

## Development of a Web-Based Arabic Competency Test for Indonesian Speakers as a Reference for Measuring Arabic Proficiency Levels of Undergraduate and Graduate Students in Arabic Language Education at Universitas Negeri Jakarta

Fatwa Arifah<sup>1</sup>, Muhammad Kamal<sup>2</sup>, Shafruddin Tajuddin<sup>3</sup>, Ihwan Rahman Bahtiar<sup>4</sup>, Siti Masyitoh<sup>5</sup>, Achmad Yani<sup>6</sup>

Universitas Negeri Jakarta, Indonesia<sup>1,2,3,4,5</sup>

Universiti Islam Sultan Sharif Ali, Brunei Darussalam<sup>6</sup>

[fatwa.arifah@unj.ac.id](mailto:fatwa.arifah@unj.ac.id)<sup>1</sup>, [m.kamal.ah@unj.ac.id](mailto:m.kamal.ah@unj.ac.id)<sup>2</sup>, [Shafruddin.ta@unj.ac.id](mailto:Shafruddin.ta@unj.ac.id)<sup>3</sup>,  
[ihwanrb@unj.ac.id](mailto:ihwanrb@unj.ac.id)<sup>4</sup>, [siti.masyitoh@unj.ac.id](mailto:siti.masyitoh@unj.ac.id)<sup>5</sup>, [achmad.yani@unissa.edu.bn](mailto:achmad.yani@unissa.edu.bn)<sup>6</sup>

---

### Cite this article:

Arifah, Fatwa, Muhammad Kamal, Shafruddin Tajuddin, Ihwan Rahman Bahtiar, Siti Masyitoh, and Achmad Yani. (2025). Development of a Web-Based Arabic Competency Test for Indonesian Speakers as a Reference for Measuring Arabic Proficiency Levels of Undergraduate and Graduate Students in Arabic Language Education at Universitas Negeri Jakarta. *Arabiyatuna: Jurnal Bahasa Arab*, 9(2), 651-666. doi: 10.29240/jba.v9i2.13241

Received: 19-05-2025

Revised: 27-10-2025

Accepted: 11-11-2025

---

### Abstract

This study aims to establish Arabic language proficiency standards through a TOAFL-based competency evaluation model that is recognised both nationally and internationally. The research employed a Research and Development (R&D) approach following the Gall and Borg model. Data analysis ensured validity and reliability through prolonged engagement, persistent observation, and triangulation. The prototype developed in this study measures three key aspects of Arabic proficiency: listening comprehension, structure and written expression, and reading comprehension. Each aspect consists of several sections that assess different linguistic skills through 150 multiple-choice questions. The study also produced a question bank containing structured Arabic test materials designed for standardised assessment. Test scores are processed using a tabulation system that converts correct and incorrect responses into numerical values to generate total scores. The outcome of this research is a prototype of a standardised web-based Arabic language competency test model for Indonesian speakers, which can be used

as a valid instrument for assessing Arabic language proficiency at the undergraduate and master's levels.

**Keywords:** Language competency test; web-based; Arabic language proficiency measurement

## Introduction

Arabic holds a unique dual status as a sacred and global language, serving as the language of the Qur'an and Islamic rituals while also functioning as an international medium of communication, diplomacy, and education recognized by the United Nations.<sup>1</sup> In Indonesia, Arabic plays a vital role in religious, cultural, and educational contexts, particularly within Islamic institutions. However, despite its widespread teaching across schools and universities, a crucial gap remains: the absence of a standardized and validated Arabic proficiency test specifically designed for Indonesian speakers. This gap highlights the urgent need to develop a comprehensive web-based Arabic competency test aligned with international standards and adapted to the linguistic characteristics of Indonesian learners.

At Universitas Negeri Jakarta (UNJ), the Arabic Language Education Study Program at both undergraduate and master's levels aims to produce qualified Arabic teachers with strong linguistic competence. However, there is no valid data indicating the actual level of graduates' Arabic proficiency, as current evaluations are limited to individual courses and do not reflect overall competence. The absence of a standardized web-based test leaves institutions without a reliable measure of students' mastery, making it difficult to assess whether four years of study have effectively produced qualified graduates.

