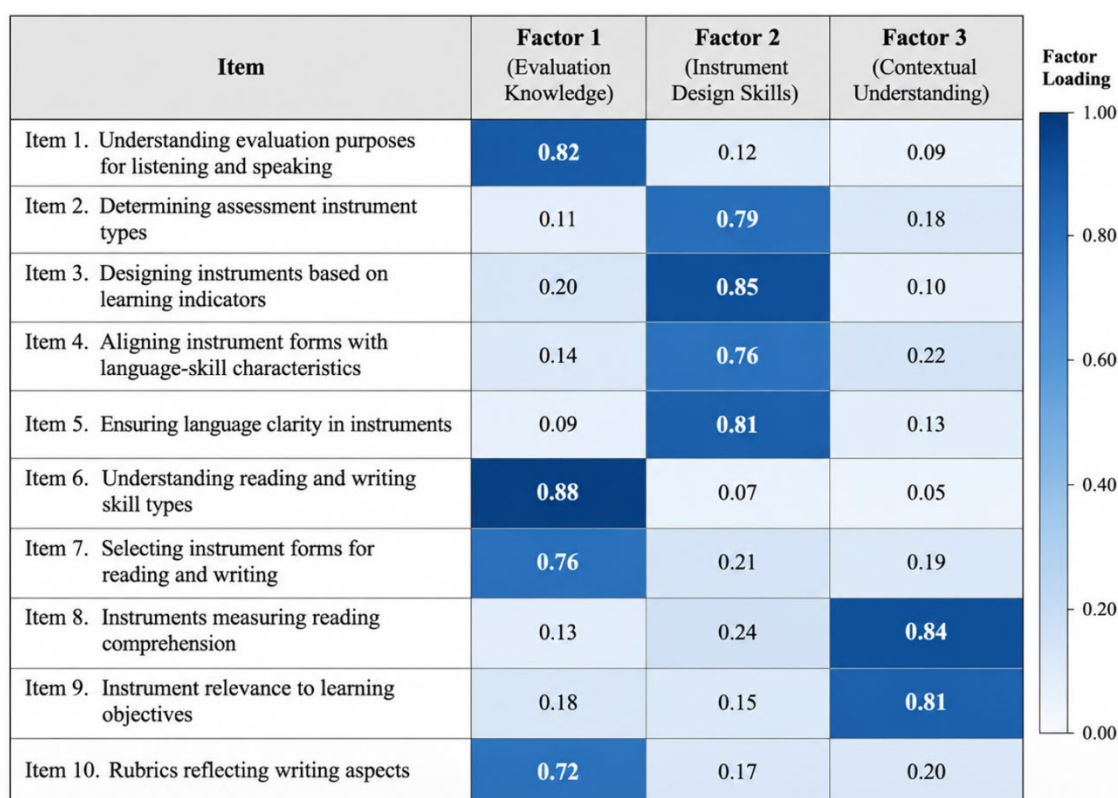


Figure 3. Factor Loading Pattern Across the Ten Assessment Competence Items

Figure 3 presents a clearer picture of the factor structure by demonstrating the concentration of the strongest loadings in particular columns of factors. Items 1, 6, 7, and 10 have the highest loadings on Factor 1, Items 2–5 have high loadings on Factor 2, and Items 8 and 9 have the highest loadings on Factor 3. This relatively clear grouping of items provides additional support to the interpretation of pre-service teachers' assessment competence as knowledge of assessment, instrument development skills, and ability to take into account context-related issues (Lazarević, 2018). However, since the study was based on only 25 participants, the observed structure should be considered preliminary exploratory evidence and not the conclusive confirmation of a stable factor structure.

The first dimension was interpreted as Evaluation Knowledge. This factor includes Items 1, 6, 7, and 10, and is characterized by the following loadings: 0.82, 0.88, 0.76, and

0.72, respectively. The content of these items is related to understanding the purpose of assessment, distinguishing language-skill characteristics, selecting assessment format, and recognizing important components of writing assessment. Conceptually, the dimension represents declarative knowledge, or knowledge about what assessment is and why certain assessment decisions are needed (Yüksel, 2014). The findings of qualitative analysis provided in the previous section support this interpretation because pre-service teachers showed the ability to articulate fundamental principles of assessment.

The second dimension identified in the analysis is Instrument Design Skills. It consists of Items 2, 3, 4, and 5, whose loadings range from 0.76 to 0.85. The items are related to determining appropriate assessment formats, developing instruments according to learning indicators, aligning assessment formats with language-skill

characteristics, and ensuring clarity of assessment language. As opposed to Evaluation Knowledge, this factor represents procedural dimension of competence because it concerns how the knowledge about assessment is implemented in concrete instrument-development decisions. This distinction corresponds to the conception of teacher cognition developed by [Farmasari et al. \(2023\)](#), according to which the knowledge is reflected in decisions that teachers make during instructional practice. The findings also correspond to the instrument development research suggesting the importance of coherent and evidence-based competency dimensions ([Leandro Cruz & do Nascimento, 2020](#)).

The third dimension, Contextual Understanding of Language Skills, is characterized by the inclusion of Items 8 and 9 with loadings of 0.84 and 0.81. The items are related to assessment instruments measuring reading comprehension correctly and relevance of the instruments to reading and writing learning objectives. The factor thus reflects the context-sensitive dimension of assessment competence involving ability to determine appropriateness of an instrument in the given context and situation ([Köş & Celik, 2023](#)). This interpretation is consistent with the research indicating that teachers' assessment literacy and their training needs vary depending on educational and contextual factors ([Vogt & Tsagari, 2020](#)).

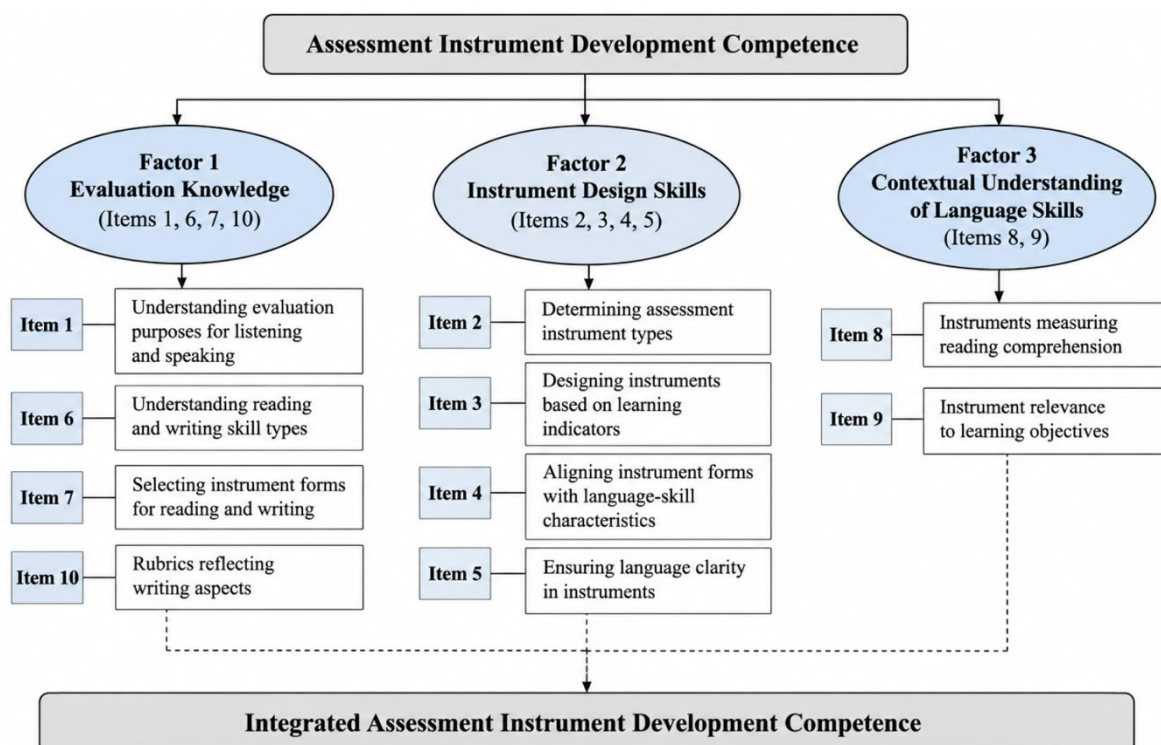


Figure 4. Three-Factor Structure of the Assessment Instrument Development Competence

