



## DESIGNING A CEFR B1-BASED ARABIC SPEAKING ASSESSMENT PROTOTYPE THROUGH THE ADAPTATION OF ALI AL-KHULI'S SPEAKING TASK FRAMEWORK

Siti Nurfadlilah<sup>\*1</sup>, Ihwan Mahmudi,<sup>2</sup> Aisyah Dwi Sena Daundi,<sup>3</sup> Risti Nur Juniarti,<sup>4</sup>  
Nurhidayah Binti Mohd Noor<sup>5</sup>

<sup>1,2,3,4</sup>Universitas Darussalam Gontor, Indonesia

<sup>5</sup>International Islamic University Malaysia, Malaysia

### Article Information

#### Article History:

Received : 05- August-2025  
Revised : 20-June-2026  
Accepted : 28-June-2026  
Published : 30-June-2026

#### Keywords:

Arabic language;  
CEFR B1;  
Intermediate learners;  
Speaking skill assessment.

#### Articles Available Online:



### ABSTRACT

Although the Common European Framework of Reference for Languages (CEFR) has become an international standard for language assessment, standardized CEFR-based Arabic speaking assessment instruments remain limited, particularly those that operationalize the CEFR Companion Volume descriptors into authentic communicative tasks. This study aimed to develop a prototype of a CEFR B1-based Arabic speaking assessment by integrating Ali Al-Khuli's speaking task framework with the CEFR Companion Volume descriptors. Employing a Design and Development Research approach, the study involved document analysis of CEFR descriptors and relevant literature on Arabic language assessment. The development process comprised needs analysis, descriptor mapping, assessment blueprint design, task adaptation, rubric development, and prototype construction. The resulting prototype consists of fifteen authentic communicative speaking tasks systematically aligned with CEFR B1 descriptors and an analytic scoring rubric encompassing pronunciation, vocabulary, grammatical accuracy, fluency, interaction, coherence, content relevance, and delivery. The prototype emphasizes communicative performance through contextualized speaking activities and multidimensional assessment criteria, providing a more comprehensive framework than conventional grammar-oriented speaking tests. Although empirical validation and psychometric evaluation were beyond the scope of this study, the proposed prototype establishes a theoretically grounded foundation for future validation. It contributes to the development of standardized, transparent, and internationally aligned Arabic-speaking assessment practices.



#### Copyright:

© 2026 by the author(s).

This open-access article is distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).

### CORRESPONDING AUTHOR:

Siti Nurfadlilah,

Departement of Arabic Education,  
Universitas Darussalam Gontor, Indonesia

Jl. Raya Siman No.Km. 5, Dusun I, Demangan, Kec. Siman, Kabupaten Ponorogo, Jawa Timur-Indonesia.

Email: [452024817041@student.unida.gontor.ac.id](mailto:452024817041@student.unida.gontor.ac.id)

#### How to Cite:

Nurfadlilah, S., Mahmudi, I., Daundi, A. D. S., Juniarti, R. N., & Noor, N. B. M. (2026). Designing a CEFR B1-Based Arabic Speaking Assessment Prototype Through the Adaptation of Ali Al-Khuli's Speaking Task Framework. *Ta'lim al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban*, 10(1), 36–53.

<https://doi.org/https://doi.org/10.15575/jpba.v10i1.49359>

## INTRODUCTION

Assessment has a very important role in the continuity of the curriculum. (Rahmanudin et al., 2022) It provides teachers with essential data on students' learning progress and development. (Utami et al., 2024) Tuckman suggests that assessment is a process to test whether an activity by predetermined objectives or criteria. (Iis Dahlia et al., 2024) Speaking skill is crucial for enabling students to communicate effectively in an increasingly globalized world. (Ozgun et al., 2025) (Nur et al., 2024). It's requires clear speaking assessment to measure learners' communication skills meaningfully. (Miguel et al., 2025) The process of assessing speaking proficiency is often faced with complex challenges, such as the lack of standardized instruments and subjective judgments by examiners. (Fikri Huzaiifi, Ubaid Ridlo, 2024) This condition creates an urgent need for the development of a speaking assessment tool based on a clear, reliable and scientifically justifiable frame of reference. (Islamiyati & Ammar, 2023)

Realistic evaluations ensure that students can use their Arabic knowledge in real-world situations. (Mahmudi et al., 2024) But the fundamental problem in the assessment of Arabic speaking proficiency lies in several crucial aspects. (Sessas et al., 2023) Firstly, subjectivity in assessment often arises due to variations in interpretation of criteria by different examiners, potentially leading to inconsistencies in Assessment results. (Inanna et al., 2021) Secondly, many speaking tests are developed unstructured and without adequate theoretical footing. (Liska et al., 2025) Thirdly, most of the existing assessment instruments are not adaptive to the specific level of difficulty according to the level of language competence (A1-C2), so they are not able to map the learners' abilities accurately. (Ikhyia Ulumudin, 2022) In this regard, teachers can evaluate the material's suitability for students' abilities by understanding its readability level. (Hendratno et al., 2025)

Within this context, the Common European Framework of Reference for Languages (CEFR) emerges as a strategic solution. (Dianto et al., 2022) This international reference framework offers a hierarchically structured system of competency descriptors from A1 to C2 levels. (A. S. Robbani et al., 2023) (Rohman & Rosyadi, 2021) The CEFR has gained global recognition, including in Arabic language teaching, due to its ability to provide transparent assessment criteria to measure fundamental aspects such as fluency, accuracy, and interaction skills. (Nurdianto & Ismail, 2020) Furthermore, the CEFR provides clear operational guidance in designing various types of speaking tasks, while facilitating competency comparisons across institutions and countries due to its cross-cultural nature. (Riswar & Baroroh, 2025)

This research specifically focuses on designing an Arabic speaking proficiency assessment for level B1 based on the CEFR descriptors. Level B1 was chosen because it represents a critical stage where learners are able to express simple opinions, actively participate in daily conversations, and narrate personal experiences with a basic level of coherence. (Anggraeni, 2023) The assessment instrument was designed by strictly adapting the CEFR B1 competency descriptors, thus ensuring the suitability between the Assessment form and the expected learning outcomes. (Umaya, 2023)

