

PROBLEMS ENCOUNTERED BY FOURTH-YEAR ENGLISH-TEACHING MAJORS AT HANOI OPEN UNIVERSITY IN DESIGNING LISTENING COMPREHENSION TESTS

Le Phuong Thao¹, Nguyen Thi Quynh Trang¹, Do Thi Trang¹, Do Quynh Trang¹
Email: lephuongthao@hou.edu.vn, ORCID: 0009-0002-8860-7159

Received: 12/09/2025. Revised: 03/03/2026. Accepted: 19/03/2026
DOI: 10.59266/houjs.2026.1194

Abstract: *This article investigates the difficulties that fourth-year English-Teaching majors at Hanoi Open University face in designing listening comprehension tests and examines the factors that cause them. Based on the CIPP evaluation model, the study focuses on three main factors affecting pre-service teachers' listening test design competence: context, input, and process. Using a mixed-method approach, data were collected from 105 fourth-year English-Teaching majors and analyzed in detail. The findings reveal that despite students' general understanding of listening assessment principles, they still face considerable challenges in selecting appropriate materials, aligning test tasks with learners' levels, targeting listening sub-skills, and ensuring test validity and reliability. The results show that limited access to suitable materials and a shortage of practice opportunities make students struggle with test design. Accordingly, the study proposes increased guided practice, improved access to listening resources, and more systematic feedback to improve the listening test design competence of the fourth-year English-Teaching majors at Hanoi Open University.*

Keywords: *evaluation model, listening test design, fourth-year English-Teaching majors*

I. Introduction

Listening comprehension is widely recognized as a core skill in English language teaching; however, its assessment remains underexplored compared to other language skills. Effective listening assessment is expected to capture learners' ability to interpret spoken language in context and reflect real-world communicative demands rather than surface-level comprehension (Bizunesh et al., 2021). Consequently, listening assessment plays a crucial pedagogical role in informing teaching practices, curriculum

decisions, and learners' development (Ockey & Wagner, 2018).

For pre-service English teachers, the ability to design listening comprehension tests constitutes an essential professional competence. In many teacher education programs in Vietnam, including those for fourth-year English-Teaching majors, students are required to complete coursework on language assessment and to design classroom-based tests as part of their training. These tasks are intended to prepare future teachers to assess learners' listening competence in authentic classroom contexts.

¹ Hanoi Open University, Hanoi, Vietnam

However, evidence from the Hanoi Open University suggests that possessing theoretical knowledge of assessment does not necessarily translate into effective test-design practices. Fourth-year English-Teaching majors frequently encounter difficulties in selecting appropriate listening input, aligning tasks with learners' proficiency levels, and ensuring fundamental test qualities such as validity and reliability. These challenges point to a persistent gap between assessment theory and practical application, which may undermine the quality of student-designed listening tests. In response to this issue, the present study investigates the common difficulties fourth-year English-Teaching majors at Hanoi Open University face in designing listening comprehension tests and proposes pedagogically relevant recommendations to improve assessment practice in teacher education.

Research questions

To achieve the objectives of the study, a survey was administered to 105 fourth-year English-Teaching majors at Hanoi Open University to answer the following research questions:

1. What common difficulties do fourth-year English-Teaching majors encounter when designing listening comprehension tests?
2. What factors are associated with the difficulties in terms of context, input, and process?

II. Literature review

2.1. Previous studies

2.1.1. Studies on language test design principles

Previous research has consistently emphasized that language test development should be grounded in clear principles, particularly when undertaken by student

teachers who typically have limited experience in assessment. Pandey (2025) argues that test design is better understood as a continuous and interconnected process, moving from planning and task construction through to administration and evaluation, rather than as a series of separate technical stages. The study further points out that test tasks are often not sufficiently aligned with learning objectives, which may result in assessments that fail to capture learners' true language ability.

Comparable findings are reported by Sitepu et al. (2025), who observe that tests designed by pre-service teachers frequently exhibit problems in item construction, a lack of authenticity, and limited practicality, suggesting that many student teachers still struggle to apply assessment principles in actual test design.