Figure 4 illustrates multidimensional nature of the construct in terms of conceptualization. Evaluation Knowledge represents conceptual knowledge, Instrument Design Skills represent its procedural

implementation, and Contextual Understanding is the ability to adjust the decisions to particular learning objectives and language-skill characteristics ([Öztürk, 2021](#)). These three dimensions should not be

interpreted as independent competencies. Instead, based on qualitative findings, it can be stated that successful instrument development involves their integration. In particular, a participant can know the principles of assessment but still face problems with selecting appropriate instrument or deciding about the appropriateness of a certain instrument in a given situation.

The reported EFA solution revealed three components explaining 78.3% of the variance, which means that the extracted

dimensions account for a considerable portion of variance of the 10-item instrument. Thus, the structure can be considered useful for the conceptualization of the construct of interest. At the same time, due to relatively small sample size, it is necessary to exercise caution in interpreting the magnitude and stability of the components (Karacan & Can, 2026). The three-factor structure should therefore be regarded as exploratory and require replication in future research with much larger sample sizes.

Table 4. Summary and Interpretation of the Extracted Competency Dimensions

Factor	Items	Dominant Loading	Reported Variance Contribution*	Construct Interpretation
Factor 1	1, 6, 7, 10	0.88	32.7%*	Evaluation Knowledge understanding what assessment is and why it is used
Factor 2	2, 3, 4, 5	0.85	26.8%*	Instrument Design Skills knowing how to construct assessment instruments
Factor 3	8, 9	0.84	18.8%*	Contextual Understanding knowing when, where, and for which learning objectives assessment should be adapted
Total	10 items	-	78.3%*	Multidimensional assessment instrument development competence

The variance percentages are consistent with those presented in the data of the study. The eigenvalues should be checked against the original statistical output of the study before publication because the previously reported eigenvalues (4.21, 3.18, and 2.17) do not mathematically correspond to the stated percentages (32.7%, 26.8%, and 18.8%) for a 10-item PCA. The final manuscript should report the values from the original EFA/PCA output.

Table 4 provides a summary of conceptual meaning of the extracted three dimensions and clarifies their logical progression from conceptual knowledge towards application in specific contexts. Evaluation Knowledge can be considered the foundation because teachers must first understand the purposes of assessment and language-skill characteristics before

developing the instruments (Borg & Edmett, 2019). Instrument Design Skills reflect this knowledge through making decisions concerning formats, indicators, and language clarity. Contextual Understanding extends the two above-mentioned dimensions by requiring participants to evaluate whether the instrument properly represents the learning objective and learner needs (Lázaro Cantabrana et al., 2019). This interpretation is compatible with the concept of development of teacher competence resulting from the interaction of knowledge, cognition, and contextual practice (Borg, 2023; Li, 2020).

However, the order of these dimensions should not be considered a statistically established developmental hierarchy. Instead, the logical progression from knowing what and why to knowing how and

further to knowing when, where, and for whom provides an interpretive framework for interpreting the qualitative findings. This distinction is important because exploratory factor analysis is used to reveal the patterns of correlation between variables while establishing a developmental hierarchy requires longitudinal or experimental evidence (Sumarni, 2023). Thus, the current study uses three dimensions as a conceptualization of the competency organization but not as a claim that participants necessarily develop through these stages in a certain sequence.

The findings also correspond to the sociocultural perspective on teacher cognition, suggesting that advanced professional learning develops through interaction, mediation, guided practice, feedback, and participation in socially situated activities rather than through passive acquisition of assessment knowledge. According to the theory, professional knowledge is developed through interaction, mediation, and participation in socially situated activities and not merely as a result of passive information acquisition (Vygotsky, 1980). The relatively clear factor structure found in the present study can thus be interpreted as evidence of recognition by participants of different dimensions of assessment competence, while the qualitative findings indicate that ability to integrate these dimensions depends on opportunities for guided practice, feedback, reflection, and interaction with more knowledgeable others (Li, 2020). This interpretation is consistent with research emphasizing the importance of training and practicum experience in translation of theoretical knowledge into practice (Borromeo Ferri, R., Greefrath, 2026; Nguyen et al., 2026a).

Overall, exploratory factor analysis provides the preliminary empirical support to the three-dimensional conceptualization of

assessment instrument development competence, including Evaluation Knowledge, Instrument Design Skills, and Contextual Understanding of Language Skills. The first dimension reflects conceptual knowledge, the second – its implementation, and the third – contextual application (Kazaz & Alagözlü, 2020). Most importantly, the relatively low perceived competence found in previous analysis of Statements 8 and 9 is consistent with the emergence of these items as the unique indicators of the third factor. Such convergence between descriptive and factor-analytic findings provides additional justification of the interpretation that contextual understanding represents an important area for further investigation (Jia et al., 2016). At the same time, the small sample size implies that the factor structure should be considered preliminary and requires replication with larger samples and more sophisticated methods in future studies. The next section thus analyzes how reflective practice may explain the relationships among the identified three dimensions and transition from assessment knowledge to assessment practice.

c. Reflective Practice as a Mechanism for Integrating Assessment Competence

However, it would be erroneous to draw conclusions about the competence of assessment instrument development from either self-perceptions of the participants or the three-factor structure obtained with the help of EFA. The results of interviews and document analysis reveal another dimension of participants' competence that explains how they translate the existing assessment knowledge into practice (Guevarra et al., 2024). First of all, the findings show that effective translation includes a consistent

distinction between the knowledge of assessment principles and their flexible application in specific instructional settings. This finding makes reflective practice the key mechanism for bridging knowledge of assessment evaluation, design, and contextual features of instruction. In this study, reflective practice is thus defined as a mechanism for bridging theory and practice rather than as a statistically tested mediating variable.

As seen in the qualitative data, the participants generally possessed the basic knowledge of the purposes and principles of language assessment. They could usually explain why language assessment should be aligned with instructional goals and why

different language skills require different approaches (Alaa et al., 2019). However, some of the participants admitted that their knowledge was mostly declarative. While they could describe what an appropriate assessment should include, it seemed more difficult for them to explain how particular decisions about assessment should be made in a specific instructional situation. The above findings agree with Borg's (2023) distinction between knowledge and practices related to it among teachers. In addition, the findings also support the sociocultural assumption that professional knowledge is developed through interaction rather than through knowledge transmission (Vygotsky, 1980).

