

Bridging Theory and Practice Through Real-World Language Assessment Design

Desi Surlitasari Dewi,^{1*} Sulastri Manurung,² Natalia Manuhutu,³ Isnaini Eddy Saputro,⁴ Widhiyanto,⁵ Sri Wahyuni,⁶ Mega Aulia Ardhi⁷

^{1,2,7}English Education Department, Universitas Riau Kepulauan, Batam, Kepulauan Riau – Indonesia,

³English Literature Department, Universitas Masumus, Merauke, South Papua – Indonesia,

⁴Universitas Muhammadiyah Sorong – Indonesia ^{5,6}Language Education Doctoral Program, Universitas Negeri Semarang, Central Java – Indonesia

Article Information

Received: December 20, 2024

Revised: October 17, 2025

Accepted: January 24, 2025

Published online: November 20, 2025

Abstract

Language assessment is a crucial component of effective language teaching. However, many pre-service teachers lack sufficient training in assessment practices, resulting in a gap between theoretical knowledge and practical application. This study explores how pre-service teachers perceive the effectiveness of a project-based approach to designing and administering real-world language assessments and examines the challenges they face. Combining a closed questionnaire and semi-structured interviews, the study gathers quantitative and qualitative data from twenty-five pre-service teachers. The findings revealed that hands-on involvement in assessment design enhances teachers' understanding of key assessment principles, increases confidence, and bridges the gap between theory and practice. However, participants also faced challenges, particularly in ensuring the validity and reliability of their assessments. These challenges were manageable with the support provided in the course, indicating the value of structured guidance and expert collaboration in developing assessment literacy. The findings underscore the value of integrating theory with practice in teacher training, providing important insights for language assessment education. Future research could examine larger samples, long-term impacts of such training, and the role of digital tools in developing assessment literacy.

Keywords: language assessment; project-based teachers; assessment literacy; real-world tasks; teacher education

Introduction

Language assessment is a vital component of effective language teaching. However, many pre-service teachers struggle to apply

theoretical concepts, such as validity and reliability, to real-world contexts (Kianinezhad, 2023; Toquero, 2021). Research has highlighted that a lack of assessment literacy among educators, particularly in understanding and

***Corresponding Author:** Desi Surlitasari Dewi
(desisurlitasari@gmail.com) Jl. Pahlawan No.99,
Bukit Tempayan, Batu Aji, Batam, Riau, 29425

©2025 by the Authors, published by Vision: Journal for
Language and Foreign Language Learning
<https://journal.walisongo.ac.id/index.php/vision>

implementing effective assessment practices, hinders their ability to accurately evaluate students' language abilities and provide meaningful feedback (Deluca & Klingebiel. Furthermore, traditional training methods often focus on theoretical knowledge rather than practical skills, leaving pre-service teachers ill-prepared to design, administer, and analyze assessments in diverse classroom settings (Coombe et al., 2020; Holmes, 2011; Islam et al., 2021). As a result, pre-service teachers frequently graduate with a limited ability to design, administer, and evaluate effective language tests (Coombe et al., 2020; Islam et al., 2021; Kalay et al., 2020; Rezai et al., 2022; Vogt et al., 2020). Language assessment is a fundamental component of effective language teaching, as it enables educators to evaluate learners' skills and progress while informing instructional practices (Bachman & Palmer, 2010; Bailey & Heritage, 2014). However, pre-service teachers often struggle with the complexities of designing reliable and valid assessments (Audet, 2014; Lee, 2005; Monsalve, 2023). The ability to create assessments that accurately measure students' listening, speaking, reading, and writing skills is critical, particularly in diverse educational contexts (Coombe et al., 2020; Hughes, 2003; Islam et al., 2021; Rezai et al., 2022). Given the growing emphasis on evidence-based teaching and accountability in education, equipping pre-service teachers with practical language assessment skills has become increasingly important.

The role of language assessment in supporting effective language education has been widely acknowledged. It not only informs instructional practices but also serves as a critical

tool for evaluating student progress and abilities across key language skills, including listening, speaking, reading, and writing (Coombe et al., 2020; Hughes, 2003). However, research consistently reveals significant gaps in assessment literacy among both pre-service and in-service teachers, which undermines their ability to design, administer, and interpret assessments effectively (Hidayat et al., 2023; Islam et al., 2021; Rezai et al., 2022; Vogt et al., 2020). These deficiencies pose challenges across diverse educational contexts, where assessments must address varied learner needs while maintaining validity, reliability, and fairness (Alonzo & Teng, 2023).

Assessment literacy (AL) is recognized as a foundational competency for language teachers, encompassing knowledge of assessment principles, design skills, and the ability to interpret results for instructional purposes (Inbar-Lourie, 2008; Vogt et al., 2020). Recent studies emphasize that without adequate AL, teachers may struggle to create assessments that align with learning objectives, accurately measure students' abilities, and promote positive washback (Islam et al., 2021; Mirza & Gottardo, 2022; Rezai et al., 2022). Coombe et al. (2020) also highlight that while theoretical frameworks for assessment are available, many pre-service teachers lack hands-on experience in designing and administering tests, which limits their practical application skills.

