

## Measuring Listening Comprehension Skills: Review of Arabic Language Proficiency Test

Syihabuddin<sup>1</sup>, Hikmah Maulani<sup>2\*</sup>

Universitas Pendidikan Indonesia<sup>1,2</sup>

e-mail : <sup>1</sup>syihabuddin@upi.edu, <sup>2\*</sup>hikmahmaulani@upi.edu

**Abstract:** Listening is one of the essential elements in communication and language comprehension, and listening tests are an effective way to measure and develop this skill. Listening tests help assess an individual's listening comprehension and provide feedback that can be used for improvement. The research method used is a quantitative approach with ex post facto methods. Data sources are answer sheets and *Ikhtibarat Maharah Lughah Arabiyah (IMALAH)* questions in the listening section. Data source is 250 respondents who are users of the *Ikhtibarat Maharah Lughah Arabiyah (IMALAH)*. Response analysis uses item analysis, which consists of 50 listening items. The results obtained are related to the difficulty level of the questions. The category range is 0.00-0.32 = difficult; 0.33-0.66 = moderate; 0.67-1.00 = easy. Out of the 50 items, 5 are categorized as difficult, 23 as moderate, and 22 as easy. Analysis of test user responses helps identify strengths and weaknesses of participants, measure the level of material comprehension, and identify areas where improvement is needed. The results of this analysis can be used to make decisions about test development and improvement, as well as provide useful feedback to participants.

**Keywords:** Arabic, Listening, Proficiency, Test.

### Introduction

Listening tests are an important component in measuring a person's language skills, especially in a second or foreign language (Acat et al., 2016)(Al-Sawalha, 2016)(Mulyadi, 2018). In its implementation, many researchers carry out evaluation tests on proficiency tests, the focus being on the components of the test product and also the test process (Zubaidi & Indonesia, 2020). The listening test is designed to measure a person's ability to listen and understand information conveyed through sound or speech. This test is often used to measure listening comprehension in a foreign language, but can also be used in a variety of other contexts. One of the most critical areas where considerable attention is required in teaching a foreign language involves listening skills. Listening plays an essential role in both communication and learning, and it is one of the most crucial necessary language skills in the field of language learning (Hayati, 2009). It is well accepted that listening is an integral skill for academic and professional achievement (Suwannasit, 2019).

In its role, listening tests are a valuable tool in measuring listening comprehension abilities, and are important in a variety of contexts, including education, workforce

recruitment, language testing, and others (Abobaker, 2017). This test helps assess a participant's ability to understand information conveyed orally, which is an important skill in communication and language comprehension (Abdelbaset & Jabari, 2020).

Foreign language listening tests have a variety of benefits, including as an accurate evaluation tool, measuring progress in learning a foreign language, and selecting suitable candidates for jobs that involve cross-cultural communication. This test also helps develop listening and speaking skills in a foreign language, which is important in today's era of globalization.

In measurements, reviewing the tests used is necessary (Imaduddin et al., 2022). Testing the questions makes it possible to measure the extent to which the questions predict the participant's actual performance on the test. Testing questions helps identify consistency problems in questions, such as ambiguous or unclear instructions or questions. Furthermore, the evaluation test results are used to develop high-quality questions that can measure abilities or knowledge in accordance with the measurement objectives.

In terms of its relation to language proficiency tests, Measurement tests are important tools that help us measure, evaluate and understand various aspects in various contexts. It plays an important role in decision making, quality improvement, understanding, and development of individuals and organizations. This is no exception in foreign languages.

As a focus, we specify that the foreign language in the context of this research is Arabic. In practice in Indonesia, the available Arabic language proficiency test options can vary depending on the purpose of use, the level of skill being measured, and the context in which the test is used. This test helps in measuring and verifying one's Arabic language skills, which is important in various situations such as studies, work or travel. In addition, the use of Arabic proficiency tests can help improve the standard of teaching and understanding of Arabic in Indonesia.

Test administration in Indonesia is widely carried out and developed, especially by organizers in universities. Like TOAFL (*Test of Arabic as a Foreign Language*) organized by UIN Jakarta and the Pare language village, IKLA (*Ikhtibar Kafaatul Lughoh Al-Arabiyah*) organized by UIN Yogyakarta, and IMALAH (*Ikhtibarat Maharat Lughah Arabiyah*) organized UPI. There were several reasons for the number of organizers, including the existence of Arabic in Indonesia.