2.1.2. Studies on language assessment literacy of pre-service teachers

Interest in language assessment literacy has increased in recent years, but research on pre-service teachers remains limited (Moraes et al., 2024). A common problem reported in the literature is that many student teachers understand assessment concepts at a basic level but struggle to apply this knowledge in real classroom tasks. Lam (2015) indicates that although assessment is taught in teacher education programs, student teachers often feel unsure when they are required to design tests or make assessment-related decisions by themselves.

This issue is partly related to how assessment is handled in teacher education programs. Oo, Alonzo, and Asih (2022) point out that assessment is often given limited attention in the curriculum, which may help explain why many student teachers lack confidence in this area.

Moraes et al. (2024) also observe that these difficulties become more noticeable when pre-service teachers move from learning about assessment to actually completing practical tasks such as constructing tests.

2.1.3. Studies on difficulties in designing language tests

Previous research has consistently reported difficulties in designing language tests across different educational contexts. Studies on general test construction show that teachers often struggle to develop assessments that adequately reflect intended language skills and assessment principles (Tewachew et al., 2024). In their listening assessment study, Nurhayati et al. (2024) found that EFL teachers experienced challenges in selecting appropriate audio materials, ensuring validity, and managing time during test development. Similar difficulties have also been identified among pre-service teachers. Tagle et al. (2024) reported that pre-service EFL teachers tended to design listening items that primarily targeted lower-level comprehension skills, a tendency associated with limited assessment experience and cognitive load.

Although previous studies have examined difficulties in language test design, most have been conducted outside Vietnam and tend to focus on individual factors. Therefore, limited attention has been paid to how contextual and instructional factors influence pre-service teachers' listening test design. To address this gap, the present study investigates the perceived difficulties of fourth-year English-Teaching majors at Hanoi Open University using the CIPP framework.

2.2. Theoretical background

2.2.1. Listening comprehension

Listening comprehension is no longer viewed as a passive process of

receiving sounds but as an active and complex construction of meaning. Vandergrift and Goh (2022) conceptualize listening as a strategic process in which learners actively deploy cognitive and metacognitive strategies, such as inferencing, monitoring, and evaluating understanding in real time. Similarly, Rost (2011) defines listening comprehension as the ability to integrate linguistic input with contextual and pragmatic knowledge to derive meaning from spoken discourse.

Although general principles of language test design apply to all skills (Brown & Abeywickrama, 2010), the design of listening tests involves skill-specific considerations. For instance, if input for reading is available and learners are allowed to process the text at their own pace (Alderson, 2000), listening requires learners to process transient spoken input in real time (Buck, 2001). In addition, writing and speaking are productive skills measured by language output (Weigle, 2002; Luoma, 2004), whereas listening is an input skill in which a person's ability to understand spoken language is usually assessed (Ockey & Wagner, 2018). These variations necessitate that test designers focus on audio input, task design, and test conditions when producing listening tests.

2.2.2. Principles for designing listening comprehension tests

The present study adopts validity and reliability as key principles in evaluating student-designed listening tests (Brown & Abeywickrama, 2010). Validity is the extent to which test tasks adequately represent the intended listening construct, while reliability is the consistency of test scores across administrations. In listening assessment, both are influenced by task design and audio characteristics (Ockey & Wagner, 2018).

2.2.3. *Language Assessment Literacy (LAL)*

Language Assessment Literacy (LAL) refers to teachers' knowledge and skills in designing, implementing, and interpreting language assessments (Davies, 2008). In listening assessment, limited LAL may constrain teachers' ability to define and measure listening constructs consistently, given the complex and less observable nature of listening skills (Buck, 2001). Language Assessment Literacy (LAL) in test design involves knowledge of assessment principles, skills in constructing test tasks, and awareness of appropriate assessment use in specific contexts (Fulcher, 2012).

