

Development of Arabic Language Exam Questions Based on CEFR and SKKNI Through The LMS Website

Marsekal Rahman Hakiem*

Digital Business, Universitas Madani, Yogyakarta Special Region, Indonesia

Muhammad Irfan Faturrahman

Arabic Language and Literature, Ahmad Dahlan University, Yogyakarta Special Region, Indonesia

Ahmad Luqman Hakim

Arabic Linguistic, King Abdulaziz University, Jeddah, Saudi Arabia

Ach. Afrian Auliansyah

Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia

Lutfi Bangun Lestari

Digital Business, Universitas Madani, Yogyakarta Special Region, Indonesia

*Correspondence author: marsekalrh@umad.ac.id

DOI: <http://dx.doi.org/10.18592/tarbiyah.v14i2.17696>

Key Words:

language evaluation
CEFR
SKKNI
LMS moodle
arabic

Abstract:

This study aims to develop Arabic language test questions based on the Common European Framework of Reference for Languages (CEFR) and the Indonesian National Work Competency Standards (SKKNI), and to present them online via the Learning Management System (LMS). This study is part of an effort to develop an Arabic language evaluation instrument that is more communicative, accountable, and aligned with global and national needs. The instrument development process was carried out using the ADDIE model across the analysis, design, development, implementation, and evaluation stages. The test was then administered to 212 students. The evaluation revealed a decline in CEFR level achievement from pre-test to post-test, indicating a more accurate measurement of Arabic language proficiency. Interviews with 57 participants confirmed that the test items were perceived as more challenging and that they reflected real Arabic language use. These findings demonstrate the success of integrating the CEFR and SKKNI to produce a more targeted and high-quality Arabic language evaluation instrument.

Received : 14 July 2025

Revised : 14 September 2025

Accepted : 15 September 2025

Published : 15 December 2025

To cite this article: Hakiem. M.R., Faturrahman. M.I., Hakim. A.L., Auliansyah. A.F., Lestari. L.B (2025). Development of Arabic Language Exam Questions Based on CEFR and SKKNI Through The LMS Website. *Tarbiyah: Jurnal Ilmiah Kependidikan*. 14(2), 265-280.

This is an open access article under the CC-BY License
(<https://creativecommons.org/licenses/by/4.0/>)



Introduction

Improving the quality of Arabic language learning can begin with developing exam questions. This development is urgently needed for non-Arabic speakers such as the Indonesian people. The use of Arabic in Indonesia continues to increase due to religious, cultural, economic, and educational encouragement (Yasmadi et al., 2024; Mahmudah, 2018). However, there is no standardized, structured, competency-based evaluation instrument for Arabic language learning in Indonesia.

Most Arabic language tests in Indonesia have so far been dominated by old-fashioned testing that emphasizes memorization rather than communicative aspects. This makes learning Arabic boring, and its success rate cannot be adequately measured. Developing Arabic language test questions is important to meet the demands of 21st-century language education competencies, which assess abilities applicable in real-world contexts (Ibrahim & Alhosani, 2020; Wahba et al., 2014).

At the national level, the development of Arabic language questions in Indonesia is still limited to conventional formats dominated by memorization and structural grammar approaches. The Indonesian Arabic Language MGMP Forum has developed Arabic Language Olympiad questions designed to hone students' language skills and critical thinking (Muradi et al., 2021). However, the majority of evaluation instruments still use traditional approaches. The development of curriculum and teaching materials that meet students' needs in the global era, along with the use of information technology, are the main focuses for improving the quality of Arabic language teaching in Indonesia, according to the Indonesian Arabic Language Teaching Policy. The development of Arabic language questions in Indonesia is still fragmented between formal and non-formal educational institutions, such as Islamic boarding schools, schools, and universities, without clear standards (Taufiqurrochman & Lubna Farah, 2024).

Meanwhile, the development of Arabic language questions has adopted a more comprehensive and standardized approach at the international level. Schools, universities, and organizations have used the Arabic Language Proficiency Test (ALPT) worldwide to assess Arabic language proficiency for placement, monitoring, and examination purposes (ALPT) (Pirooz & Saber, 2023). Various international institutions, such as the King Salman Global Academy for Arabic Language, also offer competency tests in reading, writing, speaking, and listening that assess Arabic language proficiency according to the CEFR benchmark called the Hamzah Test (Shaan, 2017), while the Eton Institute developed the Arabic Language Competency Test (ALCT) with a band-score system similar to the IELTS model (Rakhlin et al., 2021). The development of international questions also includes the Certificat International de Maîtrise en Arabe (CIMA), developed by the Arab World Institute as an international certificate of Arabic language proficiency (Ruhendi & Kosim, 2022). This development shows the importance of adopting an international framework such as CEFR in developing Arabic language questions that can measure communicative ability comprehensively and standardize.