The preparation of teachers to meet the demands of effective assessment has been identified as a significant gap in teacher education programs (Villa Larenas & Brunfaut, 2023). Previous studies (Dewi, 2015; Ghanbari & Nowroozi, 2021; Islam et al., 2021; Kianinezhad,

2023) report that many educators, especially in developing contexts, lack access to well-structured training programs that emphasize practical skills, such as portfolio, peer, and reflective assessment, and feedback. Similarly, Kalay et al. (2020) argue that integrating process-based assessments into the curriculum requires substantial teacher training to ensure alignment with frameworks such as the CEFR. However, insufficient preparation often leads teachers to rely on outdated, summative practices that do not fully capture student progress or foster deeper learning (Díaz et al., 2024; Nkonkanya Mpuangnan, 2024).

Recent research advocates for experiential learning approaches to bridge the gap between theoretical knowledge and practical application in assessment literacy. Firoozi et al. (2019) underscore the value of collaborative workshops where educators engage in real-world assessment tasks to build their skills. Similarly, Kalay et al. (2020) suggest that hands-on projects, as well as test design and administration, enable teachers to critically analyze their practices and incorporate concepts such as validity and reliability into their work. Ghanbari (2021) further stresses the importance of professional development that incorporates technology, allowing teachers to explore innovative methods for assessing language skills.

Despite growing recognition of the importance of AL, significant gaps persist in how pre-service teachers are prepared for assessment tasks. Coombe et al. (2020) and Rezai et al. (2022) note that many training programs fail to address the nuanced requirements of assessing diverse student populations, including considerations for cultural sensitivity and equitable evaluation.

Additionally, research by Islam et al. (2021) indicates that inadequate assessment practices in contexts such as Bangladesh not only affect teaching effectiveness but also impede student outcomes. These findings suggest an urgent need for teacher education programs to incorporate more comprehensive and practice-oriented assessment training.

While existing studies underscore the importance of assessment literacy and identify effective strategies to enhance teacher preparation, there is limited research on how experiential, project-based approaches affect pre-service teachers' understanding and readiness to conduct assessments. Most studies focus on in-service training (Ghanbari & Nowroozi, 2021; Rezai et al., 2022) or highlight theoretical gaps (Coombe et al., 2020; Vogt et al., 2020), leaving the specific challenges and perceptions of pre-service teachers underexplored. Studies have highlighted a persistent knowledge gap in how pre-service teachers develop the skills to connect theoretical concepts, such as test validity and reliability, to practical implementation. This gap limits their ability to critically analyze and apply assessment practices in real classrooms. Furthermore, little is known about how practical, project-based approaches to teaching language assessment influence pre-service teachers' perceptions and readiness to undertake such tasks in their future teaching careers.

To address this knowledge gap, this study explores the perceptions of pre-service teachers who engaged in real-world tasks as part of their Language Assessment course. Specifically, it investigates how designing, administering, and evaluating language assessments in authentic

school contexts impacts their understanding, skills, and attitudes. This study is grounded in the principles of Language Assessment Literacy (LAL) and in the core theoretical frameworks of language assessment, with an emphasis on the intersection of theory and practice. Language assessment serves as a tool for measuring language proficiency and a mechanism for driving effective teaching and learning. To create effective language assessments, educators must understand and apply foundational principles, which include validity, reliability, practicality, fairness, and washback (Hughes, 2003; Inbar-Lourie, 2008). The LAL framework informs the experiential learning tasks assigned to pre-service teachers. These tasks include designing tests, consulting with experts for validity checks, implementing tests in real schools, and conducting reliability analyses. By engaging with these principles in practical contexts, pre-service teachers can develop the competencies needed to integrate assessment effectively into their teaching practice. By doing so, the research seeks to provide insights into how practical, experiential learning can enhance the preparation of future language educators. Thus, this study aims to explore pre-service teachers' perceptions of the effectiveness of designing and administering real-world language tests in enhancing their understanding of assessment principles and practices. Additionally, it seeks to identify the challenges they encounter in designing, validating, and administering language assessments and to examine how manageable these challenges are with course support. Accordingly, this study addresses the following research objectives: to explore pre-service teachers' perceptions of the effectiveness of designing and administering real-world language assessments in enhancing

their assessment literacy, and to examine the challenges they encounter during the process and how they manage these challenges with support provided in the course.

Method

This study employs a convergent mixed-methods case study design to explore pre-service teachers' perceptions of designing and administering language assessments as part of a Language Assessment course (Creswell & Plano Clark, 2018). A mixed-methods approach was chosen because it allows researchers to integrate quantitative data from a closed questionnaire and qualitative insights from semi-structured interviews, thereby yielding a more comprehensive understanding of the research problem (Schoonenboom & Johnson, 2017). A case study approach was chosen. After all, it enables an in-depth exploration of participants' experiences in a real-life educational setting, focusing on a bounded system to understand complex phenomena in context (Merriam & Tisdell, 2016; Yin, 2018).

Participants

The participants in this study were 52 pre-service teachers enrolled in a Language Assessment course in the English Education Study Program at Universitas Riau Kepulauan. All participants were preparing for careers as English teachers and had completed the course activities, which involved designing, administering, and evaluating various types of language tests for real-world use in schools. A purposive sampling technique was employed to ensure that the participants possessed relevant knowledge and experience aligned with the study's objectives (Etikan, Musa, & Alkassim, 2016). Purposive sampling is widely used in

educational research because it allows the selection of individuals who can provide rich, meaningful data through their direct involvement in the learning process (Palinkas et al., 2015). From this group, 12 participants were further selected for semi-structured interviews to gain deeper insights into their perceptions and experiences. The selection aimed for maximum variation in gender, teaching experience, and engagement levels in the project activities to capture a diverse range of perspectives. The focus on this group was driven by the increasing demand for future teachers who not only understand language assessment theory. However, it can also be applied effectively in authentic classroom contexts.

