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 - 2024**

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 2024

**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**

In recent decades, global educational reforms have emphasized regular assessments to enhance student learning (Inbar-Lourie, 2008; Davison & Leung, 2009). This shift has expanded teachers' roles, requiring them to integrate varied assessment methods into teaching, involve students actively, use assessment data to inform instruction, and align reports with external standards (Inbar-Lourie, 2008; Rea-Dickins, 2008). Consequently, assessments now consume a significant portion of teachers' time, about one-quarter to one-third (Cheng et al., 2004).

Assessment Literacy (AL), defined as the ability to conduct valid and reliable assessments, is increasingly recognized as essential for effective teaching. Strong AL allows teachers to align assessments with instructional goals, while inadequate AL risks unreliable assessments, potentially misleading stakeholders like students, parents, and administrators (DeLuca & Klinger, 2010; Xu, 2019b). AL also enhances teaching effectiveness by enabling teachers to interpret and use assessment data to meet diverse student needs (Pastore & Andrade, 2019).

In language education, Language Assessment Literacy (LAL) is particularly crucial for EFL teachers, who act as both instructors and assessors. LAL includes the specific knowledge and skills necessary for assessing language learners effectively (Inbar-Lourie, 2013; Vogt et al., 2020). High LAL correlates with improved language instruction, as teachers can make informed decisions to support students' progress (Harding & Kremmel, 2016; Hakim, 2015).

Despite its importance, reliable tools to measure LAL, especially for EFL contexts, remain scarce. Many language teachers possess limited AL, leading to misalignment between assessments and learning goals (Inbar-Lourie, 2008; Xu & Brown, 2016). Inadequate focus on assessment training in teacher education further exacerbates this issue (Stabler-Havener, 2018).

In Vietnam’s tertiary EFL education, this gap is pronounced, with little research on LAL among Vietnamese EFL lecturers. Rapid expansion of English education, driven by socio-economic reforms (Ngo, 2018, 2021), heightens the need for competent language assessors. Despite initiatives like the National Foreign Languages Project 2008 - 2020, comprehensive LAL frameworks or tools for tertiary teachers are lacking.

This study addresses these gaps by developing and validating a tailored LAL measurement scale for Vietnamese higher education, aiming to enhance teacher training programs and improve English language teaching quality in Vietnam.

**1.2. Research problem**

The field of LAL has gained significant attention since Brindley (2001) integrated AL into language assessment. However, despite this focus, many challenges related to LAL remain unresolved (Giraldo, 2018). Theoretical and empirical studies on LAL are still limited (Fulcher, 2012; Giraldo, 2018; Hakim, 2015), indicating that LAL is in its developmental stages and requires further exploration (Yan & Fan, 2020). While Harding and Brunfaut (2020) argue that LAL research has advanced, a consensus on its definition has not yet been achieved, resulting in diverse studies and discussions about its nature.

The complexity of assessing language skills, influenced by numerous factors affecting the reliability and validity of assessment processes, necessitates specialized LAL components suited to specific educational and social contexts. One significant gap in the literature is the absence of validated instruments for measuring LAL, particularly when compared to the abundance of AL inventories in general education. Studies reveal that LAL is a multifaceted, contextualized concept with poorly defined dimensions (Coombe et al., 2020; Giraldo, 2021). There is an urgent need to better define EFL teachers’ LAL and develop suitable measurement tools.

In the Vietnamese higher education context, the push for English language proficiency as a key component of students’ employability has increased the demand for accurate language assessment (Bodewig & Badiani-Magnusson, 2014; Tran, 2019). However, there is a lack of comprehensive frameworks outlining the necessary competencies for Vietnamese EFL teachers in this area (Nguyen, 2017; Pham, 2017). Although various reforms, such as Project 2020, have aimed to enhance English language teaching and assessment in Vietnam, the assessment practices at the tertiary level remain largely summative and test-oriented (Nguyen, 2013; Ngo, 2018). Teachers’ limited LAL, combined with material constraints and a lack of agency, has hindered the adoption of innovative assessment methods (Phung & Tran, 2017; Vu, 2017). Furthermore, there is a lack of valid, context-sensitive tools to measure the LAL of Vietnamese EFL teachers, particularly in higher education settings (Ngo, 2021; Giraldo, 2018).

Despite efforts to conceptualize LAL from the perspectives of both researchers and pre-service teachers (Butler et al., 2021), the focus remains predominantly on in-service teachers, with varying experience and expertise in language assessment. Investigating how in-service EFL teachers in higher education settings conceptualize LAL is of particular interest. Moreover, studies suggest that LAL is demonstrated through observable and measurable behaviors, reflecting teachers’ competence in language testing and assessment (Griffin, 2014, 2018; Masters, 2003).

Given the absence of a well-articulated validity argument and validated instruments, particularly in Vietnam, this study seeks to address the gap by developing and validating a Language Assessment Literacy Scale (LALS) specifically designed for Vietnamese EFL lecturers. It will contribute to the growing need for reliable LAL measurement tools and provide a foundation for future studies on LAL development in Vietnam and similar educational contexts.