Common European Framework of Reference for Languages (CEFR) is an international standard used to describe language skills at six levels (A1-C2), with a focus on listening, reading, speaking, and writing skills (Beinhoff, 2019; Fawaid et al.,

2024). Since its introduction in the early 21st century CEFR has become an international framework of reference that influences language learning, teaching, and assessment. Implementing the CEFR across countries such as Indonesia and Malaysia faces challenges, including limited empirical research, the dominance of English, and obstacles at the educational institution level (Robbani et al., 2023). Similar challenges are also found in Malaysia, such as low teacher awareness and knowledge, lack of professional training, and limited learning resources provided by CEFR (Nawai & Said, 2020; Runnels, 2019).

Implementing the CEFR is often a challenge that affects syllabus, curriculum, and assessment design, as well as the development of teaching materials. This is because much language teaching is still centered on the teacher (Xu & Li, 2024). CEFR updates were also published in 2018 and 2020 under CEFR Companion Volume (CEFR-CV) (Kremmel et al., 2023; Nagai et al., 2020). This update emphasizes the importance of pluricultural competence, namely the presence and interaction of different cultures within an individual or society in response to global social change (Demirdöven, 2025). However, research on the relationship between CEFR and technology remains minimal and is needed.

SKKNI Arabic Language stipulated by the Decree of the Minister of Manpower of the Republic of Indonesia Number 125 of 2022 (Kementerian Ketenagakerjaan Republik & Indonesia, 2022) is a competency standard that regulates Arabic language skills for work purposes, especially in the fields of translation, interpretation, and cross-cultural communication (Reswari, 2025). Arabic language learning for Indonesian workers has developed alongside society's growing recognition of the importance of foreign language skills, especially for workers who will work in the Middle East. This Arabic Language SKKNI establishes competency units that include translating documents, communicating orally, understanding Arabic texts, and writing in a professional context, in line with industry and the world of work (Sari, 2021).

Integrating SKKNI Arabic with the CEFR framework is crucial to creating an internationally recognized assessment standard while meeting national work competencies. SKKNI, which focuses on work competencies, can be mapped with the CEFR proficiency level (A1-C2) to provide a clearer picture of the level of Arabic language skills needed in the world of work (Yuanti & Vardhani, 2021; Oktapianti, 2024).

Therefore, the urgency of this research lies in the need to develop an Arabic language evaluation instrument that is not only standardized and competency-based but also able to integrate two major frameworks: the CEFR as an international reference and the SKKNI as a national competency standard. The unification of these two frameworks through the development of technology-based questions (Subari & Siswanto, 2024), primarily via the Learning Management System (LMS) platform, is expected to create an assessment model that meets global demands and is contextual to national needs. This research offers a strategic contribution to addressing pedagogical and professional challenges in Arabic language learning by presenting a communicative, applicable, and accountable evaluation system, thereby strengthening

the competitiveness of Indonesian students in education and the international world of work.

Research Method

This study uses a Research and Development (R&D) approach, with the ADDIE (Analysis, Design, Development, Implementation, Evaluation) model, to develop an Arabic language test system based on the CEFR and SKKNI through a web-based Learning Management System (LMS). The ADDIE model was chosen because it has proven effective for developing systematic, measurable digital learning products (Maulani et al., 2021; Wang, 2024).

The ADDIE model is the most widely used framework in instructional design because its process is logical, flexible, and allows for continuous revision (Dick et al., 2015). Each ADDIE phase is interrelated and provides feedback for continuous improvement. This model is very suitable for developing a language test system because it provides a clear structure to ensure compliance with international and national competency standards.

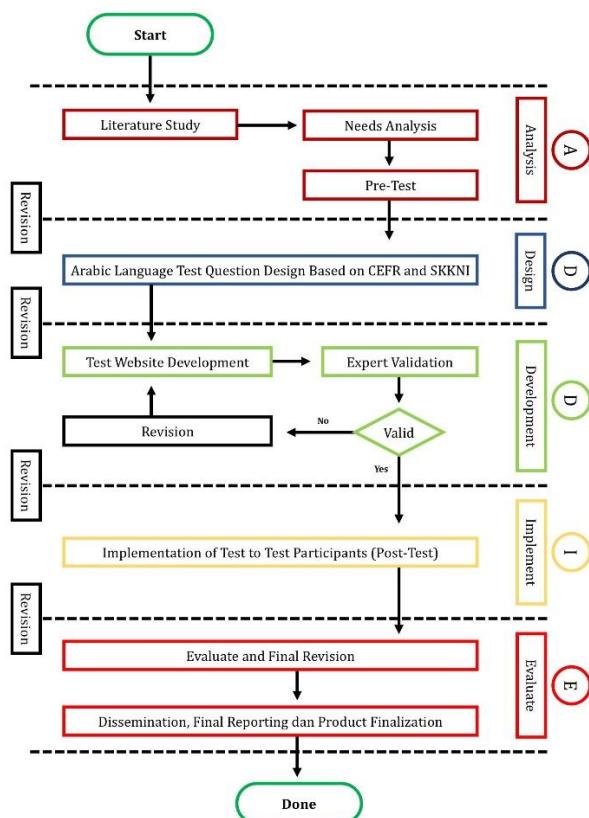


Figure 1. Research Flow Chart

Analysis

The analysis stage is a needs analysis process that determines the test's purpose, identifies problems, analyzes the results of exam questions that apply the old model, and analyzes student needs (Maulani et al., 2021). It was found that the problem was related to the need for a design model of Arabic language exam questions that test speaking and writing skills.