In Indonesia, Arabic is one of the foreign languages taught in formal educational institutions, such as Islamic Elementary Schools, Islamic Middle Schools, Islamic High Schools, Senior High Schools, and various other Islamic and regular universities. In universities, Arabic is also taught in non-formal educational institutions, such as Islamic boarding schools, and various other Islamic educational institutions (Purwanto et al., 2019).

In the context of this research, the researcher took samples from the Arabic language proficiency test held by UPI, namely IMALAH (*Ikhtibarat Maharat Lughah Arabiyah*) in the focus of listening comprehension or *istima'*. Much research has been

carried out on the analysis of proficiency questions in the last ten years, including: (Prihartini & Wahyudi, 2018) researching The Development of Integrated Learning Model To Improve Language Skills at Arabic Language at UIN STS Jambi, (Buska et al., 2020) researching the Management of Language Laboratory in Improving Students Arabic Competence, (Imaduddin et al., 2022) testing the Test the Validity and Reliability of Arabic Learning Questions on questions used in elementary schools. Research on second and foreign language acquisition reveals that both cognitive and affective factors have a significant effect on the listening process and strategies employed by students as well as the outcome of their listening (Al-Sawalha, 2016).

Furthermore, continuing from various previous research results, the researcher limited the topic to analyzing the level of difficulty in the listening questions used in the *Ikhtibarat Maharah Lughah Arabiya* (IMALAH) proficiency test in 2023. This research aims to test the accuracy of the questions used by users from the category Education, namely students, also examines user perceptions of the test.

## Method

This research uses a quantitative approach with the post facto method. Ex post facto research in question analysis can help in understanding the factors that influence test results, as well as in improving test design, test implementation, and the learning process (Sappaile & Makassar, 2020). It is a useful tool for evaluating the effectiveness of teaching and testing and for ensuring fairness and accuracy in the measurement of participants' abilities or knowledge.

The data source for this research is the analysis of the answers to the *Ikhtibarat Maharah Lughah Arabiyah* (IMALAH) listening section. Participants in this research were 250 respondents who were users of the *Ikhtibarat Maharah Lughah Arabiyah* (IMALAH).

Data collection involves analyzing pre-existing documents or archives. This can include reports, notes, or historical data that are relevant to the research variables, namely the *Ikhtibarat Maharah Lughah Arabiyah* test. Researchers use interviews to gain an in-depth understanding of participants' experiences or perceptions after taking the test. The interview data analysis procedure adopted a written one (Widodo, 2014), the interview analysis steps are as follows: First, listen to the recorded interview data repeatedly to find the themes or important points needed; second, write the interview transcript in the right format so that it is easy to classify the data; third, interpreting interview data; interpret every word and sentence conveyed by the participant, communicate the participant's views, opinions or points of view (inner voices); The fourth step is to produce reliable data by providing opportunities for participants to provide feedback on the results of data interpretation.

The data analysis used includes descriptive statistical methods such as mean, median, mode and standard deviation to summarize and describe the basic characteristics of the data. It provides an overview of the data distribution. Although ex post facto research cannot establish causality directly, analyzes can attempt to evaluate

the extent to which there is evidence to support a cause-and-effect relationship between certain variables.

## Result and Discussion

Analysis of test responses in foreign language proficiency tests helps test administrators and instructors understand participants' level of proficiency, identify areas for improvement, and assess the test's effectiveness in measuring their language proficiency. This analysis is important for future test development and refinement as well as providing useful feedback to participants.

Test response analysis in Arabic language proficiency tests involves a process carried out to understand and evaluate how test takers respond to the questions or tasks asked in the test. This analysis helps to gain insight into participants' Arabic language skills and assess the extent to which they can understand, use and master Arabic.

For listening tests, analysis includes evaluating the extent to which participants can understand the audio material, identify details, and respond to questions related to the audio they are listening to.