Data Collection and Instruments

Data for this study were collected using a closed questionnaire and semi-structured interviews to gather both quantitative and qualitative insights into pre-service teachers' experiences with designing, validating, and administering language assessments. A closed questionnaire was developed to gather quantitative data on pre-service teachers' perceptions of the effectiveness of designing and administering real-world language assessments, as well as the challenges they faced throughout the process. The questionnaire consisted of 10 statements covering various aspects of language assessment, including motivation, understanding of assessment principles, confidence in assessment practices, and perceived effectiveness of course support. Respondents were asked to rate each statement on a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). To ensure content validity, the questionnaire was

reviewed by two experts in language assessment and educational measurement. Their feedback led to minor revisions in the wording and item sequence for clarity and relevance. A pilot test was conducted with five pre-service teachers who were not part of the main study to evaluate item clarity and response consistency. The instrument's reliability was assessed using Cronbach's Alpha ($\alpha = 0.87$), indicating high internal consistency.

In addition to the closed questionnaire, semi-structured interviews were conducted with six pre-service teachers to gain a deeper understanding of their experiences, perceptions, and challenges in designing, validating, and administering language assessments. The semi-structured format allowed for flexibility in questioning, enabling participants to express their views in detail while ensuring that all relevant topics related to the research questions were covered. The interviews were guided by a set of open-ended questions that encouraged participants to reflect on their experiences, including the challenges they encountered, the support they received from the course, and how the project impacted their understanding of assessment principles. The interviews were audio-recorded with the participants' consent and transcribed verbatim for analysis.

Data Analysis

The data collected from the closed questionnaire were analyzed quantitatively using the mean and standard deviation. These statistical measures were used to determine the central tendency and variability of respondents' perceptions of the effectiveness of designing and administering real-world language tests, as well as the challenges they encountered. The

mean values provided an overall sense of the participants' views on each statement. At the same time, the standard deviation indicated the degree of variability in their responses. The results were then analyzed to identify trends and insights into how pre-service teachers perceive the assessment design process and the support they received.

For the interview data, a thematic analysis was conducted, following the procedure outlined by Braun & Clarke (2006). First, the interviews were transcribed verbatim, and the data were read repeatedly to familiarize the researcher with the content. During this stage, the researcher noted initial impressions such as "connecting theory to practice," "increased motivation," and "confidence after real-world experience." Next, initial codes were generated by identifying significant patterns and relevant responses related to the research questions. For example, statements like "Designing and validating the test made me realize how critical it is to ensure that every item measures the intended skill" were coded as enhanced understanding of validity, while "Administering my test in a real school setting was an invaluable experience" was coded as application of assessment principles in authentic contexts. Similarly, excerpts such as "Collaborating with another rater for reliability analysis was challenging but beneficial" were coded as collaborative skill development, and "Seeing how the students responded to my test made me feel proud and motivated" as increased motivation and ownership. These codes were then examined for similarities and differences and grouped into broader themes. The themes were reviewed and refined to accurately capture the participants' experiences and perspectives.

Finally, the themes were labeled and described in detail, with supporting excerpts from the interview transcripts provided to illustrate the findings. This combined approach enabled a comprehensive analysis of both quantitative and qualitative data, providing a well-rounded understanding of pre-service teachers' perceptions of the effectiveness of real-world language assessments and the challenges they encountered.

Findings and Discussion

Finding

Pre-service Teachers' Perceived Impact of Designing and Administering Real-world Language Assessments

The first research question was addressed through a closed questionnaire and semi-structured interviews with six pre-service teachers. While the closed questionnaire results were analyzed quantitatively, their responses were analyzed and grouped into recurring themes that provide insights into their perceptions. The results of the closed questionnaire are displayed in Table 2.

The data in Table 2 indicate that pre-service teachers generally perceived the project positively, with an overall mean score of 4.24 (SD = 0.64), suggesting strong agreement across most items. The highest mean score was obtained for the statement "Administering the tests in a real school setting helped me connect theory to practice" (M = 4.50, SD = 0.52), indicating that students found authentic engagement in real classrooms to be the most valuable learning component. It suggests that experiential exposure to real assessment contexts effectively strengthened their ability to

translate theoretical knowledge into practice. The next highest ratings were for "I was excited to see how students performed on the tests I created" ($M = 4.46$, $SD = 0.59$) and "Designing tests for listening, speaking, reading, and writing improved my understanding" ($M = 4.38$, $SD = 0.60$). These results demonstrate that participants found the project intrinsically motivating and intellectually rewarding, implying that practical involvement can foster both emotional engagement and cognitive understanding in learning assessment.

Statements related to interaction with experts and evaluation of test quality also received high ratings, such as "Checking the content and construct validity with experts helped me better understand how to evaluate test quality" ($M = 4.32$, $SD = 0.65$) and "Interacting with experts to validate my tests was a valuable learning experience" ($M = 4.28$, $SD = 0.71$). It indicates that professional feedback was instrumental in deepening participants' awareness of assessment validity and reliability.