**1.3. Aims of the study**

This study seeks to address gaps in the existing literature on LAL by developing and validating a comprehensive model that captures the core competencies of Vietnamese EFL lecturers in language assessment. It will contribute to improving teacher education programs and professional development in Vietnam by providing insights into the current state of LAL among EFL lecturers.

The study has two primary objectives:

- To develop an instrument (the LALS) to comprehensively assess the LAL of Vietnamese EFL lecturers.

Building on previous research and empirical data from Vietnamese lecturers, the first aim is to construct and calibrate a framework representing the LAL competencies specific to the context of Vietnamese higher education (HE). This instrument will focus on the critical knowledge, skills, and abilities required for effective language assessment in Vietnamese EFL lecturers.

- To validate the developed LALS using an argument-based validation framework.

The second aim is to validate the interpretations of the scores using Kane’s (2006) argument-based validation framework, expanded by Chapelle et al. (2008). This process will gather both qualitative and quantitative evidence to evaluate four main validity inferences: Domain inference, Evaluation inference, Generalization inference, AND Explanation inference. These inferences provide a comprehensive framework for evaluating the validity and reliability of the LALS in measuring the language assessment literacy of Vietnamese EFL lecturers. The findings from this study will inform the design of curriculum and professional development programs aimed at enhancing language assessment competencies in Vietnamese higher education.

**1.4. Research questions**

In alignment with the primary aims of this study - to develop and validate a scale for measuring Vietnamese EFL lecturers’ LAL according to argument-based validation approach - the following research questions guide the empirical investigation:

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

RQ2: To what extent is the developed LAL scale a valid and reliable instrument for assessing LAL of Vietnamese EFL lecturers?

RQ2.1. To what extent do the criteria in the LALS accurately represent the LAL construct for Vietnamese EFL lecturers, as validated by expert feedback and content relevance? (Domain description inference)

RQ2.2. To what extent do the criteria of the LALS for Vietnamese EFL lecturers demonstrate satisfactory psychometric properties to ensure reliable and valid measurement? (Evaluation inference)

RQ2.3. How generalizable are the LALS scores across different demographic groups and levels of experience among the studied Vietnamese EFL lecturers? (Generalisation inference)

RQ2.4. How well do the Vietnamese EFL lecturers’ self-rating on the LALS reflect the underlying construct of LAL, and to what extent do the items effectively differentiate their levels of LAL across varying competency levels? (Explanation inference).

**1.5. Methodological considerations**

This study utilized an exploratory sequential mixed methods design (Creswell & Plano Clark, 2018) to develop and validate the LALS for Vietnamese EFL lecturers. The process began with qualitative data collection through semi-structured interviews and expert feedback, which informed the subsequent quantitative phase. This approach allowed for an in-depth exploration of LAL components before transitioning to measurement, with data integration following an argument-based validity framework.

The qualitative phase focused on gathering insights from EFL teachers on the clarity, relevance, and applicability of LALS items. Expert feedback was also gathered to evaluate the items’ appropriateness for representing LAL competencies. The quantitative phase then involved scoring teachers’ responses to the trialed scale, providing psychometric evidence to support the validity of the instrument.

The development of the LALS included a multi-stage process beginning with a literature review to draft the scale. This draft underwent a focus group review for qualitative feedback on its relevance and accuracy, leading to revisions based on participants’ suggestions. Following this, two rounds of expert paneling refined the scale further, Round 1 focused on clarity, relevance, and representativeness, while Round 2 emphasized developmental alignment and logical item progression.

In the final quantitative phase, the revised scale was trialed with EFL teachers, whose responses were analyzed for psychometric validity. Some qualitative findings were translated into quantitative variables, enabling a comprehensive analysis that merged data from both phases. This method ensured that initial qualitative insights meaningfully informed the quantitative assessment, aligning with established practices in construct validation (e.g., Kane, 2006; Messick, 1995).

The instrument design incorporated construct mapping (Wilson, 2011; Wolfe & Smith, 2007a), defining LAL as a continuum of skills. A developmental model arranged items to reflect increasing proficiency, supported by the partial credit model (Masters, 1982) for response analysis. Rasch analysis provided structural validity by mapping item difficulty and teacher ability on a single scale.

The systematic process combined theoretical and psychometric approaches, establishing a criterion-referenced framework that breaks down LAL into core competencies, observable behaviors, and performance criteria. Through this rigorous, multi-stage methodology, the LALS provides a valid, context-sensitive tool for assessing the LAL competencies of Vietnamese EFL lecturers, supporting enhanced AL within Vietnam’s higher education context.

**1.6. Significance of the study**

This study holds significant theoretical, methodological, and practical contributions to the field LAL research, particularly in the context of Vietnamese EFL lecturers.

Theoretically, the study addresses an important research gap by clarifying the conceptualization of LAL from the perspective of in-service tertiary EFL teachers, a group often overlooked in existing literature. Previous studies have largely focused on pre-service teachers (Mertler & Campbell, 2005) and have primarily been conducted in Western contexts. By focusing on in-service EFL university instructors in Vietnam, this study extends the scope of LAL research, offering a more holistic understanding of its components. This contributes to the broader body of LAL knowledge by shedding light on how LAL is conceptualized and operationalized within a non-Western, higher education context. The insights from this study will provide a valuable framework for future research and help advance efforts to improve language education and assessment practices.