2.2.4. *The CIPP evaluation model as an analytical framework*

The CIPP evaluation model, proposed by Stufflebeam (2003), conceptualizes educational evaluation in terms of four interrelated dimensions: context, input, process, and product. Rather than focusing solely on outcomes, the model highlights the systematic examination of conditions, resources, and implementation processes that shape educational practices.

In the present study, the CIPP model is employed as an analytical framework to examine the difficulties pre-service English teachers encounter in designing listening comprehension tests. The analysis focuses on context-related factors (e.g., course requirements and institutional expectations), input-related factors (e.g., instructional resources and materials), and process-related factors (e.g., learning activities and feedback practices), which closely align with the research questions and the scope of the collected data.

III. Research methods

3.1. *Design of the study*

A mixed-method approach was employed in the study to investigate the difficulties that fourth-year English-

Teaching majors face in designing listening comprehension tests. This approach focused on examining a real, specific group within an educational context to systematically assess students' perceived challenges, verify the evidence, and propose solutions for further improvement. Data were collected through a questionnaire containing closed-ended items to obtain quantitative data on students' perceptions of challenges related to context, input, and process factors.

3.2. *Pedagogical setting and participants*

The research was conducted in the course Testing and Assessment in English Language Teaching, held at the Faculty of English, Hanoi Open University. The course runs for one semester and covers in-depth theoretical and practical aspects of language assessment. Students learn about a few major criteria underlying a good test, such as reliability, validity, authenticity, and the different approaches to language testing. Besides presenting general principles, the course also covers modules on language testing skills, including listening, speaking, reading, and writing. The part of the module on listening assessment focuses on the types of listening, micro- and macro-listening skills, and the preparation of listening assessment tasks. For listening, as part of the requirements, students are required to design classroom-based language tests, including listening comprehension tests using common formats such as multiple-choice and gap-filling.

105 students from the Faculty of English at Hanoi Open University participated in this study. They included fourth-year English-Teaching majors from class K29, who were in their final year and had completed an English-teaching and assessment course. They were required to have a foundational understanding of test

design and assessment techniques as pre-service instructors, especially in listening comprehension. Their involvement gave important insights into the challenges they face when creating listening assessments.

3.3. Data collection

The questionnaire was designed based on the CIPP evaluation model, with a primary focus on context-, input-, and process-related factors influencing students' ability to design listening comprehension tests. In addition, a separate section was included to identify specific difficulty areas encountered by students during the test design process. Another section was added to explore students' perceptions of potential pedagogical solutions to enhance their competence in designing listening tests.

3.4. Data analysis

The data were collected through a questionnaire administered to fourth-year English-Teaching majors at Hanoi Open University. The responses were analyzed using descriptive statistics,

including frequencies and percentages, in Microsoft Excel. Likert-scale items were summarized to examine overall trends in students' perceptions of context, input, and process-related factors affecting listening test design. The questionnaire items were grouped according to the CIPP framework to identify prominent patterns and areas of difficulty. The results were then presented and interpreted to inform the discussion and pedagogical implications of the study.

IV. Research results and discussion

4.1. Research results

The objective of this study is to examine the current situation of listening test design competence among fourth-year English-Teaching majors; analyze the main difficulties as well as the context-related, input-related, and process-related factors affecting the effectiveness of listening test design; and propose solutions and recommendations to improve the effectiveness of listening test design competence.

4.1.1. Core difficulties in listening test design

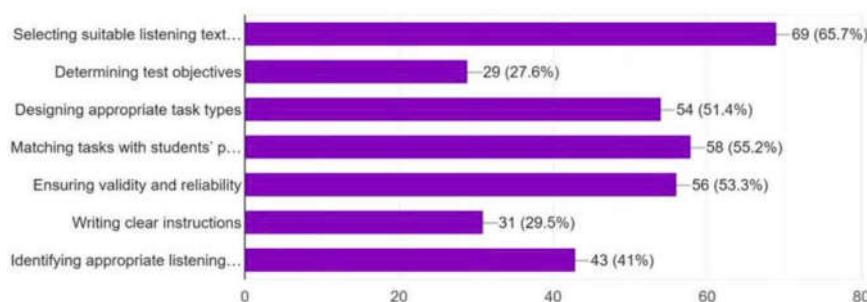


Figure 1. Core difficulties in designing listening comprehension tests among fourth-year English-Teaching majors ($N = 105$).