Table 5. Qualitative Evidence of Reflective Practice Across Competency Dimensions

Competency Dimension	Evidence from Interviews	Evidence from Assessment Documents	Reflective Challenge	Required Mediation
Evaluation Knowledge	Participants generally understood assessment purposes and principles	Basic assessment components were generally present	Translating principles into independent assessment decisions	Guided practice, modelling, and collaborative discussion
Instrument Design Skills	Participants expressed confidence in selecting formats and developing instruments	Most rubrics contained essential assessment aspects	Some reliance on routines or templates without sufficient construct consideration	Feedback, redesign, and authentic assessment practice
Contextual Understanding	Participants experienced greater difficulty adapting instruments to specific learning situations	Only 40% of rubrics contained clear and differentiated descriptors	Translating general principles into context-sensitive decisions	Authentic cases, mentoring, simulation, and structured reflection
Reflective Practice	Deeper reflection involved questioning practices and considering alternative approaches	Assessment decisions provided opportunities for evaluation and revision	Developing self-regulation and professional judgment	Pre-active, interactive, and post-active reflection

As demonstrated in Table 5, the three competency dimensions not only differ in terms of their conceptual content but also require a different type of support for their development. The evaluation knowledge seems to be rather well-developed because the participants were able to articulate the

basic principles of assessment. Similarly, the instrument design skills were also apparent from the participants' ability to produce assessment instruments with essential components (Hoang et al., 2025). Finally, the contextual understanding seemed to be somewhat more complex since the

participants had to evaluate whether a particular instrument was appropriate for specific learners, objectives, language skills, and classroom conditions. This finding is supported by the research claiming that the assessment literacy and training needs of

teachers are affected by the educational context (Vogt & Tsagari, 2020).

The relationship among the competency dimensions can be seen as moving from declarative knowledge to contextualized competence.

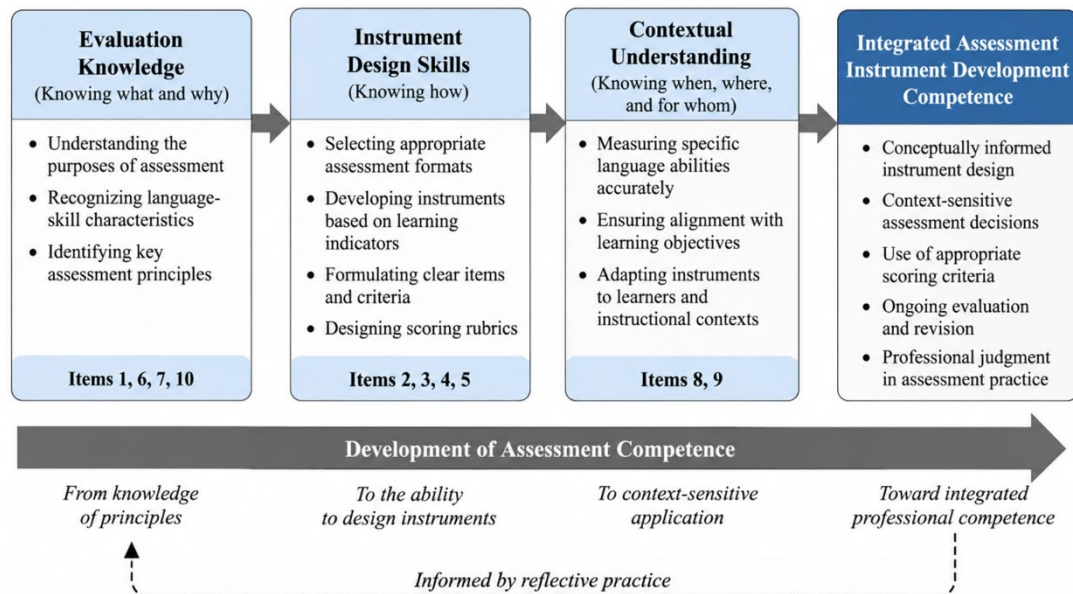


Figure 5. From Declarative Knowledge to Contextual Assessment Practice

Figure 5 reveals an interpretation of the factor structure in connection with the qualitative findings. As it can be seen, the evaluation knowledge constitutes the conceptual base from which the assessment decisions are formed. The instrument design skills reflect the operationalization of that knowledge through the selection of assessment format, the formulation of assessment indicators, creation of assessment items, and development of scoring criteria (Ölmezer-Öztürk & Aydin, 2018). The contextual understanding continues that process and requires the pre-service teachers to evaluate the appropriateness of the resulting instrument for particular learners, objectives, skills, and conditions. The described sequence is not supposed to reflect any statistical order of the development of the competence dimensions.

The qualitative data also show that reflection takes place throughout the process of assessment development. The participants have to think about the purpose of an assessment before its design, evaluate the appropriateness of the decisions made while creating or implementing an instrument, and finally, evaluate and revise it (Zhang & Wang, 2023). Such a process can be considered the part of the broad definition of reflective practice as examination of one's professional decisions and their consequences. The research on reflective practice also shows that such a reflection can support the development of professional awareness and self-monitoring among pre-service teachers (Cirocki & Hallet, 2024; Nguyen et al., 2026).

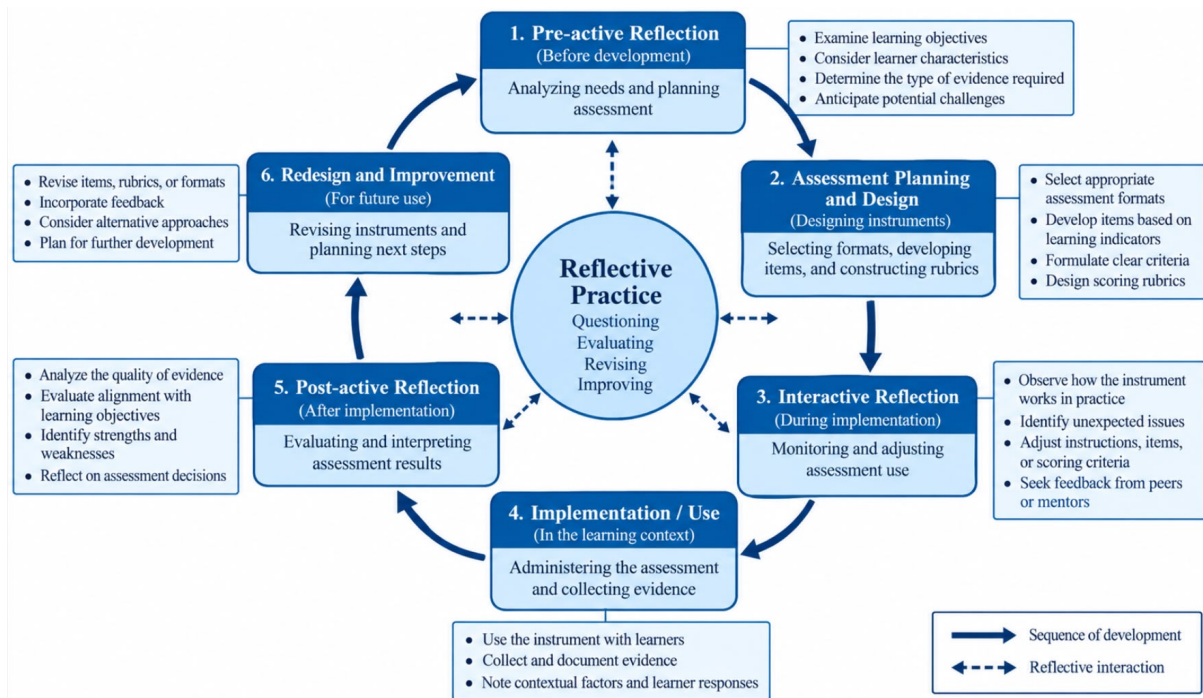


Figure 6. Reflective Practice Throughout the Assessment Development Process

Figure 6 stresses that reflection is not the activity that occurs at the end of the assessment development process. It can occur at various stages of the process. For instance, during the pre-active reflection, the participants consider the learning objectives, learner characteristics, and the type of evidence needed before creating an assessment instrument. The interactive reflection allows them to rethink the decisions made while designing or implementing the instrument (Cirocki et al., 2025). Finally, the post-active reflection involves the evaluation of the instrument regarding how well it captures the desired competence and what changes are necessary. Such an interpretation is compatible with the reflective cycles used in teacher education programs and evidence that reflective writing can progress from description to analysis of professional experiences (Nguyen et al., 2026; Colwell & Ince, 2021).

