VIETNAM NATIONAL UNIVERSITY, HANOI

**UNIVERSITY OF LANGUAGES AND INTERNATIONAL STUDIES**

**NGUYỄN THỊ TỐ LOAN**

**THE DEVELOPMENT AND ARGUMENT-BASED VALIDATION OF**

**A SCALE FOR MEASURING VIETNAMESE EFL LECTURERS’ LANGUAGE ASSESSMENT LITERACY**

(Thiết kế và xác trị thang đo năng lực đánh giá ngôn ngữ của giảng viên

dạy tiếng Anh ở Việt Nam theo hướng dựa trên các lập luận)

Major: English Language Teaching Methodology

Code: 9140231.01

**SUMMARY OF DOCTORAL THESIS**

**HÀ NỘI - 2025**

The study has been conducted at

**University of Languages and International Studies, VNU**

Supervisors: **1. Assoc. Prof. Dr. Nguyen Van Trao**

**2. Dr. Duong Thu Mai**

Reviewer 1: …………………………………………

Reviewer 2: …………………………………………

Reviewer 3: …………………………………………

Luận án sẽ được bảo vệ trước Hội đồng chấm luận án tiến sĩ họp tại:

Trường Đại học Ngoại ngữ, Đại học Quốc gia Hà Nội

vào hồi giờ ngày tháng năm 2025

**Có thể tìm hiểu luận án tại:**

- Thư viện Quốc gia Việt Nam.

- Trung tâm Thông tin - Thư viện, Đại học Quốc gia Hà Nội.

**CHAPTER 1: INTRODUCTION**

The Introduction chapter presents the reasons for conducting the research, the significance of the research and the research context in Vietnam related to the thesis. This chapter also sets out the research questions, limits the scope of the research and introduces the structure of the thesis.

**1.1. Rationale**

Global educational reforms have expanded teachers' responsibilities to include diverse assessment methods, making Assessment Literacy (AL) essential for ensuring valid and reliable evaluations. Teachers spend significant time on assessment-related tasks, and inadequate AL can lead to misleading results, negatively affecting stakeholders. For EFL teachers, Language Assessment Literacy (LAL) is particularly critical, as it equips them with the knowledge and skills to assess language learners effectively. However, many teachers lack sufficient AL due to limited assessment training in pre-service and in-service programs. In Vietnam, this gap is even more pronounced in tertiary EFL education, where research and resources for LAL development remain scarce despite initiatives like the National Foreign Languages Project 2020. This study addresses this gap by developing and validating a LAL measurement scale for Vietnamese university EFL lecturers, providing a reliable tool to assess their assessment knowledge and skills, ultimately supporting teacher education and improving English language teaching quality in Vietnam.

**1.2. Research problem**

Despite increasing attention since Brindley (2001) integrated AL into LAL, many challenges remain unresolved. Research on LAL is still theoretically and empirically limited, lacking a clear consensus on its definition and validated measurement instruments, especially compared to AL frameworks in general education.

In Vietnam’s higher education (HE) system, the emphasis on English proficiency for employability has heightened the need for effective language assessment. However, summative, exam-focused assessment dominates, limiting the adoption of formative methods. Institutional constraints, resistance to change, and a lack of professional development further hinder LAL implementation, forcing many lecturers to rely on intuition rather than evidence-based assessment practices. Most LAL research focuses on in-service teachers, with limited attention to higher education. Studies suggest that LAL manifests through observable behaviors reflecting teachers’ competence, but without a validated instrument tailored to Vietnam, assessing lecturers’ LAL remains challenging. To address this gap, this study develops and validates a Language Assessment Literacy Scale (LALS) for Vietnamese EFL lecturers (V-EFLLs). It identifies key competencies, provides empirical evidence for professional development, and enhances language assessment practices in Vietnamese higher education. The findings contribute to reliable LAL measurement tools and establish a foundation for future research in Vietnam and similar contexts.

**1.3. Aims and objectives of the study**

The study has two main objectives:

**- Developing the LALS** – This instrument will assess LAL competencies specific to Vietnamese higher education, ensuring it reflects the critical skills required for effective language assessment.

**- Validating the LALS using an argument-based validation framework** – Following Kane’s (2006) validation approach, expanded by Chapelle et al. (2008), the study will evaluate four key validity inferences:

**- Domain inference:** Ensures the LALS represents essential LAL competencies.

**- Evaluation inference:** Confirms that performance scores are reliable and accurately observed.

**- Generalization inference:** Assesses score stability across different assessment contexts.

**- Explanation inference:** Verifies that the LALS effectively measures LAL and distinguishes proficiency levels among lecturers.

By ensuring the validity and reliability of the LALS, this study will provide a structured framework for assessing V-EFLLs’ LAL. The findings will contribute to curriculum design and professional development programs, ultimately strengthening language assessment competencies in Vietnamese higher education.

**1.4. Research questions**

The following research questions guide the empirical investigation:

RQ1: What are the criteria for a scale that measures the language assessment literacy (LAL) of Vietnamese EFL lecturers (V-EFLLs)?

RQ2: To what extent is the developed LAL scale a valid and reliable instrument for assessing the LAL of V-EFLLs?

RQ2.1. To what extent do the criteria in the LALS accurately represent the LAL construct for V-EFLLs, as determined by expert feedback and content relevance?

RQ2.2. To what extent do the criteria of the LALS for V-EFLLs demonstrate satisfactory psychometric properties to ensure reliable and valid measurement?

RQ2.3. How generalizable are the LALS scores across different demographic groups and levels of experience among the studied V-EFLLs?

RQ2.4. How well do V-EFLLs’ self-ratings on the LALS reflect their actual LAL competencies, and to what extent do the items effectively differentiate their proficiency levels?

**1.5. Methodological considerations**

This study employed an exploratory sequential mixed methods design (Creswell & Plano Clark, 2018) to develop and validate the LALS for V-EFLLs. Qualitative data from semi-structured interviews and expert feedback informed the scale’s refinement, ensuring alignment with an argument-based validity framework. The process began with a literature review, followed by focus group reviews and two rounds of expert paneling to assess clarity, relevance, and developmental alignment. The quantitative phase trialed the revised scale with EFL lecturers, analyzing responses using Rasch measurement and the partial credit model (Masters, 1982) to ensure psychometric validity. Construct mapping (Wilson, 2011; Wolfe & Smith, 2007a) defined LAL as a progressive continuum of skills.

By integrating theoretical and psychometric approaches, this study established a criterion-referenced framework for assessing LAL in Vietnamese higher education. The LALS provides a valid and context-sensitive tool, supporting enhanced assessment literacy among lecturers.

**1.6. Scope of the study**

This study develops and validates the LALS for Vietnamese EFL university lecturers, ensuring a contextually relevant assessment framework. It focuses on 177 in-service lecturers across Vietnam, excluding primary and secondary teachers due to differing assessment practices. Conducted between 2023 and 2024, the study aligns with recent educational reforms, including updates to the National Foreign Languages Project 2020. For rigorous validation, Kane’s (2006) argument-based framework assesses domain description, evaluation, generalization, and explanation. Rasch measurement theory (Wright & Masters, 1982) and Classical Test Theory (Hambleton & Swaminathan, 1985) establish reliability and validity, while a developmental perspective (Griffin, 2018; Wilson, 2011; Masters, 1982) conceptualizes LAL as an evolving competence. Although focused on university lecturers, the study’s applicability to other educational contexts requires further exploration. The reliance on self-reported data may introduce bias, mitigated through triangulation via expert reviews, focus groups, and psychometric analyses.