According to the survey, fourth-year English-teaching majors encountered several fundamental challenges when creating listening comprehension assessments. Selecting audio texts that align with test objectives and learners' proficiency levels was reported as difficult by 65.7% of respondents. Maintaining

test validity and reliability (53.3%) and aligning tasks with students' core listening skills (55.2%) were also commonly reported difficulties, both of which require the practical use of assessment expertise. Moreover, more than half of the participants (51.4%) expressed uncertainty when designing task types targeting specific

listening sub-skills. Difficulties related to defining test objectives and writing clear instructions were reported less frequently; only 27.6% of participants said they were unsure how to define the objectives, and 29.5% said the directions were unclear.

4.1.2. Context-related factors affecting listening test design

The results show that the listening comprehension test design was challenging for fourth-year English-Teaching majors, with almost 70.5% of respondents agreeing or strongly agreeing that the requirements exceeded their current level. Despite these

difficulties, a majority of respondents (73.3%) reported understanding the objectives of the listening test design. Nevertheless, this comprehension was not consistent across particular competencies. A notable proportion (33.4%) voiced hesitation or disagreement, although somewhat more than half of the respondents (55.2%) said they could distinguish between listening for the gist and listening for specific information. In a similar vein, only 50.5% of participants expressed confidence in applying fundamental criteria of listening test design, including validity and reliability.

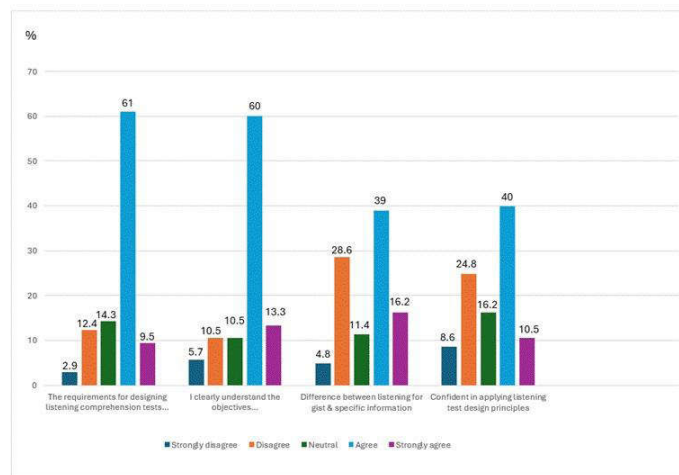


Figure 2. Distribution of students' responses to context-related factors in designing listening comprehension tests ($N = 105$).

4.1.3. Input-related factors affecting listening test design

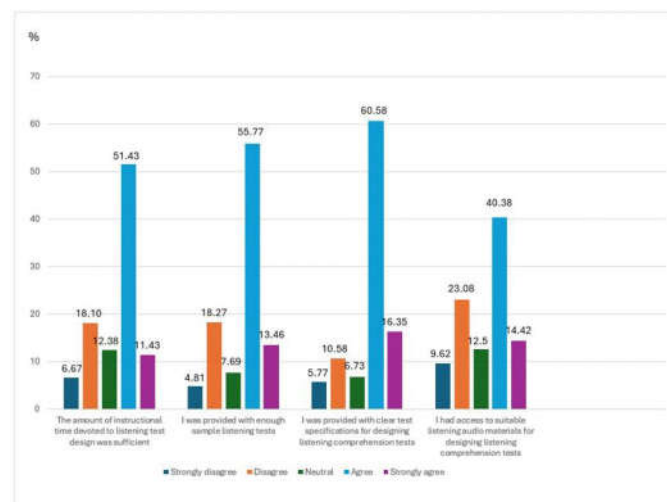


Figure 3. Distribution of students' responses to input-related factors in designing listening comprehension tests ($N = 105$).

