



Submitted : May, 10th, 2025

Revised : May, 15th, 2025

Accepted : May, 15th, 2025

Published : June, 11th, 2025

**A QUALITATIVE OBSERVATIONAL STUDY ON STUDENTS'
ENGAGEMENT AND CHALLENGES IN LEARNING ENGLISH
ASSESSMENT: EXPLORING ENGLISH LANGUAGE TEST
DEVELOPMENT**

Muhammad Yasminto¹

antoenyaz83@gmail.com

Abstract

This study investigates students' engagement and challenges in learning English assessment, with a particular emphasis on the process of test development. Conducted through a qualitative observational approach, the research draws from a lecturer's classroom experience in teaching a language assessment course to English education students at the university level. Observation notes, reflective teaching journals, and student interactions served as primary data sources. The findings reveal that while students exhibit high levels of curiosity and participation during hands-on activities, such as item writing, rubric development, and peer assessment, they often encounter difficulties in understanding abstract concepts related to validity, reliability, and test standardization. These challenges stem from unfamiliar terminology, limited prior exposure to assessment design, and a theoretical overload without sufficient contextual grounding. However, when assessment theory is integrated with experiential learning, such as collaborative projects, simulations, and peer feedback, students show improved comprehension and a deeper sense of assessment literacy. The study highlights the value of scaffolded learning, reflective teaching, and interactive engagement in fostering students' ability to design and analyze language tests effectively. The research contributes to pedagogical strategies for improving language assessment instruction in teacher education programs and suggests that student-centered, practice-based approaches are essential to bridging the gap between theory and application in assessment learning

Keywords : English Assessment, Students' Engagement, Test Development

¹ Lecturer of Tadris Bahasa Inggris, IAI Hasanuddin Pare



A. INTRODUCTION

In the field of English language education, assessment plays a pivotal role not only in measuring student performance but also in shaping instructional practices and learning outcomes. Effective language assessment requires a careful balance between theoretical knowledge and practical application. Therefore, equipping students of English education study program with the ability to design, analyze, and reflect on language tests is an essential component. However, despite its importance, the process of learning English assessment, particularly test development, is often perceived as complex and cognitively demanding by students.²

In addition, over the past decade, English language assessment has become an essential component in language education, especially in higher education institutions where language proficiency is often a key requirement for academic and professional advancement. However, based on researcher's experiences as a lecturer teaching language assessment to undergraduate students of sixth semester on English education study program, Islamic Institute of Hasanuddin Pare, it has become increasingly evident that many students face both cognitive and affective challenges in engaging with the subject. These challenges often stem from the abstract nature of test development concepts, unfamiliar technical terminology, and the perceived disconnect between theory and practical application.

Language assessment is frequently viewed by students as a demanding and rigid subject. During classroom observations and interactions, he noticed varying levels of student engagement from enthusiastic participation during hands-on test design activities to visible frustration during discussions on complex frameworks such as reliability, validity, and washback. These observations raise crucial questions about how students perceive the learning process of assessment and what factors contribute to or hinder their understanding.

In the field of English language education, the development of effective and reliable assessment tools is a critical component of both teaching and learning. As language proficiency becomes increasingly important in academic, professional, and global contexts, there is a growing need for well-constructed language tests that can accurately evaluate learners' abilities across various skill domains. English language test development is not a linear task; it requires

² Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Routledge.



a rigorous and reflective process grounded in sound assessment principles. Among these, validity and reliability serve as the two fundamental pillars that determine the overall quality and usefulness of a test. Validity ensures that the test truly measures the intended language construct, while reliability guarantees the consistency of test results across time, tasks, and raters. Without these qualities, test outcomes may be misleading or unfair, potentially affecting learners' academic or professional opportunities. Therefore, understanding and applying these principles is essential for educators, language testers, and curriculum designers who aim to create meaningful, fair, and effective assessments in the English language classroom.

Much of the existing research has centered on teachers' knowledge and perceptions of assessment practices, often relying on self-reported data through surveys and questionnaires.³ These studies, while valuable, tend to overlook how students engage with language assessment concepts during actual instructional practices, especially in the context of developing language test items.

Moreover, prior studies have primarily emphasized theoretical knowledge rather than experiential or practical learning. There is a notable absence of classroom-based qualitative studies that explore how students internalize and apply test development principles, particularly in Indonesian EFL settings. Little is known about the real-time challenges students face, how they respond to feedback, and how collaborative work affects their learning during the construction of assessment tools.