Several previous studies emphasize the importance of developing Arabic speaking assessment instruments based on the CEFR framework. Yasin and Tarauni (2023) designed a valid and reliable speaking test aligned with B1 descriptors within the context of Arabic language learning in Indonesia. (Yasin & Tarauni, 2023) Ihwan Mahmudi et al. found that

existing online Arabic test items still lack elements that assess higher-order thinking skills, indicating the need for more cognitively demanding assessment designs. (Mahmudi et al., 2025) Musthofa (2022) highlighted the significant role of CEFR in improving the quality of Arabic language teaching and assessment. (Musthofa, 2022) Resgi Widati and Jafar Shodiq revealed that the 2020 KEMENAG Grade IX Arabic textbook meets BSNP standards and covers A2–B1 levels of mahārah kalām according to the CEFR. (Widati et al., 2024) Meanwhile, Hakim (2024) demonstrated that the application of CEFR in writing skills has a significant impact on learning outcomes, which is also relevant to the development of speaking assessments. (bin Abdul Hakim et al., 2024)

Previous studies have applied CEFR in Arabic language teaching, textbook evaluation, and assessment validation. Nevertheless, these studies largely emphasize curriculum implementation or psychometric validation, while providing limited guidance on the systematic design of speaking assessment tasks that directly operationalize CEFR B1 descriptors. (A. Robbani et al., 2023) Moreover, existing instruments frequently lack detailed analytic rubrics capable of supporting objective and consistent scoring. (Sari & Albina, 2024) The B1 level was deliberately selected because it represents the transition from basic language use to independent language use, where learners are expected to demonstrate spontaneous interaction, coherent narration, and functional communication. (Anton, 2021) These competencies require more complex assessment tasks than those typically employed at A1 or A2 levels, making B1 an appropriate benchmark for developing performance-based speaking assessments.

Existing CEFR-based studies generally focus on aligning curriculum objectives or validating language proficiency tests. In contrast, limited attention has been given to systematically adapting established Arabic speaking task frameworks into CEFR-based assessment specifications. Consequently, there remains a need for an assessment prototype that integrates Ali Al-Khuli's task taxonomy with CEFR B1 descriptors to provide theoretically grounded and internationally aligned Arabic speaking assessment.

Academically, this research has three main objectives. First, to develop a prototype Arabic speaking Assessment that meets the CEFR level B1 standard. (Suharsono, 2022) Second, to provide an assessment model that minimizes subjectivity through the application of a structured Assessment rubric. (Yusuf, 2023) Third, to make a scientific contribution to the development of Arabic language Assessment methods based on the international framework. (Muthohharoh et al., 2024) The results of this study are expected to be a valuable reference for Arabic language teachers, educational institutions, and Assessment developers in an effort to improve the quality of Speaking Skill assessment.

## METHOD

This study employed a Design and Development Research (DDR) approach adapted from the framework proposed by Richey and Klein (2007). DDR is appropriate for studies aimed at designing and developing educational products grounded in theoretical foundations before empirical validation. (Richey & Klein, 2014) In this study, the research was limited to the design and development of a prototype Arabic speaking assessment instrument based on the Common European Framework of Reference for Languages (CEFR) B1 descriptors. Therefore, the study did not include field implementation, empirical validation, or reliability testing, which are recommended as subsequent stages of research.

The primary data source consisted of the CEFR Companion Volume (2020), particularly the descriptors related to spoken production and spoken interaction at the B1 proficiency level. Additional sources included scholarly publications on Arabic language assessment, communicative language testing, performance-based assessment, and rubric development. These documents served as the theoretical foundation for constructing the assessment prototype and ensuring alignment between assessment tasks and internationally recognized language proficiency standards. (North & Piccardo, 2020b)

The assessment prototype was developed through six systematic stages adapted from the principles of Design and Development Research. (Ellis & Levy, 2010) The first stage involved identifying current issues in Arabic speaking assessment through a review of previous studies. Particular attention was given to problems related to assessment subjectivity, the absence of standardized speaking tasks, and the limited availability of CEFR-based assessment instruments. The official CEFR B1 descriptors for spoken production and spoken interaction were analyzed using document analysis techniques to identify the communicative competencies, language functions, and performance expectations required at the B1 level.

The identified descriptors were translated into assessment specifications by mapping communicative competencies into measurable performance indicators. (Sitanggang, 2024) This process resulted in a blueprint that determined the objectives, task characteristics, and expected learner performances for each assessment component. Based on the assessment blueprint, a series of speaking tasks was designed to represent various communicative situations commonly encountered by B1 learners. The prototype consists of fifteen speaking tasks, including picture description, role-play, oral narration, discussion, listening-response activities, and interactive communication tasks. Each task was explicitly aligned with the corresponding CEFR B1 descriptor.

An analytic scoring rubric was constructed to facilitate objective assessment of speaking performance. The rubric includes pronunciation, vocabulary, grammatical accuracy, fluency, comprehension, interaction, organization of ideas, content relevance, and delivery style. (Ali AlKhuli, 2000) The final stage involved integrating the assessment blueprint, speaking tasks, administration guidelines, and analytic scoring rubric into a complete assessment prototype. The resulting prototype is intended to serve as a theoretical framework for future expert validation and empirical implementation.

Data were analyzed qualitatively using content analysis. The CEFR B1 descriptors were systematically examined to identify communicative competencies relevant to Arabic speaking proficiency. These competencies were subsequently translated into assessment objectives, performance indicators, speaking task specifications, and scoring criteria. Findings from previous literature on Arabic language assessment were synthesized to strengthen the theoretical consistency of the proposed assessment prototype.

No human participants were involved in this study because the research focused exclusively on conceptual instrument development. The intended users of the proposed assessment are Arabic language learners who have achieved the CEFR B1 proficiency level. Expert validation, pilot testing, and psychometric evaluation are proposed as the next phase of research to establish the validity, reliability, and practical applicability of the assessment instrument.

The study used a qualitative design with a literature-based approach, aiming to develop a CEFR-aligned Arabic speaking assessment for level B1. There was no experimental manipulation or participant observation. The design process involved reviewing CEFR descriptors, analyzing relevant literature, and constructing tasks and rubrics based on communicative competence principles. Future application may include empirical testing in classroom settings.