The finding indicates that although certain instructional supports were provided, students still perceived several limitations. While the majority of students (62.86%) agreed or strongly agreed that they perceived sufficient instructional input devoted to the listening test, more than 24.77% responded in the opposite direction. Similarly, more than half of the respondents (69.23%) agreed or strongly agreed that they were provided with enough

sample listening tests. Additionally, over three-quarters of the students (76.93%) reported that clear test specifications for designing listening comprehension tests were provided, suggesting strong institutional support in this regard. However, access to suitable listening materials remained a noticeable concern, as more than 32% of learners expressed disagreement or strong disagreement about having adequate access.

4.1.4. Process-related factors affecting listening test design

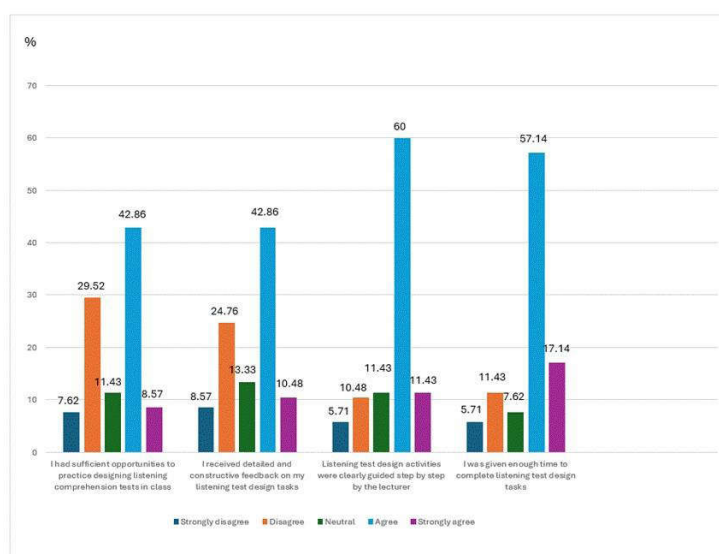


Figure 4. Distribution of students' responses to process-related factors in designing listening comprehension tests ($N = 105$).

The finding reveals several challenges related to classroom processes. Figure 4 shows that a considerable proportion (37.14%) of students disagreed or strongly disagreed that they had sufficient opportunities to practice designing listening comprehension tests in class, suggesting limited hands-on practice. Similarly, a significant number of respondents (33.33%) reported receiving insufficient detailed and constructive feedback on their listening test design tasks. Although a majority of students (71.43%) strongly agreed or agreed that the listening test design activities were

clearly guided, step by step, by the lecturer, time pressure remained a significant concern, with almost 24.76% of students expressing disagreement and uncertainty about this issue

4.2. Discussion

The results demonstrate that although fourth-year English-Teaching majors at Hanoi Open University have a fundamental understanding of listening test design principles, they face significant challenges when putting this knowledge into practice. Contextual, input-, and process-related factors all contribute to these difficulties, particularly

the cognitive demands involved in transferring theoretical knowledge into practice. In addition, the instructional guidelines and sample materials provided appear insufficient to support effective audio selection and task customization to learners' proficiency levels. The lack of practical experience and feedback further hampers students' ability to translate assessment knowledge into effective listening test design.

The findings are generally consistent with previous studies on test design and language assessment literacy. In line with Lam (2015) and Moraes et al. (2024), students in the current study demonstrated a significant gap between their theoretical understanding of assessment principles and their ability to apply them in practice. Similarly, Tagle et al. (2024) and Nurhayati et al. (2024) reported difficulties in determining test objectives, selecting appropriate listening materials, and ensuring validity and reliability. This similarity may be explained by the limited practical training and assessment experience in teacher education programs. However, unlike earlier research that mainly focused on individual competence, this study highlights the influence of contextual, input-, and process-related factors. Therefore, the CIPP framework provides a more comprehensive explanation by identifying systemic causes rather than attributing the problem solely to learner limitations.