Therefore, this research addresses the urgent need for a standardized, web-based Arabic competency test model tailored to Indonesian speakers. The development of such a model not only fills a significant gap in Arabic language assessment but also strengthens institutional capacity to ensure that graduates possess the desired competencies: knowledge (to know), professional skills (to do), academic identity (to be), and ethical collaboration (to live together).

Language competence is a multidimensional construct that integrates language skills (i.e., listening, speaking, reading, writing) and linguistic

---

<sup>1</sup> Noza Aflisia and Partomuan Harahap, "Perbandingan Pengajaran Keterampilan Berbicara Bahasa Arab Dan Bahasa Inggris Di Sekolah Tinggi Agama Islam Negeri Curup," *Arabiyatuna : Jurnal Bahasa Arab* 1, no. 2 (December 29, 2017): 153, <https://doi.org/10.29240/jba.v1i2.323>.

components (i.e., sound systems, vocabulary, grammar).<sup>2</sup> In Arabic learning, these elements are inseparable,<sup>3</sup> as true proficiency requires mastery of communicative and structural aspects of the language. According to Tarigan, grammatical competence encompasses knowledge of vocabulary, morphological and syntactic rules, semantics, and orthography, demonstrating that linguistic competence forms the foundation of overall language ability.<sup>4</sup> Therefore, to accurately measure Arabic proficiency, an assessment must comprehensively evaluate both language performance and structural knowledge. This theoretical foundation supports the implementation of an Arabic Language Competency Test to assess and validate the linguistic and communicative proficiency of students, particularly graduates of the undergraduate and master's programs in Arabic Language Education.

Evaluation in language education is a systematic process of measurement, judgment, and decision-making aimed at determining the effectiveness of learning and the achievement of intended outcomes.<sup>5</sup> According to Hamalik, evaluation requires making judgments based on clear and accountable criteria and involves three key components: consideration, description of the object being assessed, and measurable standards.<sup>6</sup> In language learning, testing serves as an essential tool to objectively measure learners' mastery of the target language. A language proficiency test, in particular, evaluates a learner's ability to use the language regardless of the teaching materials, making it crucial for assessing genuine communicative competence.

Evaluation plays a vital role not only in assessing student performance but also in improving program quality, informing educational policy, and ensuring alignment with learning outcomes.<sup>7</sup> Within this process, measurement provides

---

<sup>2</sup> Khor Ching Pey, Low Hui Min, and Lee Lay Wah, "Relationship between Oral Reading Fluency and Reading Comprehension among ESL Students," *GEMA Online Journal of Language Studies* 14, no. 3 (2014): 19–32, <https://doi.org/10.17576/gema-2014-1403-02>.

<sup>3</sup> Noza Aflisia et al., "Pemanfaatan Aplikasi Kahoot Untuk Meningkatkan Penguasaan Unsur Bahasa Arab," in *Al-Mu'tamar Ats-Tsanawi Li Al-Lughah Al-'Arabiyah*, vol. 1 (Prodi Pendidikan Bahasa Arab IAIN Curup, 2020), 1–17, <http://prosiding.iaincurup.ac.id/index.php/musla/article/view/8>.

<sup>4</sup> H.G. Tarigan, *Pengajaran Kompetensi Bahasa* (Bandung: Angkasa, 1990).

<sup>5</sup> Eliana Pangestika, R. Umi Baroroh, and Salma Lulu Maknuna, "Innovation Discrete Approach Based on Arabic Language Test Assessment," *Arabiyatuna: Jurnal Bahasa Arab* 8, no. 1 (May 27, 2024): 395, <https://doi.org/10.29240/jba.v8i1.8992>.

<sup>6</sup> O. Hamalik, *Evaluasi Kurikulum* (Bandung: PT. Remaja Rosdakarya, 1993).