**1.7. Significance of the study**

This study contributes to LAL research by addressing theoretical, methodological, and practical gaps, particularly for V-EFLLs. Theoretically, it clarifies LAL conceptualization among in-service university instructors, expanding research beyond pre-service teachers and Western contexts. By focusing on Vietnam’s higher education system, it offers a broader understanding of LAL’s components and applications. Methodologically, the study enhances LAL research by integrating the developmental approach for scale construction with Kane’s argument-based validation framework (Kane, 2004, 2013). This structured approach strengthens the reliability and validity of LAL assessment. Practically, the study provides a validated LALS for pedagogical and research use. It supports self-assessment and professional development for EFL instructors, informs teacher education programs, and offers insights to guide policy decisions, ultimately advancing LAL development in Vietnam and similar educational contexts.

**1.8. Structure of the study**

The thesis consists of nine chapters, beginning with Chapter 1 which introduces the study. Chapter 2 defines the construct of LAL, followed by Chapter 3 which outlines the development of the LAL scale. Chapter 4 discusses the approaches used for scale validation, and Chapter 5 presents the research methodology. Chapters 6, 7, and 8 focus on the results of the scale validation for V-EFLLs, addressing domain definition, evaluation, generalization, and explanation inferences. Finally, Chapter 9 offers conclusions and implications, highlighting the study's impact and future research directions.

**CHAPTER 2: THE CONSTRUCT OF LANGUAGE ASSESSMENT LITERACY**

This chapter explores the key elements of language assessment literacy (LAL) from both conceptual and methodological perspectives. It defines “assessment” and “literacy,” tracing the evolution of literacy from basic skills to broader competencies across domains, while reviewing key LAL models that highlight various aspects of LAL. Through sociocultural, constructivist, and psychometric perspectives, the chapter lays the groundwork for developing a tailored LAL framework for Vietnamese EFL educators.

**2.1. Understanding assessment literacy in education**

Literacy has evolved from basic reading, writing, and numeracy to include broader competencies such as critical thinking and digital proficiency (Taylor, 2013). In education, AL, as defined by Stiggins (1991), enables educators to design and implement effective assessments that enhance learning.

Two perspectives shape literacy: a dichotomous view, categorizing individuals as literate or illiterate, and a continuum approach, which sees literacy as a range of context-specific proficiencies. AL comprises assessment, measurement, testing, and evaluation, each serving a distinct role in education (Brown & Abeywickrama, 2010). Assessment broadly monitors and improves learning, measurement assigns quantitative values to outcomes, testing evaluates competencies through specific instruments, and evaluation interprets results to refine instruction. Together, these elements equip educators to apply AL effectively across diverse educational contexts.

**2.2. Ontological and epistomological assumptions in defining LAL**

Defining LAL requires a comprehensive approach that integrates both ontological and epistemological dimensions. Framed through sociocultural, constructivist, and psychometric theories, this section conceptualizes LAL as an evolving, contextually embedded competence that balances complexity with systematic measurement:

***2.2.1. Sociocultural theory: LAL as a socially constructed competence***

Rooted in Vygotsky’s sociocultural theory, LAL develops through social interactions within professional communities. The Zone of Proximal Development (ZPD) highlights how educators expand their LAL through collaboration, mentorship, and experience, adapting assessments to linguistic and cultural contexts.

***2.2.2. Constructivist theory: LAL as an adaptive, reflective competence***

Following Piaget (1977) and Dewey (1938), constructivist theory sees LAL as actively developed through reflection and experience. Educators refine their assessment practices based on student learning outcomes, making LAL flexible and adaptable to evolving classroom needs.

***2.2.3. Psychometric theory: LAL as a measurable competence***

Psychometric models like Rasch and Item Response Theory (IRT) quantify LAL, aligning skill development with task complexity. These models ensure objective assessment of educators’ competence, allowing institutions to track growth systematically while maintaining assessment reliability.

***2.2.4. Integrated perspective: Developmental models in LAL***

Developmental models synthesize sociocultural, constructivist, and psychometric approaches, framing LAL as a progressive competence rather than a static skill. Educators advance from basic to expert, acquiring increasingly complex assessment abilities (Griffin et al., 1997, 2004, 2007, 2018). Ontologically, LAL is both latent and observable, shaped by experience and social interaction (Masters, 2003; Larter, 1991). Epistemologically, it merges quantitative measurement with experiential learning. Griffin, Francis, & Robertson (2018) link this framework to Glaser’s (1981) criterion-referenced approach, balancing structured assessment with reflective adaptation. For V-EFLLs, developmental models provide a context-sensitive framework that tracks measurable growth while incorporating institutional and cultural influences. By integrating technical skills with reflective practice, LAL fosters lifelong professional development in assessment literacy.

**2.3. Approaches to define language assessment literacy (LAL)**

Competence-listed models, such as those by AFT, NCME & NEA (1990), define LAL through specific competencies required for effective assessment, including method selection and ethical considerations. Initially focused on summative assessment, this approach was later expanded by Brookhart (2011) to include formative assessment, acknowledging the need for a teaching-centered skill set.

Three-component models, like Davies’ (2008) Knowledge, Skills, and Principles framework, broaden LAL beyond technical competencies to include ethical and contextual dimensions. Fulcher (2012) proposed a hierarchical structure, integrating foundational skills, guiding principles, and educational awareness, reinforcing that effective language assessment requires more than technical expertise.

Scaled models view LAL as a developmental continuum, with educators progressing from basic to advanced proficiency. Pill & Harding’s (2013) “literacy ladder” categorizes skills across progressive literacy levels, while Taylor’s (2013) framework tailors competencies to different stakeholders, recognizing that LAL requirements vary by professional role.

The negotiation approach, proposed by Xu & Brown (2016), positions LAL as a socially constructed competence shaped through interaction and professional identity. It balances technical skills with context-specific adaptation, emphasizing that educators refine their assessment literacy through engagement with teaching environments rather than static skill acquisition.

This study defines LAL as a dynamic, evolving competence developed through continuous learning, reflection, and adaptation to social, political, and cultural contexts. Unlike static frameworks, a process-based approach sees LAL as an iterative practice, shaped by institutional constraints, pedagogical advancements, and technological shifts (Inbar-Lourie, 2008; Fulcher, 2012; Xu & Brown, 2016). In Vietnam’s top-down, high-stakes assessment system, this approach fosters gradual integration of formative, learner-centered practices, preparing educators for future developments in digital assessment and global education standards. It lays the groundwork for a context-sensitive LAL framework in Vietnamese HE.

**2.4. Previous studies on language assessment literacy development**

***2.4.1. International studies on LAL development***

International research on LAL has explored various assessment methodologies, but reliable, language-specific instruments remain scarce (Stabler-Havener, 2018). Many existing tools focus on general assessment literacy rather than the unique demands of language assessment (Gotch & French, 2014), with U.S.-centric measures showing limited reliability in international contexts (Xu & Brown, 2016). Efforts to develop context-sensitive LAL tools include scenario-based assessments (Muhammad & Bardakçi, 2019; Tao, 2014; Xu & Brown, 2017) and skill-based measures like the Language Assessment Knowledge Scale (LAKS) (Ölmezer-Öztürk & Aydin, 2018), though validation remains limited. Researcher-developed instruments address some gaps, such as Zolfaghari & Ashraf’s (2015) 50-item inventory, Farhady & Tavassoli’s (2018) adaptation of Fulcher’s (2012) questionnaire, and Kremmel & Harding’s (2020) survey, later expanded into the Critical Language Assessment Literacy (CLAL) scale (Tajeddin et al., 2022). Despite these advances, limited cross-context validation highlights the ongoing need for adaptable, reliable LAL instruments that reflect the complexity of language assessment in diverse educational settings.