Based on these findings, teacher education programs should strengthen practical support to improve students' listening test design skills. More opportunities for guided practice, feedback, and revision should be incorporated into training courses, along with clearer instruction on listening subskills and test design principles. In addition, enriching listening test samples,

providing appropriate audio resources, and creating practice tasks that reflect real teaching contexts may help bridge the gap between theory and practice.

V. Conclusion

In summary, this study examines the challenges fourth-year English-Teaching majors at Hanoi Open University face in designing listening comprehension tests, the factors that contribute to these difficulties, and possible pedagogical solutions. The findings first reveal a significant discrepancy between theoretical knowledge and practical implementation. Although most students reported gaining a foundational understanding of assessment objectives and principles, they still found it challenging to construct an assessment. The core difficulties included selecting appropriate listening materials, aligning tasks with learners' proficiency levels, designing tasks focusing on specific listening sub-skills, and ensuring test validity and reliability. Using the CIPP framework, the study also concludes that while the context (curriculum goals) is clear, the input (access to diverse audio resources) and process (opportunities for practice and feedback) require improvement to better support pre-service teachers. In addition, despite the insights gained, this study has certain limitations. Firstly, because the primary reliance on questionnaire data reflects students' perceptions, the findings lack a full capture of actual test-design difficulties in performance. To address this limitation, future studies should embrace mixed-method approaches, such as combining questionnaires with interviews or analyzing student-designed listening tests, thereby allowing researchers to compare perceived difficulties with actual performance. Secondly, the participants were limited to fourth-year English-teaching majors at

Hanoi Open University, which restricts the applicability of the findings to all pre-service teachers in Vietnam. Therefore, future studies should expand to a larger, more diverse group of attendees to help provide more comprehensive perspectives of pre-service teachers' listening assessment competence. Finally, the study did not examine the actual student-designed listening tests, as it focused solely on perceived difficulties rather than on assessment outcomes. This constraint can be addressed by evaluating the quality of student-designed listening tests, using clear, criterion-referenced rubrics. The use of criterion-referenced rubrics aligned with validity and reliability principles would provide more concrete evidence of students' strengths and weaknesses in the design of listening tests.

References

- Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.
- Buck, G. (2001). *Assessing listening*. Cambridge University Press.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). Pearson.
- Bizunesh, W. L., Alemu, A., & Ashenafi, B. (2021). *EFL teachers' perception and practice of assessing listening skills, and the challenges they face: Secondary schools in West Jimma in focus* (Master's thesis). Jimma University.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347. <https://doi.org/10.1177/0265532208090156>.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. <https://doi.org/10.1080/15434303.2011.642041>.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197. <https://doi.org/10.1177/0265532214554321>.
- Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.
- Moraes, I. T., Furtoso, V. B., & Giraldo, F. (2024). Developing pre-service teachers' language assessment literacy: An analysis of classroom tasks. *Ilha do Desterro*, 77, 96898. <https://doi.org/10.5007/2175-8026.2024.e96898>.
- Nurhayati, N., Setiawaty, P. W., & Nur, S. (2024). EFL teachers' challenges in designing assessment material for students' listening skills. *English Franca: Academic Journal of English Language and Education*, 8(2), 409-422.
- Oo, C. Z., Alonzo, D., & Asih, R. (2022). Acquisition of teacher assessment literacy by pre-service teachers: A review of practices and program designs. *Issues in Educational Research*, 32(1), 352-373. <http://www.iier.org.au/iier32/oo.pdf>.
- Ockey, G. J., & Wagner, E. (2018). *Assessing L2 listening*. Cambridge University Press.
- Pandey, G. P. (2025). Designing effective language tests: A framework for integrating stages, processes, and best practices. *Jo-ELT (Journal of English Language Teaching)*, 12(1), 62-74. <https://doi.org/10.33394/jo-elt.v12i1.15274>.
- Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Pearson Education.
- Sitepu, D. N., Situmeang, S. L., Sembiring, R. V., Saragih, D. C., & Silalahi, D. E. (2025). Effectiveness of language assessment tests: An observation based on reliability, validity, authenticity, practicality, and washback principles. *Innovative: Journal of Social Science Research*, 5(4), 4711-4722. <https://doi.org/10.31004/innovative.v5i4.20752>.
- Stufflebeam, D. L. (2003). The CIPP model for evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 31-62). Kluwer Academic Publishers.