Therefore, this study addresses a critical gap by observing students' behaviors, interactions, and difficulties during their engagement with English language test development tasks. By using a qualitative observational approach, this research aims to provide a deeper understanding of how language assessment knowledge is experienced and negotiated in a real educational context.

Research has highlighted the value of active engagement in assessment education, particularly through hands-on experiences such as peer review, test construction, classroom

³ Mohammad Farid, "An Analysis of Students' Perception Toward English Language Assessment in Indonesian EFL Classrooms," *Jurnal Pendidikan Bahasa dan Sastra* 20, no. 1 (2020): 13–24, https://doi.org/10.17509/bs_jpbasp.v20i1.24679; in A. Setiawan, "Enhancing Language Assessment Literacy of Pre-Service English Teachers through Reflective Teaching Practice," *Indonesian Journal of Applied Linguistics* 9, no. 1 (2019): 130–138, <https://doi.org/10.17509/ijal.v9i1.15201>



simulations, and authentic assessment tasks.⁴ However, relatively few studies have examined how students actually engage with these activities in real classroom contexts and what kinds of challenges they face in navigating the learning process. Understanding these aspects is crucial for improving the design of English assessment course in English education study program as the outcome of the study is to be either an English teacher or instructor.

This study seeks to explore these dynamics through a qualitative observational approach, focusing on students' engagement and the challenges they encounter in the context of English language test development. By drawing from classroom interactions, reflective teaching journals, and informal dialogues with students, the study aims to provide deeper insights into how language assessment is internalized by learners and how pedagogical practices can be improved to foster more meaningful and effective learning experiences.

This study contributes to the growing body of literature on language assessment literacy by offering a classroom-based perspective on how students engage with the process of English language test development. Unlike previous research that largely emphasizes teachers' theoretical understanding or uses survey-based methods, this study provides in-depth qualitative insights derived from direct observation of student behavior, collaboration, and problem-solving in real-time instructional settings.

The foundation of this study is grounded in the intersection between language assessment literacy (LAL) and experiential learning theory. As English teacher education increasingly emphasizes the importance of assessment competence, there is a growing need to understand how students internalize assessment principles not only conceptually but also through active, reflective participation in test development tasks.

Simultaneously, the concept of language assessment literacy informs the knowledge, skills, and beliefs needed to design and evaluate language assessments. By observing students' engagement, confusion, strategies, and interactions, this study explores how assessment literacy is formed, challenged, and supported within the classroom.⁵

⁴ DeLuca and Don A. Klinger, "Assessment Literacy Development: Identifying Gaps in Teacher Candidates' Learning," *Assessment in Education: Principles, Policy & Practice* 17, no. 4 (2010): 419–438, <https://doi.org/10.1080/0969594X.2010.516643>.

⁵ Lynda Taylor, "Communicative Assessment: Implications for Language Testing," in *The Companion to Language Assessment*, ed. Antony John Kunnan (Hoboken, NJ: Wiley Blackwell, 2013), 1–14; Glenn Fulcher, "Assessment Literacy for the Language Classroom," *Language Assessment Quarterly* 9, no. 2 (2012): 113–132, <https://doi.org/10.1080/15434303.2011.642041>



Within this framework, student engagement is not viewed as a passive indicator but as an active cognitive, emotional, and behavioral process. The observation-based method allows researchers to trace how students make meaning of complex assessment concepts and navigate real-time challenges, revealing the pedagogical gaps and potential improvements in how language assessment is taught.

B. RESEARCH METHOD

This study employed a qualitative observational research design to explore students' engagement and challenges in learning English language assessment, specifically within the topic of test development. The choice of qualitative observation was driven by the researcher's dual role as both lecturer and observer, allowing for in-depth insights through prolonged engagement in the classroom setting. This method aligns with the goals of naturalistic inquiry, which seeks to understand participants' behaviors, attitudes, and learning experiences within an authentic educational context (Creswell, 2013).⁶

The participants consisted of 19 undergraduate students enrolled in a sixth-semester Language Assessment course in an English education study program at Islamic Institute of Hasanuddin, Pare. The course was taught over a 14-week semester, with weekly sessions focused on assessment principles, test construction, and practical applications. Students engaged in both theoretical discussions and collaborative test-design projects.

Data were collected through three primary methods: 1) Non-participant and participant observation: The researcher documented student behavior, interaction patterns, participation levels, and responses during lectures, group work, and presentations. Field notes and reflective teaching journals were kept after each session; 2) Student artifacts: These included students' test development projects, self-reflections, and worksheets related to formative assessment design, item analysis, and rubric creation. These materials were analyzed to identify areas of understanding and confusion; 3) reflective dialogues: The researcher conducted brief, informal conversations with students before or after class to capture their

⁶ John W. Creswell, *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*, 3rd ed. (Thousand Oaks, CA: SAGE Publications, 2013)



thoughts about the course content, their motivation, and perceived difficulties. Notes from these dialogues were thematically coded.