***2.4.2. Studies on LAL development in Vietnam***

LAL research in Vietnam is limited, with studies highlighting systemic challenges in its implementation at the tertiary level. Global LAL models often overlook Vietnam’s summative-driven assessment culture and the lack of formative assessment training in teacher education. This misalignment underscores the need for a locally adapted framework that reflects V-EFLLs’ assessment responsibilities. Project 2020 aimed to improve English proficiency but reinforced high-stakes testing rather than fostering comprehensive LAL (Nguyen, 2017; Pham, 2017). Standardized exams dominate, restricting lecturers' autonomy in assessment design, while professional training remains test-focused (Ngo, 2021). Institutional policies further limit the adoption of process-based assessment approaches (Tran, 2019). Existing LAL tools, such as those adapted by Tu (2023), fail to address lecturers’ limited agency and the policy-driven nature of Vietnamese higher education. A Vietnam-specific LAL framework must integrate formative assessment training, practical application, and policy awareness, equipping lecturers with essential skills while addressing institutional constraints.

This chapter evaluates LAL models, highlighting their limitations. These gaps underscore the need for a Vietnam-specific LAL scale that reflects policy-driven assessment, institutional constraints, and limited professional development in higher education. Existing models fail to address the technical, social, and political dimensions shaping LAL. Framing LAL as an evolving competence, this study integrates sociocultural, constructivist, and psychometric perspectives to ensure contextual relevance and measurement validity. The next chapter details the methodology for designing and validating this empirically grounded LAL scale for V-EFLLs.

**CHAPTER 3: THE DEVELOPMENT OF A SCALE FOR MEASURING EFL LECTURERS’ LANGUAGE ASSESSMENT LITERACY**

This chapter outlines the development of the Scale (LALS) designed to measure the language assessment literacy (LAL) of V-EFLLs. It details the theoretical foundations, practical steps, and psychometric processes used to ensure the scale’s validity, reliability, and contextual relevance.

**3.1. Theoretical approaches to LAL scale development**

LAL scale development is informed by several approaches, each contributing unique insights:

**Capabilities approach** emphasizes transferable skills, such as adaptability and critical thinking, essential for managing diverse assessment scenarios. However, its lack of structured benchmarks for skill acquisition requires integration with other approaches for concrete application. **Functional analysis** focuses on task-specific competencies, breaking down professional roles into distinct skills needed for effective assessment, like designing test items and interpreting results. While it provides clear benchmarks, it may lack the adaptability of other approaches. **Consensus approach** gathers expert feedback to define core competencies. Though resource-intensive, it fosters relevance across educational settings, incorporating interdisciplinary perspectives. **Developmental multiple-methods approach** integrates elements of each approach, balancing concrete competencies with flexibility, supporting both immediate teaching needs and long-term professional growth. This comprehensive approach is ideal for developing a dynamic LAL framework suited to complex educational demands.

**3.2. LAL scale development process**

**Table 3.1:** *Methods employed in the development of the draft framework for V-EFLLs’ LALS*

|  |  |
| --- | --- |
| **Methods** | **Purposes** |
| **Literature review** | * To generate a draft framework for V-EFLLs’ LAL |
| **Focus group** | * To examine the appropriateness and relevance of the draft competency framework in Vietnamese higher education settings. |
| **Expert paneling Round 1** | * To review and agree upon the dimensions, competencies and performance with SMEs indicators identified from the literature, examined at and amended following the focus group.   (A panel of professional experts first reviewed and rated these items for relevance to LAL, providing feedback on clarity and coverage. Based on their feedback, the items were revised.) |
| **Expert paneling Round 2** | * To improve the conciseness of and review the wording and ordering of the quality criteria for each performance indicator.   (A second panel of professional experts then reviewed the revised items, providing suggestions for the conciseness of the items, leading to further refinement.) |

The LAL scale development followed a structured process, beginning with a construct definition and mapping LAL as a progression of skills (Wilson, 2011; Wolfe & Smith, 2007a). Validity was assessed through Rasch analysis, comparing expected and actual performance (Woods & Griffin, 2013), with expert input ensuring comprehensive representation. A developmental model illustrated LAL growth, arranging items from foundational to advanced skills. The Partial Credit Model (Masters, 1982) enabled nuanced proficiency measurement with fewer items while maintaining interpretability. This approach aligned theoretical predictions with observed competence, ensuring substantive validity. To create measurable indicators, Griffin, Gillis, and Calvitto’s (2004) framework on observable behaviors structured LAL into capabilities, behaviors, and proficiency criteria, following Glaser’s (1981) criterion-referenced assessment. The LALS was developed through drafting, focus groups, and two rounds of expert paneling, refining the framework through iterative feedback.

The drafting phase began with a literature review to define LAL and identify key competencies, drawing from frameworks by Brindley (2001), Davies (2008), and Fulcher (2012). This informed a hierarchical framework for Vietnamese teachers, structured around three dimensions: preparing for assessment, conducting assessment, and revising assessment. Each dimension included 13 competencies, totaling 65 performance indicators. Quality criteria were developed based on rubric design principles (Griffin, Nguyen, & Gillis, 2004) and the Dreyfus and Dreyfus (1986) skill acquisition model, ensuring distinct proficiency levels for structured self-assessment.

The expert paneling phase refined the scale through two rounds of review. Experts in language assessment, educational measurement, and Vietnamese higher education evaluated items for clarity, relevance, and completeness. The first round emphasized digital assessment competencies, given the increasing role of online and formative assessments. Revisions addressed digital literacy and accessibility for lecturers in resource-limited areas. The second round refined the developmental progression, simplifying language and ensuring items accurately represented a continuum from foundational to advanced competencies. This iterative process produced a streamlined, logically ordered scale.

The trial phase tested the scale’s psychometric properties with 177 V-EFLLs from diverse institutions, using stratified and cluster sampling for broad representation. Participants completed the LALS as a self-assessment, structured as a developmental continuum of LAL proficiency. Classical Test Theory (CTT) confirmed high internal consistency, while Rasch modeling validated the developmental scaling by aligning item difficulty with expected lecturer progression. Fit statistics affirmed unidimensionality, and item-person maps showed an appropriate difficulty range, confirming the scale’s ability to distinguish between competence levels.3.3. Construct mapping of LAL from literature review

Based on prior research, the construct map frames LAL as a dynamic, context-sensitive competence, divided into three stages: Preparing for Assessment (WHAT), Conducting Assessment (HOW), and Revising Assessment (WHY). Each stage builds on competencies defined by foundational LAL models, providing V-EFLLs with a scaffolded approach to developing assessment skills.

**Preparing for Assessment (WHAT)**: Involves understanding key assessment principles and setting clear, measurable targets aligned with educational standards. **Conducting Assessment (HOW)**: Encompasses the practical steps of administering, scoring, and interpreting assessments, ensuring alignment with learning objectives and fairness. **Revising Assessment (WHY)**: Focuses on ongoing evaluation and refinement of assessment practices to ensure relevance and effectiveness.

This process-based framework supports continuous professional growth, allowing educators to contextualize LAL within their unique teaching environments and reflect on their evolving role as assessors.

**3.4. Paneling**

The paneling phase involved two rounds of expert reviews, refining the LALS to ensure its relevance, clarity, and alignment with V-EFLLs' needs. Experts in English language assessment, educational measurement, and Vietnamese higher education evaluated the scale, focusing on its technical accuracy and contextual appropriateness. In the first round, experts assessed the initial draft of 65 items across three dimensions: preparing, conducting, and revising assessments. Feedback highlighted the need for greater emphasis on digital assessment and formative practices, leading to the inclusion of new items and revisions to improve clarity, particularly for rural educators. The second round focused on developmental progression, simplifying language to distinguish competence levels and reordering items to enhance logical flow. By the end, the scale reflected a well-structured and contextually relevant tool for V-EFLLs.

**3.5. Trialing**

The trialing phase tested the LALS with 177 V-EFLLs from diverse teaching contexts across the country. Using stratified, cluster, and convenience sampling, the study captured data from institutions with varying technological resources. Participants completed a self-assessment of LAL proficiency, and data analysis combined Classical Test Theory and Rasch modeling. High internal consistency (Cronbach’s alpha > .80) and unidimensionality confirmed the scale's reliability and alignment with LAL competencies. Rasch analysis demonstrated a broad range of item difficulty, effectively differentiating lecturers’ abilities, while responses across demographics indicated consistency and adaptability. Feedback prompted minor adjustments to digital competencies, ensuring inclusivity. The findings validated the LALS as a reliable, contextually relevant measure, capable of supporting self-assessment, professional development, and institutional evaluation for V-EFLLs.