<sup>7</sup> Nur Kholis and M. Arif Mustofa, "Development of Competency-Based Arabic Language Curriculum in Traditional Islamic Boarding Schools," *Arabiyatuna: Jurnal Bahasa Arab* 8, no. 2 (November 1, 2024): 827–48, <https://doi.org/10.29240/jba.v8i2.10819>.

quantitative data on students' language mastery, while assessment interprets these results against established benchmarks to determine levels of success. Therefore, careful planning, implementation, and interpretation of test results are critical to supporting effective teaching and learning.<sup>8</sup>

Within this theoretical framework, language testing is understood as a structured set of tasks designed to measure specific language skills and competencies.<sup>9</sup> Assessment data are also used by stakeholders to make decisions related to educational policies. Consequently, the planning, implementation, scoring, and interpretation of assessment results must be conducted meticulously.<sup>10</sup> Based on this, at least three aspects must be considered in learning evaluation: (1) evaluation activities, (2) data related to the object being evaluated, and (3) learning objectives.<sup>11</sup> To create effective learning, all developmental efforts must begin with the establishment of an appropriate assessment system, so that the assessment results can function as a guide in the teaching and learning process.<sup>12</sup>

Language tests consist of questions or tasks that students must answer or complete to measure their mastery of specific language skills, explain their progress, and compare their performance with that of their peers.<sup>13</sup> Accordingly, the Arabic Language Competency Test is designed using a proficiency-based measurement approach, one that assesses a person's Arabic language ability independently of classroom instruction.<sup>14</sup> This is an essential and urgent need that

---

<sup>8</sup> Muhammad Firdaus, Syahrudin Usman, and Kamaluddin Abunawas, "Evaluation of the Arabic Language Learning Implementation and Challenges Faced in the Foreign Language Intensification Program (PIBA)" 9, no. 1 (2025), <https://doi.org/10.29240/jba.v9i1.12505>.

<sup>9</sup> Lim Hui Woan and Lee Soon Tat, "Assessing Children's Native Language in Mandarin Using the Adapted New Reynell Developmental Language Scales-Mandarin (NRDLS-M)," *GEMA Online Journal of Language Studies* 17, no. 2 (2017): 123–45, <https://doi.org/10.17576/gema-2017-1702-08>.

<sup>10</sup> Ima Fitriyah, Fitriatul Masitoh, and Utami Widiati, "Classroom-Based Language Assessment Literacy and Professional Development Need between Novice and Experienced EFL Teachers," *Indonesian Journal of Applied Linguistics* 12, no. 1 (2022): 126–36, <https://doi.org/10.17509/ijal.v12i1.46539>.

<sup>11</sup> Ubaid Ridlo, "Multiple Intelligences in Evaluation of Arabic Learning at Islamic Boarding School," *Arabiyatuna: Jurnal Bahasa Arab* 6, no. 1 (April 11, 2022): 125, <https://doi.org/10.29240/jba.v6i1.3798>.

<sup>12</sup> Vismaia Sabariah Damaiani, Yunus Abidin, and Rosita Rahma, "Higher Order Thinking Skills-Based Reading Literacy Assessment Instrument: An Indonesian Context," *Indonesian Journal of Applied Linguistics* 10, no. 2 (2020): 513–25, <https://doi.org/10.17509/ijal.v10i2.28600>.

<sup>13</sup> Woan and Tat, "Assessing Children's Native Language in Mandarin Using the Adapted New Reynell Developmental Language Scales-Mandarin (NRDLS-M)."

<sup>14</sup> I.R.N. Hula, *Evaluasi Dan Tes Bahasa Arab : Tinjauan Teori Abstrak* (Language Development Center, 2021).



*Kompetensi Bahasa Arab untuk Penutur Indonesia* (UKBA-PI), that can be implemented at the undergraduate and master's levels in the Arabic Language Education Study Program at Universitas Negeri Jakarta (UNJ).

Based on this background, this study aims to develop and validate a comprehensive web-based Arabic competency test for Indonesian speakers. The main research question is: "How can a comprehensive web-based Arabic language competency test instrument be developed in terms of form, type, and linguistic domain?" This question is elaborated into two sub-questions: (1) What is the model of the web-based Arabic language competency test (UKBA-PI)? and (2) How are test items for Arabic linguistic and language skills designed in a web-based format to meet comprehensive and standardized criteria suitable for undergraduate and master's students in the Arabic Language Education Study Program?