Methodologically, this study applied both the developmental approach for scale development and Kane’s argument-based validation framework (Kane, 2004, 2013). While these frameworks have been extensively used in language testing to ensure robust test design and quality, their application in LAL research has been limited. This study’s use of these frameworks contributes to the methodological rigor of LAL research, offering a structured approach to developing and validating a context-specific scale for assessing LAL.

Practically, the study offers a ready-to-use, validated LALS that can be employed for pedagogical and research purposes. Vietnamese EFL university instructors, as well as instructors in similar contexts, can use the scale for self-assessment or professional development, identifying gaps in their language assessment knowledge and setting relevant goals. The findings from this research can inform the design of both pre-service teacher education programs and in-service professional development initiatives, addressing the specific LAL needs of instructors. Furthermore, this research provides valuable insights into current LAL levels among Vietnamese EFL lecturers, which can inform policy decisions and contribute to the broader development of LAL in the region.

**1.7. 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 Vietnamese EFL lecturers, 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**

Historically, literacy referred to foundational skills in reading, writing, and numeracy. Over time, literacy has expanded to include competencies like critical thinking, problem-solving, and digital proficiency (Taylor, 2013). In education, assessment literacy (AL), as Stiggins (1991) defined it, equips educators to understand, design, and implement assessments that enhance learning outcomes. Two perspectives on literacy dominate: a dichotomous view, labeling individuals as literate or illiterate, and a continuum approach, which considers literacy a range of proficiencies relevant to specific cultural or educational contexts.

AL consists of assessment, measurement, testing, and evaluation—each with a distinct purpose in the educational landscape (Brown & Abeywickrama, 2010). Assessment is a broad concept, encompassing processes to monitor and improve learning, while measurement assigns quantitative values to learning outcomes. Testing refers to specific instruments, like exams, that evaluate student competencies, and evaluation interprets these results to refine instructional practices. Together, these components enable educators to apply their skills effectively across contexts.

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

LAL, a specialized subset of AL, integrates technical, contextual, and ethical skills in language assessment. Three major theoretical perspectives contribute to our understanding of LAL:

***2.2.1. Sociocultural theory and contextualized competence***

Based on Vygotsky’s sociocultural theory, competence is co-constructed through social interactions within specific contexts. For language assessment, sociocultural theory suggests that educators develop LAL by engaging in professional communities and learning from peers. Vygotsky’s Zone of Proximal Development (ZPD) underscores that LAL grows through guided interactions, where educators expand their competencies through support and collaboration. This framework emphasizes LAL’s context-sensitive nature, especially in diverse settings where assessments need to be tailored to students’ linguistic and cultural backgrounds.

***2.2.2. Constructivist theory and experiential learning***

Constructivist theory, advocated by Piaget (1977) and Dewey (1938), views knowledge construction as an active, experiential process. Constructivist principles suggest that educators build LAL through reflective practice, continually refining their assessment approaches based on real-world classroom experiences. This reflective approach emphasizes that LAL is flexible and adaptable, allowing educators to adjust their practices as they gain insight from their students’ needs and learning outcomes.

***2.2.3. Psychometric theory and measurability of competence***

Psychometric theory provides a structured, measurable framework for LAL by applying models like Rasch and Item Response Theory (IRT). These models enable educators to evaluate their competencies incrementally, quantifying LAL as it develops. For example, Rasch modeling aligns individual skill levels with task complexities, allowing institutions to track educators’ growth systematically. This approach supports LAL as an evidence-based, measurable competence, ensuring reliability and consistency in assessments.

These theories collectively frame LAL as a developmental, context-sensitive competence that combines technical skills with socio-cultural awareness, adaptable to varied educational environments.

**2.3. Approaches to defining the construct of LAL**

Several frameworks inform the structure of LAL, highlighting it as both a technical skill set and an evolving competence:

***2.3.1. Competence-listed approach***

Competence-listed models, like those by AFT, NCME & NEA (1990), define LAL through discrete competencies needed for effective assessment. These include skills like selecting assessment methods and recognizing ethical considerations, but they have been critiqued for focusing heavily on summative assessment. Brookhart (2011) later adapted the model to include formative assessment, recognizing the need for an adaptable, teaching-centered skill set in LAL.

***2.3.2. Three-component approach***

Three-component models, such as Davies’ (2008) Knowledge, Skills, and Principles framework, expand LAL beyond technical competencies to include ethical and contextual dimensions. Fulcher (2012) also proposed a hierarchical structure, suggesting that LAL integrates foundational skills, guiding principles, and an awareness of broader educational impacts. This model underscores that effective language assessment must consider not only technical skills but also ethical and contextual awareness.

***2.3.3. Scaled approach***

Scaled models conceptualize LAL as a continuum, with educators progressing through developmental stages from foundational to advanced proficiency. Pill and Harding’s (2013) “literacy ladder” offers a structured progression, categorizing skills across levels of literacy. Taylor’s (2013) framework adds dimensions for stakeholders with differing literacy requirements, emphasizing that different roles within education require different LAL competencies.

***2.3.4. Negotiation approach***

The negotiation approach, exemplified by Xu and Brown’s (2016) model, situates LAL as a dynamic, socio-cultural skill developed through interaction within teaching environments. This model positions LAL as a competence shaped by educators’ professional identities and reflects a continuum that balances technical skills with context-specific adaptation. Xu and Brown’s framework highlights the role of professional identity and contextual sensitivity in shaping effective language assessment practices.