Chapter 3 presents the LALS development as a structured, empirically validated approach to measure LAL for V-EFLLs. The scale, underpinned by expert feedback and rigorous psychometric analysis, offers a reliable, contextually relevant tool for assessing and fostering LAL competencies. This scale supports self-assessment, professional development, and institutional evaluation, aligning with the specific needs of EFL educators in diverse Vietnamese educational settings. The next steps involve refining this scale’s application to varied instructional contexts and leveraging it for targeted LAL improvement programs across Vietnam.

**CHAPTER 4: APPROACHES FOR VALIDATING LANGUAGE ASSESSMENT LITERACY SCALE**

This chapter outlines the validation of the Language Assessment Literacy Scale (LALS) using an argument-based framework (Kane, 1992, 2013; Chapelle et al., 2008) supported by Rasch modeling. Validation focuses on evidence-backed score interpretations rather than the test itself, following two stages. First, the Interpretive and Use Argument (IUA) defines intended score interpretations and uses. Second, a validity argument evaluates empirical support for these interpretations.

**4.1. Approaches to assessment validation**

***4.1.1. Contemporary perspectives of validity and validation***

The concept of validity has evolved from simple definitions to integrative approaches (AERA, APA, & NCME, 1985, 2014; Kane, 2013; Sireci, 2013). Traditional views classified validity into three primary types: content, criterion-related, and construct validity. Content validity evaluates if test items accurately represent the domain; criterion-related validity focuses on test score correlations with external benchmarks, while construct validity, introduced by Cronbach and Meehl (1955), involves ensuring the test measures an intended theoretical trait. Later, Messick’s (1989) unified theory consolidated these into a single framework, viewing validity as an evaluative judgment supported by diverse evidence types. This unified model, reflected in the Standards for Educational and Psychological Testing, underscores a holistic view of validity, considering test content, structure, and social consequences.

***4.1.2. Development of modern validation frameworks in language assessment***

Modern validation frameworks, including Messick’s unified validity theory, Bachman and Palmer’s (2010) assessment use argument (AUA), Kane’s (1992, 2006, 2013) argument-based approach, Mislevy and Yin’s (2012) evidence-centered design (ECD), and Weir’s (2005) socio-cognitive framework, provide comprehensive approaches that integrate theoretical and empirical evidence for validating test score interpretations and uses.

Messick’s Unified Theory of Validity integrates all validity forms, though its abstract nature limits practical guidance. Bachman and Palmer’s AUA structures validation through an argument-based model focused on assessment use, making it valuable for practical applications. Kane’s Argument-Based Approach organizes validation through inferences and supporting evidence, emphasizing practical and consequential aspects. This model, adaptable to various testing contexts, is particularly suited for high-stakes and classroom assessments. Mislevy and Yin’s ECD framework embeds evidentiary needs within test design, aligning tasks with constructs from the start, although it is more applicable to standardized contexts. Weir’s Socio-Cognitive Framework integrates cognitive and social considerations, beneficial for high-stakes language assessments where task authenticity is prioritized. Kane’s approach, with its inference-focused validation, is particularly useful for educational assessments and language testing. The flexibility of his framework supports structured validation arguments, making it well-suited to LALS validation in varied contexts.

**4.2. Building the study’s validation framework**

***4.2.1. Rationale for framework selection***

Kane’s argument-based approach (1992, 2013) is adopted for its structured, inference-driven process, systematically linking test scores to their interpretations and uses. It provides a clear framework for constructing an IUA and a Validity Argument, ensuring logical coherence and empirical support for key inferences like domain description, evaluation, and generalization. Kane’s adaptability makes it ideal for validating the LALS, which assesses the multi-dimensional construct of LAL across diverse Vietnamese educational settings. Unlike traditional models, it prioritizes practical implications, aligning with the socio-cultural nature of LAL and ensuring meaningful, actionable results for stakeholders. By integrating Kane’s framework with insights from Bachman, Palmer, and Chapelle, this study establishes a rigorous, context-sensitive validation process, ensuring the LALS is a reliable tool for assessing LAL among V-EFLLs.

***4.2.2. The argument-based construction for LALS validation***

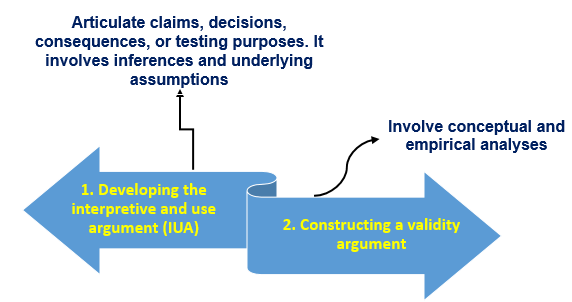
*4.2.2.1. Toulmin’s argument model*

Toulmin’s argument model (1958) underpins the argument-based validation framework, offering a structured method for linking claims to evidence. It consists of six components: data, claim, warrants, backing, qualifiers, and rebuttals, ensuring logical coherence between data and claims while addressing counterarguments. Widely influential across disciplines, Toulmin’s model has been adapted in educational measurement, notably by Kane (2013) for test validation. Kane’s adaptation emphasizes both practical and theoretical coherence in interpreting and using test scores, making it a foundational approach in assessment validation.

*4.2.2.2. The argument-based framework in this study*

Kane’s argument-based approach conceptualizes validity as a property of the interpretation and use of test scores, not the test itself. Validation involves constructing two arguments: The IUA, which specifies the intended interpretation and use of scores, and the Validity Argument, which evaluates the IUA’s coherence and plausibility with empirical evidence. This framework was extended by Bachman, Palmer, and Chapelle to include pragmatic, cognitive, and consequential aspects, ensuring alignment with real-world language tasks and contexts. The argument-based approach integrates both theoretical constructs and practical considerations, using quantitative and qualitative evidence to create a comprehensive validation process.

**Figure 4.2:** *The argument-based approach framework in the study*



***4.2.3. Procedural steps in framework application***

*4.2.3.1. Interpretive and use argument*

The IUA structures the logical sequence from observed test performances to claims about EFL teachers’ LAL competencies. This involves specifying assumptions about LALS’s domain alignment, scoring accuracy, score consistency, and construct reflection. Each inference is backed by warrants - statements justifying inferences based on assumptions about the assessment process. For instance, the domain description inference is supported by evidence that LALS items align with essential LAL components for EFL teachers.

*4.2.3.2. Validity argument*

The validity argument, the second stage, evaluates the coherence of the IUA, collecting evidence to support each inference. This process assesses if LALS scores accurately represent the LAL construct, are reliably scored, and generalize across diverse educational contexts. The LALS’s validity argument includes empirical support through Rasch modeling, expert reviews, and psychometric analyses, ensuring a comprehensive validation framework.