The sociocultural approach also does not limit reflective practice to individual reflection. On the contrary, reflection can be stimulated and supported by the interaction with instructors, mentors, and other peers. In the present case, the process of pre-service teachers' movement from the knowledge of assessment principles to their application and adaptation can be seen as the Zone of Proximal Development (ZPD) proposed by Vygotsky (1980). According to ZPD, the tasks that initially seem beyond the capacity of an individual can be accomplished due to modeling, scaffolding, dialogue, feedback, and collaboration. Similar ideas are presented in the literature on teacher cognition that emphasizes the influence of social and environmental factors on teachers' knowledge, beliefs, and interpretations (Cai & Zheng, 2020; Borg, 2023).

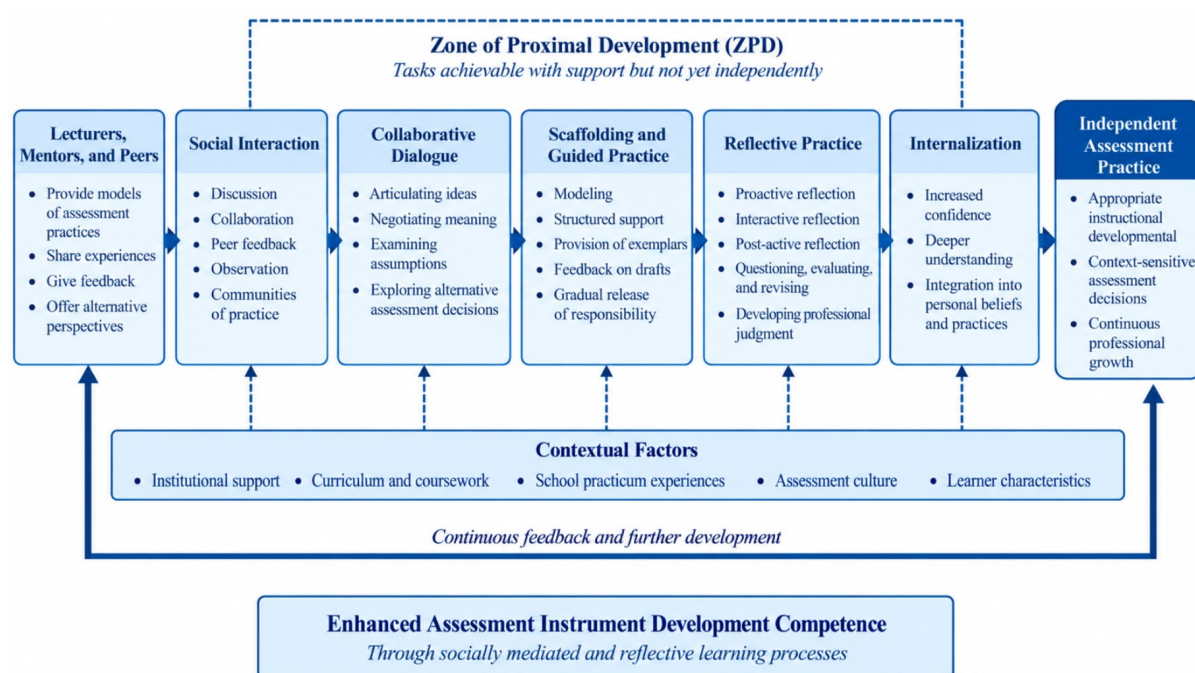


Figure 7. Sociocultural Mediation of Assessment Competence Development

Figure 7 illustrates how the process of the integration of the three competency dimensions is facilitated by social mediation. Instructors and experienced mentors can provide models of assessment decision-making, while peers can offer alternative interpretations and criticism. Dialogue helps pre-service teachers to articulate the assumptions that might otherwise remain unnoticed, and scaffolding lets them perform the assessment-related tasks that they might initially find difficult to accomplish on their own (Dy & Sumayao, 2023; Harmadi et al., 2025). With the repeated participation and reflection, the initially externally mediated assessments can gradually be internalized. Such an interpretation agrees with the findings about the effect of social interaction on teachers' knowledge, beliefs, and concepts (Cai & Zheng, 2020).

The findings of the current study also show some differences in the degree of the participants' reflective engagement. In the first case, the participants could just say whether an assessment instrument was appropriate or not. In a more complex case,

reflection included questioning of the reasons for choosing a particular format, considering other possible assessment approaches, assessing the alignment of the instrument with the learning objectives, and identifying necessary revisions (Hernández-López et al., 2024). Such a distinction is essential since reflective practice is not equal to just thinking about an experience. Reflection requires a systematic examination of assumptions, evidence, and consequences and consideration of alternative actions (Angela et al., 2025). Such reflective practice can promote the development of professional judgment, especially when the pre-service teachers have to face the situations that are not covered by standardized procedures or templates.

In turn, the relationship between reflection and assessment competence can also be explained through Borg's (2023) framework of teacher cognition. The assessment decisions are based on the knowledge, beliefs, and interpretations of the teachers. The reflective practice provides an opportunity for those elements to become

explicit and to be reassessed. Such an interpretation is supported by the findings about the positive role of reflective practice for the professional development of teacher educators and context-sensitive pedagogical decisions (Amalia et al., 2020). The same can be said about the critical incident approaches, which can trigger reflection among pre-service teachers, and action research that can promote the reflective cycles of inquiry (Anselmann, 2023; Ayubayeva & McLaughlin, 2023).

Summarizing the above findings, it can be stated that the three competency dimensions are interconnected through the reflective practice and sociocultural learning process. The evaluation knowledge provides the conceptual basis for the assessment decisions, instrument design skills help to operationalize this knowledge into concrete assessment instruments, and the contextual understanding enables the adaptation of these instruments to particular learning conditions (König et al., 2020). Reflective practice links these dimensions through encouragement of their evaluation and revision.

Such an interpretation of the findings seems particularly relevant in view of the identified theory-practice gap. According to the current findings, the main problem is not the lack of assessment knowledge but its insufficient integration with design decisions and contextual judgments. From the sociocultural perspective, such integration depends on the mediation, scaffolding, social interaction, and internalization (Giraldo & Murcia, 2019). From the perspective of teacher cognition, it also requires a consistent consideration of the role of knowledge and beliefs in making professional decisions (Borg, 2023). Thus, reflective practice can be seen as a bridging mechanism for increasing the integration of assessment knowledge.

To conclude, this section has shown that the three-factor structure obtained through EFA cannot be viewed as three separate components of the participants' assessment competence. Instead, the findings of the qualitative stage have revealed that their effective combination is achieved through reflection and socially mediated learning (Kazykhankyzy & Alagözlü, 2019). This interpretation becomes the basis for the following discussion of pedagogical and theoretical implications of the findings for language teacher education.

d. Pedagogical and Theoretical Implications for Language Teacher Education

The findings of this study can be applied to designing teacher education curricula regarding the teaching of assessment competence in terms of its assessment, practice, reflection, and evaluation. The three-dimensional structure revealed in the exploratory analysis (Evaluation Knowledge, Instrument Design Skills, and Contextual Understanding) suggests that assessment competence transcends conceptual knowledge acquisition (Hidayat et al., 2025). Qualitative findings reveal that pre-service teachers may know the basic principles of assessment while having difficulty applying them in context (Lazarević, 2018). Therefore, teacher education needs to create learning activities combining assessment theory, instrument design, context application, and reflection. These implications are coherent with the research stressing the importance of targeted assessment training in the development of assessment competence of pre-service teachers (Yüksel, 2014) and with research focused on the importance of context in teachers' assessment literacy and training needs (Vogt & Tsagari, 2020).