**2.4. Previous studies on LAL development**

Research on LAL reveals a variety of methodologies, each reflecting unique educational needs. Internationally, studies often use scenario-based instruments to simulate classroom assessments, though these tools may lack the language-specific focus required for EFL contexts. In Vietnam, Tu (2023) developed the Language Assessment Literacy–Revised Vietnam (LAL-RV), adapted to meet the cultural and educational needs of Vietnamese EFL instructors. This adaptation highlights the importance of context-sensitive LAL instruments, particularly in multilingual, multicultural environments. Despite advances in LAL measurement, the limited validation of tools across varied contexts underscores the need for reliable, adaptable instruments tailored to specific educational settings.

This chapter examined LAL as a multifaceted competence that integrates technical skills, ethical considerations, and socio-cultural understanding. It explored core models and theoretical perspectives, emphasizing LAL as a dynamic and context-dependent skill set. The discussion laid the foundation for creating a responsive LAL framework tailored to Vietnamese EFL educators, addressing both local needs and professional standards. The next chapter will focus on developing a customized LAL scale for Vietnamese EFL educators, incorporating rigorous validation methods to ensure its relevance, reliability, and alignment with LAL's theoretical and practical demands.

**CHAPTER 3: SCALE DEVELOPMENT**

This chapter outlines the development of the Language Assessment Literacy Scale (LALS) designed to measure the language assessment literacy (LAL) of Vietnamese EFL lecturers. 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**

The LAL scale development followed a systematic approach, beginning with a construct definition and construct map. Drawing from Wilson (2011) and Wolfe & Smith (2007a), this construct map outlined LAL as a progression of skills, reflecting varying levels of assessment competence. Validity was examined through Rasch analysis, comparing hypothesized LAL progression with actual performance, as suggested by Woods & Griffin (2013). Additionally, input from subject matter experts was integrated to ensure comprehensive representation of the construct.

A developmental model was selected to illustrate the evolution of LAL competencies over time, with items arranged from foundational to advanced skills. The Partial Credit Model (Masters, 1982) was chosen for scoring to capture a broad range of proficiency with fewer items, allowing nuanced measurement of skill levels while maintaining interpretability. This model helped establish substantive validity by aligning theoretical predictions of competence with observed results.

To create measurable indicators, Griffin, Gillis, and Calvitto's (2004) guidelines on observable behaviors were employed. The framework delineated the LAL domain into capabilities, behaviors, and proficiency criteria, creating a criterion-referenced assessment rooted in Glaser’s (1981) work. Development of the LALS proceeded through four stages: drafting, focus groups, and two rounds of expert paneling, which refined the framework through iterative feedback.

***3.2.1. Drafting item pool***

The drafting phase began with a literature review to define the construct and identify specific LAL components. Sources included prominent LAL frameworks by scholars such as Brindley (2001), Davies (2008), and Fulcher (2012). This review informed a draft framework for Vietnamese teachers, with competencies organized hierarchically as suggested by Griffin et al. (2007).

Three dimensions of LAL emerged from this framework: preparing for assessment, conducting assessment, and revising assessment. Each dimension comprised 13 competencies, totaling 65 performance indicators derived from various LAL models. Quality criteria for each indicator were crafted based on principles of rubric design by Griffin, Nguyen, and Gillis (2004) and the Dreyfus and Dreyfus model of skill acquisition (1986), ensuring that each criterion represented distinct proficiency levels without overlap. This developmental scaling approach allowed teachers to self-assess their competence in a structured, progressive format.

***3.2.2. Expert paneling***

The paneling phase involved two rounds of expert review, critical to refining and validating the scale’s content. Experts in language assessment, educational measurement, and Vietnamese higher education reviewed items for relevance, clarity, and completeness.

In the first round, experts evaluated item relevance and emphasized the inclusion of digital assessment competencies, given the rise of online and formative assessment practices. Items were revised to incorporate digital literacy and ensure clarity, especially for lecturers in resource-limited areas. The second round focused on refining developmental progression within quality criteria, simplifying complex language, and ensuring that items accurately represented a continuum from foundational to advanced competencies. This iterative process resulted in a streamlined, logically ordered set of items suitable for developmental scoring.

***3.2.3. Trialing the scale***

The trial phase empirically tested the scale’s psychometric properties using a sample of 177 Vietnamese EFL lecturers from diverse regions. Stratified and cluster sampling strategies ensured representation across institutions, capturing a variety of educational settings.

Participants completed the LALS as a self-assessment, with items reflecting a developmental continuum of LAL proficiency. Classical Test Theory (CTT) and Rasch modeling analyzed reliability and construct validity. CTT showed high internal consistency, while Rasch’s Partial Credit Model validated the developmental scaling by aligning item difficulty with expected progression in lecturer competence. Fit statistics indicated unidimensionality, affirming that items collectively represented LAL competencies, and item-person maps confirmed an appropriate range of difficulty to distinguish lecturer abilities.

**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 Vietnamese EFL lecturers 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 Vietnamese EFL lecturers' 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 Vietnamese EFL lecturers.