### *Test Question Difficulty*

Teaching languages in general and teaching Arabic to non-native speakers, in particular, have had two attempts to create a standardized test to assess language proficiency in Arabic for non-native speakers (Abdelbaset & Jabari, 2020).

Question difficulty tests in the context of evaluation or assessment are a method for measuring the level of difficulty of a question or task in a test. Determining the level of difficulty can help test administrators or instructors understand the extent to which test takers have mastered the material or skills being tested.

The following table 1 and fig. 1 contain an overview of the results of the comprehension test on the listening section in the IMALAH category.

Table 1. Range

<b>Difficulty Level Range</b>	<b>Category</b>
0,00 - 0,20	Very Difficult
0,20 - 0,40	Hard
0,40 - 0,60	Currently
0,60 - 0,90	Easy
0,90 - 1,00	Very easy

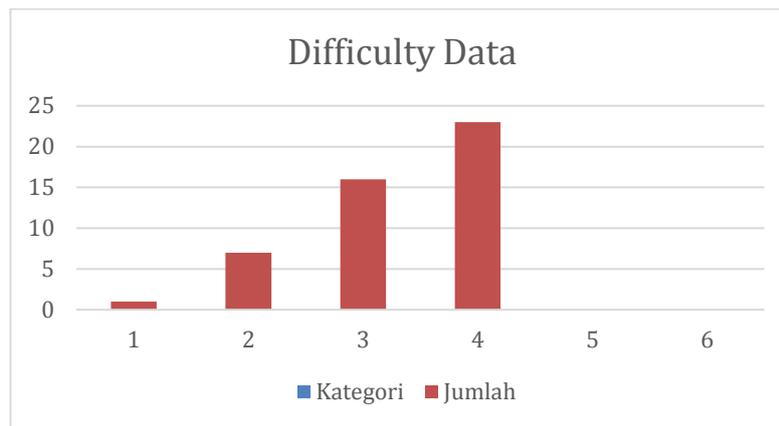


Figure 1. Question difficulty data

The level of difficulty of a question can be expressed in terms of the percentage of participants who answered correctly. Questions that are considered easy are usually answered correctly by most participants, while questions that are considered difficult have a lower level of correctness.

The difficulty index (P) is a statistical measure that describes the extent to which test takers successfully answer a question. The formula is the number of participants who answered correctly divided by the total number of participants. The difficulty index ranges from 0 (very difficult questions) to 1 (very easy questions).

The item difficulty test is an important step in the development and refinement of an evaluation instrument to ensure that the test provides accurate and relevant results related to the stated measurement objectives.

Difficulty criteria may vary depending on the objectives and standards set by the test organizer. Some tests may have strict standards for determining difficulty, while others may be more relative.

The results of the difficulty analysis can be used to revise and develop questions in the future. The ultimate goal is to create a test that has a level of difficulty that matches the participant's abilities and can provide meaningful information.

#### *Test Response Analysis*

For an effective language program, the factors responsible are the learner, teacher, syllabus, method, materials, setting and testing. Richards states, "listening skills are a core component of second language proficiency, and also reflect the assumption that if listening is not tested, teachers will not teach it" (Jeklin et al., 2016).

The research context is different from previous research such as (Abubakar & Mohammed, 2022) which carried out the description using a quantitative approach and differential statistical calculations, but this research was deepened using a qualitative approach with the aim of describing the situation and perception of using the iBT-based Arabic language proficiency test more naturally. To obtain perception data from lecturers, a qualitative approach was carried out using structured interviews, participants as subjects were free to fully express their thoughts and report their experiences regarding the use of the Arabic language proficiency test

called IMALAH. The use of qualitative context in research on user perceptions in the educational realm has been carried out previously by (Abilities et al., 2013) about research on student perceptions of PJBL projects.

The findings show that providing access to an iBT-based Arabic language proficiency test functions with significant value and is considered a useful technological innovation but also does not significantly assess the test content as interesting and up-to-date as seen from the responses regarding the suitability of the question content, obtained by 33.3% of respondents who assess that the content is considered interesting and contemporary according to the technological context. On the contrary, it is very important because the technology is related to the information contained in the test content, as well as the attraction of interested users.