Design

This stage includes designing the exam questions that will be used. This design process includes creating a question matrix and framework that includes the distribution of questions in each aspect and sub-aspect of CEFR and SKKNI.

Development

This development stage includes creating exam questions that integrate CEFR and SKKNI standards and then entering them into the Learning Management System (LMS) website. Question validation involves experts in Arabic language education. The validity analysis technique used to assess the validity of the content of the questions developed uses the formula Aiken's V,

$$V = \frac{\sum s}{n(c-1)}$$

"V" Refers to the validator agreement index in terms of the validity of the test items; "s" is the validator's assessment score minus the lowest assessment score; "n" refers to the number of validators; "c" is the number of categories that the validator can select. All test items are valid if Aiken's V index is 0.37-1.00 (Retnawati, 2016)(Widyaningsih et al., 2021). The Aiken's V value of each test item is calculated based on the test items assessed by each validator. After obtaining the results of the content validity calculation, to find out which category the validity coefficient is in, it is known based on the following table,

Table 1. Content Validity Coefficient Criteria

| Coefficient | Validity |
|--------------------|-------------------------------|
| 0,80 - 1,00 | Content validity is very high |
| 0,60 - 0,79 | Content validity is high |
| 0,40 - 0,59 | Content validity is moderate |
| 0,20 - 0,39 | Content validity is low |
| 0,00 - 0,19 | Content validity is very low |

The test content instrument is considered valid if the validity coefficient is ≥ 0.40 , as this indicates that the majority of items already reflect the content/material to be measured. However, the best validity is above 0.60, or even > 0.80 , indicating very high validity (Azwar, 2017). At this stage, an evaluation process is also carried out, namely, revising the questions with the validator's direction and correction.

Implementation

The next stage is to test the Arabic language test questions developed for 212 students, and university students who taught Arabic. The questions are based on the ITMAM Test, which assesses listening, grammar, and reading skills. In the developed test, students will be assessed in listening, reading, writing, and speaking.

Evaluation

Evaluation is a process for determining whether the Arabic language test questions developed meet the needs and expectations. This evaluation stage is carried out after testing both tests (before and after development) by comparing the results. In

developing Arabic language tests based on the CEFR and SKKNI, the evaluation does not merely measure increases or decreases in scores. However, it assesses the accuracy and validity of measuring students' linguistic abilities (Bachman & Palmer, 2010).

The interpretation of evaluation results is based on the principle of accurate, bias-free measurement. Suppose the results of the second test show a significant increase in scores. In that case, this indicates a potential problem in developing test questions that require revision, such as a level of difficulty that is too low, too explicit instructions, or a positive bias in item construction (Messick, 1989). Conversely, suppose the test results show a decrease in scores. This phenomenon indicates that the new evaluation instrument provides a more accurate and realistic measurement of the student's abilities. According to the Classical Test Theory proposed by Lord and Novick (Lord & Novick, 1968), a valid test must measure the actual score without bias, so that a decrease in scores can indicate the elimination of measurement error and overestimation that occurred in the previous test.

In addition to quantitative scoring, questionnaire-based interviews were conducted with test participants to obtain qualitative data on the difficulty of the questions and to compare students' perceptions of the pre-test and post-test questions. This questionnaire contains indicators such as clarity of instructions, complexity of discourse, diversity of questions, and suitability for Arabic communicative skills. The results of these interviews provide subjective perspectives from test participants that can complement quantitative data and help evaluate the extent to which changes in question format or content influenced their perceptions and experiences.

Results and Discussion

Analysis

Needs analysis was conducted in two stages: the initial stage involved a pre-test, namely testing students with ITMAM test questions, and the second stage involved distributing questionnaires to evaluate the tests conducted and analyze students' needs for developing Arabic language tests in the future. The following results were obtained based on the ITMAM test carried out by 212 participants at the Madrasah Aliyah (MA) Islamic Center Bin Baz and Sekolah Tinggi Ilmu Tarbiyah Madani (STITMA) Yogyakarta.