C. RESULTS AND DISCUSSION

This section presents the main findings of the study, organized into two core themes: (1) Student Engagement in Language Assessment Learning, and (2) Challenges in Understanding and Applying Test Development Concepts. Each theme is supported by observational data, student artifacts, and informal dialogues with learners.

Observation Sheet (With Sample Answers)

- Title : Observation on Student Engagement and Challenges in Learning Language Assessment
- Course : English Assessment
- Observer : Lecturer – Muhammad Yasminto
- Date : **May, 3rd 2025**

a) General Classroom Information

Table 1. General Classroom Information

Item	Description
Total students present	19 Students
Learning Topic	Understanding Validity and Reliability in Language Testing
Teaching method used	☒ Lecture ☒ Group Work ☒ Authentic Test Sample Analysis
Materials used	☒ Slide ☒ Worksheet ☒ Sample Grammar Test Items

b) Student Engagement Indicators

Table 2. Student Engagement Indicators



Engagement Indicator	Observation Notes
Students ask relevant questions	3 students asked clarifying questions about “construct validity” and “test fairness.”
Students actively respond to lecturer’s prompts	About 36,8 % atau 7 students of the class (19) responded when asked about the difference between reliability and validity.
Participation in group work or discussions	All groups were actively discussing how to evaluate the sample test.
Students show interest during explanation/test activity	Most students appeared focused when discussing the example READING and SPEAKING items.
Non-verbal cues (note-taking)	Around 79 % or 15 students took notes and nodded during explanation. 4 students looked disengaged.
Volunteering to present or lead group	2 students volunteered to present their group’s findings on poor test item design.



c) Learning Challenges Indicators

Table 3. Learning Challenges Indicators

Challenge Observed	Observation Notes
Confusion over key concepts	Several students confused “content validity” with “face validity” and with “construct validity.”
Difficulty applying theory to test design	Some groups listed surface features (e.g., neat format) as criteria for test validity.
Misunderstanding instructions	Group 2 missed the instruction to analyze items based on <i>reliability</i> as well.
Reliance on peers	About three students copied answers without understanding.
Anxiety or frustration	A few students expressed frustration over complex definitions, saying “ <i>I don’t get the difference between them.</i> ”

d) Observer’s Reflective Notes

- **What worked well in today’s session?**

The use of real reading test items increased engagement. Students showed curiosity and actively discussed validity-related issues in groups.

- **What challenges did students most commonly face?**

Many struggled to distinguish between types of validity and to apply these concepts to real test questions. Terminology was a major barrier.

- **What should be adjusted or improved in future lessons?**

More scaffolding is needed, perhaps using a comparison chart or visual aid for types of validity. In addition, consider using Indonesian explanations for key terms.

- **Emerging Themes or Patterns:**

Students are more engaged when analyzing real test content. Theoretical content needs to be broken into simpler, digestible parts

e) Additional Comments



Students seem to appreciate collaborative work, but still require clearer scaffolding when moving from theory to practice. Peer-led presentations may help reinforce abstract concepts

1. Student Engagement in Language Assessment Learning

Observational data indicated varied levels of student engagement throughout the course. In the initial weeks, students appeared hesitant and passive during discussions of foundational assessment concepts such as validity, reliability, and test types. However, engagement increased during practical activities, particularly during collaborative test design sessions where students were required to create reading and speaking test items.

This pattern suggests that students responded more positively to hands-on, constructivist learning tasks than to lectures focused on theory. This finding aligns with Scarino (2013), who argued that language assessment instruction becomes more meaningful when learners can apply abstract concepts in authentic ways.⁷

Moreover, informal dialogue revealed that students appreciated assignments that involved peer feedback and revision. One student remarked, “I started to enjoy the course when we had to make our own test. It felt like I was learning something I could really use as a future teacher.” This reinforces Taylor’s (2009) claim that assessment literacy is best developed through reflective practice and experiential learning.⁸

2. Challenges in Understanding and Applying Test Development Concepts

Despite increased engagement during applied tasks, many students continued to struggle with the theoretical depth of test development. Specifically, concepts such as construct validity, washback, and test fairness were frequently misunderstood. Analysis of students’ test

⁷ Angela Scarino, “Language Assessment Literacy as Self-Awareness: Understanding the Role of Interpretation in Assessment and in Teacher Learning,” *Language Testing* 30, no. 3 (2013): 309–327, <https://doi.org/10.1177/0265532213480128>.