*"I wish that IMALAH can also be accessed by non-Arabic academics. In the future, it is possible to develop both the content of the questions and the test constructions further. Also, it can be used for public purposes such as for those who wish to work in the Middle East. Overall, I think the contents of the IMALAH iBT are quite convenient and easy to use"* (Nal#4 Whatsapp interview).

*So far, IMALAH iBT has presented technology that relies on the internet. This has been a substantial progress in the Arabic language proficiency test situated in Indonesia, which is still relatively rare. I hope that it can be redeveloped using new technological features and it covers proficiency test for both academic and non-academic purposes* (Hik#11 Zoom interview).

Therefore, in planning and developing the construct of the IMALAH test questions as an Arabic language proficiency test, system developers must emphasize development in terms of test question content that can be in line with innovations made in technology. The new technological model innovation was previously proposed by (Ajzen & Fishbein, 1977) In technological development, whether for companies or individuals, this is also related to the determinants of user behavior (Davis & Venkatesh, 1996). Previous research results explained that perceived usefulness and perceived ease of use are related to attitudes towards accepting new technology, ultimately influencing user attitudes (Shen & Chiou, 2010).

Furthermore, other responses were found from the analysis of the difficulty of the questions carried out, including several important functions in the context of test development, evaluation and improvement. The following are some of the main functions of item difficulty analysis in foreign language proficiency tests:

*First*, evaluate the quality of the questions: Difficulty analysis helps evaluate the quality of the questions. By identifying the level of difficulty, the test administrator can determine whether the questions meet the established standards and whether they can provide an accurate measure of the participant's language abilities.

*Second*, Understanding Participants' Abilities: Analysis of question difficulty provides insight into the extent to which test takers are able to answer questions correctly. This helps in assessing their level of foreign language skills.

*Third*, Improving Question Design: By knowing the level of difficulty, test

administrators can refine the question design to suit the target participant's ability level. This helps increase the validity and reliability of the test.

*Fourth, Selection of Appropriate Questions:* By knowing the level of difficulty, the test organizer can select questions that suit the participant's target ability level. This is important to produce adequate and fair tests.

*Fifth, Identify Participant Limitations:* Through difficulty analysis, test administrators can identify limitations or difficulties that participants may experience in answering questions. This can provide insight into improving teaching methods or curriculum preparation.

*Sixth, Development of Better Tests:* Analysis of item difficulty helps in the development of better tests by constructing questions that can provide an accurate picture of the participant's language abilities. This improves the quality and reliability of the test.

By systematically analyzing the difficulty of questions, test administrators can optimize test design, understand participant characteristics, and improve the overall quality of the Arabic language proficiency measurement process.

## Conclusion

Analysis of item difficulty helps increase the validity of the test by ensuring that the questions prepared provide an accurate measure of the participant's language abilities. Test validity is very important to ensure that the test results reflect the participant's true ability in Arabic. Item difficulty analysis is a critical step in the development and evaluation of foreign language proficiency tests. The urgency of this analysis brings great benefits in improving the accuracy, fairness, and reliability of the test, as well as contributing to the development of participants' language skills. With this conclusion, the urgency of analyzing the difficulty of listening questions in foreign language proficiency tests becomes increasingly apparent, considering the importance of listening skills in effective foreign language communication.

## References

- Abdelbaset, T. E. M., & Jabari, N. A. M. (2020). Arabic language teaching evaluation process for non-native speakers. *International Journal of Education and Learning*, 2(2), 105-110. <https://doi.org/10.31763/ijele.v3i1.157>
- Abilities, V. I., Aptitude, A., Doughty, C. J., Listening, A., Wagner, E., Literacy, A., Weigle, S. C., Grammar, A., Purpura, J. E., Pronunciation, A., Isaacs, T., Speaking, A., Sullivan, B. O., Katz, A. M., Armstrong, P., & Kim, S. H. (2013). *Full Table of Contents*. 1, 9-16.
- Abobaker, R. (2017). Improving ELLs' Listening Competence Through Written Scaffolds. *TESOL Journal*, 8(4), 831-849. <https://doi.org/10.1002/tesj.339>