Table 2. Pre Test Results

| CEFR Level | Description | Number of Exam Participants |
|-------------------|--------------------|------------------------------------|
| A1 | Beginner | 0 |
| A2 | Elementary | 0 |
| B1 | Intermediate | 101 |
| B2 | Upper Intermediate | 111 |
| C1 | Advanced | 0 |
| C2 | Proficient | 0 |

The ITMAM Test Results showed that 111 students scored at level B2 and 101 students scored at level B1. The Questionnaire Results from 57 Respondents showed that 75.4% of students needed exam questions to assess their ability to write Arabic, and 73.7% needed exam questions to assess their ability to speak Arabic.

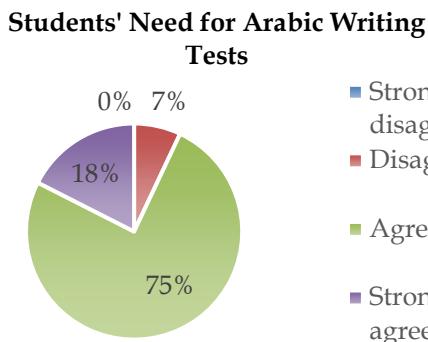


Figure 2. Analysis of Students' Needs for Arabic Writing Tests

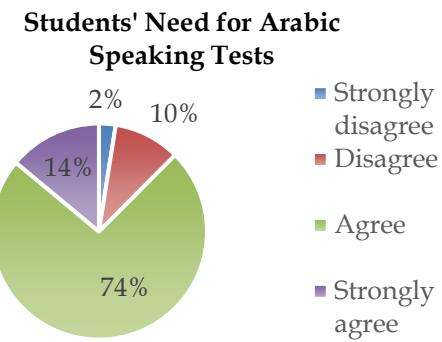


Figure 3. Analysis of Students' Needs for Arabic Speaking Tests

Design

In the design stage, the test instrument is designed based on the analysis results in the first stage. The test instrument is designed as a question matrix and outline, aligned with the SKKNI achievement curriculum and CEFR levels, using the IELTS question model in English. The exam consists of multiple-choice listening and reading tests, a writing test in the available column with two questions presented as pictures and instructions, and a speaking test with two command questions and an answer model in the form of a recording. The question matrix can be seen in the table below.

Table 3. Arabic Language Question Matrix Based on CEFR and SKKNI

| Language Skills | SKKNI Standards | CEFR Standards | Number of Questions | Question Type |
|------------------|--|---|---------------------|--|
| | Main Functions | Basic Functions | | |
| Listening skills | Listening, exploring and remembering simple sounds | Capturing sound symbols in simple sentences Describe the differences in sound symbols Characterize voice intonation | A2-C2 | 40 Questions Multiple Choice (Questions in the form of recordings and text) |
| | Responding to sound symbols in simple sentences | Show a response to the use of language heard according to the context Respond to the interlocutor according to the context of the use of sound symbols | | |
| Reading skills | Read written text aloud | Pronounce words and numbers according to the rules Say simple sentences according to the rules | A2-C2 | 40 Questions Multiple Choice (Text questions) |
| | Responding to written text | Finding the meaning of words in sentences Finding messages in written text | | |

Development of Arabic Language Exam Questions ...

| | | | | | |
|------------------------|--|------------------------------------|--------------|-------------------|--|
| Speaking skills | Producing sound symbols and words | Pronouncing sound symbols | A2-C2 | 1 Question | Instructions and Audio Recordings |
| | Pronounce words according to usage | | | | |
| Writing skills | Produce verbal expressions according to the context of use | Expressing simple sentences | A2-C2 | 1 Question | Instructions and Writing Columns |
| | Using expressions according to the situation and conditions | | | | |
| Writing skills | Delivering writing in the form of letters, words, numbers, figures and simple sentences | Writing letters and words | A2-C2 | 1 Question | Instructions and Writing Columns |
| | Writing numbers and figures | | | | |
| Writing skills | Conveying simple information in writing | | | | |

Development

The development of Arabic language exam questions is based on the designed question matrix and framework. In addition, the questions are formulated online via Moodle LMS for e-learning. The images below show all parts of the Arabic language questions.

The screenshot shows a Moodle quiz interface. On the left, a sidebar displays 'Question 1' (Not yet answered, Marked out of 1.00) with options to 'Flag question', 'Edit question', and 'View details'. The central area shows a question with a text box containing Arabic text and five multiple-choice options (a, b, c, d, e) with corresponding labels. On the right, a 'Quiz navigation' grid shows question numbers 10 through 36, with question 14 currently selected. A 'Finish attempt ...' button is at the bottom right of the grid.

Figure 4. Image of Reading Section (*al-Qira'ah*) Arabic Questions

The screenshot shows a Moodle quiz interface. On the left, a sidebar displays 'Question 1' (Not yet answered, Marked out of 1.00) with options to 'Flag question', 'Edit question', and 'View details'. The central area shows a question with a media player (play, pause, stop) and five multiple-choice options (a, b, c, d, e) with corresponding labels. On the right, a 'Quiz navigation' grid shows question numbers 10 through 36, with question 14 currently selected. A 'Finish attempt ...' button is at the bottom right of the grid.