The first implication is the need to provide more contextualized assessment training for pre-service teachers. Statements 8 and 9 had the relatively low perception of competence, and qualitative evidence indicates that pre-service teachers need real-life application of assessment principles. Teacher education should provide opportunities for the use of authentic

instructional contexts rather than just replicating assessment forms (Farmasari et al., 2023). Authentic case analysis can enable pre-service teachers to realize that the instrument can be assessed as appropriate only if its format is technologically correct and if the evidence generated by the instrument corresponds with the instructional situation and intended construct.

Table 6. Pedagogical Implications Derived from the Study Findings

Empirical Finding	Interpretation	Pedagogical Response	Expected Development
High perceived competence in foundational assessment principles	Declarative assessment knowledge is relatively well established	Use application-oriented assessment tasks	From knowing principles to applying them
Strong perceived instrument-design competence	Participants can develop basic assessment instruments	Introduce construct alignment, item analysis, and redesign activities	More principled instrument design
Lower perceived competence in contextual application	Context-sensitive assessment judgment remains challenging	Use authentic cases, simulations, and mentoring	Stronger contextual assessment decisions
72% of rubrics contained essential aspects, but only 40% had clear differentiated descriptors	Basic rubric construction was stronger than performance-level specification	Provide explicit rubric-development and evaluation practice	Improved scoring criteria
Interviews revealed a theory–practice discrepancy	Knowledge is not always translated into practice	Embed structured reflection throughout assessment development	Stronger theory–practice integration
Reflection involved questioning and revising assessment decisions	Reflection can support integration of competency dimensions	Implement systematic reflective activities and portfolios	Greater professional judgment and self-regulation

Table 6 provides concrete pedagogical implications from the findings of the study. It should be noted that these recommendations are not provided solely on the basis of high questionnaire results but are derived from the convergence of evidence from the questionnaire, interviews, and documents. As 72% of rubrics contain the necessary components while 40% include clear and differentiated descriptors, it suggests that teacher education should address the quality of performance criteria instead of requiring students to develop rubrics. This is in line with the research indicating that the development of rubrics

improves teachers' understanding of the assessment process and leads to better learning outcomes (Nurhayati et al., 2019). Similarly, the analysis of assessment instrument feasibility should consider both material and language issues in order to enable appropriate measurement of learners' abilities (Listeani et al., 2021).

The second implication is the need to integrate reflective practice into assessment courses systematically. Rather than treating reflection as an add-on exercise following an assignment, reflection can be conducted in pre-active, interactive and post-active stages of assessment instrument development (Kös

& Celik, 2023). This recommendation is consistent with the reflective process revealed in this study and described above: pre-active reflection, interactive reflection, and post-active reflection. Research on reflective practice and lesson study indicates the usefulness of structured reflection for the

connection of theoretical knowledge and teaching decisions (Colwell & Ince, 2021). The evidence about the reflection using e-portfolios reveals the possibility of professional identity development and self-monitoring (Cirocki & Hallet, 2024).

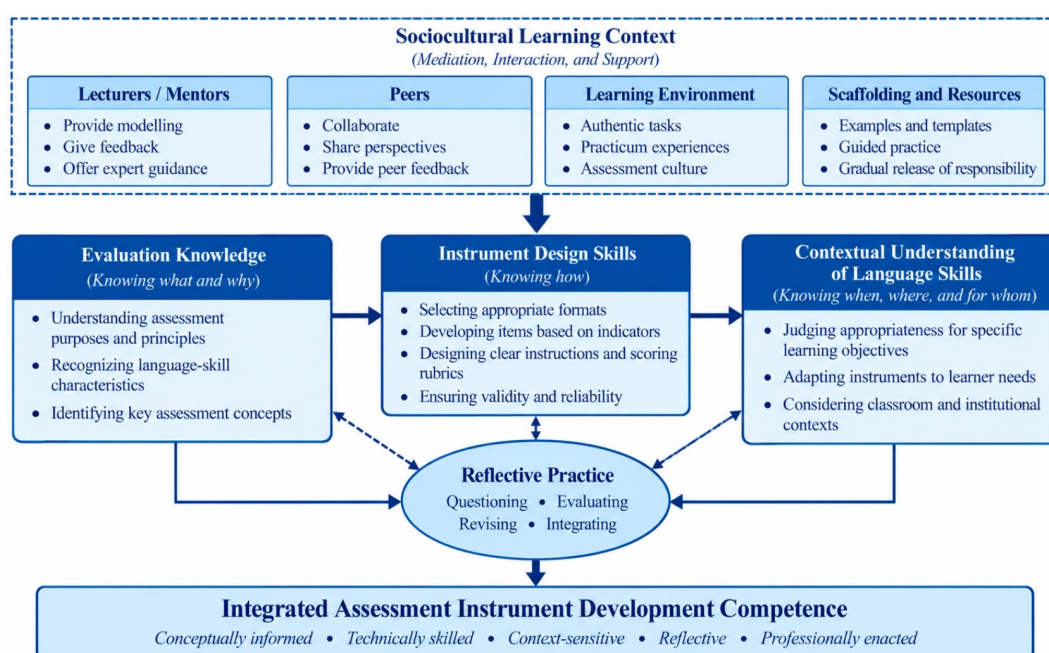


Figure 8. Integrated Model of Assessment Instrument Development Competence

Figure 8 summarizes the main findings and theoretical implications of the study. The model does not see three dimensions as independent constructs. Instead, the Evaluation Knowledge provides the conceptual background, the Instrument Design Skills transform knowledge into the product, and the Contextual Understanding is responsible for adapting the product to particular learning objectives and instructional situations (Öztürk, 2021). The Reflective Practice is the integrative mechanism since qualitative findings suggest that questioning, evaluation, and revision of decisions related to assessment can be helpful in the transition between dimensions. Thus, the model is the interpretation of the

present findings rather than the causal statistical model.

The third implication involves the explicit addressing of the theory-practice gap in assessment education. Results show that pre-service teachers can learn assessment concepts but conceptual understanding alone does not guarantee appropriate decisions. Teacher educators therefore need to model the decision-making process related to assessment rather than presenting assessment instruments as finished products (Karacan & Can, 2026). Students should be required to justify the choice of the particular assessment format, to explain how the item presents the indicator of learning, to discuss the appropriateness of the criterion and modification of the instrument in new learner

or classroom conditions. Such an approach is consistent with [Vygotsky's \(1980\)](#) view of mediation and internalization and with [Borg's \(2023\)](#) concept of teacher cognition.