Table 1

The Perceived Impact of the Project

No.	Statement	Mean	SD
1	The project of designing and administering tests made the course more engaging.	4.02	0.57
2	Creating real-world language tests motivated me to learn more about assessment principles.	4.18	0.61
3	I was excited to see how students performed on the tests I created.	4.46	0.59
4	Designing tests for listening, speaking, reading, and writing improved my understanding.	4.38	0.60
5	Administering the tests in a real school setting helped me connect theory to practice.	4.50	0.52
6	Checking the content and construct validity with experts helped me better understand how to evaluate test quality.	4.32	0.65
7	Working with another rater for inter-rater reliability gave me valuable insight.	4.00	0.76
8	Analyzing the validity and reliability of my tests taught me to evaluate assessment quality.	4.14	0.67
9	Interacting with experts to validate my tests was a valuable learning experience.	4.28	0.71
10	Collaborating with another rater improved my ability to work as part of a professional team.	4.22	0.69
11	I was able to effectively communicate with peers and experts during this project.	4.16	0.64

Meanwhile, the relatively lower mean for "Working with another rater for inter-rater reliability gave me valuable insight" ($M = 4.00$, $SD = 0.76$) suggests that while collaboration was

beneficial, some participants may have found it more challenging than individual work. Nonetheless, related items such as "Collaborating with another rater improved my

ability to work as part of a professional team" ($M = 4.22$, $SD = 0.69$) and "I was able to effectively communicate with peers and experts during this project" ($M = 4.16$, $SD = 0.64$) show that most participants viewed teamwork and communication as positive components of the learning process.

In summary, the results reveal that the project was effective in enhancing pre-service teachers' engagement, motivation, and understanding of assessment principles, especially through real-world application and professional collaboration. The consistently high mean scores indicate that practical, experience-based assessment projects can serve as a powerful approach to developing assessment literacy among future educators.

Meanwhile, to support the results of the closed questionnaire, interviews with six pre-service teachers were conducted, revealing several key themes. Participants consistently reported that the process deepened their theoretical understanding of assessment principles, particularly in validity, reliability, and alignment with learning objectives. Respondent 1 shared, "Before this project, I only had a vague idea about what validity meant. Designing and validating the test made me realize how critical it is to ensure that every item measures the intended skill." Respondent 4 noted, "Working on real tests helped me connect the theories I learned in class with their practical applications. It was eye-opening to see how these principles guide every step of the process." Compared to learning only through lectures and class assignments, participants found that the hands-on experience provided a richer, more meaningful understanding of assessment principles.

Participants also highlighted that administering the tests in real classrooms bridged the gap between theory and practice, giving them a clearer sense of the practical challenges and considerations in assessment. As Respondent 3 remarked, "Administering my test in a real school setting was an invaluable experience. It taught me how to adjust my approach based on the classroom environment and student responses." This sentiment was echoed by Respondent 6, who stated, "I realized that even a well-designed test can face unexpected issues during administration. This experience made me appreciate the importance of preparation and adaptability." The opportunity to apply what they learned in real-world settings was seen as far more impactful than completing tasks in a controlled classroom environment, where such challenges are not encountered.

Designing and administering real-world language tests also motivated participants to become more involved in language assessment, as they saw the tangible impact of their efforts. Respondent 2 stated, "Seeing how the students responded to the tests I created made me feel proud and motivated to learn more about making assessments even better." Respondent 5 added, "I felt like I was contributing something meaningful to the learning process. This experience inspired me to explore more about how assessments can drive student progress." Participants emphasized that working on a real-world project was more engaging and motivating than merely completing theoretical tasks in class.

The project required participants to collaborate with peers and experts, fostering teamwork and critical analysis of their work.

Respondent 1 mentioned, "Discussing my test items with my peers and getting feedback from experts made me realize the value of multiple perspectives. It helped me refine my test and understand my own biases." Meanwhile, Respondent 3 shared, "Collaborating with another rater for reliability analysis was challenging but extremely beneficial. It taught me how to evaluate tests more objectively." Participants appreciated the opportunity to work with others and noted that the collaborative aspects of the project offered a level of learning and reflection that individual classroom tasks could not.

Finally, participants reported that the project significantly boosted their confidence in designing and administering assessments, making them better prepared for future teaching roles. Respondent 4 explained, "After completing this project, I feel much more confident about creating my own assessments. I know how to align them with learning goals and evaluate their quality." Respondent 6 shared a similar experience, stating, "This was the first time I created something like this from scratch. By the end, I felt capable of not just designing a test but also explaining why it was effective." Several other respondents noted that this practical experience was far more effective in building their confidence than merely studying theory or completing assignments in the classroom. They felt that facing real challenges in authentic settings prepared them better for future professional tasks.

The interview findings reveal that pre-service teachers perceive designing and administering real-world language tests as an effective way to enhance their understanding of assessment principles and practices. Themes such as

improved theoretical knowledge, practical skill development, increased motivation, collaboration, and confidence-building emerged as central to their experiences. Participants emphasized that this project was more impactful than traditional classroom-based learning, as it allowed them to directly apply theory in authentic contexts, face real-world challenges, and develop a stronger sense of preparedness for their teaching careers. It highlights the value of experiential learning in teacher education, as it bridges the gap between theory and practice and fosters professional growth.

Pre-service Teachers' Challenges in Designing and Administering Real-world Language Assessments

To address the second research question, a thorough analysis is provided by combining questionnaire and interview data. The results of the closed questionnaire are displayed in Table 2.