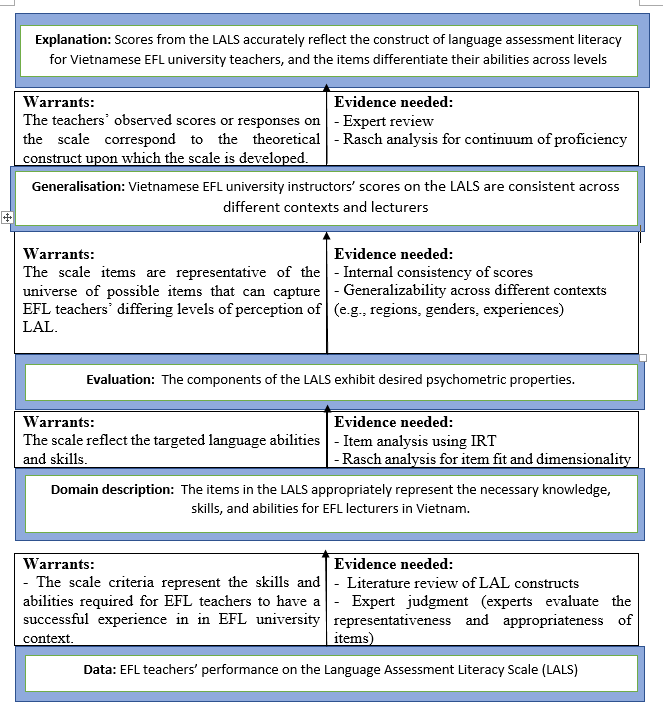
**4.3. Evidence required for inferences in argument-based validation**

The argument-based validation approach requires specific evidence to support each inference. Domain description evidence ensures content relevance to LAL through expert reviews and alignment analyses. Evaluation evidence establishes scoring accuracy using inter-rater reliability and Rasch modeling. Generalization evidence confirms score consistency across different contexts through reliability studies, including Cronbach’s alpha. Explanation evidence validates the construct by using factor analyses and Wright maps to verify the alignment between LALS scores and underlying LAL competencies.

**4.4. Articulating interpretive and validity arguments for the LALS**

This chapter details the argument-based validation of the LALS, focusing on four key inferences: domain description, evaluation, generalization, and explanation. Following Kane’s (1992, 2013) framework, the study develops an IUA to define intended score interpretations and a validity argument to assess empirical support. This targeted approach ensures a rigorous validation process while reserving extrapolation and utilization inferences for future research. The IUA establishes how LALS scores are interpreted and used, linking observed performances to theoretical claims. Each inference is supported by warrants and assumptions, guiding the types of empirical backing required. Domain description defines LAL competencies using literature and expert reviews, ensuring content relevance. Evaluation verifies that scoring rubrics and item statistics effectively measure LAL skills through psychometric analysis. Generalization ensures score consistency across diverse participant samples via Rasch analysis and internal consistency measures. Explanation examines whether LALS scores align with theoretical expectations, validating item difficulty and ability differentiation through Wright maps and thematic analysis. A structured validation approach integrates expert feedback, empirical data, and psychometric evidence, demonstrating the LALS’s reliability and practical applicability for assessing LAL among Vietnamese EFL lecturers.

**Figure 4.5:** *Validation framework for the LALS*



Chapter 4 describes the validation of the LALS using Kane’s argument-based framework, emphasizing the IUA and Validity Argument construction. By focusing on core inferences and gathering relevant evidence, this approach establishes the LALS as a reliable, valid tool for assessing LAL among Vietnamese EFL teachers. This evidence-driven validation framework, with additional inferences reserved for future research, confirms that the LALS aligns with its intended applications in diverse educational contexts.

**CHAPTER 5: RESEARCH METHODOLOGY**

This chapter details the methodology for creating and validating the Language Assessment Literacy Scale (LALS) for V-EFLLs within the higher education (HE) system. The methodology encompasses the research context, research questions, design, sampling, data collection instruments, analysis methods, and validation procedures. This structured approach lays a foundation for the LALS’s practical relevance and robust psychometric properties.

**5.1. Research context**

The study is set within Vietnam’s HE system, where recent education reforms and an emphasis on English for global integration have heightened the need for Language Assessment Literacy (LAL) among EFL lecturers. Despite MOET’s policies and standards, assessment practices often rely on high-stakes testing with a focus on memorization over communicative skills, creating a test-centric system. Furthermore, institutional constraints - such as large class sizes, resource disparities, and heavy teaching loads - have limited opportunities for formative assessment, especially in rural areas. Culturally, Vietnam’s Confucian heritage emphasizes a hierarchical student-teacher relationship, which challenges more formative, student-centered approaches. EFL teachers often have limited training in varied assessment techniques, with a strong reliance on traditional, summative methods. These combined factors emphasize the need for a localized LAL framework to guide Vietnamese EFL educators in adopting more balanced and effective assessment practices.

**5.2. Research questions revisited**

The study addresses two main objectives:

Developing the LALS: To assess critical competencies for V-EFLLs’ LAL, focusing on knowledge, skills, and abilities in assessment, and validating the LALS: Using Kane’s (2006) argument-based validation framework, the study validates four inferences - domain, evaluation, generalization, and explanation:

Domain inference: To what extent do the criteria in the LALS accurately represent the LAL construct for V-EFLLs, as determined by expert feedback and content relevance?

Evaluation inference: To what extent do the criteria of the LALS for V-EFLLs demonstrate satisfactory psychometric properties to ensure reliable and valid measurement?

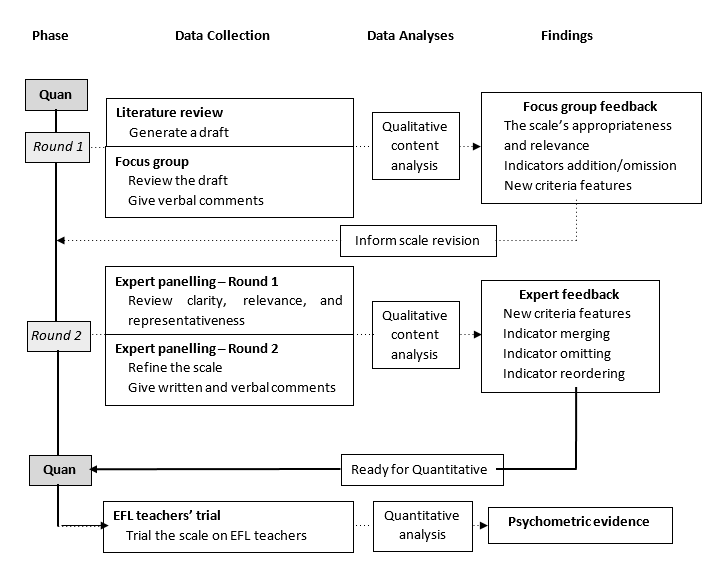
Generalization inference: How generalizable are the LALS scores across different demographic groups and levels of experience among the studied V-EFLLs?

Explanation inference: How well do V-EFLLs’ self-ratings on the LALS reflect their actual LAL competencies, and to what extent do the items effectively differentiate their proficiency levels?

**5.3. Research design**

An exploratory sequential mixed-methods design was applied, integrating qualitative and quantitative data through an argument-based validation framework. The qualitative phase, including focus groups and expert feedback, informed the subsequent quantitative phase. This design allowed for a detailed exploration of LAL before transitioning to measurement, and it facilitated the alignment of qualitative insights with quantitative evaluation.

**Figure 5.1.** *Procedural diagram of the scale development and validation stages*



**5.4. Research participants and sampling**

***5.4.1. Qualitative phase:*** Purposive sampling was used for focus groups and expert reviews. In the focus group, 12 participants, including EFL lecturers and language assessment experts, provided insights into the framework's clarity and applicability. Six expert reviewers, with extensive backgrounds in language assessment, further refined the framework to ensure it aligned with the Vietnamese HE context.

**5.4.2. Quantitative phase:** A stratified sampling strategy across Vietnam’s northern, central, and southern regions ensured a representative sample for quantitative validation. The final sample included 177 participants, meeting requirements for Rasch analysis and allowing for reliable item parameter estimates. This sampling approach balanced logistical feasibility with comprehensive coverage across Vietnam’s HE landscape.

**5.5. Data collection instruments**

The study used three main instruments:

**Focus group protocol**: A guide with 12 open-ended questions facilitated in-depth discussions on the framework’s content relevance. The protocol, field-tested with high school teachers, provided a structured yet flexible format for gathering participant feedback.

**Semi-structured interviews for expert panelling**: In two rounds, six experts provided detailed feedback on the framework's clarity, relevance, and structure. Their insights helped refine indicators and align the framework with developmental principles in LAL.