**3.5. Trialing**

The trialing phase tested the LALS with 177 Vietnamese EFL lecturers 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 Vietnamese EFL lecturers.

Chapter 3 presents the LALS development as a structured, empirically validated approach to measure LAL for Vietnamese EFL lecturers. 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 IN SCALE VALIDATION**

This chapter details the validation approach for the Language Assessment Literacy Scale (LALS), focusing on an argument-based framework (Kane, 1992, 2013; Chapelle et al., 2008) supported by Rasch modeling. The validation emphasizes evidence-backed score interpretations rather than the test itself, proceeding in two stages. First, the Interpretive and Use Argument (IUA) is developed to outline intended score interpretations and uses. Second, a validity argument is constructed to evaluate empirical support for these interpretations. Core inferences, domain description, evaluation, generalization, and explanation, are used to build evidence supporting the LALS. This structure provides a rigorous, evidence-driven approach to LALS validation, setting a foundation for addressing further extrapolation and utilization inferences in future studies.

**4.1. Approaches to assessment validation**

***4.1.1. Contemporary validity perspectives***

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***

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 as the foundation for this study due to its structured, inference-driven process, which systematically links test scores to their interpretations and uses. It provides a clear framework for constructing an Interpretive Use Argument (IUA) and a Validity Argument, ensuring logical coherence and empirical support for each inference, such as domain description, evaluation, and generalization.

Kane’s adaptability to various assessment contexts makes it suitable for validating the LALS, which aims to assess the complex, multi-dimensional construct of LAL across diverse Vietnamese educational settings. Unlike traditional validity models, Kane’s approach emphasizes practical implications, aligning with the socio-cultural nature of LAL and ensuring test results are meaningful and actionable for stakeholders. This emphasis supports the LALS’s relevance to real-world teaching and assessment practices. By combining Kane’s framework with insights from Bachman, Palmer, and Chapelle, this study ensures a rigorous, context-sensitive validation process, establishing the LALS as a reliable tool for assessing LAL among Vietnamese EFL lecturers.

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

*4.2.2.1. Toulmin’s argument model*

Toulmin’s (1958) argument model forms the foundation of the argument-based validation framework, providing a structured approach for linking claims to evidence. It consists of six components: data, claim, warrants, backing, qualifiers, and rebuttals. This model emphasizes the logical connection between data and claims, supported by warrants and their backing, while addressing counterarguments through rebuttals. Toulmin’s framework has been widely influential in various fields, including educational measurement, where Kane (2013) adapted it for test validation, emphasizing practical and theoretical coherence in the interpretation and use of test scores.

*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 Interpretive Use Argument (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.

***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 requirements for inferences**

The argument-based approach specifies evidence types to substantiate each inference:

- Domain Description Evidence: Expert reviews and alignment analyses establish content relevance to LAL.

- Evaluation Evidence: Scoring consistency is demonstrated through inter-rater reliability and Rasch modeling, confirming LALS’s scoring accuracy.

- Generalization Evidence: Reliability studies, including Cronbach’s alpha, support score consistency across different contexts.

- Explanation Evidence: Construct validity is confirmed through factor analyses and Wright maps, verifying alignment between LALS scores and underlying LAL competencies.

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

The LALS validation prioritizes the domain description, evaluation, generalization, and explanation inferences, aligning empirical evidence with each to substantiate the LALS’s intended score interpretations. The IUA specifies assumptions, inferences, and supporting evidence, while the validity argument collects empirical data to confirm the IUA’s logic. For instance, the explanation inference involves evidence from factor analyses and Rasch fit statistics to verify that LALS scores reflect the intended LAL construct.

Chapter 4 describes the validation of the LALS using Kane’s argument-based framework, emphasizing the Interpretive and Use Argument 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 Vietnamese EFL lecturers 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 Vietnamese EFL lecturers’ LAL, focusing on knowledge, skills, and abilities in assessment.

Validating the LALS: Using Kane’s (2006) argument-based validation framework, the study validates four inferences—domain, evaluation, generalization, and explanation:

RQ1: Domain inference: How well do the LALS items represent essential LAL competencies for Vietnamese EFL lecturers?

RQ2: Evaluation inference: Do LALS components demonstrate strong psychometric properties?

RQ3: Generalization inference: How generalizable are LALS scores across demographics and experience levels?

RQ4: Explanation inference: How accurately do self-ratings on the LALS reflect LAL competency levels among Vietnamese EFL lecturers?

**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.

**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 Vietnamese EFL lecturers. 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 higher education.

**6.1. Initial validity evidence from scale development**

The LALS development for Vietnamese EFL lecturers followed four steps to ensure relevance to the teaching context. The competencies span three primary domains, each aligned with critical assessment stages: preparing, conducting, and revising assessments.

**\* Draft generation**: Using literature and contextual analysis (from Chapters 3 and 4), this phase outlined the three assessment stages:

**Preparing for assessment**: Knowledge of competence-based frameworks, assessment logistics, and task construction.

**Conducting assessment**: Task administration, scoring, feedback provision, and reporting.

**Revising assessment**: Continuous evaluation to refine assessment quality based on feedback and performance data.

These stages correspond to the Dreyfus model of skill acquisition (Dreyfus & Dreyfus, 1986), which categorizes competence progression from novice to expert. This structured approach ensures teachers have foundational to advanced assessment knowledge, relevant to LAL competencies in Vietnamese EFL classrooms.