Tagle, T., Etchegaray, P., Díaz, C., Ortiz, M., Ramos, L., & Quintana, M. (2024). Learning to assess foreign language listening: Analyzing the assessment items designed by Chilean pre-service EFL teachers and the cognitive load they experience. *Edelweiss Applied Science and Technology*, 8(6), 9328-9344.

Tewachew, A., Shiferie, K., & Tefera, E. (2024). Practices of EFL teachers in test construction. *Cogent Education*, 11(1), 2412496. <https://doi.org/10.1080/2331186X.2024.2412496>.

Vandergrift, L., & Goh, C. C. M. (2022). *Teaching and learning second language listening: Metacognition in action* (2nd ed.). Routledge.

Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.

KHÓ KHĂN CỦA SINH VIÊN NĂM THỨ TƯ NGÀNH SƯ PHẠM TIẾNG ANH TẠI TRƯỜNG ĐẠI HỌC MỞ HÀ NỘI TRONG QUÁ TRÌNH THIẾT KẾ BÀI KIỂM TRA NGHE HIỂU

Lê Phương Thảo¹, Nguyễn Thị Quỳnh Trang¹, Đỗ Thị Trang¹, Đỗ Quỳnh Trang¹

Tóm tắt: Bài báo này nghiên cứu những khó khăn mà sinh viên năm thứ tư chuyên ngành Sư phạm tiếng Anh Trường Đại học Mở Hà Nội gặp phải trong việc thiết kế bài kiểm tra nghe hiểu và xem xét các yếu tố gây ra những khó khăn này. Dựa trên mô hình đánh giá CIPP, nghiên cứu tập trung vào 3 yếu tố chính ảnh hưởng đến năng lực thiết kế bài kiểm tra nghe hiểu của các giáo viên tương lai: bối cảnh, đầu vào và quy trình. Bằng việc sử dụng phương pháp hỗn hợp, dữ liệu được thu thập từ 105 sinh viên năm thứ tư chuyên ngành Sư phạm tiếng Anh và được phân tích chi tiết. Kết quả cho thấy, mặc dù sinh viên có hiểu biết chung về các nguyên tắc đánh giá kỹ năng nghe, họ vẫn gặp phải những thách thức đáng kể trong việc lựa chọn tài liệu phù hợp, điều chỉnh các hoạt động kiểm tra phù hợp với trình độ của người học, xác định các kỹ năng nghe cụ thể và đảm bảo tính hợp lệ và độ tin cậy của bài kiểm tra. Kết quả cũng cho thấy việc hạn chế tiếp cận với các tài liệu phù hợp và thiếu cơ hội thực hành, sinh viên vẫn gặp khó khăn trong việc thiết kế bài kiểm tra. Do đó, nghiên cứu đề xuất tăng cường thực hành có hướng dẫn, cải thiện khả năng tiếp cận các nguồn tài liệu nghe và phản hồi có hệ thống hơn để nâng cao năng lực thiết kế bài kiểm tra nghe của sinh viên năm thứ tư chuyên ngành Sư phạm tiếng Anh tại Trường Đại học Mở Hà Nội.

Từ khóa: mô hình đánh giá, thiết kế bài kiểm tra nghe, sinh viên năm thứ tư chuyên ngành Sư phạm tiếng Anh

¹ Trường Đại học Mở Hà Nội, Hà Nội, Việt Nam