The data in Table 2 shows participants' perceptions of the challenges and support they experienced while designing, validating, and administering language tests, as well as their overall evaluation of the project's effectiveness. Participants acknowledged encountering challenges during the process (mean = 3.96, SD = 0.77), indicating that while the project was rigorous, these difficulties were manageable. The slightly lower score for the manageability of challenges with course guidance (mean = 3.76, SD = 0.51) suggests that although participants generally felt supported, some may have required additional resources or guidance to overcome obstacles effectively.

Despite the challenges, the course instructor's support was highly rated (mean = 4.04, SD = 0.72), reflecting the importance of mentorship and structured support in helping participants succeed. The project itself was perceived as an effective way to learn about designing and administering assessments, as

evidenced by a strong mean score of 4.16 (SD = 0.73). Moreover, participants expressed confidence that the experience prepared them for real-world teaching, with the highest mean score in the table (4.20, SD = 0.75), indicating that they felt equipped to apply these skills in professional settings.

Table 2

Perceived Challenges of the Project

Statement	Mean	Standard Deviation
I encountered challenges in designing, validating, or administering the tests.	3.96	0.77
The challenges I faced were manageable with the guidance provided in the course.	3.76	0.51
The course instructor provided sufficient support to help me complete the project.	4.04	0.72
This project was an effective way to learn about designing and administering assessments.	4.16	0.73
Completing this project has prepared me to create and administer tests in real teaching.	4.20	0.75

To complement the quantitative findings, interviews were conducted with 12 pre-service teachers who participated in the project to design and administer real-world language assessments. The interviews revealed five major themes: (1) deepened theoretical understanding, (2) development of practical assessment skills, (3) motivation and professional identity, (4) collaboration and reflective engagement, and (5) challenges and tensions in applying theory to practice. Together, these themes offer a comprehensive picture of how authentic assessment tasks influence pre-service teachers' learning experiences and professional growth.

Deepened Theoretical Understanding

Nearly all participants emphasized that their involvement in designing and validating

language tests enabled them to move beyond textbook definitions and gain a more grounded understanding of fundamental concepts such as validity, reliability, and fairness. Respondent 1 explained, "Before doing this project, I thought validity was just about making good questions. But now I understand it means making sure every item actually measures the skill it's supposed to." Similarly, Respondent 8 reflected, "When I compared my scoring with another rater, I finally understood why reliability matters, it's about being consistent and fair, not just about getting the same result." Several participants shared that working through the process of aligning test items with learning objectives helped them recognize how theory underpins every design decision. Respondent 3 noted, "At first, I didn't see why test blueprints were so important. But when I started writing

items, I realized how they ensure that every question connects to the course goals." Through hands-on experience, participants internalized the meaning of assessment principles that previously seemed abstract. They began to see assessment as a system guided by logic and evidence, rather than a mechanical classroom task.

Development of Practical Assessment Skills

A central outcome of the project was the development of practical competence in creating, administering, and analyzing assessments. Participants repeatedly described how conducting their tests in real classrooms revealed the complexities of the assessment process. Respondent 10 explained, "When I gave my test to students, I noticed that some instructions were confusing, even though they looked fine on paper. I had to clarify things on the spot." Respondent 2 echoed this, stating, "I learned that designing a test is one thing, but making sure students understand it and feel comfortable is another challenge." These experiences allowed participants to link theoretical understanding to authentic teaching conditions. Respondent 5 described, "Administering the test taught me how to manage time, observe students, and adapt when something unexpected happens. It's very different from just talking about it in class." Several participants commented that calculating validity and reliability helped them see how quantitative data could inform qualitative judgment. Respondent 12 said, "After analyzing my test results, I could see which items didn't perform well. That was the moment I really understood how theory guides improvement." The participants' comments reflect a growing

sense of agency as they learned to handle complex, real-world situations. The project turned theoretical assessment literacy into practical competence, enabling them to make informed decisions in designing and evaluating their own tests.

Motivation and Professional Identity

Many participants expressed that the project fostered a stronger sense of motivation and professional identity. Seeing their work applied in real school contexts gave them a sense of purpose and ownership. Respondent 6 remarked, "It was exciting to see the students take my test. I felt like a real teacher for the first time." Similarly, Respondent 9 said, "When I saw that my test actually helped students reflect on their learning, I felt proud. It made me more confident about becoming a teacher." For several participants, the experience reshaped their view of assessment. Rather than seeing it as a routine requirement, they began to perceive it as a meaningful part of teaching that influences learning outcomes. Respondent 11 stated, "Before, I thought tests were just for grading. Now I see them as tools to understand students better and improve instruction."

However, not all participants felt equally confident throughout the process. Respondent 4 shared, "I was nervous about whether my test would really work. I worried that it might not be good enough." This uncertainty, though challenging, pushed participants to engage more deeply with the principles they were applying. The combination of pride and self-doubt highlighted how authentic experiences can

both challenge and strengthen the formation of professional identity.

Collaboration and Reflective Engagement

Another recurring theme was the role of collaboration and reflection in learning. Participants worked closely with peers, co-raters, and expert reviewers throughout the project. Respondent 7 described, "Discussing my test items with my peers made me realize how differently people interpret the same question. It helped me think more carefully about wording." Respondent 3 added, "Getting feedback from an expert was really valuable. Sometimes their comments were hard to accept, but they made me improve the quality of my test." Working with another rater to ensure reliability also prompted participants to reflect on their objectivity and decision-making. Respondent 10 stated, "At first, my scoring was different from my partner's. We had to negotiate and agree on criteria. That discussion helped me see how subjective assessment can be." Respondent 2 agreed, saying, "Collaboration forced me to question my assumptions and be more consistent."