This research employed a Research and Development (R&D) design following Gall and Borg's model<sup>18</sup>, which aimed to develop and validate educational products such as learning materials, evaluation models, and testing instruments. The research process included needs analysis, product design, expert validation, limited trials, and revisions to produce a valid and reliable prototype of the web-based Arabic competency test. To ensure data validity, qualitative verification techniques such as extended engagement, persistent observation, and triangulation were applied. The final product of this study was the prototype of the Test of Arabic Competence for Indonesian Speakers (TOAFL), consisting of three major components: (1) Listening Comprehension, (2) Structure and Written Expression, and (3) Reading Comprehension.

## Findings and Discussion

### Arabic Competency Test Approach

Arabic examinations or evaluations are always connected to the teaching and learning process, as they are essential for objectively determining students' abilities in the subjects being tested. Lexicologically, the word *test* refers to (1) an examination, the result of an evaluation, or a trial; and (2) a tool used to measure the quality of something (such as intelligence, ability, learning outcomes, etc.). The term *test* can also mean a written, oral, or interview-based assessment used to determine a person's knowledge, abilities, talents, and personality.<sup>19</sup>

---

<sup>18</sup> M.D. Gall, J.P. Gall, and W.R. Borg, *Educational Research: An Introduction. Recording for the Blind & Dyslexic* (Boston: Allyn & Bacon, 2003).

<sup>19</sup> M. Pateda, *Linguistik Terapan* (Flores: Nusa Indah, 1991) p. 200.

In Arabic language learning, where innovation and evaluation through progressive, modern approaches are required, Arabic language tests must be thoughtfully designed to promote meaningful innovation and comprehensive assessment.<sup>20</sup>

The Arabic test innovation developed in this study used the proficiency test approach. Referring to the *TOEFL Bulletin 2002–2003*, the TOEFL test is a proficiency test that measures a person's English ability without being directly tied to the teaching and learning process.<sup>21</sup> Thus, in the context of Arabic language assessment, a proficiency test model measures students' Arabic language skills without linking them to specific courses or instructional backgrounds.

Within Arabic language evaluation, this approach is considered progressive, as it allows the test to be taken by anyone regardless of their learning background. In addition, the test does not focus on a single field but instead covers various domains of language use.

This proficiency-based approach aligns with the long-established practices of the Center for Applied Linguistics at Stanford University in evaluating individuals' English proficiency.<sup>22</sup> Therefore, the innovation developed for the Arabic competency test is called the Test of Arabic as a Foreign Language (TOAFL).

TOAFL is designed to evaluate students' ability to use and understand Arabic in academic contexts. It measures Arabic language mastery among individuals whose daily language is not Arabic.

### **Arabic Language Test Strategies and Materials**

The test strategy focuses on understanding the meaning of Arabic texts, both spoken and written, implicitly through various language styles and through inferential comprehension. For example: "*A father speaks to his son in Indonesian, while his mother speaks Javanese.*" The inference from this statement is that the child can speak two languages. This type of inferential understanding is developed within the TOAFL test.

---

<sup>20</sup> Faliqul Isbah, M. Fairuz Rosyid, and Nilna Uswah Tamama, "Evaluation Implementation of the Arabic Language Intensification Program (PIBA) in Islamic Higher Education," *Arabiyatuna: Jurnal Bahasa Arab* 8, no. 2 (October 22, 2024): 731–56, <https://doi.org/10.29240/jba.v8i2.11129>.

<sup>21</sup> Muhammad, *Memahami Struktur Dan Tata Bahasa Toefl* (Yogyakarta: LINGUA Publishing, 2008).

<sup>22</sup> Shafqat Zaidi et al., "Grammatical Knowledge of English Cleft Constructions among Pakistani ESL Learners across L2 Proficiency Levels and Learning Styles," *GEMA Online Journal of Language Studies* 23, no. 3 (2023): 76–97, <https://doi.org/10.17576/gema-2023-2303-05>.