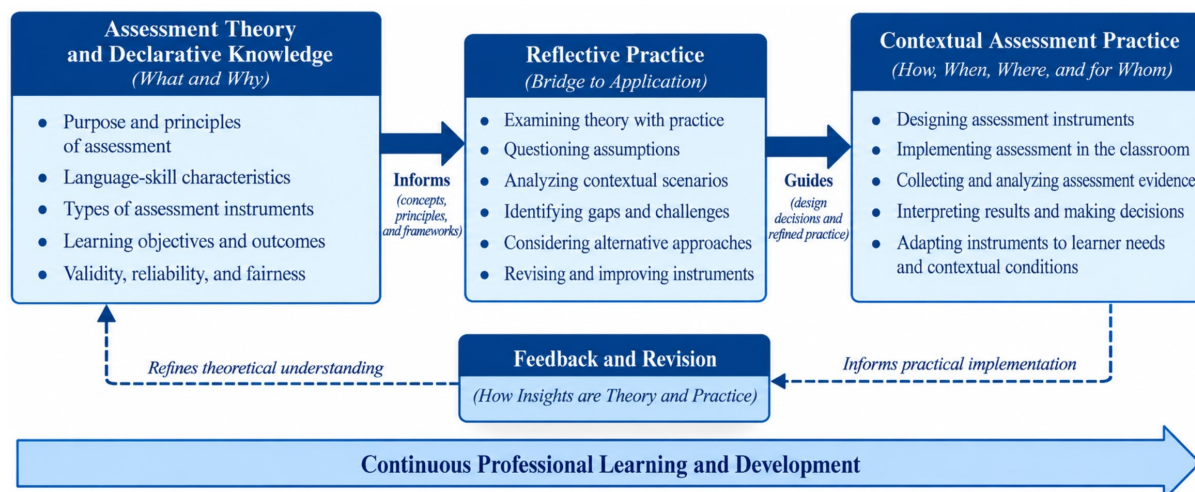


Figure 9. Theory-Practice Integration Framework

Figure 9 suggests that the relationship between theory and practice should be seen as an iterative process rather than the one-directional transfer of knowledge. Assessment theory provides principles and conceptual resources, whereas practical experience gives pre-service teachers information about context-specific limitations and unexpected assessment problems ([Borg & Edmett, 2019](#)). Reflective practice connects these two dimensions by requiring students to compare theoretical assumptions with their assessment experience, to identify discrepancies, to consider alternative solutions and to revise their decisions. This interpretation is compatible with sociocultural accounts of teacher development in which professional knowledge is mediated through interaction and internalized gradually ([Li, 2020](#); [Cai & Zheng, 2020](#)). It is also compatible with research on the usefulness of reflective lesson planning for contextualization of pedagogical decisions ([Amalia et al., 2020](#)).

The fourth implication concerns the development of the continuous assessment competence cycle in teacher education. Instead of viewing assessment instrument development as a single course assignment, it can be organized as an iterative process of learning, planning, design, implementation, evaluation, reflection, and redesign. This approach will give pre-service teachers the opportunity to repeatedly face authentic assessment problems and to refine their professional judgement progressively ([Lázaro Cantabrana et al., 2019](#)). The cycle can involve portfolio documentation, peer review, mentor feedback, and empirical analysis of the assessment instrument. These approaches are consistent with the evidence suggesting the usefulness of reflective practice, critical incident approach, and action research as mechanisms for systematic reflection ([Anselmann, 2023](#); [Ayubayeva & McLaughlin, 2023](#)).

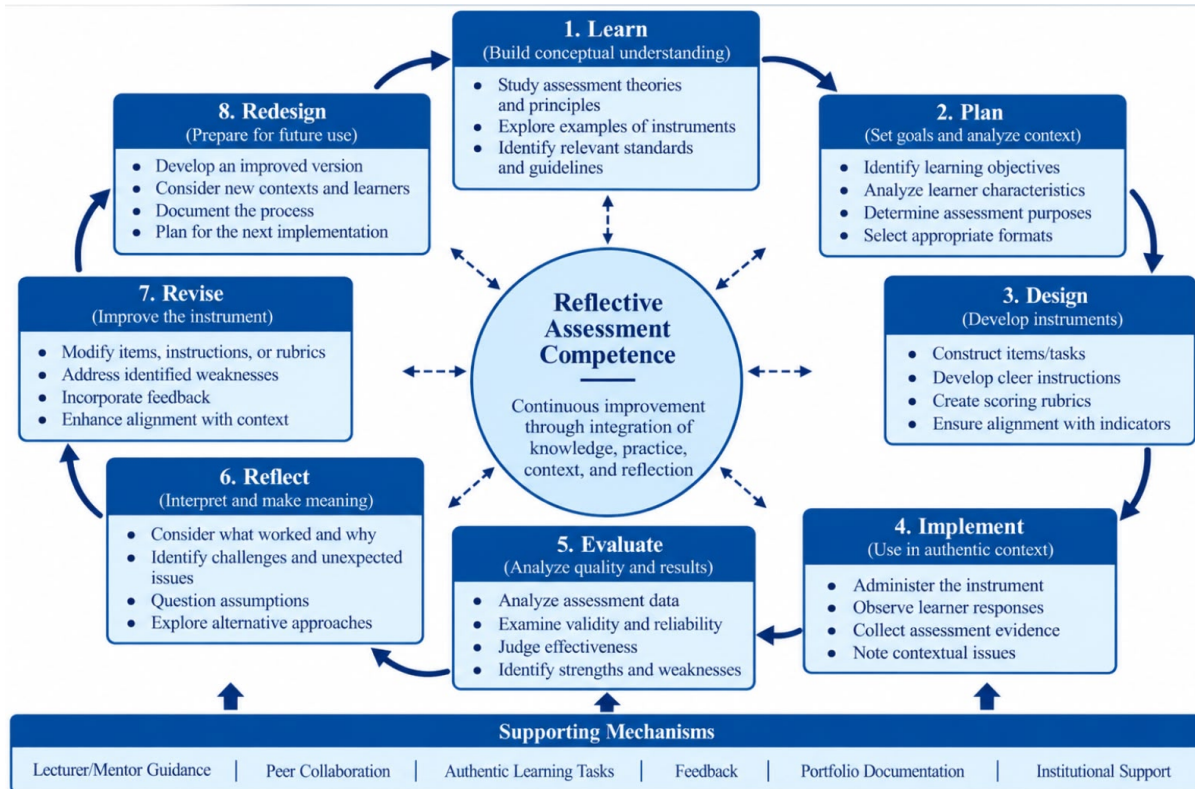


Figure 10. Reflective Assessment Competence Development Cycle

Figure 10 operationalizes the pedagogical implications of the study as a professional learning cycle. The model emphasizes that competence develops through repeated interaction between knowledge and experience rather than through one-time exposure to assessment theory (Sumarni, 2023). Evaluation provides information about the quality of the instrument, reflection is needed to interpret the information and to revise the instrument on this basis. This cycle is especially relevant for the contextual dimension revealed in the study since the contextual judgement can require repeated exposure to different assessment situations. The portfolio-based documentation can make the process of development visible by allowing pre-service teachers to monitor changes in their assessment decisions in the course of iterative process (Cirocki & Hallet, 2024; Nguyen et al., 2026).

From the theoretical point of view, the findings contribute to the integrated interpretation of Sociocultural Theory and Teacher Cognition. Sociocultural Theory explains how the assessment competence develops through interaction, mediation, scaffolding, and internalization (Vygotsky, 1980; Li, 2020; Cai & Zheng, 2020), whereas Borg's (2023) framework explains how the knowledge, beliefs and cognitive processes influence the decisions. The combination of these theories provides a useful conceptual tool to understand the reason for which pre-service teachers may have good declarative knowledge but require mediated experience for turning this knowledge into professional practice (Kusumaningtyas et al., 2024). The three-factor structure reveals in this study provides an empirical basis for distinction of conceptual, procedural, and contextual dimensions of assessment competence, although the preliminary character of the

findings requires verification with larger samples.

Overall, the pedagogical implications suggest an assessment education model that combines conceptual knowledge, technical instrument development skills, contextual application and reflection. Instead of viewing these elements as separate instructional goals, teacher education programs can integrate them through authentic assessment tasks, collaborative learning, mentoring, reflective portfolios, empirical analysis of assessment instruments, and iterative redesign, thereby creating progressive learning environments that connect theory, practice, reflection, and continuous improvement. Such an approach is consistent with the research indicating that assessment competence develops through interaction of knowledge, practice, context, and reflection (Kazaz & Alagözlü, 2020). The suggested framework should be seen as the theoretically based pedagogical model derived from the findings of the present study, which should be used for future empirical research and refinement.