Although most found collaboration beneficial, some participants admitted that managing group work was challenging. Respondent 12 mentioned, "It was sometimes difficult to coordinate with my peer because we had different working styles." Nonetheless, participants recognized that these interpersonal challenges mirrored real-world professional experiences. Through collaboration, they learned that effective assessment requires

not only technical skill but also open communication and teamwork.

Challenges and Tensions in Applying Theory to Practice

Despite overall positive experiences, participants also encountered difficulties that tested their patience and adaptability. Several noted the workload and time pressure involved in completing each phase of the project. Respondent 5 stated, "It was exhausting to handle everything; designing, validating, administering, and analyzing, all within one semester." Respondent 8 shared a similar experience: "I sometimes felt overwhelmed by how many details we had to check. It was easy to get lost in the technical parts." Others reflected on the tension between theoretical expectations and real-world conditions. Respondent 9 explained, "Sometimes the classroom situation didn't match the ideal procedures we learned. I had to make quick adjustments." These experiences revealed that while theory provides a foundation, practice often demands flexibility and critical judgment. Interestingly, a few participants expressed frustration when their tests did not perform as expected. Respondent 11 remarked, "I was disappointed when some students misunderstood the tasks, even though I thought my instructions were clear." Nevertheless, such moments became valuable learning opportunities, prompting deeper reflection about communication, fairness, and learner diversity.

The interview findings indicate that engaging in real-world language assessment projects transformed how pre-service teachers understood, practiced, and valued assessment. They learned to translate theoretical knowledge

into concrete decisions, developed practical competence, built confidence as future educators, and cultivated reflective habits through collaboration. While participants acknowledged challenges such as workload, uncertainty, and differing perspectives, they viewed these difficulties as integral to their professional growth. It suggests that authentic assessment experiences not only enhance conceptual understanding but also shape the attitudes, identities, and reflective capacities necessary for effective teaching practice.

Discussion

This study explored pre-service teachers' perceptions of the effectiveness of designing and administering real-world language tests (RQ1) and the challenges they encountered in this process (RQ2). The findings provide insights into how such experiential learning opportunities enhance assessment literacy and address gaps in teacher training, aligning with and extending existing literature on language assessment practices and challenges. Unlike prior research that primarily focuses on theoretical assessment training (Coombe et al., 2020; Islam et al., 2021; Rezai et al., 2022), this study contributes empirical evidence on how direct engagement with test development and implementation influences pre-service teachers' understanding of key assessment principles, motivation, and professional preparedness.

The results revealed that pre-service teachers perceive designing and administering real-world language tests as highly effective in enhancing their understanding of assessment principles and practices. Participants reported improved knowledge of key concepts such as validity, reliability, and fairness, which are essential components of assessment literacy

(Inbar-Lourie, 2008). The project's practical nature helped bridge the gap between theoretical knowledge and its application, supporting findings from prior studies on experiential learning (Firoozi et al., 2019; Kalay et al., 2020). However, our study extends these findings by emphasizing that direct involvement in test administration, beyond test design alone, is a crucial factor in reinforcing assessment concepts. While pre-service teachers initially grasped theoretical principles during the test design phase, it was the process of administering these assessments that solidified their understanding of validity, reliability, and fairness.

In contrast to Rütli-Joy (2024), who identified only the lack of assessment literacy as the ultimate barrier, this study found that participants encountered real-time challenges, including adapting test conditions to diverse student needs, ensuring consistent scoring criteria, and managing unexpected variables such as student anxiety or time constraints. These experiences compelled them to apply assessment principles dynamically rather than in a controlled, hypothetical context, thereby deepening their comprehension. Unlike previous studies that focus primarily on pre-service teachers' conceptual knowledge of assessment (Islam et al., 2021; Vogt et al., 2020), our findings indicate that experiential learning through test administration fosters a more applied understanding of assessment literacy. Without this hands-on experience, pre-service teachers might struggle to anticipate and address practical assessment issues in their future classrooms.

One of the most significant contributions of this study is the identification of challenges pre-

service teachers face in this process, particularly the difficulty of reconciling expert feedback with their own understanding of assessment principles. While prior research has highlighted a general lack of assessment literacy among teachers (Islam et al., 2021; Vogt et al., 2020), it has not sufficiently addressed the cognitive and emotional struggles that arise when pre-service teachers engage with expert validation. In this study, pre-service teachers were required to engage with expert evaluations of their test designs and scoring rubrics, often receiving conflicting feedback from different raters. While this process exposed them to diverse assessment viewpoints, it also led to uncertainty in decision-making, as participants struggled to determine which feedback to prioritize and how to reconcile discrepancies in scoring. Some participants reported that while expert critiques improved their awareness of validity and reliability issues, they felt overwhelmed when expected to integrate differing opinions into a cohesive revision of their assessments. Without explicit guidance on evaluating and synthesizing multiple inputs, pre-service teachers may experience cognitive overload rather than clarity. Therefore, teacher education programs should not only provide expert feedback but also include structured reflection and decision-making strategies to help pre-service teachers navigate conflicting evaluations effectively.