Meanwhile, the test materials are comprehensively designed to measure students' Arabic language skills. This ensures alignment between the test materials and the strategies for understanding the implicit meaning of the tested Arabic texts. In Arabic language testing, the criteria for developing test materials are based on providing insights into Arabic as a tool for developing the skills required in Arabic language study programs. In comprehensive testing, the distinct characteristics of the Arabic study program and its test content are adjusted to match the types of linguistic and language skills emphasised during learning; learning is the mastery of knowledge that is carried out consciously and formally.<sup>23</sup> TOAFL test materials and their development are as follows:

*First*, test materials based on educational principles.

The principle of education refers to aspects related to educational theory in exam implementation, meaning that an exam must contain concepts, empirical considerations, knowledge, understanding, and skills.

*Second*, test materials based on linguistic principles.

This principle requires attention to the language tested, including linguistic elements (sounds, vocabulary, and structure) and language skills (listening, speaking, reading, and writing).

*Third*, test materials based on multi-social and multi-disciplinary principles.

This principle requires test materials to cover various social and disciplinary fields. This is inseparable from the broader scope of Arabic, which has become an international language of communication and is no longer limited to Islamic cultural and scientific contexts.

*Fourth*, test materials developed according to Arabic language standardisation.

This involves using standard Arabic, including appropriate language varieties, word choice, sentence structure, and coherent paragraph development. In this context, the Arabic used in the test materials is *fushha*.

*Fifth*, test materials developed adaptively.

Adaptive development occurs in two ways: (1) adaptation to test material substance, and (2) adaptation to ongoing development. *First*, adapting to test material substance involves assimilating social and cultural aspects of Arabic to the cultural background of non-Arab learners, particularly in Indonesia. This minimises inappropriate language transfer from the mother tongue. For example, the Indonesian mindset may treat "animal" as a singular category, while Arabic speakers typically use the plural term *حديقة الحيوانات* to indicate a zoo. *Second*,

---

<sup>23</sup> S. Pringgawidagda, *Strategi Penguasaan Berbahasa* (Yogyakarta: Adicipta Karya Nusa, 2002) p. 17.



adaptation to development concerns the scope of test material content. The themes of the test materials extend beyond predominantly religious topics and include a wider range of universal themes that are accessible to Arabic and non-Arabic speakers, including Indonesian learners.

## Arabic Language Test Prototype Design

### *First, Arabic Language Test Model*

The prototype of the Arabic language competency evaluation model (TOAFL) is a standardized Arabic-language proficiency test designed to assess an individual's mastery of Arabic as a foreign language. The skills evaluated in this model include three components: (1) Listening Comprehension, which tests the participant's ability to understand spoken texts; (2) Structure and Written Expression, which measures proficiency in standard Arabic grammar, morphology, and syntax; and (3) Reading Comprehension, which assesses receptive skills, including the ability to understand written passages within a limited time.

The listening section consists of four parts: (1) listening to short expressions, (2) listening to short conversations, (3) listening to simple conversations, and (4) listening to read texts. In this test, participants hear each expression, conversation, or text only once. Each segment is followed by one or more questions, depending on the length of the material. These questions assess the participant's ability to identify the main idea, key details, implications, interrelated ideas, coherence of information, and the speaker's purpose and attitude. The skills tested include: (a) identifying main ideas, key points, facts, essential details, and vocabulary in context; (b) recognizing the speaker's purpose and the subject being discussed; (c) drawing and explaining inferences; and (d) recognizing topic shifts within the speech.