**Online questionnaire for trialling:** This included demographic questions and self-assessment items based on the LALS framework, field-tested to ensure clarity. Participant feedback helped to finalize items, ensuring accurate self-assessment among a larger sample.

**5.6. Data analysis**

The study employed both qualitative and quantitative analysis techniques to address the research questions and support validity inferences.

***5.6.1. Qualitative analysis***

Thematic analysis, following Braun and Clarke’s (2012) six-phase method, identified patterns and themes from focus groups and expert reviews. Coding through MAXQDA allowed for refining themes, particularly related to the Domain Description inference, ensuring the LALS framework accurately reflected key competencies in Vietnamese HE.

***5.6.2. Quantitative analysis***

**Rasch analysis:** The Partial Credit Model (PCM) was used to calibrate items and determine item-person interactions, providing a nuanced understanding of item difficulty and discrimination.

**Reliability analysis:** Both internal consistency (Cronbach’s alpha) and Rasch-based indices confirmed reliability. Person and item separation indices supported the scale’s ability to distinguish varying competency levels.

**Model fit and item discrimination**: Fit statistics (INFIT and OUTFIT) measured item alignment with the LAL construct, with mean-square values and t-statistics guiding item refinement. The item separation index and point-biserial correlations confirmed effective item discrimination.

**Item characteristic curves (ICCs) and Wright maps**: ICCs displayed how well items differentiated competency levels, while Wright maps visualized item distribution, confirming the continuum of LAL skills from basic to advanced.

This chapter presented a comprehensive approach to developing and validating the LALS through a mixed-methods design grounded in argument-based validation. The research process systematically addressed each validity inference, aligning with best practices for educational assessment in Vietnamese HE. This methodology, combining qualitative insights and quantitative rigor, supports the LALS’s applicability and robustness as a tool for evaluating LAL among V-EFLLs. The next chapter will further analyze the LALS’s psychometric properties, presenting validation outcomes that reinforce its reliability and practical relevance.

**CHAPTER 6: THE RESULTS OF VIETNAMESE EFL LECTURERS’ LANGUAGE ASSESSMENT LITERACY SCALE VALIDATION - DOMAIN DEFINITION INFERENCE**

This chapter details the validation process for the LALS through the domain definition inference, focusing on establishing the scale’s relevance and applicability within Vietnamese HE.

**6.1. Validity evidence 1: Scale development**

The development of the LALS for Vietnamese EFL lecturers was guided by a literature review and contextual analysis. The LALS was structured around three key assessment stages: Preparing for Assessment, Conducting Assessment, and Revising Assessment. These stages align with established models such as the Dreyfus and Dreyfus skill acquisition framework (1986) and competence-based assessment theories. The draft framework was iteratively refined based on expert feedback, ensuring relevance to Vietnamese higher education.

Sixty-five performance indicators were identified to capture key LAL competencies across assessment stages. These indicators, informed by literature and local assessment practices, were designed to reflect fundamental assessment principles such as validity, reliability, and fairness. Key competencies included understanding assessment frameworks, setting learning targets, planning assessment methods, designing and implementing assessment tasks, and interpreting assessment results. The indicators were aligned with international and local standards, ensuring contextual applicability.

Quality criteria were established using the Dreyfus and Dreyfus skill acquisition model (1986) to represent varying levels of competence. Performance indicators were structured along a continuum from basic to expert levels, ensuring clear progression in assessment literacy development. This structure was refined to improve clarity, practicality, and consistency with partial credit scoring models used in Rasch measurement.

Focus group discussions with Vietnamese EFL lecturers resulted in substantial refinements to the draft LALS. A total of 380 primary codes were initially extracted, later condensed into 260 open codes and 44 subcategories. Nine overarching categories emerged, including framework revisions, cultural relevance, practical application, and usability enhancements. Teachers emphasized the need to reduce redundancy, refine performance indicators, and ensure alignment with Vietnamese assessment standards. Key adjustments included integrating formative assessment components, simplifying criteria, and improving clarity in descriptor language.

**6.2. Validity evidence 2: Expert paneling data**

***6.2.1. First round of expert review***

The first round of expert paneling focused on refining the scale’s clarity, relevance, and applicability. Six experts provided feedback on comprehensibility, structural balance, external alignment, and generalizability. Through thematic coding, key revisions included clarifying assessment-related terminology, improving logical flow, and ensuring alignment with Vietnamese and international competency standards. Experts recommended strengthening digital literacy and formative assessment components, refining rubric descriptions, and streamlining redundant indicators. These modifications ensured that the LALS accurately reflected assessment competencies required in Vietnamese higher education.

***6.2.2. Second round of expert review***

The second round of expert paneling finalized the scale, focusing on refining specificity, logical flow, and usability. Competencies were reordered to enhance coherence, and sub-indicators were introduced to provide greater depth while maintaining usability. Experts validated the scale’s alignment with national and international frameworks, ensuring clarity in terminology and practical applicability. Minor revisions were made to enhance readability, streamline indicator wording, and refine examples to facilitate user comprehension. Experts agreed that the LALS was ready for field testing, confirming its robustness as a valid measurement tool for LAL assessment.

**6.3. Overall assessment of the domain inference**

The domain inference was assessed based on two key assumptions: that essential LAL competencies could be systematically derived from literature and that scale items could effectively serve as indicators of these competencies. Empirical validation supported both assumptions, demonstrating that the LALS accurately captures the required skills and knowledge for EFL lecturers in Vietnamese universities. Expert paneling and focus group findings confirmed that the scale is contextually relevant, theoretically grounded, and practically applicable.

Table 6.4 summarizes the validation process, highlighting key warrants, evidence, and the degree of support for the domain inference. The iterative refinement process ensured that the scale met rigorous validity criteria, positioning it as a reliable tool for assessing LAL in Vietnamese higher education.

**Table 6.4:** *Summary of warrants, evidence and degree of support for the domain definition inference*

|  |  |  |
| --- | --- | --- |
| **Warrant** | **Evidence** | **Degree of support** |
| The scale criteria represent the skills and abilities required for EFL lecturers to have a successful experience in an EFL university classroom context. | A comprehensive review of relevant documents and empirical studies to identify components of LAL in this context. | Fully supported |
| The identification of items from previous studies that represent each component, and the refinement of items to better suit the context in which the scale is administered. | Fully supported |

This chapter documented the validation of the LALS domain inference, detailing the development process, expert paneling, and empirical findings. The iterative design ensured that the scale reflects critical LAL competencies while remaining contextually relevant. The refined LALS is now prepared for field testing, marking a significant step in the development of a robust, evidence-based tool for assessing language assessment literacy among Vietnamese EFL lecturers.

**CHAPTER 7: THE RESULTS OF VIETNAMESE EFL LECTURERS’ LANGUAGE SCALE VALIDATION - EVALUATION AND GENERALIZATION INFERENCES**

This chapter presents findings supporting the evaluation and generalization inferences within the validation framework for the LALS, aiming to assess V-EFLLs’ language assessment competencies.

**7.1. Evaluation inference**

***7.1.1. Validity evidence 1: Scale development and expert consensus***

The development of the LALS followed a rigorous expert consensus process to ensure the scale accurately captured the targeted language assessment abilities essential for V-EFLLs. Supporting details from the data provided demonstrate the rigorous procedures involved in scale development. The LALS was initially developed through an extensive review of scholarly literature on LAL, integrating key theoretical models and frameworks. These models formed the foundation of the initial scale, with a focus on identifying the core competencies necessary for language assessment. As detailed in Chapter 5, Section 5.1.1, the items in the scale were constructed by drawing on both theoretical insights and focus groups with SMEs, ensuring the scale was grounded in the relevant constructs of LAL for V-EFLLs.