**\* Competency and indicator development**: Sixty-five performance indicators were identified based on literature and contextual needs. Indicators in *Preparing for Assessment* include understanding assessment concepts, setting objectives aligned with university curricula, and selecting appropriate assessment methods. *Conducting Assessment* indicators cover practical execution skills like scoring consistency, providing effective feedback, and communicating assessment results. In *Revising Assessment*, teachers are expected to evaluate assessment task quality based on practical classroom realities, as outlined by scholars like Davies (2008) and Fulcher (2012).

**\* Quality criteria development**: The Dreyfus model also informed quality criteria, which are organized by increasing proficiency from basic recall to application in practical teaching contexts. For example, teachers progress from defining fundamental terms to evaluating and implementing language assessment frameworks in a university context. This approach supports a multidimensional, structured assessment for teachers’ LAL.

**\* Focus group findings**: Analysis of focus group data revealed several themes:

*+ Framework revisions*: Experts recommended simplifying indicators for clarity and reordering them to better align with practical assessment stages.

*+ Cultural relevance*: The framework integrated Vietnamese educational standards, enhancing its applicability within local contexts.

*+ Competency structure*: The three-stage structure ensures a comprehensive representation of core LAL processes for EFL lecturers.

*+ Practicality of indicators*: Performance indicators were refined to reflect classroom realities, including additions for diagnostic, formative, and summative assessment types.

These focus group insights confirmed the LALS’s relevance and facilitated revisions for more efficient usage among Vietnamese EFL lecturers.

**6.2. Expert paneling for validation**

Expert reviews were conducted in two rounds, focusing on item clarity, relevance, and representativeness.

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

This round provided insights into several validation aspects, such as comprehensibility, representativeness, structure, external alignment, applicability, and generalizability.

**Comprehensibility**: Experts found most descriptors clear, though they recommended additional practical examples to support item comprehension.

**Relevance and representativeness**: Experts validated that the competency groups adequately covered essential LAL components, though they emphasized digital literacy and formative assessment practices.

**Structural aspects**: Suggested refinements included clearer connections between indicators and streamlined language to avoid redundancy.

**External structure**: Experts recommended aligning terminology with international standards and adapting content to reflect Vietnamese-specific requirements.

**Applicability**: Suggestions to clarify vague terms and ensure observable performance indicators increased practical utility for lecturers.

**Generalizability**: Experts highlighted the need for differentiated competencies to accommodate lecturers with varied qualifications and institutional contexts, emphasizing adaptability for both theoretical and practical knowledge.

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

Further refinement focused on clarity, usability, and overall readiness for field testing.

**Clarity and specificity**: Revisions clarified indicators, especially terms related to language competence components, improving practical applicability.

**Logical flow and usability**: Competencies were reordered for a logical sequence, beginning with foundational skills, which facilitated intuitive navigation.

**Depth and representativeness**: Additional sub-indicators enhanced depth, while experts cautioned against overly complex structures that might reduce practicality.

**Terminological consistency**: Standardized terminology, aligning with international frameworks, improved cross-context clarity and application.

**Readiness for field testing**: The instrument was deemed ready for broader implementation, with minor clarifications suggested to further enhance real-world applicability.

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

The validation of domain inference rested on two key assumptions, each of which was supported by evidence from expert feedback and empirical data.

**Representation of skills and knowledge**: Evidence from literature reviews and expert input confirmed that the LALS effectively represents the essential LAL competencies required for Vietnamese EFL lecturers. Assumptions about necessary LAL knowledge and skills were upheld, confirming the scale’s foundation in empirical research.

**Items as indicators of LAL dimensions**: Iterative expert reviews and focus group discussions refined the items to ensure they accurately reflect each LAL domain. This process demonstrated that the scale’s items effectively capture the multidimensional aspects of LAL in a contextually relevant manner.

These findings validate the domain description inference, confirming that the LALS serves as a robust, culturally adapted tool for measuring language assessment literacy among Vietnamese EFL lecturers.

This chapter presented the validation process for the domain description inference in the LALS, establishing a strong foundation for assessing LAL among Vietnamese EFL lecturers. Expert feedback and focus group discussions enabled iterative refinements, enhancing the scale’s comprehensibility, representativeness, and cultural relevance. The domain definition inference is fully supported, affirming the LALS’s alignment with the specific educational needs of Vietnamese EFL lecturers. Further validation through field testing will provide additional insights into the scale’s reliability and applicability, with subsequent chapters exploring the evaluation and generalization inferences.

**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 Language Assessment Literacy Scale (LALS), aiming to assess Vietnamese EFL lecturers' language assessment competencies.

**7.1. Evaluation inference**

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

The LALS was developed through an iterative, expert-informed process to ensure it aligns with the targeted competencies. Major steps in scale development included:

**Theoretical foundations**: The LALS was grounded in extensive literature, synthesizing essential competencies for language assessment from established theoretical models.

**Expert review and refinement:** Focus groups provided feedback on alignment with Vietnamese EFL needs, ensuring relevance, clarity, and the representativeness of items.