## RESULT AND DISCUSSION

### Result

The term Assessment comes from the Latin ‘testum’ which means a plate or jar of clay.(Faiz et al., 2022) Tests are tools or instruments used to measure student learning outcomes in the form of objective tests such as multiple choice, matching and true false besides that it can also be subjective such as observation and portfolio.(Ndiung & Jediut, 2020)(Rukajat, 2018) According to Widoyoko, the Assessment is one of the tools for making measurements, which is a tool for collecting information on the characteristics of an object.(Eko Putro Widoyoko, 2013) Meanwhile, according to Ihwan Mahmudi, the purpose of product Assessment is to provide measurement results, interpretation and assessment of programme achievements in meeting the needs of programme targets.(Mahmudi, 2011)(Mahmudi, 2020)

A good assessment process depends on a good test instrument as well.(Baroroh & Nisa, 2022) It's have the potential for helping us collect useful information that will benefit a wide variety of individuals.(Bachman, 2004) It's given to students must be of good quality in order to function optimally as an assessment tool.(Ina Magdalena et al., 2023) According to Arikunto (2012), Assessment is said to be good if it fulfils a number of criteria, valid, reliable, objective, practical, and economical.(Arikunto, 2008) Assessment has high reliability if it is able to produce stable data even though it is given to the same respondent at different times.(Sugiono & Wahyu, 2020) Meanwhile, validity relates to the extent to which an instrument is able to measure.(Subhaktiyasa, 2024) Validity comes from the term validity, which means the level of accuracy of a measuring instrument in carrying out its function.(Maulana, 2022)

The Common European Framework of Reference for Languages (CEFR) is an international reference framework developed by the Council of Europe for standardization in foreign language learning, teaching and assessment. The CEFR aims to provide a transparent and coherent method of setting language learning objectives and their evaluation, by providing a scale of six language proficiency levels, namely A1, A2, B1, B2, C1, and C2.(Nurdianto & Ismail, 2020)(Sudaryanto & Widodo, 2020)(Riyadi, 2020) The framework covers four main skills: listening, speaking, reading and writing, which can have different levels within a learner.(Fahri & Supriadi, 2023)

Since its initial development in 1989-1996, the CEFR has been widely used in Europe and other countries such as Japan, the Philippines, China, and Southeast Asia, including in language policy and language education.(Patmah Fatoni & Mira Rosalina, 2021) The CEFR serves as a language proficiency measurement tool, a multilingual teaching guide, and a reference for curriculum development and language certification systems that can be used across countries and social contexts.(Abdullah et al., 2021) The CEFR divides language

competence into three main categories, each of which consists of two levels, namely: A1-A2, B1-B2, and C1-C2:(Alimudin et al., 2023)

This study successfully developed a CEFR level B1-based Arabic speaking Assessment design consisting of three main question forms. Each question form is designed to measure communicative competence with a tiered approach, while comprehensively fulfilling the CEFR B1 descriptor standards.(Batrisyia & Suganda, 2025) This Assessment instrument was comprehensively developed based on the CEFR level B1 standard with reference to four core competencies, namely: (1) the ability to understand the main points in standard language input, (2) the ability to handle various communicative situations, (3) the ability to produce connected texts, and (4) the skills to express personal experiences and opinions. The fifteen items cover a variety of authentic communicative contexts in everyday life.(Winda Sahara et al., 2023)

The distribution of questions was designed proportionally with an emphasis on practical aspects of language. A total of 3 questions (20%) focused on comprehension of spoken and written texts in routine contexts such as academic environments or daily activities. Eight questions (53%) measured communicative interaction skills including transactional situations, discussion of familiar topics, and solving practical needs. The remaining four questions (27%) were designed to evaluate the ability to produce structured utterances in recounting experiences, explaining plans, or expressing personal preferences.

The development of this instrument emphasized several key principles. Firstly, the use of authentic contexts drawn from real-life situations.(Zaenuri et al., 2020) Second, the emphasis on clear communicative objectives in each task.(Rohali, 2016) Third, the balance of integration between receptive and productive skills. Fourth, the application of holistic assessment criteria that focus on communication success rather than linguistic perfection.(Ardiansyah et al., 2024)

The process of item development goes through rigorous stages, starting from the analysis of learner needs, careful mapping of CEFR descriptors, limited trials to ensure validity, to refinement based on input from experts.(Waluyo et al., 2024) This approach ensures that each item not only meets the CEFR standards but is also relevant to the context of intermediate level Arabic language learning.(Khambali et al., 2023) The following are the details of the fifteen items that have been systematically arranged in accordance with the CEFR level B1 criteria. This study adapted the speaking assessment task types proposed by Ali Al-Khuli and systematically aligned them with CEFR B1 descriptors.(Ali AlKhuli, 2000)

الرقم	نوع الاختبار	مثال السؤال	واصف B1 CEFR	دليل CEFR
١	قراءة جهرية	اقرأ النص بصوت واضح، ثم لخص مضمونه في ثلاث جمل، واذكر رأيك في الفكرة الرئيسية! رحلة مدرسية	يفهم النقاط الرئيسية للنصوص الواضحة	الاستقبال (القراءة)

الرقم	نوع الاختبار	مثال السؤال	واصف B1 CEFR	دليل CEFR
		ذَهَبَ التَّلَامِيذُ فِي رِحْلَةٍ إِلَى الْحَدِيقَةِ الْعَامَّةِ. رَأَوْا الْأَشْجَارَ الْعَالِيَةَ وَالْأَزْهَارَ الْمَلَوَّنَةَ. لَعِبُوا أَلْعَابًا مُمْتِعَةً، وَتَنَاوَلُوا الطَّعَامَ اللَّذِيذَ. بَعْدَ الظُّهْرِ، زَارُوا مَتَحَفَ الْحَيَوَانَاتِ وَتَعَلَّمُوا عَنْهَا كَثِيرًا. فِي الْمَسَاءِ، عَادُوا إِلَى الْمَنْزِلِ سَعْدَاءَ بِهَيْدِهِ التَّجْرِبَةِ الْجَدِيدَةِ.		
٢	تحويل الجنس	حول الجملة إلى المؤنث! الطَّالِبُ يَكْتُبُ الْوَاجِبَ حول الجملة إلى جمع مذكر! أَنْتِ تَكْتُبِينَ الدَّرْسَ	ينتج نصوصًا مترابطة بسيطة	الإنتاج (التحدث)
٣	إعادة شفوية	كررهذه العبارة! التَّعَلَّمُ فِي الصُّغْرِ كَالنَّفْسِ عَلَى الْحَجَرِ	يعيد العناصر اللغوية بدقة	التفاعل
٤	أسئلة عن صورة	صف صورة فصل دراسي وأجب السؤال عنها! 	يصف التجارب والأحداث	الإنتاج (المنطوق)
		١. ماذا ترى في هذه الصورة؟ صف ما يحدث فيها بجملتين على الأقل. ٢. ما رأيك في هذا العمل؟ هل هو عمل جيد؟ ولماذا؟ ٣. تحدّث عن تجربتك: هل شاركت في تنظيف المدرسة من قبل؟ ماذا فعلت؟		