4. Conclusion

The current study demonstrates that assessment instrument development competence among pre-service language teachers is a multidimensional construct comprising Evaluation Knowledge, Instrument Design Skills, and Contextual Understanding of Language Skills, which together provide a foundation for advanced, contextualized, and progressive assessment learning. While generally being competent, the participants exhibited a disconnect between their knowledge base and practical competencies, especially in terms of differentiating performance criteria and adaptation of assessment tools to the particular learning and instruction situation. The use of reflective practice proved to be a useful way to integrate the

aforementioned competencies into coherent activities in planning, implementing, evaluating, and revising assessments. The results suggest that teacher education programs should move beyond the transfer of theoretical assessment knowledge by providing authentic assessment experiences, context-based practice, mentoring, feedback, collaborative learning, and structured reflection that promote progressive and practice-oriented professional learning. An integrated perspective combining assessment knowledge, technical design skills, context-sensitive judgment, and reflective practice is therefore proposed as a foundation for progressive and sustainable professional learning in teacher education. Due to exploratory nature of the factor analysis and the sample size, however, the multi-dimensional construct should be regarded as tentative and deserves further exploration using larger samples.

5. References

- Aalaa, M., Albakri, I. S. M. A., Singh, C. K. S., Hammed, H., Zaidan, A. A., Zaidan, B. B., ... Jasim, A. N. (2019). Assessment and ranking framework for the English skills of pre-service teachers based on fuzzy Delphi and TOPSIS methods. *IEEE Access*, 7, 126201–126223.
- Aini, S. W., & Mukhlis, M. (2022). Analisis taksonomi pada soal Asesmen Kompetensi Minimum literasi membaca di SMK Negeri 3 Pekanbaru. *Diglosia: Jurnal Kajian Bahasa, Sastra, dan Pengajarannya*, 5(4), 933–948. <https://doi.org/10.30872/diglosia.v5i4.513>
- Amalia, F., Mustikawati, D., & Nuraeni, S. (2020). Reflective practice on lesson planning among EFL teacher educators. *Indonesian Journal of Applied Linguistics*, 10(1), 145–156. <https://doi.org/10.17509/ijal.v10i1.23545>
- Angela, L., Handican, R., & Casanova, A.

- (2025). Linking belief to thought: A structural equation modeling analysis of self-efficacy and critical thinking among Indonesian pre-service teacher. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 393–410.
- Anselmann, V. (2023). Promoting pre-service teachers reflections through critical incident training. *Teaching and Teacher Education*. <https://doi.org/10.1016/j.tate.2022.103998>
- Ayubayeva, N., & McLaughlin, C. (2023). Action research as a school development approach: A systematic review. *Educational Action Research*, 31(2), 245–263. <https://doi.org/10.1080/09650792.2022.2039254>
- Borg, S. (2023). Teacher cognition in language teaching. In C. J. Hall & R. Wicaksono (Eds.), *The Routledge handbook of English language teaching* (pp. 211–225).
- Borg, S., & Edmett, A. (2019). Developing a self-assessment tool for English language teachers. *Language Teaching Research*, 23(5), 655–679.
- Borromeo Ferri, R., & Greefrath, G. (2026). Enhancing pre-service teachers assessment competence through targeted training. *Teaching and Teacher Education*. <https://doi.org/10.1016/j.tate.2025.104832>
- Braun, V., & Clarke, V. (2022). Conceptual and design thinking for thematic analysis. *Qualitative Psychology*, 9(1), 3–26. <https://doi.org/10.1037/qup0000196>
- Cai, C., & Zheng, X. (2020). A sociocultural perspective on language teacher cognition: A review of the literature. *Frontiers of Education in China*, 15(3), 419–440. <https://doi.org/10.1007/s11516-020-0021-2>
- Chen, J., Li, Z., & Zhang, J. (2026). Teacher cognition and its influence on practice: A systematic review of the last decade. *Teaching and Teacher Education*. <https://doi.org/10.1016/j.tate.2026.104918>
- Cirocki, A., Anam, S., Drajiati, N. A., & Soden, B. (2025). Assessment literacy among Indonesian pre-service English language teachers: A mixed-methods study. *Iranian Journal of Language Teaching Research*, 13(1), 23–44.
- Cirocki, A., & Hallet, W. (2024). Reflective practice in English language teaching: A systematic review. *Language Teaching Research*, 28(1), 89–112. <https://doi.org/10.1177/13621688231174523>
- Colwell, J., & Ince, C. (2021). Lesson study in initial teacher education: A systematic review. *International Journal for Lesson and Learning Studies*, 10(2), 143–158. <https://doi.org/10.1108/IJLLS-11-2020-0093>
- Dy, A. S., & Sumayao, E. D. (2023). Influence of the pre-service teachers' language proficiency to their teaching competence. *AJELP: Asian Journal of English Language and Pedagogy*, 11(1), 1–21.
- Farmasari, S., Wardana, L. A., Herayana, D., & Suryaningsih, H. (2023). Pre-service EFL teachers' language assessment literacy satisfaction and assessment preparedness. *International Journal of Language Education*, 7(4), 646–660.
- Giraldo, F., & Murcia, D. (2019). Language assessment literacy and the professional development of pre-service language teachers. *Colombian Applied Linguistics Journal*, 21(2), 243–259.
- Guevarra, J. X. L., Bendicio, M. K. T., Alerosa, J. J., & Ilao, A. T. (2024). Assessment skills of pre-service teachers: Basis for the development of an assessment skills development guide. *International Research Journal of Education and Technology*, 6(7), 21–31.
- Haerazi, H., S., Wandu, S., & Nurkamto, J. (2023). Diagnostic assessment in the Merdeka Curriculum: A case study of

- Indonesian language teachers. *Indonesian Journal of Applied Linguistics*, 12(3), 567–581. <https://doi.org/10.17509/ijal.v12i3.56423>
- Hani, R. (2024). Reflective practice in assessment instrument development. *Journal of Language Testing and Assessment*, 41(1), 78–95. <https://doi.org/10.1080/02602938.2023.2278845>
- Harmadi, F., Maryani, I., Sukirman, S., & Montano, E. C. N. (2025). Digital transformation: Exploring the relationship between literacy, motivation, and TPACK in elementary education. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 295–311.
- Harizon, R., Putri, A., & Sari, D. (2023). Sequential explanatory design in language education research. *Educational Methodology Review*, 18(3), 212–229. <https://doi.org/10.1080/1743727X.2023.2196789>
- Hasim, M., Yusop, F. D., & Adnan, M. (2025). Traditional versus innovative assessment practices in Malaysian language education. *Assessment & Evaluation in Higher Education*, 50(2), 156–173. <https://doi.org/10.1080/02602938.2024.2391452>
- Hernández-López, E. M., Núñez-Mercado, P., & Bórquez-Morales, L. S. (2024). Assessing intercultural competence: The case of EFL pre-service teachers in the Mexican context. *Journal of Intercultural Communication*, 24(4), 84–96.
- Hidayat, M. L., Abdurahman, S. G., Astuti, D. S., Prabawati, R., Anif, S., Hariyatmi, H., & Zannah, F. (2025). Pilot study of digital competency mapping of Indonesian preservice teachers: Rasch model analysis. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 100–116.
- Hoang, L. P., Le, P. A., Le, H. T., Nguyen, D. T., Phan, T. C., Le, H. T., ... Tran-Duong, Q. H. (2025). Evaluating educational assessment competence of pre-service teachers: Extended standards in the context of digital classroom assessment transformation. *Education and Information Technologies*, 30(12), 16347–16374.
- Jamil, H. (2025). Formative assessment practices in Malay language speaking skills: Teachers' feedback and implications for secondary education in Malaysia. *Journal of Public Administration and Governance*, 15(3), 1. <https://doi.org/10.5296/jpag.v15i3.23475>
- Jia, Y., Oh, Y. J., Sibuma, B., LaBanca, F., & Lorentson, M. (2016). Measuring twenty-first century skills: Development and validation of a scale for in-service and pre-service teachers. *Teacher Development*, 20(2), 229–252.
- Karacan, C. G., & Can, T. (2026). Assessing teacher digital competences in pre-service language teachers: Adaptation and validation of the DigCompEdu scale. *Journal of Research on Technology in Education*, 58(5), 1114–1133.
- Kazakhykyzy, L., & Alagözlü, N. (2019). Developing and validating a scale to measure Turkish and Kazakhstani ELT pre-service teachers' intercultural communicative competence. *International Journal of Instruction*, 12(1), 931–946.
- Kazaz, I., & Alagözlü, N. (2020). Evaluation of teaching practicum for pre-service English language teachers: A scale development study. *Journal of Language and Linguistic Studies*, 16(3), 1562–1593.
- Köş, İ. E., & Celik, H. (2023). Global competence scale: An adaptation to measure pre-service English teachers' global competences. *International Journal of Assessment Tools in Education*, 10(4), 813–838.