Moreover, this study highlights logistical and situational challenges that pre-service teachers encountered while administering their tests. Prior research (Díaz et al., 2024; Nkonkanya Mpuangnan, 2024) has emphasized the need for teachers to adapt assessments for diverse student populations. This study goes further by underscoring the additional stressors

posed by real-world constraints, such as scheduling difficulties and unpredictable student behavior. It revealed that pre-service teachers encountered significant real-world difficulties when administering their assessments, including logistical constraints such as scheduling conflicts, student disengagement, and unexpected classroom dynamics. While they demonstrated improved understanding of assessment principles such as validity and reliability, many participants struggled to adapt their test administration plans in unpredictable circumstances. Several participants expressed frustration over having to modify test conditions on the spot without clear guidelines, highlighting a gap in their preparation. Therefore, teacher education programs should not only strengthen theoretical and technical training but also integrate structured problem-solving exercises and case-based learning to help pre-service teachers develop the practical resilience needed for effective assessment implementation.

Collaboration was another area where this study both aligns with and challenges previous findings. Prior studies (Firoozi et al., 2019; Ghanbari & Nowroozi, 2021) have emphasized the benefits of collaborative learning in fostering AL. The finding supports this claim, as indicated by the high mean score (4.24, SD = 0.71) for the statement "Collaborating with another rater improved my ability to work as part of a professional team." However, it also reveals that collaboration is not always straightforward. The slightly lower mean (3.92, SD = 0.80) for "Working with another rater for inter-rater reliability gave me valuable insight." Qualitative data indicating struggles with inconsistent feedback suggest that collaboration can also

create cognitive challenges when pre-service teachers must reconcile conflicting evaluations. This finding extends the discussion on collaborative learning by suggesting that effective teamwork in assessment training requires explicit guidance on synthesizing differing perspectives. In real-world projects, pre-service teachers may experience cognitive overload, making it difficult for them to confidently apply assessment principles. Kalay et al. (2020) have emphasized the importance of structured training in process-based assessment. Moreover, the study reveals that unstructured can lead to confusion rather than clarity.

The study also highlights the importance of structured support mechanisms in mitigating these challenges. Previous studies focused on professional development for in-service teachers (Rezai et al., 2022). However, they did not specify the types of support most beneficial for pre-service teachers. This study highlights that pre-service teachers may experience higher levels of uncertainty and anxiety when engaging with assessment practices for the first time. Therefore, teacher education programs should offer hands-on mentorship opportunities that help pre-service teachers build confidence in their assessment literacy while managing the cognitive and emotional demands of the process. In this study, structured support mechanisms such as expert feedback sessions, guided peer discussions, and instructor-led reflections played a crucial role in helping pre-service teachers navigate the complexities of test design and administration. These supports provided clarity in resolving assessment discrepancies, alleviated anxiety associated with

decision-making, and reinforced theoretical concepts through practical application.

Despite these contributions, this study has limitations. First, it relies on self-reported data, which may be subject to social desirability bias or individual differences in self-perception. Future research could incorporate observational data or performance-based assessment to validate findings. Second, while our study was conducted in a pre-service teacher education program, findings may not be fully generalizable to in-service teachers or educators in different cultural contexts. Comparative studies across various educational settings would provide further insight into how assessment literacy develops over time. Additionally, the study did not examine long-term retention of AL skills. Future research could explore how pre-service teachers apply these skills in real teaching contexts after graduation.

Conclusion

This study examined pre-service teachers' perceptions of designing and administering real-world language assessments, highlighting both the benefits and challenges of this experiential learning approach. The findings demonstrate that engaging in real project-based approach tasks significantly enhances assessment literacy by strengthening understanding of key concepts such as validity, reliability, and fairness. Importantly, the study shows that moving beyond theory to practical application allows pre-service teachers to internalize assessment principles in a more meaningful and lasting way than through traditional coursework alone.

Despite these contributions, this study has limitations that should be acknowledged. The research was conducted within a specific

teacher education context, which may affect the generalizability of the findings to different institutional settings or teacher training models. Additionally, while the study captured self-reported perceptions and qualitative insights, further research incorporating longitudinal observations or experimental designs could provide a more comprehensive understanding of how pre-service teachers internalize assessment literacy over time. The implications of these findings underscore the need for teacher education programs to integrate structured, hands-on assessment training with clear support mechanisms. Providing expert mentorship, scaffolding collaborative discussions, and equipping pre-service teachers with strategies to navigate real-world assessment challenges can ensure that they develop the confidence and competence required for effective classroom assessment practices. Future research could explore the long-term impact of such training on in-service teachers or investigate how digital tools and assessment technologies might further support the development of assessment literacy in teacher education. By addressing these areas, teacher education programs can better prepare future educators to implement valid, reliable, and fair assessments that support student learning.

References

- Alonzo, D., & Teng, S. (2023). Trustworthiness of Teacher Assessment and Decision-Making: Reframing the Consistency and Accuracy Measures. *International Journal of Instruction*, 16(3), 1075–1094. <https://doi.org/10.29333/iji.2023.16357a>
- Audet, K. K. (2014). Assessing Pre-Service Teacher Beliefs: Development of the Effective Teacher Efficacy survey. In *Montana State University* (Issue November). Montana State University.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice : developing language assessments and justifying their use in the real world*. Oxford University Press. <https://cir.nii.ac.jp/crid/1130282272591332352.bib?lang=en>
- Bailey, A. L., & Heritage, M. (2014). The role of language learning progressions in improved instruction and assessment of english language learners. *TESOL Quarterly*, 48(3), 480–506. <https://doi.org/10.1002/tesq.176>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy : what do we need to learn, unlearn, and relearn? *Language Testing in Asia*, 10(3), 1–16. <https://doi.org/10.1186/s40468-020-00101-6>
- Deluca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher Candidates' learning. *Assessment in Education: Principles, Policy and Practice*, 17(4), 419–438. <https://doi.org/10.1080/0969594X.2010.516643>
- Dewi, D. S. (2015). Corrective Feedback in Class. *Anglo-Saxon*, 8, 1–12.
- Díaz, B., Nussbaum, M., Greiff, S., & Santana, M. (2024). The role of technology in reading literacy: Is Sweden going back or moving forward by returning to paper-based