Expert focus groups reviewed the scale and provided critical feedback on its alignment with the skills and competencies required for V-EFLLs. These experts identified areas for improvement, ensuring the scale covered relevant language assessment abilities. Throughout the iterative process, as documented in the analysis, experts provided feedback on the relevance, clarity, and representativeness of the items. For instance, specific feedback from SMEs helped refine the items to better reflect the competencies outlined in the LAL framework. The scale underwent multiple iterations of revisions based on expert input. After each round of feedback, the scale was revised to enhance its alignment with both theoretical frameworks and the practical needs of the Vietnamese EFL teaching context. Chapter 6 details two rounds of trials in which the scale was reviewed and refined based on expert feedback. The experts’ comments were integrated into the revisions, leading to improved clarity and appropriateness of the items, particularly in terms of their relevance to language assessment skills.

The consensus-building process among experts was systematically documented, providing transparency throughout the scale development. Experts offered detailed justifications for their suggested changes, and these were used to ensure the scale reflected the competencies required for EFL lecturers in the Vietnamese context. Appendices F.1, F.2, and F.3 present a clear representation of the revised items, showing how expert feedback shaped the final scale. The alignment between competency descriptions and proficiency levels was strengthened through this iterative review process, enhancing the accuracy and applicability of the scale.

The final version of the LALS was the product of this rigorous and iterative development process, with expert consensus confirming the scale’s appropriateness for assessing the targeted language assessment abilities. The experts validated the scale’s content, ensuring it reflected the competencies needed by V-EFLLs. The final 42-item behaviorally anchored rating scale, developed after expert reviews, demonstrates the careful revision of items based on expert evaluations. Recommendations from the experts, such as refining descriptors to improve clarity and ensuring the items matched the proficiency levels of V-EFLLs, were integrated into the final version of the scale, ensuring it was a reliable tool for assessing LAL.

***7.1.2. Validity evidence 2: Psychometric properties***

The LALS, as presented in the previous chapter, was constructed based on the rubrics of the framework for V-EFLLs’ language competence developed through an extensive review of the scholarly literature on LAL, focus groups, and expert reviews (see Chapter 5). This 42-item behaviorally anchored rating scale was developed for the participants to self-assess their competence in language assessment. Based on the descriptive statistics of the scale, 177 valid responses were collected from 213 V-EFLLs who took part in the self-assessment of their test item writing according to the scale. Thirty-six responses were excluded because the respondents were non-major English lecturers and did not meet the participant criteria. Therefore, 177 cases were included in the IRT estimation procedure.

The results of the calibration for the finalized 42-item LALS generated from both Classical and Rasch analyses are presented in Table 7.2. For each item, the summary statistics are displayed as follows: The item Thurstonian thresholds (ɣ), reported in Logit value, followed by the standard error in relation to each item difficulty estimate (SE). The Thurstonian thresholds for each item category in the scale illustrated in Table 7.1 show increasing levels of difficulty. Next, the item-rest correlation is reported, which refers to the relationship between the item score and the scale score after removing the item score from the total test score. Finally, the INFIT mean square (INFIT MNSQ) and t-statistics are presented, showing the extent to which the pattern of item responses fits the Rasch model. The item-level estimates that the LALS captures increasing cognitive demand with progressive ability levels. The standard deviation of the item difficulty level was 0.685, with relatively small standard errors for each item. The high item separation reliability (r = 0.945) confirms that the scale effectively differentiates between competence levels, aligning with the psychometric expectations of an assessment tool designed for this purpose.

Further evidence supporting the psychometric quality of the LALS includes the evaluation of item discrimination indices and category fit statistics. The item-rest correlations ranged from 0.29 to 0.62, with most items displaying strong discrimination indices. While a few items demonstrated marginally lower discrimination, they still met the acceptable threshold for inclusion in the scale. This confirms that the LALS functions as an effective instrument for differentiating varying levels of LAL among Vietnamese EFL lecturers.

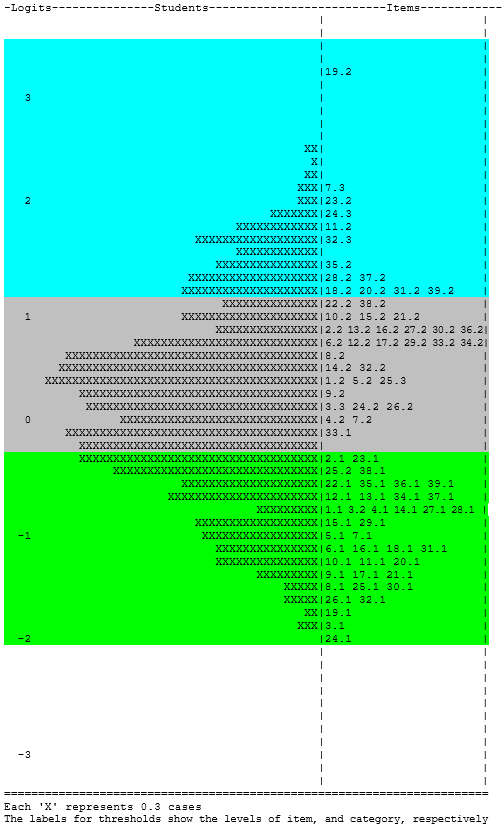
This chapter confirmed the LALS’s validity through expert-informed development, psychometric properties, and consistency across diverse groups. The next chapter will address the explanation inference, integrating the results from prior inferences.

**CHAPTER 8: THE RESULTS OF VIETNAMESE EFL LECTURERS’ LANGUAGE ASSESSMENT LITERACY SCALE VALIDATION - EXPLANATION INFERENCE**

This chapter examines the explanation inference, evaluating whether the observed scores of V-EFLLs align with the theoretical constructs underlying the LALS. Two key assumptions guide this validation: (1) scale items must accurately capture intended competencies, and (2) the scale should differentiate competence levels effectively. Psychometric analyses, including Rasch modeling and Wright maps, are used to verify these assumptions.

**8.1. Validity evidence 1: Variable map interpretation**

**Figure 8.1**. *Interpretability of the variable map for the LALS*



The variable map visualizes item difficulty and lecturer competence on a logit scale. It confirms that the mean teacher ability (0.21 logits) is well-matched with item difficulty, supporting the scale’ s effectiveness in assessing LAL levels. The Wright map highlights a hierarchical progression of competencies, from basic to expert levels, showing how lecturers develop assessment literacy over time. The partial credit model (PCM) confirms the expected probabilistic nature of this progression, ensuring that higher proficiency levels encompass lower-level competencies. Items were clustered into four developmental levels, revealing a coherent structure aligned with existing skill acquisition models. Level 1 (Basic) reflects foundational assessment knowledge, while Level 4 (Expert) involves refining assessments, implementing digital methods, and ensuring cultural alignment. The competence descriptions derived from the item clusters provide a structured framework for assessing V-EFLLs.

**8.2. Validity evidence 2: Comparison of empirical and intended continua**

Comparing empirical and intended continua confirms alignment with theoretical proficiency levels (Basic to Expert) while highlighting key adjustments: Competence placement: Advanced tasks like stakeholder engagement (Item 33.1) were elevated, while foundational items (3.2, 25.2) were reassigned to lower levels for accuracy. Advanced proficiency emphasis: Critical skills such as performance level consistency (Item 23.2), identifying improvement areas (Item 11.2), and scorer training for reliability (Item 35.2) were placed at Level 4 to reflect their importance in high-stakes assessment. Technology integration challenges: Digital methods in assessments (Item 19.2) emerged as the most demanding skill, reinforcing the need for further training.

The empirical scale, validated through Rasch analysis, aligns with Dreyfus’s (1986) skill acquisition model and integrates modern assessment strategies such as technology-enhanced methods, annotated feedback, and feedforward mechanisms. The findings emphasize digital literacy gaps, underscoring the importance of professional development in technology-assisted assessment practices.