Figure 5. Image of Listening Section (*al-Istima'*) Arabic Questions

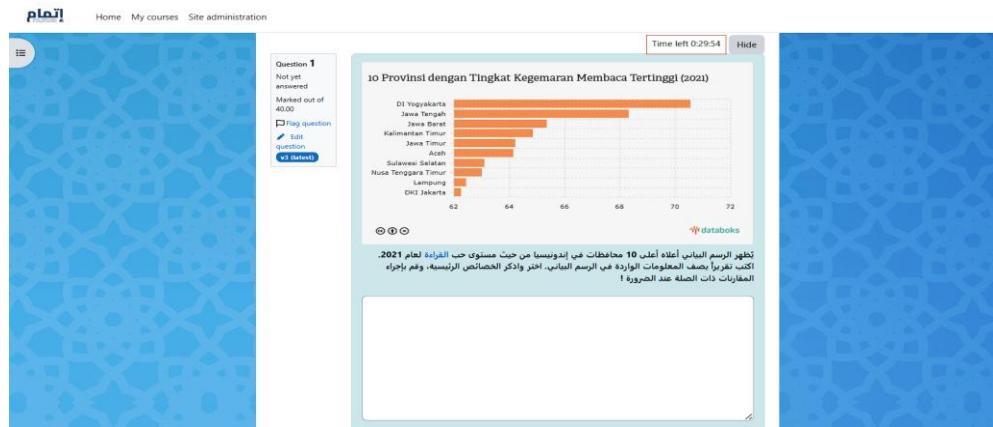


Figure 6. Image of Writing Section (*al-Kitabah*) Arabic Questions

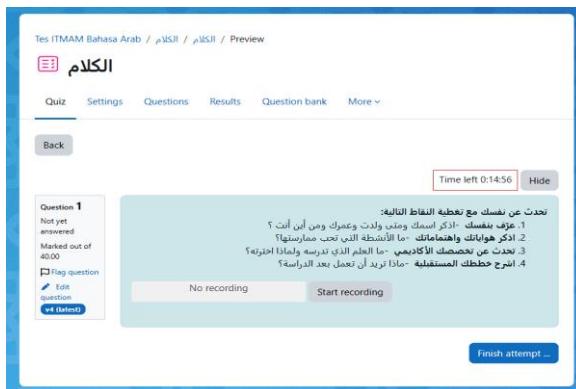


Figure 7. Image of Speaking Section (*al-Kalam*) Arabic Questions

The development stage aims to produce Arabic language test instruments based on CEFR and SKKNI that experts and practitioners have validated. Product validation assesses a product designed, or, in this case, an Arabic language test instrument, based on the CEFR and SKKNI. Product validation is carried out by involving two validators – Arabic language experts and practitioners. The instrument validity test includes material, question construction, language, online test form, innovation, and feasibility of use. The analysis of the validity of the questions assessed by the validator yielded an Aiken's V value of 0.79, indicating high content validity. The questions that experts and practitioners have validated are then revised according to the corrections and suggestions.

Implementation

The implementation stage in this study is a product test. Arabic language questions in the Moodle LMS were administered online to 212 participants, comprising 85 high school students and 127 college students at the research location. Test participants worked on the questions online through e-learning using the accounts that had been shared with them after completing all parts of the Arabic language questions in the form of 40 questions that test listening skills (*istima'*), 40 questions that test reading skills (*qiblah*), 1 question that tests writing skills (*kitabah*) and 1 question that tests speaking skills (*kalam*). The test participants' implementation results are shown in the next stage.

Evaluation

Table 4. Post Test Results

| CEFR Level | Description | Number of Exam Participants |
|-------------------|--------------------|------------------------------------|
| A1 | Beginner | 3 |
| A2 | Elementary | 17 |
| B1 | Intermediate | 131 |
| B2 | Upper Intermediate | 61 |
| C1 | Advanced | 0 |
| C2 | Proficient | 0 |

Based on pre-test and post-test results from 212 participants, Arabic proficiency decreased after the application of the developed questions. In the pre-test, 101 participants were at level B1, and 111 were at level B2, with no participants below level B1. However, in the post-test, the distribution shifted significantly: the number of participants at level B2 decreased to 61. In comparison, the number of participants at level B1 increased to 131. Some participants achieved levels A2 (17 people) and A1 (3 people), which were not previously observed in the pre-test.

These data show that the post-test questions developed by integrating CEFR and SKKNI present a higher challenge, resulting in a distribution of scores that appears to decrease. However, this decline does not necessarily indicate the failure of the instrument but rather indicates that the developed test is more accurate and rigorous in measuring students' linguistic abilities in real terms, as stated by the Classical Test Theory ([Lord & Novick, 1968](#)), which states that the decline in scores can reflect the elimination of overestimation and error measurement from the previous test.