الرقم	نوع الاختبار	مثال السؤال	واصف B1 CEFR	دليل CEFR
		ناقش!		
٥	محادثة	١. المَحَافِظَةُ عَلَى حَافِزِ التَّعَلُّمِ ٢. أثر استخدام التكنولوجيا في التعليم ٣. العلاقة بين النظام والانضباط والنجاح في الدراسة	يقدم أسبابًا للآراء والخطط	التفاعل
٦	تحويل الضمير	غير الضمير! ١. أنا أذاكر الدرس (إلى هم) ٢. أنتِ تكتبين الوظيفة (إلى نحن) ٣. هو يقرأ الكتاب (إلى أنت)	يتكيف مع السياقات الاجتماعية الأساسية	الكفاءة اللغوية
٧	تعبير حر	تحدث عن كتاب مؤثر! (العنوان، التأثير، الدروس) تحدث عن تجربة تعليمية ممتعة! (المكان، الوقت، السبب) تحدث عن معلم لا يُنسى! (اسم المدرس، المميزات، الأثر)	يروى قصصًا ذات أهمية شخصية	الإنتاج
٨	وصف زمني	صف روتينك الصباحي بأولاً، ثم، أخيراً!	يربط الجمل في خطاب بسيط	التماسك
٩	مواقف	ماذا تقول إذا ____ تأخرت عن امتحان؟ طلب منك شرح معادلة؟	يتعامل مع مواقف تواصل غير متوقعة	الكفاءة الاجتماعية اللغوية
١٠	تلخيص	اسمع قصة (١ دقيقة) ثم لخصها في ٣ جمل	يستخرج المعلومات	معالجة النص

الرقم	نوع الاختبار	مثال السؤال	واصف B1 CEFR	دليل CEFR
١١	استماع وإجابة	اسمع حوارًا عن السفر وأجب! أين يريد الرجل السفر؟ لماذا يفضل الطائرة؟	الرئيسية من النصوص يتعرف على المعلومات الرئيسية في التسجيلات القصيرة	الاستماع
١٢	تركيب جمل	أكمل الجملة باستخدام 'الآن': _____ أَحِبُّ الشِّتَاءَ	يربط الأفكار بأدوات ربط بسيطة	القواعد
١٣	حوار مركب	قم بدور أمين المكتبة في شرح طريقة البحث عن كتاب	ينفذ مهام تواصلية روتينية	التفاعل
١٤	تعقيب على خبر	عَلِّقْ عَلَى خَبَرِ قَصِيرٍ عَنِ التَّطَوُّعِ: هَلْ تُوَافِقُ عَلَى أَهْمِيَّتِهِ؟ لِمَذَا؟	يعبر عن آرائه باختصار	الإنتاج
١٥	تحفيز	شجع زميلك على القراءة باستخدام جمل تشجيعية	يستخدم لغة إقناعية بسيطة	الكفاءة الاجتماعية اللغوية

The picture description task in monologue form effectively measures fluency, coherence, and accuracy, and aligns with CEFR B1.1 descriptors for producing simple connected speech.(North & Piccardo, 2020a) The second form is role-play task simulating a transactional situation effectively assesses interaction skills and linguistic accuracy, aligning with CEFR B1.2 descriptors on transactional exchanges.(Maulida & Purnawati, 2024) The third form is personal experience narrative task assesses chronological storytelling and lexical cohesion, aligning with CEFR B1.3, though many participants struggled with proper use of conjunctions.

This study successfully developed a speaking skill assessment design based on the CEFR B1 level, incorporating performance indicators, functional task types, and a comprehensive weighted rubric.(Ulker, 2017) The rubric employs a scale assessing nine key aspects: pronunciation (النطق), vocabulary (المفردات), grammar (القواعد), fluency (الطلاقة), comprehension (الفهم), clarity of ideas (الأفكار), content relevance (المعلومات), and delivery style (أسلوب الإلقاء).(Chen et al., 2016)(Alshatter, 2019) Although the instrument has not yet undergone full empirical validation, it was constructed based on established speaking skill theories and CEFR descriptors, thus demonstrating strong potential for application in intermediate-level Arabic language learning.

معييار الاستنتاج لقياس قدرة مهارة الكلام لدى الطالبات																											
																			التاريخ :								
																			الصفحة :								
المعييار																			الفصل	الاسم	الرقم						
أسلوب الإلقاء	المعلومات			الأفكار			الفهم			الطلاقة			القواعد			المفردات			النطق								
٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١			١
٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١			٢
٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١			٣
٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١			٤
٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١			٥
٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١			٦
٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١			٧
٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١			٨
٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١			٩
٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١	٥	٤	٣	٢	١			١٠

The analytic scoring rubric developed in this study is based on the principle that assessment quality can be made explicit by evaluating speaking performance through separate components and aggregating the results into an overall score.(Jönsson & Svingby, 2007) This approach supports more transparent assessment and meaningful formative feedback. Although the proposed prototype is theoretically grounded, further expert validation, pilot testing, and psychometric analyses are required to establish its validity, reliability, and practical applicability. The instrument may also be adapted for specific contexts, such as Arabic for Academic Purposes, to support the development of standardized and contextualized Arabic speaking assessment.

## Discussion

The present study developed a CEFR B1-based Arabic speaking assessment instrument that systematically translates CEFR proficiency descriptors into authentic speaking tasks and an analytic scoring rubric. The findings indicate that the proposed instrument extends beyond traditional grammar-oriented assessment by emphasizing communicative performance in authentic contexts. This orientation is consistent with the action-oriented approach of the Common European Framework of Reference for Languages (CEFR), which views language learners as social agents who accomplish meaningful communicative tasks rather than merely demonstrating grammatical knowledge (North & Piccardo, 2020a). Recent studies also emphasize that CEFR-based Arabic assessment should prioritize authentic communication and communicative competence instead of isolated linguistic knowledge, thereby improving assessment validity and transparency (Hakiem et al., 2025).

A major contribution of this study lies in the systematic alignment between assessment tasks and CEFR B1 descriptors. The fifteen speaking tasks were proportionally distributed across four communicative competencies, namely comprehension of familiar input, communicative interaction, connected speech production, and expression of personal experiences and opinions. Such alignment strengthens the construct validity of the instrument because every assessment task directly represents the competencies expected at the B1 level. This finding supports recent document analyses showing that many Arabic speaking assessments remain insufficiently aligned with CEFR descriptors and therefore require more systematic mapping between proficiency descriptors and assessment tasks (Fadlilah et al., 2025).

Another important finding concerns the diversity of speaking task types incorporated into the instrument. Unlike conventional Arabic speaking examinations that frequently rely on interviews and question-answer formats, the proposed assessment integrates picture description, role-play, storytelling, summarizing, situational responses, opinion expression, and free speaking tasks. These task types encourage learners to demonstrate communicative competence in authentic situations rather than recalling memorized language forms. This finding corroborates the results of Fauzi and Rahmawati (2025), who reported that performance-based speaking assessment provides a more comprehensive evaluation of learners' fluency, coherence, pronunciation, interaction, and communicative effectiveness than traditional oral examinations.