Table 1. Listening Comprehension Section

القسم الأول: فهم النصوص المسموعة (40 سؤالاً) (30 دقيقة)				
الوحدة	أنواع فهم النصوص المسموعة	عدد الأسئلة	عدد الكتاب	الكفاءة التي تُمتَحَن
الأولى	فهم العبارة القصيرة	10 items (1-10)	4	(a) Understanding the main idea, key points, essential facts and details, and vocabulary in context related to the main idea.
الثانية	فهم المحادثة القصيرة	10 items (11-20)	3	

(b) Recognising the speaker's speech or the purpose and main point of the speech. (c) Drawing and describing inferences. (d) Recognising shifts in topic within the speech.	3	10 items (31-40)	فهم المحادثة البسيطة	الثالثة
	3	10 items (21-30)	فهم النصوص المستمعة	الرابعة

The structure and written expression section consists of three parts: (1) understanding proper language usage, including vocabulary and sentence patterns; (2) identifying correct stylistic structures; and (3) assessing the accuracy of Arabic grammatical structures. Participants are required to select the correct Arabic structures related to vocabulary and sentence formation. The questions in this section are designed to measure participants' mastery of standard Arabic grammar. The abilities assessed include: (a) recognizing correct forms and patterns of Arabic vocabulary; (b) identifying appropriate usage of Arabic vocabulary; (c) recognizing grammatically correct Arabic sentence structures; and (d) assessing the accuracy of Arabic sentence construction based on sentence elements.

Table 2. Structure and Written Expression Section

القسم الثاني: فهم التراكيب العربية والتعبير الكتابي (60 سؤالاً) (50 دقيقة)				
الوحدة	أنواع فهم التراكيب	عدد الأسئلة	عدد الكتاب	الكفاءة التي تُمتحن
الأولى	الاستعمال اللغوي	15 items (1-15)	4	(a) Recognizing the correct forms and patterns of Arabic vocabulary.
الثانية	الكلمات المناسبة / الأساليب الصحيحة	15 items (16-30)	4	(b) Recognizing the correct usage of Arabic vocabulary.
الثالثة	تعيين الكلمة المخطئة في الجملة	15 items (31-45)	5	(c) Recognizing Arabic sentence structures correctly and appropriately.
الرابعة	التعبير الكتابي	15 items (46-60)	4	(d) Recognizing the accuracy of Arabic sentence structure in the use of sentence elements.

The reading comprehension section consists of two parts: (1) vocabulary assessment, which includes identifying synonyms, antonyms, and appropriate usage; and (2) reading several passages followed by questions related to each text. The reading topics in this section are universal, and the questions are designed to be general so that participants are not limited or constrained by a specific theme. The skills assessed include the ability to understand, interpret, and analyse written

texts across various topics.

Table 3. Reading Comprehension Section

القسم الثالث : فهم المفردات و النصوص المقروءة (50 سؤالاً) (40 دقيقة)				
الوحدة	أنواع فهم المقروءة	عدد الأسئلة	عدد الكتاب	الكفاءة التي تُمتَحَن
الأولى	فهم المفردات			Comprehending, interpreting, and analyzing reading texts on a wide variety of topics, both on the main idea of the reading, as well as certain specific information contained in the reading, or the conclusion of a particular statement in the reading.
	الجزء الأول	10 items (1-10)	3	
	الجزء الثاني	10 items (11-20)	3	
	الجزء الثالث	10 items (21-30)	5	
الثانية	فهم النصوص المقروءة البسيطة / الطويلة	20 items (31-50)	3	

### *Second*, Types and Forms of Arabic Language Tests

The test developed in this study was an objective test, meaning that its scoring can be carried out objectively. Objective assessment refers to an evaluation method that produces consistent, stable results, even when administered repeatedly or assessed by different evaluators.<sup>24</sup>

This Arabic language test was designed based on TOAFL standards. Broadly, the TOAFL test is divided into two parts. *First*, in terms of test scope, the Arabic language competency test consists of two domains: language skills and language components. The receptive language skills tested include listening and reading, while the language components assessed include vocabulary, morphology, and syntax. *Second*, in terms of question format, the Arabic competency test uses objective, multiple-choice questions.

All items in the Arabic competency test follow TOAFL standards and use the multiple-choice format. A multiple-choice test consists of a problem and several answer options. The problems are presented as an interrogative sentence, a simple question, a specific statement, an instruction to complete an incomplete

<sup>24</sup> M.S. Djiwandono, *Tes Bahasa Dalam Pengajaran* (Bandung: ITB Bandung, 1996) p. 28.

sentence, or an instruction to identify an incorrect linguistic element. The answer options include words, numbers, symbols, clauses, or sentences. Each item provides four choices, with only one correct answer.