- reading? *Computers & Education*, 213, 105014.
<https://doi.org/10.1016/j.COMPEDU.2024.105014>
- Firoozi, T., Razavipour, K., & Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, 9(2), 1–14.
<https://doi.org/10.1186/s40468-019-0078-7>
- Ghanbari, N., & Nowroozi, S. (2021). The practice of online assessment in an EFL context amidst COVID-19 pandemic: views from teachers. *Language Testing in Asia*, 11(27), 1–18.
- Hidayat, M. L., Hariyatmi, Astuti, D. S., Sumintono, B., Meccawy, M., & Khanzada, T. J. S. (2023). Digital competency mapping dataset of pre-service teachers in Indonesia. *Data in Brief*, 49(109310), 1–12.
<https://doi.org/10.1016/j.dib.2023.109310>
- Holmes, S. R. (2011). *Teacher Preparedness for Teaching and Assessing Depth of Knowledge* (Issue May). The University of Southern Mississippi.
- Hughes, A. (2003). *Testing for Language Teachers Second Edition*. Cambridge University Press.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402.
<https://doi.org/10.1177/0265532208090158>
- Islam, S., Hasan, K., Sultana, S., Karim, A., & Rahman, M. M. (2021). English language assessment in Bangladesh today: principles, practices, and problems. *Language Testing in Asia*, 11(1), 1–21.
<https://doi.org/https://doi.org/10.1186/s40468-020-00116-z>
- Kalay, S., Büyükkarci, K., & English, K. (2020). English Language Teachers' Views on Teaching and Assessment of Writing Skills. *SDU International Journal of Educational Studies*, 7(2), 262–286.
<https://doi.org/10.33710/sduijes.710062>
- Kianinezhad, N. (2023). Language Assessment Literacy: The Role of Self-Reflection. *Journal of English Language and Culture*, 14(1), 40–55.
<https://doi.org/10.30813/jelc.v14i1.4662>
- Lee, H. J. (2005). Understanding and assessing pre-service teachers' reflective thinking. *Teaching and Teacher Education*, 21(6), 699–715.
<https://doi.org/10.1016/j.tate.2005.05.007>
- Mirza, A., & Gottardo, A. (2022). The effects of second language literacy instruction on first language literacy: a comparison between Hindi–English and Urdu–English Canadian bilinguals. *Journal of Cultural Cognitive Science*, 6(2), 229–248.
<https://doi.org/10.1007/s41809-022-00100-4>
- Monsalve, E. F. R. (2023). English Assessment Course as Meaningful Learning for the Construction of Teacher Identity Elkin. In *Universidad Católica Luis Amigó*. Universidad Católica Luis Amigó.
- Nkonkonya Mpuangnan, K. (2024). Determining skills of basic schoolteachers in test item construction. *Review of Education*, 12(2), 1–19. <https://doi.org/10.1002/rev3.3478>
- Rezai, A., Namaziandost, E., Miri, M., & Kumar, T. (2022). Demographic biases and assessment fairness in classroom: insights from Iranian university teachers. *Language Testing in Asia*, 12(8), 1–20.

- <https://doi.org/https://doi.org/10.1186/s40468-022-00157-6>
- Rütti-Joy, O. (2024). Applying Peer Rubric Feedback in Foreign Language Teacher Education: Fostering Pre-service Teachers' Oral Feedback Skills Through Facilitating Reflective Moments BT - Tools, Techniques and Strategies for Reflective Second & Foreign Language Teacher Educat. In P. Voerke, M. A. Vaz Ferreira, & N. Drescher (Eds.), *Tools, Techniques and Strategies for Reflective Second & Foreign Language Teacher Education*. (pp. 31–52). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-68741-3_3
- Toquero, C. M. D. (2021). "Real-world:" pre-service teachers' research competence and research difficulties in action research. *Journal of Applied Research in Higher Education*, 13(1), 126–148. <https://doi.org/10.1108/JARHE-03-2019-0060>
- Villa Larenas, S., & Brunfaut, T. (2023). But who trains the language teacher educator who trains the language teacher? An empirical investigation of Chilean EFL teacher educators' language assessment literacy. *Language Testing*, 40(3), 463–492. <https://doi.org/10.1177/02655322221134218>
- Vogt, K., Tsagari, D., Csépes, I., Green, A., Sifakis, N., Vogt, K., Tsagari, D., Csépes, I., Green, A., & Sifakis, N. (2020). Linking Learners' Perspectives on Language Assessment Practices to Teachers' Assessment Literacy Enhancement (TALE): Insights from Four European Countries Linking Learners' Perspectives on Language Assessment Practices to Teachers' Assessment Litera. *Language Assessment Quarterly*, 17(4), 410–433. <https://doi.org/10.1080/15434303.2020.1776714>