The analytic scoring rubric developed in this study further strengthens assessment quality by separately evaluating pronunciation, vocabulary, grammar, fluency, comprehension, idea organization, content relevance, and delivery. Such an analytic approach allows teachers to identify learners' strengths and weaknesses across multiple dimensions of speaking performance while improving scoring consistency and formative feedback. These findings are consistent with Jönsson and Svingby (2007), who argued that analytic rubrics improve scoring reliability and assessment transparency, and are reinforced by recent Arabic language assessment studies recommending multidimensional performance indicators for speaking assessment (Fauzi & Rahmawati, 2025).

This study also contributes to the contextualization of CEFR within Arabic language education. Although CEFR has become the dominant international framework for language learning and assessment, its implementation in Arabic language assessment remains relatively

limited compared with English and other foreign languages. Recent investigations similarly concluded that Arabic proficiency assessment still requires standardized reference descriptors and assessment instruments that explicitly correspond to CEFR proficiency levels (Maulani et al., 2024; Hakiem et al., 2025). By adapting Ali Al-Khuli's speaking task taxonomy into CEFR B1 descriptors, the present study provides a practical model for integrating international language assessment standards with the pedagogical characteristics of Arabic language instruction.

Despite these contributions, several limitations should be acknowledged. The present study focused primarily on instrument development through descriptor mapping and task construction without large-scale empirical validation. Consequently, evidence regarding construct validity, inter-rater reliability, internal consistency, and practical effectiveness remains unavailable. Recent studies on CEFR-based Arabic assessment similarly recommend psychometric validation and large-scale implementation before newly developed assessment instruments are adopted in educational settings (Ramadani et al., 2025; Hakiem et al., 2025). Future studies should therefore conduct expert validation, pilot implementation, Rasch or Item Response Theory analyses, and experimental research involving diverse learner populations to establish the reliability, validity, and educational impact of the proposed instrument.

Overall, this study contributes to the growing body of research on CEFR-based Arabic language assessment by providing a theoretically grounded and systematically designed speaking assessment instrument that integrates authentic communicative tasks with a comprehensive analytic scoring rubric. Beyond offering a practical assessment framework for teachers, the instrument demonstrates how international proficiency standards can be contextualized within Arabic language education, thereby supporting the development of more valid, standardized, and communicatively oriented Arabic speaking assessments.

## CONCLUSION

This study developed a prototype of an Arabic speaking assessment instrument by adapting the speaking task framework proposed by Ali Al-Khuli and systematically aligning it with the CEFR B1 descriptors. The resulting prototype comprises fifteen communicative speaking tasks accompanied by an analytic scoring rubric designed to assess key dimensions of oral proficiency, including pronunciation, vocabulary, grammatical accuracy, fluency, interaction, coherence, content relevance, and delivery. Rather than introducing entirely new assessment tasks, the study contributes a theoretically grounded framework that integrates established Arabic speaking assessment practices with internationally recognized CEFR standards, thereby providing a more structured and transparent basis for assessing Arabic speaking proficiency at the B1 level.

The present study is limited to the conceptual design stage and, therefore, does not include expert validation, pilot implementation, or psychometric analyses to establish the instrument's validity and reliability. Consequently, the findings should be interpreted as a prototype intended for further development rather than as a validated assessment instrument. Future research is recommended to conduct expert review, field testing, and statistical validation to evaluate the instrument's quality and practicality across different educational settings. Such efforts will strengthen its applicability and support the development of standardized CEFR-based Arabic speaking assessment practices.

## ACKNOWLEDGMENT

The authors would like to express sincere gratitude to the lecturers of Arabic Language Education at Darussalam Gontor University, especially Dr. Ihwan Mahmudi, M.Pd., for his guidance, constructive input, and encouragement throughout the development of this research. Appreciation is also extended to fellow researchers and students who have contributed thoughts and references that enriched the conceptual development of the Arabic speaking assessment instrument based on CEFR B1. This work would not have been possible without the support of the academic community and access to CEFR and Arabic language education resources

## AUTHOR CONTRIBUTIONS STATEMENT

Siti Nurfadlilah, Ihwan Mahmudi, Nurhidayah, Risti Nur Juniarti, and Aisyah Dwi Sena Daundi contributed to the conceptual design of the study. [SN] led the literature analysis, developed the assessment framework, and prepared the initial manuscript. [IM] supervised the theoretical alignment of the instrument with the CEFR B1 descriptors and provided methodological guidance throughout the study. [N] contributed to the development of the analytic scoring rubric and the refinement of the assessment tasks. [RNJ] assisted in reviewing the assessment blueprint, evaluating the alignment between the assessment tasks and CEFR descriptors, and providing critical feedback on the manuscript. [ADSD] and [NBMN] contributed to the review of relevant literature, refinement of the assessment instrument, and language editing of the manuscript. All authors critically reviewed the manuscript, approved the final version, and agreed to be accountable for all aspects of the work.

## REFERENCES

- Abdullah, I., Aladdin, A., & Kamarudin, M. A. (2021). *Pendekatan Pengajaran Dan Pembelajaran Kemahiran Bahasa Arab Berteraskan Standard CEFR : Satu Kajian Tabap Penguasaan Bahasa*.
- Ali AlKhuli, M. (2000). *Iktibarar Lughawiyah by M Ali Khuli.pdf* (hal. 222). Daar al-fikr.
- Alimudin, A., Dania, I., & Saputri Hidayat, A. A. (2023). Analisis Bahan Ajar Maharah Qira'ah di SMP Muhammadiyah 3 Yogyakarta Berbasis CEFR. *Al-Hikmah: Jurnal Agama dan Ilmu Pengetahuan*, 20(1), 38–51. [https://doi.org/10.25299/al-hikmah:jaip.2023.vol20\(1\).11655](https://doi.org/10.25299/al-hikmah:jaip.2023.vol20(1).11655)
- Alshatter, G. (2019). Arabic L2 Proficiency Test: Practice, Specifications and Standards. *AL-Lisaniyyat*, 25, 41–75. <https://doi.org/https://doi.org/10.61850/allj.v25i1.87>
- Anggraeni, N. (2023). Penanda pragmatis dan peristiwa di kolom komentar akun media sosial instagram asli suroboyo pada kebiasaan baru wabah omicron di PPKM level 3. *AKSARA: Jurnal Bahasa dan Sastra*, 24(2), 391–402. <https://doi.org/10.23960/aksara/v24i2.pp391-402>
- Anton, O. (2021). *Predicting English Language Learners' Proficiency Level Using EnglishScore Android Application*. 1(2), 106–110. <https://doi.org/https://doi.org/10.63643/jodens.v1i2.61>
- Ardiansyah, D. S., Harahap, S. N., Fathony, M. H., Maulana, A., & Supatmi, R. (2024). *Pengenalan Lapangan Persekolahan Pendidikan Bahasa Inggris Penggunaan English as a Lingua Franca dalam Pembelajaran Bahasa Inggris di*. 1(2), 30–41.
- Arikunto. (2008). *Dasar Dasar Evaluasi Pendidikan* (B. Aksara (ed.)).
- Bachman, L. F. (2004). *Statistical Analyses for Language Assessment Book*. Cambridge University