- König, J., Bremerich-Vos, A., Buchholtz, C., Fladung, I., & Glutsch, N. (2020). Planning competence of pre-service German language teachers: Conceptualization, measurement, and validation. In *Student learning in German higher education: Innovative measurement approaches and research results* (pp. 53–74). Springer Fachmedien Wiesbaden.
- Kurniawan, Y., Suryani, R., & Rahmawati, D. (2025). Portfolio-based assessment in English for Specific Purposes: A systematic review. *English for Specific Purposes*, 78, 34–49. <https://doi.org/10.1016/j.esp.2024.10.003>
- Kusumaningtyas, D. A., Manyunu, M., Kurniasari, E., Awal, A. N., Rahmaniati, R., & Febriyanti, A. (2024). Enhancing learning outcomes: A study on the development of higher order thinking skills based evaluation instruments for work and energy in high school physics. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 14–31.
- Lázaro Cantabrana, J. L., Usart Rodríguez, M., & Gisbert Cervera, M. (2019). Assessing teacher digital competence: The construction of an instrument for measuring the knowledge of pre-service teachers. *Journal of New Approaches in Educational Research*, 8(1), 73–78.
- Lazarević, N. (2018). Assessment of intercultural communicative competence in pre-service English language teachers: Challenges and practices. *Nastava i Vaspitanje*, 67(3), 471–487.
- Leandro Cruz, J., & do Nascimento, N. (2020). Exploratory factor analysis for instrument validation in educational research. *International Journal of Educational Methodology*, 6(3), 523–534. <https://doi.org/10.12973/ijem.6.3.523>
- Li, L. (2020). A sociocultural perspective on teacher cognition research: A review of the literature. *Language Teaching Research*, 24(3), 305–326. <https://doi.org/10.1177/1362168818817369>
- Likita, A. J., et al. (2024). Strengthening mixed-methods research in education: A methodological review. *Educational Researcher*, 53(2), 89–104. <https://doi.org/10.3102/0013189X231213456>
- Listeani, F. Y., Safi'i, I., & Ibrahim, N. (2021). Kelayakan instrumen evaluasi Buku Sekolah Elektronik (BSE) Bahasa Indonesia siswa SMA kelas XI. *Diglosia: Jurnal Kajian Bahasa, Sastra, dan Pengajarannya*, 4(3), 375–386. <https://doi.org/10.30872/diglosia.v4i3.216>
- Lustyantie, N., Murtadho, F., Aceh, B., Pascasarjana, S., Jakarta, U. N., & Timur, J. (2025). Development of an assessment rubric for story writing skills in Indonesian integrated with Acehese folklore for senior high school. *Journal of Language Teaching and Learning, Linguistics and Literature*, 4778, 1938–1946.
- Marlina, L., Putri, D., & Wahyuni, S. (2023). Exploratory factor analysis for instrument validation in language education. *International Journal of Educational Psychology*, 12(2), 178–195. <https://doi.org/10.17583/ijep.2023.10475>
- Mursyida, R., Nurhasanah, R., & Amri, Z. (2020). Training needs in reflective practice among Indonesian language teachers. *Teacher Development*, 24(3), 345–362. <https://doi.org/10.1080/13664530.2020.1776922>
- Nguyen, H. T., Ngo, L. T., & Pham, T. H. (2026). E-portfolios for reflective practice in pre-service teacher education: A qualitative case study. *Teaching and Teacher Education*. <https://doi.org/10.1016/j.tate.2025.104750>
- Nurhayati, Hardoko, & Warman. (2019). Pengembangan evaluasi pembelajaran

- tematik dengan menggunakan rubrik kelas IV di Gugus IV Kecamatan Samarinda Ulu. *Diglosia: Jurnal Kajian Bahasa, Sastra, dan Pengajarannya*, 2(1), 47–58. <https://doi.org/10.30872/diglosia.v2i1.17>
- Ochoa, C., Giraldo, F., & Mosquera, R. (2024). Assessment instrument development competence in language teacher education. *Language Assessment Quarterly*, 21(3), 289–306. <https://doi.org/10.1080/15434303.2024.2327891>
- Ölmezer-Öztürk, E., & Aydin, B. (2018). Toward measuring language teachers' assessment knowledge: Development and validation of Language Assessment Knowledge Scale (LAKS). *Language Testing in Asia*, 8(1), 20.
- Öztürk, E. Ö. (2021). Developing language assessment literacy of EFL pre-service teachers through classroom assessment course. *International Journal of Contemporary Educational Research*, 8(3), 13–22.
- Racero Diz, A., & Canabal Guzmán, C. (2026). Teacher beliefs and instructional decision-making in contexts of structural vulnerability. *Teaching and Teacher Education*. <https://doi.org/10.1016/j.tate.2026.104928>
- Rahmawati, D. (2023). Rubrics in writing assessment for secondary school students. *Journal of Second Language Writing*. <https://doi.org/10.1016/j.jslw.2023.100953>
- Sumarni, S. (2023). Assessing digital competence among pre-service English language teachers: Strengths and weaknesses. *Journal of English Education and Teaching*, 7(3), 613–631.
- Sutami, Hudiyono, Y., & Ilyas, M. (2020). Pengembangan instrumen asesmen higher order thinking skills (HOTS) pada mata pelajaran Bahasa Indonesia SMA dan SMK. *Diglosia: Jurnal Kajian Bahasa, Sastra, dan Pengajarannya*, 3(1), 102–113. <https://doi.org/10.30872/diglosia.v3i1.24>
- Vetere, L. (2020). A Vygotskian approach to language teacher education: Joint mediated learning in concept development. *Language Teaching Research*, 24(5), 612–632. <https://doi.org/10.1177/1362168820915678>
- Vogt, K., & Tsagari, D. (2020). Teacher training needs in language assessment literacy. *Language Testing*, 37(3), 345–367. <https://doi.org/10.1177/0265532219879876>
- Vygotsky, L. (1980). *Mind in society: The development of higher psychological processes* (M. Cole, V. Jolm-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Wahyuni, S., & Arieffiani, M. (2022). Traditional assessment practices in Indonesian schools: A case study. *Journal of Educational Practice*, 13(2), 89–106. <https://doi.org/10.1080/19415257.2022.2027891>
- Yüksel, H. G. (2014). Teachers of the future: Perceived teaching competences and visions of pre-service English language teachers. *Journal of Human Sciences*, 11(2), 27–39.
- Zhang, D., & Wang, W. (2023). The current status and development of pre-service language teachers' TLLT competence in a PjBL classroom. *Cogent Education*, 10(2), 2273174.