**Consensus documentation:** The item modifications, justifications, and final 42-item scale were based on expert recommendations, as displayed in Table 6.5.

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

Psychometric evaluation involved a series of analyses, assessing 177 valid responses via Classical Test Theory (CTT) and Rasch models. Thirty-six responses were excluded due to non-eligibility.

Thurstonian Thresholds: Items exhibited appropriate threshold estimates (ranging from -1.71 to +3.15 logits), confirming suitability for measuring varying levels of competence.

Item-rest correlation: Most items showed strong discrimination, with indices above 0.4, indicating effective differentiation among respondent abilities.

Item fit and reliability: INFIT mean square values ranged between 0.77 and 1.25, suggesting items’ consistency with Rasch model expectations.

The item-rest correlations ranged from 0.29 to 0.62, indicating adequate discrimination. Items with lower correlations (e.g., Q214 at 0.29) were noted for potential revision, though most items contributed meaningfully to the scale’s construct.

Rasch analysis yielded weighted MNSQ values between 0.6 and 1.4, with the mean INFIT at 1.02, validating most items’ alignment with the Rasch model. A few items, such as Q213 and Q215, showed slight misfit but within acceptable limits.

**7.2. Validity results of generalization inference**

This inference examines whether LALS scores are consistent across regions, genders, and experience levels. High reliability and item separation indices substantiate LALS’s robustness across samples.

***7.2.1. Validity evidence 1: Weighted mean square (MNSQ) fit statistics across aamples***

**Regions:** The majority of items showed MNSQ values within the accepted range. Slight misfits in the Centre and South regions indicated minor regional variability.

Genders: Items performed well for both genders, though Item 27 showed higher misfit in the male sample.

**Experience levels:** Items were consistent across experience groups, though items such as Q215 were noted for review in specific experience groups (6-10 years).

**Item Weighted Fit by Region, Gender, and Experience Groups for LALS:** Items 4, 5, and 27 exhibited some misfit across certain groups due to perceived irrelevance or logistical rather than competence-based tasks.

***7.2.2. Validity evidence 2: Reliability and item separation analysis (After scale adjustment)***

**Item separation reliability:** High coefficient (r = 0.892) confirms item spread across difficulty levels.

**EAP/PV reliability:** An index of r = 0.914 indicates stable competence measurement across samples.

**Cronbach’s Alpha:** A high alpha of 0.907 further supports item coherence.

***7.2.3. Overall assessment of the generalization inference***

All items generally supported the generalization inference across groups, with only minor adjustments required.

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 addresses the explanation inference, examining whether Vietnamese EFL lecturers’ observed scores align with the theoretical constructs of the LALS. The inference analysis is guided by two assumptions: first, that scale items accurately capture the intended components, verified through psychometric analyses, and second, that the scale differentiates competence levels, assessed through Rasch analysis and Wright maps.

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

The **variable map** illustrates the alignment of teacher abilities with item difficulty parameters, assessing whether item difficulties match teacher competencies. The map displays:

* **Logit scale representation**: Both teacher abilities and item difficulties are mapped on a logit scale, with competencies spanning from -2 to +3 logits for teachers and -1.74 to +3.20 logits for items. The close alignment of the item and person means (item mean = 0; person mean = 0.21) indicates that the items are well-suited to the teachers’ abilities.
* **Behavioral hierarchy**: Lower map items represent novice competencies, while higher items reflect advanced proficiency, as seen in the Wright map for the LALS. The map supports the probabilistic nature of competencies, where higher-level performers likely exhibit foundational competencies, with item fits confirmed by the partial credit model.

The hierarchical, developmental nature of the map aligns with the LALS’s theoretical foundation, outlining teachers’ progress from novice to proficient competencies, with detailed proficiency descriptions for each level provided in Tables 8.2 to 8.5.

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

Comparing **empirical and intended continua** supports interpretability by confirming alignment with theoretical levels (novice to proficient). Adjustments made include:

* **Revised competence placement**: Items such as 33.1 (stakeholder engagement) were elevated to emphasize advanced competencies, while foundational items like 3.2 and 25.2 were realigned to lower levels, reflecting foundational relevance.
* **Advanced proficiency emphasis**: Items 23.2 (consistency in performance level descriptors), 11.2 (identifying improvement areas), and 35.2 (improving reliability through scorer training) were raised to Level 4, highlighting their critical role in high-level assessment practices.
* **Technology integration challenges**: Item 19.2 (digital methods in assessments) was positioned as the highest challenge, acknowledging ongoing technological adaptation hurdles in assessment practices.

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

As shown in Table 8.8, 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 Vietnamese EFL lecturers’ 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 final chapter provides a summary of the development and validation process of the LALS for Vietnamese EFL teachers, synthesizing key findings from earlier chapters to clarify the study’s outcomes. It revisits decisions about the scale's components and highlights how supporting evidence evolved after the scale was trialed, marking the completion of the instrument development process. The chapter also evaluates the study's strengths and limitations, offering practical recommendations to improve teacher training and support while suggesting directions for future research.