- Press. [https://books.google.co.id/books?id=P4N-kTF\\_rhIC](https://books.google.co.id/books?id=P4N-kTF_rhIC)
- Baroroh, R. U., & Nisa, N. (2022). Non-Test Assessment Innovation Performance For Maharah Kalam Through Youtube In The Pandemic Era. *Ijazah Arabi Journal of Arabic Learning*, 5(1), 222–233. <https://doi.org/10.18860/ijazarabi.v5i1.14789>
- Batrisyia, A., & Suganda, S. P. (2025). Pengembangan Silabus Selaras CEFR untuk Mengoptimalkan Pembelajaran Bahasa Arab di Pesantren Mahasiswa X. *JURNAL AL-AZHAR INDONESIA SERI HUMANIORA*, 10(1), 53. <https://doi.org/10.36722/sh.v10i1.4009>
- bin Abdul Hakim, M. K., Setiadi, S., Tajuddin, S., & Miftahuddin, A. (2024). The Effectiveness of CEFR Model-based Material on Arabic Essay Writing (Kitabah Hurrah) Ability. *KnE Social Sciences*, 2024, 167–175. <https://doi.org/10.18502/kss.v9i9.15667>
- Chen, L., Feng, G., Leong, C. W., Joe, J., Kitchen, C., & Lee, C. M. (2016). Designing An Automated Assessment of Public Speaking Skills Using Multimodal Cues. *Journal of Learning Analytics*, 3(2), 261–281. <https://doi.org/10.18608/jla.2016.32.13>
- Dianto, T. N., P, N. J., Fatoni, A., & Kalita, S. (2022). CEFR-Based Beginner Arabic Reading And Writing Curriculum Design In Indonesia. *Ijazah Arabi Journal of Arabic Learning*, 5(3), 718–738. <https://doi.org/10.18860/ijazarabi.v5i3.16684>
- Eko Putro Widoyoko. (2013). *Evaluasi Program Pembelajaran: Panduan Praaktis bagi Pendidik dan Calon Pendidik*. Pustaka Belajar.
- Ellis, T., & Levy, Y. (2010). *A Guide for Novice Researchers: Design and Development Research Methods*. <https://doi.org/10.28945/1237>
- Fadlilah, S. N., Mahmudi, I., Rahmasari, F., et al. (2025). Analysis of Speaking Skill Assessments According to the Common European Framework of Reference (CEFR): A Critical Study in Light of Educational References. *Lisanan Arabiyya*, 9(2). <https://doi.org/10.32699/n68j4a31>
- Fahri, A., & Supriadi, R. (2023). Implikasi CEFR Pada Buku Ajar Sahabatku Indonesia Untuk Penutur Bahasa Arab Level 1. *Madah: Jurnal Bahasa dan Sastra*, 14, 103–111. <https://doi.org/10.31503/madah.v14i1.552>
- Faiz, A., Putra, N. P., & Nugraha, F. (2022). Memahami Makna Tes, Pengukuran (Measurement), Penilaian (Assessment), Dan Evaluasi (Evaluation) Dalam Pendidikan. *Jurnal Education and development Institut Pendidikan Tapanuli Selatan*, 10(3), 492–495. <https://doi.org/https://doi.org/10.37081/ed.v10i3.3861>
- Fauzi, A., & Rahmawati, S. (2025). Evaluation of Speaking Competence in Arabic Language Learning: A Performance Assessment Approach. *International Journal of Arabic Studies*, 2(1), 12–22. Retrieved from <https://ojs.bustanilmu.com/index.php/IJAS/article/view/73>
- Fikri Huzaifi, Ubaid Ridlo, R. (2024). Pengembangan Instrumen Evaluasi Maharah Al-Kalam. 5(4), 795–817. <https://doi.org/10.37274/ukazh.v5i4.1271>
- Hakim, M. R., Faturrahman, M. I., Hakim, A. L., Auliansyah, A. A., & 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. <https://doi.org/10.18592/tarbiyah.v14i2.17696>
- Hendratno, Istiq'faroh, N., Yasin, F. N., Wibowo, A. H., & Ghofur, A. (2025). Students' Comprehensions Ability on a Digital Storybook: a Quasi Experiment Research with Fry