- Abubakar, H., & Mohammed, H. B. (2022). Technology Acceptance Model (Tam) As a Mechanism for Predicting Internet Use for Academic Purposes Among Students of Nigeria Certificate in Education. *International Journal of Education, Technology & Science*, 2(2), 164–180.
- Acat, M. B., Demiral, H., & Kaya, M. F. (2016). *Measuring Listening Comprehension Skills of 5th Grade School Students with the Help of Web Based System 1*. 9(1). <https://doi.org/10.12973/iji.2016.9116a>
- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84(5), 888–918. <https://doi.org/10.1037/0033-2909.84.5.888>
- Al-Sawalha, A. M. (2016). Qualitative and Quantitative Study on Listening Anxiety of Jordanian Students Majoring in English Language at Jerash University. *International Journal of Humanities and Social Science*, 6(1), 82–93.
- Buska, W., Latif, M., Risnita, R., Us, K. A., & Prihartini, Y. (2020). The Management of Language Laboratory in Improving Students Arabic Competence. *Arabi : Journal of Arabic Studies*, 5(1), 51. <https://doi.org/10.24865/ajas.v5i1.249>
- Davis, F. D., & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology acceptance model: Three experiments. *International Journal of Human Computer Studies*, 45(1), 19–45. <https://doi.org/10.1006/ijhc.1996.0040>
- Hayati, A. M. (2009). The Impact of Cultural Knowledge on Listening Comprehension of EFL Learners. *English Language Teaching*, 2(3), 349–371. <https://doi.org/10.5539/elt.v2n3p144>
- Imaduddin, M. F., Maulani, H., & Taufik, I. H. (2022). Test the Validity and Reliability of Arabic Learning Questions. *Arabi : Journal of Arabic Studies*, 7(2), 198–207. <https://doi.org/10.24865/ajas.v7i2.523>
- Jeklin, A., Bustamante Farías, Ó., Saludables, P., Para, E., Menores, P. D. E., Violencia, V. D. E., Desde, I., Enfoque, E. L., En, C., Que, T., Obtenor, P., Maestra, G. D. E., & Desarrollo, E. N. (2016). 濟無No Title No Title No Title. *Correspondencias & Análisis*, 15018, 1–23.
- Mulyadi, D. (2018). Enhancing Students' Listening Proficiency through the Instruction of (Meta)-Cognitive Listening Strategy. *Getsempena English Education Journal (GEEJ)*, 5(2), 168–176. <https://doi.org/10.1088/1751-8113/44/8/085201>
- Prihartini, Y., & Wahyudi, W. (2018). The Development of Integrated Learning Model To Improve Language Skills at Arabic Language. ... *Journal of Educational Research*.
- Purwanto, Y., Ma, diyah, Fauzi, R., & Diterima, N. (2019). 16640 3 Pusat Penelitian dan Pengembangan Kementerian Agama RI. *Jl. Setu, Leuwimekar, Kec. Leuwiliang*, 17(2), 110–124.

- Sappaile, B. I., & Makassar, U. N. (2020). *KONSEP PENELITIAN EX-POST FACTO*. January.
- Shen, C. C., & Chiou, J. S. (2010). The impact of perceived ease of use on Internet service adoption: The moderating effects of temporal distance and perceived risk. *Computers in Human Behavior*, 26(1), 42–50. <https://doi.org/10.1016/j.chb.2009.07.003>
- Suwannasit, W. (2019). บทความวิชาการ ( Academic Article ) *EFL LEARNERS ' LISTENING PROBLEMS , PRINCIPLES OF TEACHING LISTENING Causes of Listening problems faced by EFL learners*. March, 345–359.
- Widodo, H. P. (2014). Methodological Considerations in Interview Data Transcription. *International Journal of Innovation in English Language*, 3(1), 102–107.
- Zubaidi, A., & Indonesia, U. I. (2020). *DEVELOPMENT OF MAHÂRAH AL-ISTIMÂ' TEST INSTRUMENT FOR ELECTRONIC BASED ARABIC STUDENT USING THE KAHOOT! DEVELOPMENT OF MAHÂRAH AL- ISTIMÂ' TEST INSTRUMENT FOR ELECTRONIC BASED ARABIC STUDENT USING THE KAHOOT! APPLICATION*. June. <https://doi.org/10.15408/a.v7i1.13002>