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

***9.1.1. Final scale***

The LALS for Vietnamese EFL teachers was developed and validated to meet Vietnam's specific educational needs. The scale assesses teacher competencies across three stages: preparing, conducting, and revising assessment, covering essential skills from understanding core concepts to improving assessment tasks. Based on Dreyfus & Dreyfus’s (1986) skill acquisition model, the scale’s competence-based approach tracks teachers’ development from novice to expert, providing observable behaviors at each stage. The scale includes 39 items, each with descriptions across four levels of proficiency, offering a comprehensive framework for teachers to self-assess their Language Assessment Literacy (LAL). This framework is illustrated in Figure 9.1 and detailed in Table 9.1.

***9.1.2. Distinctiveness of the Vietnamese LAL scale***

The Vietnamese LALS is uniquely designed for Vietnam’s educational context, integrating cultural relevance, policy alignment, and resource sensitivity. It aligns with MOET, KNLNNVN, and NFLP guidelines, ensuring coherence with national goals.

The scale addresses local constraints like large class sizes and limited technology, promoting practical assessments and ethical practices tailored to Vietnam. Its developmental model supports EFL lecturers in refining competencies and fostering intercultural competence.

By balancing modernization with feasibility and ensuring quality through psychometric validation, the LALS provides a localized, practical framework for Vietnamese educators and resource-limited contexts.

**9.2. Evidence for validity of the scale: A summary**

***9.2.1. Domain description inference***

The LALS items accurately represent essential LAL competencies needed by Vietnamese EFL lecturers, confirmed by expert feedback, content relevance, and alignment with established LAL theories. Iterative expert consultations refined items to ensure practical relevance for Vietnam’s context, addressing specific needs in language assessment literacy.

***9.2.2. Evaluation inference***

The LALS demonstrates robust psychometric properties, confirming reliable measurement. Rasch analysis validated item difficulty levels, discrimination indices, and internal consistency. Fit statistics and item-rest discrimination estimates affirmed the scale’s capacity to measure LAL consistently, with high item separation reliability (r=0.92), moderate person separation reliability (r=0.68), and an EAP/PV reliability index of 0.70.

***9.2.3. Generalization inference***

LALS scores are generalizable across Vietnamese EFL lecturers with different demographic backgrounds and experience levels. Consistency in item performance across these groups supports the scale's adaptability. Although regional differences were not fully represented due to sampling challenges, fit statistics and item separation reliability indicate the scale’s effectiveness across subgroups.

***9.2.4. Explanation inference***

The LALS effectively reflects the theoretical construct of LAL, reliably differentiating proficiency levels. Rasch modeling and Wright maps confirm strong alignment between teacher abilities and item difficulties. Psychometric properties validate the scale’s ability to capture teacher competence progression, supporting construct validity in differentiating teacher proficiency levels.

**9.3. Implications**

The findings of this study have critical implications for theory, methodology, and pedagogy in language assessment within the Vietnamese EFL context.

***9.3.1 Theoretical implications***

This study’s competence-based approach to LAL defines it as a dynamic, evolving skill set rather than a static construct. Localizing LAL for the Vietnamese context reveals unique competencies, such as adapting assessment practices to align with local policies. The findings underscore the importance of continuous professional development in LAL, incorporating both traditional assessment principles and digital innovations.

***9.3.2. Methodological mplications***

This study contributes to developmental assessment methods and expertise development. Applying the Dreyfus model of skill acquisition, it outlines progressive LAL competencies, from novice to expert, aligning with real-world applications in Vietnamese EFL education. The integration of Rasch modeling provides empirical insights into scale validity, reinforcing its potential application across other domains in education.

***9.3.3 Pedagogical implications***

*9.3.3.1. Supporting self-assessment and reflective practice*

The LALS enables Vietnamese EFL lecturers to self-assess their competencies, promoting reflective practice that enhances assessment transparency and student-centered evaluations.

*9.3.3.2. Guiding professional development*

The LALS framework can be integrated into teacher development programs, offering a roadmap for skills progression. The scale’s clear criteria enable targeted improvement strategies aligned with evolving language assessment demands.

*9.3.3.3. Enhancing teacher training programs*

The LALS has potential for pre-service and in-service training, supporting personalized mentorship, resource allocation, and continuous growth across varying competency levels.

**9.4. Limitations of the study**

The study’s geographical sampling limits its generalizability across Vietnam. Most participants were from northern Vietnam due to data collection barriers, which could restrict findings' applicability to other regions. The use of self-report measures, while valid and reliable, may not fully capture teachers' LAL in practice, warranting future research to incorporate observational and external evaluations for a more comprehensive view.

**9.5. Areas for continued research**

Further studies should explore LAL’s application in diverse educational settings and examine the influence of contextual factors, such as teacher background and regional policies. Continued validation through Rasch modeling and the inclusion of additional data sources will strengthen scale reliability, with opportunities to extend this framework to other subject areas.

**9.6. Guidelines for using the LALS for Vietnamese EFL lecturers**

The LALS offers Vietnamese EFL lecturers a developmental tool for assessing and advancing their language assessment skills, with Level 2 (Emergent) as the minimum proficiency target across assessment stages. Institutions are encouraged to integrate the scale into pre-service and in-service programs, fostering regular self-assessment and professional growth aligned with evolving educational standard.

|  |  |  |
| --- | --- | --- |
|  | **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. 235-258). 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. |  |