This quantitative finding is supported by qualitative data from questionnaire-based interviews with 57 test participants. The interview results showed that most participants felt the post-test questions were more difficult than the pre-test ones. They stated that the developed questions were longer, more complex, and more explicit, but required a deeper understanding of the communication context. In addition, several participants stated that the more communicative question format demanded productive skills such as writing and speaking actively, which they had not yet fully mastered.

This participant's opinion supports the assumption that the decline in post-test scores reflects the increasing difficulty of the questions, not a decline in student ability. This supports the principle of quality evaluation, which measures absolute score increases and assesses the instrument's validity and accuracy in representing students' abilities.

Discussion

This study uses the Moodle LMS to produce an Arabic language test instrument based on CEFR and SKKNI. The findings of this study are valid and suitable for use. The validation results from two expert Arabic language validators indicate that the developed test instrument has high content validity, as evidenced by Aiken's V index of 0.79. This shows that the test items accurately reflect the competencies being measured in terms of the suitability of the material, question construction, language used, and the online presentation format through the Learning Management System (LMS). The use of Aiken's V index is widely recognized as a robust method for

establishing content validity through expert judgment, particularly for language assessment instruments (Muliana et al., 2020). Furthermore, integrating LMS platforms such as Moodle for language testing has been shown to effectively support valid and reliable assessment processes across various educational contexts (Gamage et al., 2022). This validation also considers the suitability of the questions to the CEFR and SKKNI standards, as well as the instrument's usability in Arabic language learning for Indonesian students.

The evaluation results show a shift in the distribution of participant abilities after implementing the CEFR and SKKNI-based post-test questions. Although quantitatively, there was a decrease in the number of participants at level B2 and the emergence of participants at levels A1 and A2, this phenomenon cannot be directly interpreted as a decrease in the quality of learning but rather as a signal that the post-test instrument developed has a higher level of validity and accuracy in measuring Arabic language skills objectively and contextually. Research on CEFR-based language assessment demonstrates that more valid instruments often reveal a more accurate distribution of learner proficiency levels, which may initially appear as a shift in performance but actually represents improved diagnostic accuracy (Zhu et al., 2023). This aligns with findings that suggest effective language proficiency scales serve as benchmarks for diagnostic assessment, enabling educators to obtain accurate language profiles of students and create appropriate teaching approaches.

In addition, qualitative findings from interviews with 57 participants showed that they found the post-test questions more complex and challenging, particularly in terms of text comprehension and the demands on productive skills such as writing and speaking. Participants said that the more communicative form of questions, more natural instructions, and realistic question contexts required them to use critical thinking skills and more authentic language skills. This aligns with the characteristics of CEFR-based evaluations, which emphasize functional language skills in authentic contexts rather than simply memorizing grammatical structures. The emphasis on communicative competence and authentic language use in CEFR-aligned assessments has been consistently supported by research in language testing literature, demonstrating that such approaches better reflect real-world language proficiency (Wahyuni et al., 2024). Moreover, studies on language assessment in various contexts have shown that instruments incorporating authentic communicative tasks and realistic contexts yield more valid measures of students' practical language abilities.

This condition shows that integrating the CEFR and SKKNI in the development of Arabic language test instruments provides international standardization and improves the quality of student competency measurement more comprehensively and accurately. Thus, test developers need to understand that an instrument's success is not solely measured by high scores but by its ability to select, classify, and represent student achievement fairly and validly. This perspective is supported by contemporary validity theory, which emphasizes that test quality should be judged by the instrument's capacity to provide accurate, fair, and meaningful representations of learner abilities rather than merely producing high achievement scores (Gupta, 2023). Research on language proficiency assessment further confirms that adopting

international frameworks, such as the CEFR, enhances the comprehensiveness and accuracy of competency measurement across diverse learning contexts.

Conclusion

This study shows that developing Arabic language test questions based on the Common European Framework of Reference for Languages (CEFR) and the Indonesian National Work Competency Standards (SKKNI) through a Learning Management System (LMS) platform significantly improves the quality of Arabic language learning evaluation. The questions were designed to measure not only the structural aspects of the language but also communicative skills comprehensively, in line with global demands and the needs of the world of work.

Although the post-test scores decreased compared to the pre-test, the developed test instrument can provide a more accurate and realistic measure of the participant's abilities. This finding is reinforced by interviews with test participants, who reported that the post-test questions were more challenging. This means that the decrease in scores was not caused by learning failure but rather by an increase in the difficulty of questions that were more in line with the authentic context of language use.

Two experts validated the developed instrument, which had high content validity and an Aiken V value of 0.79. This shows that the test items meet the expected competency standards in terms of content, form, and function in evaluating Arabic language learning. Thus, developing evaluation instruments based on CEFR and SKKNI is feasible for use in Arabic language education in Indonesia. It can also serve as a model for developing tests aligned with an international framework while meeting local and national needs.