The design of the Arabic language test prototype based on the types and forms described above can be illustrated in the following examples.

- 1) Example of a listening multiple-choice test, where the question is based on a statement. Participants listen to the following statement:

اليد العليا خير من اليد السفلى

They then select the correct answer from four options:

1 ا اليد اليمنى خير من اليد اليسرى ج اليد المعطية خير من اليد الأخذ

ب اليد اليسرى خير من اليد اليمنى د اليد السائلة خير من اليد المسئولة

- 2) Example of a reading multiple-choice test, where the questions are posed at the end of the text:

أصبحت مشكلة زيادة السكان من أهم المشكلات التي تعاني منها معظم دول العالم وقد اتجهت معظم هذه الدول إلى تحديد النسل، فالصين – ألف مليون نسمة- قد حددت النسل طفلاً لأسرة الواحدة وما زاد عن ذاك فهناك عقاب صارم للأبوين ولكن الشعب الصيني مطيع للأوامر ولم تشذ عن هذا القانون إلا حالات نادرة حددها الأطباء، وقد نتج عن هذه المشكلة مشكلات اقتصادية واجتماعية كثيرة .

1 كيف حدد الصين النسل ؟ 2 ما هي المشكلة التي تعاني منها العالم؟

ا طفل واحد لكل أسرة ا مشكلة زيادة السكان  
ب طفلين للأسرتين ب المشكلة الاجتماعية  
ج طفلين للأسرة ج المشكلة الاقتصادية  
د ثلاثة أطفال للأسرة د المشكلة الزراعية

- 3) Example of a multiple-choice question assessing mastery of the morphology component, presented as an incomplete sentence:

1 ..... الطبيب الممرضة لتساعده 2 قابلت أصدقائي ..... في الحلقة

أمس

ا استدعى ا المجموعين  
ب دعا ب المجتمعين

ج استعاد ج الجامعين  
د عاد د المجمعين

- 4) Example of a multiple-choice question assessing mastery of syntax through completion of an incomplete sentence:

1 ..... في الصغر ينفعكم في الكبر 2 أخذ الناس .... القرآن ليعرفوا تعاليم الإسلام

ا إن تتعلموا ا يقرءون  
ب إن تتعلمون ب أن يقرءوا  
ج متى تتعلموا ج قراءة  
د ما تتعلموا د أن قرءوا

### Third, Scoring System

The scoring system in the TOAFL test is applied to each item and is based on several considerations. *First*, the TOAFL questions refer to the tested material and the use of standard Arabic. *Second*, the identification of TOAFL scores uses a score-tabulation model based on the Norm-Referenced Test (NRT) method, which is applied to support Arabic language proficiency assessment. In the NRT method, scoring is determined through weighted question conversion.<sup>25</sup>

The tabulation system works by converting the number of correct and incorrect responses into numerical units to determine the total accumulated test score. The following is an example of the TOAFL scoring system, which applies a conversion to each correct answer.

Table 4. Example of TOAFL Score Calculation Using the NRT Method

TOAFL Questions	Number of Questions	Number of Correct Answers
Section I: Listening Comprehension	40	25
Section II: Structure and Written Expression	60	42
Section III: Reading Comprehension	50	35
Total Items	150	
TOAFL Questions	Number of Questions	Number of Correct Answers
Section I: Listening Comprehension	25	48
Section II: Structure and Written Expression	42	58

<sup>25</sup> E. Sudarmono, *The King TOEFL: Tips & Trik Plus Bank Soal* (Wahyumedia, 2015) p. 14-15.

Section III: Reading Comprehension	35	50
Converted Score Total		156
Multiplied by 10		1560 : 3
Final TOAFL Score		520

The development of the TOAFL test provides substantial contributions to Universitas Negeri Jakarta (UNJ) and the broader academic community in Indonesia. For UNJ, this product serves as a standardised, web-based Arabic proficiency assessment tool that can be integrated into curriculum evaluation and graduate competency certification at undergraduate and postgraduate levels. For Indonesian students, the TOAFL test offers an objective measurement of Arabic language proficiency aligned with international standards, enabling them to assess their abilities more accurately and prepare for academic or professional contexts requiring Arabic competence. Moreover, this innovation strengthens Indonesia's role in Arabic language education research and promotes digital transformation in language-assessment practices.