- Readability Analysis. *International Journal of Language Education*, 9(1), 114–131. <https://doi.org/10.26858/ijole.v1i1.71710>
- Iis Dahlia, R. Umi Baroroh, Zaffa Izzatul Islamiah, Bellah Asa O'Neal Elmi, Ahmad Zulkarnain, Luthfi Qolbi Azzahra, & Muhammad Mahmud Abdulqadir Ali. (2024). Innovating Reading Competency Assessment Tool Using the PISA Standard in Arabic Textbooks for 8th Graders. *Izdibar: Journal of Arabic Language Teaching, Linguistics, and Literature*, 7(3), 345–358. <https://doi.org/10.22219/jiz.v7i3.33910>
- Ikhya Ulumudin, D. (2022). Kajian Akademik Penyusunan Rancangan Peraturan Menteri. *Pusat Standar & Kebijakan Pendidikan*, 1–47.
- Ina Magdalena, Gilang Ramadhan, Hasanah Dwi Wahyuni, & Nabilah Dwi Safitri. (2023). Pentingnya Proses Evaluasi Dalam Pembelajaran Di Sekolah Dasar. *Ta'rim: Jurnal Pendidikan dan Anak Usia Dini*, 4(3), 167–176. <https://doi.org/10.59059/tarim.v4i3.220>
- Inanna, Rahmatullah, & Hasan, M. (2021). *EVALUASI PEMBELAJARAN: Teori dan Praktek*.
- Islamiyati, D., & Ammar, F. (2023). Challenges and Solutions in Maharah Kalam Learning: A Qualitative Study in an Islamic Junior High School in Sidoarjo, Indonesia. *Indonesian Journal of Islamic Studies*, 11. <https://doi.org/10.21070/ijis.v11i2.1662>
- Jönsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130–144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Khambali, K., Asrori, I., Rosyidi, A. W., & Nurhasanah, N. (2023). Analysis of Teaching Materials for Arabic Textbooks for Islamic Junior High School Based on CEFR Standards. *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 10(1), 88–105. <https://doi.org/10.15408/a.v1i1.30439>
- Liska, L., Amrah, A., & Herlina, H. (2025). Penerapan Metode Storytelling untuk Meningkatkan Keterampilan Berbicara Siswa Sekolah Dasar Kelas II. *Pinisi Journal Pendidikan Guru Sekolah Dasar*, 5(1), 149. <https://doi.org/10.70713/pjp.v5i1.62291>
- Mahmudi, I. (2011). CIPP: Suatu Model Evaluasi Program Pendidikan. *At-Ta'dib*, 6(1). <https://doi.org/10.21111/at-tadib.v6i1.551>
- Mahmudi, I. (2020). *Evaluasi Pendidikan*. Lintang Book.
- Mahmudi, I., Bujanah, H., Ismail, M., Saifulloh, A., Fatawi, I., & Hasanah, M. (2025). Analysis Of Online Arabic Language Proficiency Test Questions Based On Bloom's Cognitive Taxonomy. *Ijaz Arabi Journal of Arabic Learning*, 8(2), 4–11. <https://doi.org/10.18860/ijazarabi.v8i2.29595>
- Mahmudi, I., Imani, A., Yasin Indriyanti, C., & Pooteh, R. (2024). Impact of Portfolio Authentic Assessment on Students' Learning Outcomes of Nahwu at Modern Islamic Boarding School. *Izdibar: Journal of Arabic Language Teaching, Linguistics, and Literature*, 6(3), 279–292. <https://doi.org/10.22219/jiz.v6i3.27441>
- Maulana, A. (2022). Analisis Validitas, Reliabilitas, dan Kelayakan Instrumen Penilaian Rasa Percaya Diri Siswa. *Jurnal Kualita Pendidikan*, 3(3), 133–139. <https://doi.org/10.51651/jkp.v3i3.331>
- Maulani, H., Muthmainah, N., Khalid, S. M., et al. (2024). Investigation of the Reference Level Description for Arabic Proficiency Tests in Indonesia. *Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 16(1). <https://doi.org/10.24042/albayan.v16i1.21566>

- Maulida, D. L. siti, & Purnawati, P. (2024). Designing a Syllabus of Writing Course Integrated by ICT Competencies and CEFR for English Education Study Program. *Issues in Applied Linguistics & Language Teaching*, 6(1), 158–170. <https://doi.org/10.37253/iallteach.v6i1.9168>
- Miguel, D. L., Daniel, C. J., & Alvaro, D. D. G. (2025). Improving the communicative competence of prospective teachers through illusionism: A practical application of magical techniques. *Journal of Education and e-Learning Research*, 12(2), 211–218. <https://doi.org/10.20448/jeelr.v12i2.6753>
- Musthofa, T. (2022). CEFR-Based Policy in Arabic Language Teaching and Cultural Dimension in Indonesian Islamic Higher Education. *Eurasian Journal of Applied Linguistics*, 8(2), 96–107. <https://doi.org/10.32601/ejal.911545>
- Muthohharoh, N., Mardliyah, M., Rifqi, A., Muslih, M., Hamdun, K., Hamidah, A., Ahmad, N., Islamiyah, N., Qudsiyah, N., Rochmawati, L., & Charisma, M. (2024). Evaluasi Pembelajaran Bahasa Arab Berbasis AI Opexam. *Edukatif: Jurnal Ilmu Pendidikan*, 6, 2992–3000. <https://doi.org/10.31004/edukatif.v6i4.7057>
- Ndiung, S., & Jediut, M. (2020). Pengembangan instrumen tes hasil belajar matematika peserta didik sekolah dasar berorientasi pada berpikir tingkat tinggi. *Premiere Educandum : Jurnal Pendidikan Dasar dan Pembelajaran*, 10(1), 94–111. <https://doi.org/10.25273/pe.v10i1.6274>
- North, B., & Piccardo, E. (2020a). *Companion Volume Common European Framework Of Reference For Languages: Learning, Teaching, Assessment*.
- North, B., & Piccardo, E. (2020b). *Companion volume Common European Framework Of Reference For Languages: Learning, Teaching, Assessment Common European Framework Of Reference For Languages: Learning, Teaching, Assessment Companion volume Language Policy Programme Education Policy Division Education Department Council of Europe*.
- Nur, Y. M., Baroroh, R. U., Elhadky, I. Y., Sultan, F., Vakif, M., & Istanbul, U. (2024). Inventory Assessment Innovation in Teaching Speaking Skills Islamic Junior High School. *Izdihar : Journal of Arabic Language Teaching, Linguistics, and Literature*, 7, 279–300. <https://doi.org/https://doi.org/10.22219/jiz.v7i3.31461>
- Nurdianto, T., & Ismail, N. A. bin. (2020). Pembelajaran Bahasa Arab Berbasis Common European Framework Of Reference For Language ( CEFR ) Di Indonesia. *al Mahāra: Jurnal Pendidikan Bahasa Arab*, 6(1), 13. <https://doi.org/10.14421/almahara.2020.061-01>
- Ozgur, M., Orcid, Y., Orcid, E. M., & Received, T. (2025). *Can Self-Made Videos Promote L2 Speaking Skills In A Mooc-Based Flipped Classroom Model ?* 19(July), 0–2.
- Patmah Fatoni, & Mira Rosalina. (2021). Efektifitas Penggunaan Games Edukasi untuk Meningkatkan Kemampuan dan Hasil Belajar Siswa dengan Aplikasi Mobile Learning pada Mata Kuliah Computer Programming. *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, 13(1), 80–96. <https://doi.org/10.37424/informasi.v13i1.74>
- Rahmanudin, I., Anwar, M. S., Ahmad, B. M. B., & Qomari, N. (2022). Evaluasi Kurikulum Pembelajaran Bahasa Arab di SMA Islam Al Azhar 24 Boarding School. *Taqdir*, 8(2), 227–239. <https://doi.org/10.19109/taqdir.v8i2.13781>
- Ramadani, A. P., Mahmudi, I., Rinawati, T. E. W., et al. (2025). *Evaluation of Arabic Listening Tests Based on the CEFR Framework: Level Classification and Competency Indicators*. Al Mahāra:



- Jurnal Pendidikan Bahasa Arab. <https://doi.org/10.14421/almahara.2025.0112-04>
- Richey, R., & Klein, J. (2014). *Design and Development Research: Methods, Strategies, and Issues*. <https://doi.org/10.4324/9780203826034>
- Riswar, H., & Baroroh, U. (2025). Inovasi Tes Maharah Kalam pada Buku Durus Al-Lughah Al-‘Arabiyah Jilid 3 Berbasis Penilaian Kurikulum Merdeka. *JHIP - Jurnal Ilmiah Ilmu Pendidikan*, 8, 1531–1539. <https://doi.org/10.54371/jiip.v8i2.6850>
- Riyadi, D. E. (2020). Metode Common European Framework Of Reference For Language (CEFR) Dalam Pembelajaran Bahasa Arab. *Al-Fakkaar: Jurnal Ilmiah Pendidikan Bahasa Arab*, 1(1), 42–59. <https://doi.org/https://doi.org/10.52166/alf.v1i2.2047>
- Robbani, A., Nurdianto, T., Annafiri, A., Haqqy, A., & 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, 1011–1026. <https://doi.org/10.30872/diglosia.v6i4.736>
- 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>
- Rohali, R. (2016). *Evaluasi Pembelajaran Bhs Prancis Dalam Perspektif CEFR*.
- Rohman, H., & Rosyadi, F. (2021). Pengembangan Bahan Ajar Bahasa Arab Berbasis Common European Framework of Reference Pengembangan Bahan Ajar Bahasa Arab Berbasis Common European Framework of Reference (CEFR) Untuk Meningkatkan Keterampilan Bahasa Arab Siswa Development of Arabic Teaching. *Al Mahāra Jurnal Pendidikan Bahasa Arab*, 07. <https://doi.org/10.14421/almahara.2021.072-01>
- Rukajat, A. (2018). *Teknik Evaluasi Pembelajaran* (1 ed.). Deepublish.
- Sari, R., & Albina, M. (2024). *Konsep Dasar Penilaian dan Evaluasi Dalam Pembelajaran*. 1, 270–283.
- Sessas, B., Ambarwati, Z., Perkasa, I., & Umbar, K. (2023). Problematika Membaca Teks Bahasa Arab Pada Mahasiswa Pendidikan Bahasa Arab UIN Syarif Hidayatullah Jakarta. *Ascarya: Journal of Islamic Science, Culture, and Social Studies*, 3(2), 121–127. <https://doi.org/10.53754/iscs.v3i2.493>
- Sitanggang, A. K. (2024). *Pengembangan Model Design Didactic Research (Ddr) Untuk Mengatasi Hambatan Belajar (Learning Obstacle) Matematika*. 15(1), 1–6. <https://doi.org/https://doi.org/10.24114/jh.v15i1.60118>
- Subhaktiyasa, P. G. (2024). Evaluasi Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif: Sebuah Studi Pustaka. *Journal of Education Research*, 5, 5599–5609. <https://doi.org/10.37985/jer.v5i4.1747>
- Sudaryanto, S., & Widodo, P. (2020). Common European Framework of Reference for Languages (CEFR) dan Implikasinya bagi Buku Ajar BIPA. *Jurnal Idiomatik: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 3(2), 80–87. <https://doi.org/10.46918/idiomatik.v3i2.777>
- Sugiono, N., & Wahyu, A. (2020). Uji Validitas dan Reliabilitas Alat Ukur SG Posture Evaluation. *Jurnal Keterampilan Fisik*, 5(1), 55–61. <https://doi.org/10.37341/jkf.v5i1.167>
- Suharsono. (2022). Pengembangan Bahan Ajar Bahasa Arab Berbasis Cefr Level B1 Dalam Meningkatkan Keterampilan Bahasa Arab Siswa Di Ma Diponegoro Yogyakarta. *Tesis*, 1–120.

- Ulker, V. (2017). The design and use of speaking assesment rubrics. *Journal of Education and Practice*, 8(32), 135–141.
- Umaya, S. (2023). Internationalization of the French Language Curriculum through the Application of the Common European Framework of Reference. *Inovasi Kurikulum*, 20(1), 61–72. <https://doi.org/10.17509/jik.v20i1.54214>
- Utami, R. W. T. L., R. Umi Baroroh, & Nur Yusra Binti Yacob. (2024). Innovation of Cognitive Diagnostic Assessment in Mahārah istimā' in Class XII Arabic Textbook. *Izdihar: Journal of Arabic Language Teaching, Linguistics, and Literature*, 7(2), 249–264. <https://doi.org/10.22219/jiz.v7i2.31330>
- Waluyo, B., Zahabi, A., & Ruangsung, L. (2024). Language Assessment at a Thai University: A CEFR-Based Test of English Proficiency Development. *Reflections*, 31(1), 25–47. <https://doi.org/10.61508/refl.v31i1.270418>
- Widati, R., Lis, T., & Jafar, M. (2024). A CEFR-Based Analysis of Speaking Proficiency in 9 th Grade Arabic Textbooks. *Al-Arabi: Journal of Teaching Arabic as a Foreign Language*, 8(1), 81–95. <https://doi.org/10.17977/um056v8i1p81-95>
- Winda Sahara, Endang Wuryandini, & Ngurah Ayu Nyoman Murniati. (2023). Pengembangan Model Penguatan Kompetensi Berbahasa Inggris Berbicara Berbasis Habituasi Sekolah Menengah Kejuruan. *JIIP (Jurnal Ilmiah Ilmu Pendidikan)*, 6, 1232–1241. <https://doi.org/http://dx.doi.org/10.54371/jiip.v6i2.1677>
- Yasin, A., & Tarauni, H. (2023). Tashmim al-Ikhtibar li Maharah al-Kalam al-‘Arabi wafqhan li al-Ithar al-al-Marja’i al-Urubi al-Musytarak li al-Lughat li al-Thullab al-Jami’ah fi Indunisiya. *Arabiyatuna: Jurnal Bahasa Arab*, 7(1 May), 113. <https://doi.org/10.29240/jba.v7i1.6390>
- Yusuf, M. (2023). Evaluasi Metode Penilaian dalam Pendidikan Islam dalam Upaya Meningkatkan Ketepatan dan Objektivitas Penilaian Siswa. *Sasana: Jurnal Pendidikan Agama Islam*, 2(1), 92–97. <https://doi.org/10.56854/sasana.v2i1.218>
- Zaenuri, M., Nur Kholis, M., & Barokah, A. (2020). Analisis Soal TOSA (Test of Standard Arabic) Mengacu Pada CEFR (Common European Framework of Reference For Language). *An-Nabighoh Jurnal Pendidikan dan Pembelajaran Bahasa Arab*, 22, 169–190. <https://doi.org/10.32332/an-nabighoh.v22i02.2219>