**8.3. Overall assessment of the explanation inference**

Both assumptions of the explanation inference, that items represent targeted components and that scale scores vary with ability levels, are supported. Statistical analyses validated the alignment of scale items with the theoretical construct, and the score variation across regions and experience levels reflects the LALS’s construct validity, capturing differing proficiency levels.

This chapter validated the explanation inference, confirming the LALS’ s construct validity and its capability to distinguish between proficiency levels. Psychometric and content analyses, along with Rasch modeling, show that the LALS accurately represents language assessment competencies and differentiates teacher abilities. The LALS progression aligns with Dreyfus and Dreyfus’s skill model, confirming reliable measurement of V-EFLLs’ language assessment literacy from novice to proficient levels. The next chapter will synthesize these validation results in the overall framework evaluation.

**CHAPTER 9: CONCLUSIONS AND IMPLICATIONS**

This chapter presents the conclusions, implications, and future research directions for the development and validation of the LALS for V-EFLs. It synthesizes key findings, evaluates the study’s strengths and limitations, and provides recommendations for teacher training, assessment practices, and future research.

**9.1. Summary of research findings and validity argument**

The study developed and validated the LALS to assess Vietnamese EFL lecturers' competencies in preparing, conducting, and revising assessments. Based on Dreyfus & Dreyfus’s (1986) skill acquisition model, it tracks four developmental stages: Basic, Emergent, Independent, and Expert across 39 items and 122 quality criteria. Psychometric analyses (CTT and Rasch modeling) confirmed its validity and reliability, with high item separation reliability (r = 0.92) and strong alignment between teacher competence (3.79 logits) and item difficulty (3.59 logits). The scale accurately differentiates proficiency levels and aligns observed scores with actual competencies. While tailored for Vietnamese EFL lecturers, its competence-based structure makes it adaptable to other contexts.

The LALS is uniquely tailored to Vietnam’s educational context, aligning with MOET policies, the KNLNNVN framework, and the NFLP. Unlike international models, it addresses local challenges such as large class sizes and limited technology while promoting fair, reliable, and transparent assessment practices. It balances digital transformation with practical constraints, ensuring realistic technology integration. Its developmental approach provides a structured path for lecturers to progressively enhance their assessment literacy.

To establish the validity and reliability of the LALS, the study examined four key inferences: domain description, evaluation, generalization, and explanation.

- Domain description was confirmed through expert feedback, theoretical frameworks, and empirical validation, ensuring the scale comprehensively represents LAL.

- Evaluation showed strong psychometric properties, with Rasch analysis confirming well-distributed item difficulty, high internal consistency, and strong construct validity, making it a reliable measure of LAL proficiency.

- Generalization demonstrated the scale’s stability across gender, experience levels, and regions, though minor item misfits led to recommended adjustments for improved consistency.

- Explanation confirmed that LALS scores accurately reflect lecturers’ competencies, with Wright maps and Rasch modeling validating its hierarchical structure and developmental progression.

Through these validations, the LALS was confirmed as a reliable, valid, and contextually relevant instrument for assessing and improving language assessment literacy in Vietnam, with potential for adaptation in similar educational settings.

**9.2. Implications**

The study has theoretical, methodological, and pedagogical implications:

Theoretically, LAL is conceptualized as a dynamic, evolving competence, integrating knowledge, skills, and adaptation to assessment contexts. The study highlights the contextual nature of LAL in non-Western settings, emphasizing localized assessment practices aligned with Vietnamese MOET, KNLNNVN, and NFLP policies. The argument-based validation framework (Kane, 2006) was applied, demonstrating a systematic, multi-step validation approach. Methodologically, the developmental assessment approach structures LAL progression along a continuum, aligning with Dreyfus’s model and Vygotsky’s ZPD. Rasch models effectively differentiated teacher abilities and validated the scale’s psychometric properties. The study refines competency-based assessment tools, emphasizing digital literacy and real-world applications. Pedagogically, the LALS empowers self-assessment and reflective teaching, encouraging continuous professional development. The scale informs teacher training programs, supporting structured development from basic to expert assessment competence. Institutions can integrate the LALS into pre-service and in-service training, using data to tailor professional development and enhance curriculum design.

**9.3. Limitations of the ttudy**

The study faced several limitations. The sample was geographically imbalanced, with most participants from northern Vietnam, limiting generalizability. Additionally, reliance on self-report data may not fully capture actual competence levels. The cut-off scores for proficiency levels require further validation to confirm their applicability across diverse lecturer populations.

**9.4. Areas for continued research**

Future research should explore regional variations in LAL by expanding data collection to central and southern Vietnam. Further validation of cut-off scores through longitudinal studies is necessary to ensure their stability. Researchers should also investigate how digital assessment literacy develops among lecturers, considering the increasing role of technology in language assessment. The LALS framework could be adapted to other educational settings, including different subjects beyond English. Additionally, future studies should incorporate multiple data sources, such as classroom observations, peer evaluations, and teacher portfolios, to provide a more comprehensive measurement of LAL competence.

The LALS represents a significant advancement in assessing V-EFLLs’ LAL competencies. It is contextually relevant, psychometrically sound, and developmentally structured, offering a practical framework for professional growth. By integrating cultural, policy, and pedagogical considerations, it provides a structured approach to teacher competence development. Future research should continue refining and expanding its application to enhance language assessment literacy and teacher training worldwide.

|  |  |  |
| --- | --- | --- |
|  | **LIST OF PUBLISHED PAPERS** |  |
| 1. | **Nguyen Thi To Loan** (2021). *Conceptualizing language teachers’ language assessment literacy: A review of literature*. Proceedings of 2021 International Graduate Research Symposium (IGRS), pp. 898-906. Vietnam National University Press, Hanoi. |  |
| 2. | **Nguyễn Thị Tố Loan** (2022). *Năng lực khảo thí ngôn ngữ của giáo viên dạy ngoại ngữ: Những khoảng trống trong nghiên cứu*. Kỷ yếu Hội thảo Ngữ học toàn quốc 2022 - Ngôn ngữ học Ứng dụng trong Xu hướng Hội nhập Quốc tế, tr. 952-959. Nhà xuất bản Đại học Quốc gia Hà Nội, Hà Nội. |  |
| 3. | **Nguyen Thi To Loan**, Duong Thu Mai (2023). Assessment literacy in language education: the story so far. *VNU Journal of Foreign Studies*, 39(4), pp. 59-72. |  |
| 4. | **Nguyen Thi To Loan**, Vu Thi Quynh Dung (2023). *Defining the language assessment literacy framework required for EFL language teachers*. Proceedings of the 11th OpenTESOL International Conference 2023, pp. 229-255. Ho Chi Minh City Open University. |  |
| 5. | Cutri, J., Yip, S.Y., **Nguyen, T. T. L.** (2024). Global Teachers on the Move: A Discussion of Teacher Identity and Intercultural Pedagogical Practices in Cross-cultural Contexts. In: Chowdhury, R., Tuan, H. A. (Eds), *Engaging with Australasia: Comparative Research on ELT and English Teacher Education* (pp. 379-403). Palgrave Macmillan, Singapore. |  |
| 6. | **Nguyen Thi To Loan**, Ta Thi Mai Huong, Ngo Thi Thanh and Ngoc-Tung Vu (2024). Vietnamese EFL Teachers’ Conceptions of Assessment: A Model to Navigate Accountability in the Digital Transformation of Education. In: Yen Hoang Phuong, Thanh-Thao Le (Eds.), *Exploring Teacher Beliefs: Insights and Perspectives* (pp. 235-257). Nova Science Publishers, New York, USA. |  |
| 7. | **Nguyen Thi To Loan**, Duong Thu Mai (2024). *Exploring EFL teachers’ assessment literacy in digital educational contexts: Opportunities and challenges*. Proceedings of the 12th OpenTESOL International Conference 2024, pp. 464-476. Ho Chi Minh City Open University. |  |