Acknowledgements

We would like to express our gratitude to the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia for funding this research through the BIMA Beginner Lecturer Research Scheme in 2025.

References

Azwar, S. (2017). *Reliabilitas dan Validitas*. Pustaka Pelajar.

Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World*. Oxford University Press.

Beinhoff, B. (2019). The Common European Framework of Reference for Languages: A European framework for foreign language speech development. *Language Value*, 6(1). <https://doi.org/10.6035/LanguageV.2014.6.5>

Demirdöven, G. H. (2025). CEFR updates (2020)-based next-gen immersive learning in 5 steps. *Frontiers in Education*, 10. <https://doi.org/10.3389/feduc.2025.1567249>

Dick, W., Carey, L., & Carey, J. O. (2015). *The Systematic Design of Instruction*. Vital Source (for Pearson) VST.

Fawaid, A., Solehah, L. Y., & Assya'bani, R. (2024). Structured Infographic Worksheet through Lesson Study in Improving Madrasa Students' Writing Skills in Indonesian Language Subject. *Tarbiyah: Jurnal Ilmiah Kependidikan*, 13(2), 273-290. <https://doi.org/10.18592/tarbiyah.v13i2.13849>

Gamage, S. H. P. W., Ayres, J. R., & Behrend, M. B. (2022). A systematic review on trends in using Moodle for teaching and learning. *International Journal of STEM Education*, 9(1), 9. <https://doi.org/10.1186/s40594-021-00323-x>

Gupta, K. (2023). Validity and Reliability of Students' Assessment: Case for Recognition as a Unified Concept of Valid Reliability. *International Journal of Applied & Basic Medical Research*, 13(3), 129-132. https://doi.org/10.4103/ijabmr.ijabmr_382_23

Ibrahim, A., & Alhosani, N. (2020). Impact of language and curriculum on student international exam performances in the United Arab Emirates. *Cogent Education*, 7(1). <https://doi.org/10.1080/2331186X.2020.1808284>

Kremmel, B., Eberharter, K., Konrad, E., & ... (2023). The CEFR Companion Volume: Opportunities and challenges for language assessment. *Didaktik Slawischer* ..., 3. <https://doi.org/10.48789/2023>

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley.

Mahmudah, M. (2018). Pengaruh Media Audio Terhadap Keterampilan Menyimak dan Berbicara Bahasa Arab. *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan*, 7(2). <https://doi.org/10.18592/tarbiyah.v7i2.2219>

Maulani, A., Ayu, S. M., Dermawan, O., & Akmansyah, M. (2021). The Development of Android-Based Islamic Education Learning Media. *Tarbiyah: Jurnal Ilmiah Kependidikan*, 10(1), 35. <https://doi.org/10.18592/tarbiyah.v10i1.4410>

Messick, S. (1989). *Validity*. In R. L. Linn (Ed.), *Educational Measurement*. Macmillan Publishing Co.

Development of Arabic Language Exam Questions ...

Muliana, N., Pada, A. U. T., & Nurmaliah, C. (2020). Content validity of conation assessment. *Journal of Physics: Conference Series*, 1460(1), 012057. <https://doi.org/10.1088/1742-6596/1460/1/012057>

Muradi, A., Mubarak, F., Permana, F., Hidayat, Y., & Wekke, I. (2021). Revitalization of the Existence of Arabic in Indonesia. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3767448>

Nagai, N., Birch, G. C., Bower, J. V., & Schmidt, M. G. (2020). *The CEFR and Practical Resources* (pp. 1-36). https://doi.org/10.1007/978-981-15-5894-8_1

Nawai, R., & Said, N. E. M. (2020). Implementation Challenges of Common European Framework Reference (CEFR) in a Malaysian Setting: Insights on English Teachers' Attitude. *International Journal of Academic Research in Business and Social Sciences*, 10(7). <https://doi.org/10.6007/IJARBSS/v10-i7/7394>

Oktapianti, S. (2024). TOEFL Score of Tadris Chemistry Students at UIN Antasari in Achievement of CEFR Standard. *Tarbiyah : Jurnal Ilmiah Kependidikan*, 13(1), 55-63. <https://doi.org/10.18592/tarbiyah.v13i1.11417>

Pirooz, H., & Saber, Khaghaninejad. M. (2023). An Investigation into the Interrelation of the First and the Foreign Languages Reading Comprehension Proficiencies: Linguistic Interdependence Versus Balance Hypotheses. *Jordan Journal of Modern Languages and Literatures*, 15(2), 699-711. <https://doi.org/10.47012/jjml.15.2.17>

R. Taufiqurrochman & Lubna Farah. (2024). Transitioning from Monodisciplinary in Arabic Education: Indonesian Insights. *EVOLUTIONARY STUDIES IN IMAGINATIVE CULTURE*, 915-924. <https://doi.org/10.70082/esiculture.vi.745>

Rakhlin, N. V., Aljughaiman, A., & Grigorenko, E. L. (2021). Assessing language development in Arabic: The Arabic language: Evaluation of function (ALEF). *Applied Neuropsychology: Child*, 10(1), 37-52. <https://doi.org/10.1080/21622965.2019.1596113>

Reswari, G. P. A. (2025). *Adopting A Service-Learning Approach in Developing English for Specific Purposes Communicative Competence in Indonesia: A Sociocultural Theory-*

Based Mediational Study Girindra Putri Ardana Reswari A thesis submitted for the degree of Doctor of Philosophy (Issue February).