⁸ Lynda Taylor, “Developing Assessment Literacy,” *Annual Review of Applied Linguistics* 29 (2009): 21–36, <https://doi.org/10.1017/S0267190509090035>.



drafts showed that many items lacked clear alignment with the stated test objectives or CEFR descriptors, indicating a gap in students' ability to translate theory into practice.

Several students also expressed anxiety when faced with rubrics and statistical terms used in item analysis. This reflects the findings of Inbar-Lourie (2013), who emphasized that learners often experience cognitive overload when confronted with technical assessment language not scaffolded appropriately.⁹

Lecturer reflections noted that traditional lecture-based sessions were insufficient in clarifying these abstract concepts. A key turning point occurred when students were introduced to real-life sample tests and asked to critique them using a checklist based on test principles. This activity appeared to foster deeper understanding, as students began to question design choices and suggest improvements. The findings here highlight the need for scaffolded learning, where theory is gradually introduced alongside concrete examples and peer collaboration (Brown & Abeywickrama, 2019).¹⁰

CONCLUSION AND SUGGESTIONS

Conclusion

This study investigated students' engagement and challenges in learning English assessment, particularly in the context of language test development. Based on classroom observations, reflective journals, and student artifacts, the findings revealed that while students gradually developed interest and participation during hands-on activities such as test design projects, they continued to face difficulties in understanding abstract theoretical constructs like validity, reliability, and washback (the influence that testing has on teaching and learning)

Student engagement increased significantly when instruction was contextualized and collaborative, affirming that practical, interactive approaches foster deeper learning and enhance assessment literacy. However, challenges persisted in aligning theory with practice, especially when students were required to justify their test items using assessment principles.

⁹ Ofra Inbar-Lourie, "Language Assessment Literacy and the Challenges of Teacher Education," in *The Companion to Language Assessment*, ed. Antony John Kunnan (Hoboken, NJ: Wiley Blackwell, 2013), 560–575.

¹⁰ H. Douglas Brown and Priyanvada Abeywickrama, *Language Assessment: Principles and Classroom Practices*, 2nd ed. (White Plains, NY: Pearson Education, 2019)



These findings highlight the complexity of teaching language assessment and the need for intentional scaffolding and reflective practice in the classroom.

Suggestions

Based on the conclusions, several pedagogical suggestions can be made for improving the teaching and learning of English assessment:

1. Integrate More Practice-Based Learning

Teachers and Instructors should design activities that allow students to apply concepts through real or simulated assessment tasks. These include item-writing workshops, peer review of test blueprints, and critiquing authentic language tests.

2. Use Scaffolding to Teach Abstract Concepts

Theoretical aspects of assessment (e.g., types of validity, test fairness) should be introduced gradually, using visual aids, real examples, and guided practice to help students internalize key ideas.

3. Promote Reflective Teaching and Learning

Encouraging students to reflect on their assessment design decisions can deepen their understanding and help them connect theory to practice. Reflective journals and group discussions can serve this purpose.

4. Foster Assessment Literacy Early in the Curriculum

Elements of language assessment should be integrated into earlier courses in the teacher education program, preparing students gradually for more advanced topics.

5. Encourage Collaborative Work and Peer Feedback

Group work in designing tests and evaluating peer products can promote active engagement, build confidence, and foster shared responsibility in learning.



REFERENCES

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Sage Publications.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419–438. <https://doi.org/10.1080/0969594X.2010.516643>
- Farid, M. (2020). An analysis of students' perception toward English language assessment in Indonesian EFL classrooms. *Jurnal Pendidikan Bahasa dan Sastra*, 20(1), 13–24. https://doi.org/10.17509/bs_jpbsp.v20i1.24679; in A. Setiawan. (2019). Enhancing language assessment literacy of pre-service English teachers through reflective teaching practice. *Indonesian Journal of Applied Linguistics*, 9(1), 130–138. <https://doi.org/10.17509/ijal.v9i1.15201>
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Routledge.
- Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment: An Advanced Resource Book*. Routledge.
- Inbar-Lourie, O. (2013). Language assessment literacy. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1–7). Wiley-Blackwell. <https://doi.org/10.1002/9781405198431.wbeal0658>
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309–327. <https://doi.org/10.1177/0265532213480128>
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36. <https://doi.org/10.1017/S0267190509090035>