Retnawati, H. (2016). Proving content validity of self-regulated learning scale (The comparison of Aiken index and expanded Gregory index). *REID (Research and Evaluation in Education)*, 2(2), 155-164. <https://doi.org/10.21831/reid.v2i2.11029>

Robbani, A. S., Nurdianto, T., Annafiri, A. Z., Haqqy, A. M., & Irfan, I. (2023). Tren penelitian dan tantangan penerapan Common European Framework of Reference for Languages (CEFR) di Indonesia. *Diglosia: Jurnal Kajian Bahasa, Sastra, Dan Pengajarannya*, 6(4), 1011-1026. <https://doi.org/10.30872/diglosia.v6i4.736>

Ruhendi, A., & Kosim, N. (2022). Developing Arabic Language Teachersâ€™ Competence and Performance through Teacher Profession Education. *Jurnal Pendidikan Islam*, 8(1), 37-50. <https://doi.org/10.15575/jpi.v8i1.18243>

Runnels, J. (2019). Impact of the Common European Framework of Reference – A bibliometric analysis of research from 1990-2017. *CEFR Journal - Research and Practice*, 1, 18-32. <https://doi.org/10.37546/JALTSIG.CEFR1-2>

Sari, I. (2021). *Error Analysis in the Work of Novice Indonesian Translators*.

Sari, Indah. 2020. Error analysis of economic terms in the work of indonesian novice translation. *Revista de lenguas para fines específicos* 26.1, pp. 122-132 . <http://dx.doi.org/10.20420/rlfe.2020.317>

Shaalan, S. (2017). Reliability and Validity of Four Arabic Language Tests: A comparison of performance of Qatari School-aged children with and without language impairment. *Arab Journal of Applied Linguistics*, 2(1 SE-), 20-48.

Subari, A. J., & Siswanto, S. (2024). Development of Accounting Learning E-Modules with a Guided Inquiry Approach to Increase Student Learning Motivation in Accounting Computer Subjects. *Tarbiyah : Jurnal Ilmiah Kependidikan*, 13(2), 209-227. <https://doi.org/10.18592/tarbiyah.v13i2.13757>

Wahba, K., Taha, Z. A., & England, L. (Eds.). (2014). *Handbook for Arabic Language Teaching Professionals in the 21st Century*. Routledge. <https://doi.org/10.4324/9780203824757>

Wahyuni, T., Rahayu, W., & Naga, D. S. (2024). Validity and reliability of the situational judgment test instrument for assessing the 'Pancasila' student profile. *Cogent Education*, 11(1), 2439157. <https://doi.org/10.1080/2331186X.2024.2439157>

Wang, J. (2024). Teaching Reform Program of Integrated Practice Course for Automotive Majors Based on ADDIE Model. *Applied Mathematics and Nonlinear Sciences*, 9(1), 20242135. <https://doi.org/10.2478/amns-2024-2135>

Widyaningsih, S. W., Yusuf, I., Prasetyo, Z. K., & Istiyono, E. (2021). The Development of the HOTS Test of Physics Based on Modern Test Theory: Question Modeling through E-learning of Moodle LMS. *International Journal of Instruction*, 14(4), 51-68. <https://doi.org/10.29333/iji.2021.1444a>

Xu, H., & Li, G. (2024). CEFR-informed learning, teaching and assessment: A practical guide. *Studies in Second Language Learning and Teaching*, 14(2), 365-370. <https://doi.org/10.14746/ssllt.32669>

Yasmadi, Y., Husna, I., Arsy, F., Rahmah, A., Kustati, M., & Perrodin, D. D. (2024). Insights from a State Islamic University on Arabic Education as a Catalyst for Religious Moderation in Indonesia. *International Journal of Language Education*, 8(3). <https://doi.org/10.26858/ijole.v8i3.66496>

Yuanti, E. E., & Vardhani, N. K. (2021). Penyusunan Kurikulum Blok Bahasa Inggris Vokasional. *Jurnal Vokasi Indonesia*, 9(1). <https://doi.org/10.7454/jvi.v9i1.219>

Zhu, A., Mofreh, S. A. M., & Salem, S. (2023). The Application of Language Proficiency Scales in Education Context: A Systematic Literature Review. *Sage Open*, 13(3), 21582440231199692. <https://doi.org/10.1177/21582440231199692>