## Conclusion

Based on the findings of this study, the researchers successfully developed a prototype of a standardised Arabic language competency evaluation model, namely the Test of Arabic as a Foreign Language (TOAFL). The model comprises three integrated components: Listening Comprehension, Structure and Written Expression, and Reading Comprehension. Each component is designed with structured sub-sections and a total of 150 multiple-choice items to comprehensively measure Arabic language proficiency. The test prototype reflects the essential skills required to master Arabic as a foreign language, covering comprehension, interpretation, and the application of linguistic and communicative competence. This model serves as a foundation for establishing a valid and reliable Arabic language proficiency test for Indonesian learners, which can later be institutionalised within the Arabic Language Education Study Program at Universitas Negeri Jakarta (UNJ).

Theoretically, this study contributes to the field of Arabic language testing and evaluation by proposing a model that integrates principles of language assessment, linguistic competence, and web-based testing design. It also enriches the literature on Arabic language pedagogy and measurement by contextualising the TOAFL model for non-Arabic-speaking learners in Indonesia. Practically, the findings provide a structured framework for implementing Arabic proficiency assessments in higher education institutions, offering educators a valid tool to measure students' language achievement objectively. In addition, the development of a question bank and an automated scoring system enhances efficiency, transparency, and accessibility in language evaluation.

This study is limited to the design and initial prototype validation of the TOAFL model and has not yet progressed to large-scale implementation or statistical validation through psychometric analysis. Future research should conduct empirical trials involving larger participant groups to examine item validity, reliability, and level calibration. Further studies may also expand the web-based model into adaptive testing systems and integrate multimedia-based question formats to increase interactivity and authenticity. These subsequent efforts are expected to ensure that the Arabic language competency test evolves into a robust, standardised, and internationally recognised instrument for Arabic language assessment in Indonesia.

## References

- Aflisia, Noza, and Partomuan Harahap. "Perbandingan Pengajaran Keterampilan Berbicara Bahasa Arab Dan Bahasa Inggris Di Sekolah Tinggi Agama Islam Negeri Curup." *Arabiyatuna: Jurnal Bahasa Arab* 1, no. 2 (December 29, 2017): 153. <https://doi.org/10.29240/jba.v1i2.323>.
- Aflisia, Noza, Asri Karolina, Eka Yanuarti, and Muhammad Raihan. "Pemanfaatan Aplikasi Kahoot Untuk Meningkatkan Penguasaan Unsur Bahasa Arab." In *Al-Mu'tamar Ats-Tsanawi Li Al-Lughah Al-'Arabiyah*, 1:1–17. Prodi Pendidikan Bahasa Arab IAIN Curup, 2020. <http://prosiding.iaincurup.ac.id/index.php/musla/article/view/8>.
- Damaianti, Vismaia Sabariah, Yunus Abidin, and Rosita Rahma. "Higher Order Thinking Skills-Based Reading Literacy Assessment Instrument: An Indonesian Context." *Indonesian Journal of Applied Linguistics* 10, no. 2 (2020): 513–25. <https://doi.org/10.17509/ijal.v10i2.28600>.
- Djiwandono, M.S. *Tes Bahasa Dalam Pengajaran*. Bandung: ITB Bandung, 1996.
- Firdaus, Muhammad, Syahrudin Usman, and Kamaluddin Abunawas. "Evaluation of the Arabic Language Learning Implementation and Challenges Faced in the Foreign Language Intensification Program (PIBA)" 9, no. 1 (2025). <https://doi.org/10.29240/jba.v9i1.12505>.
- Fitriyah, Ima, Fitriatul Masitoh, and Utami Widiati. "Classroom-Based Language Assessment Literacy and Professional Development Need between Novice and Experienced EFL Teachers." *Indonesian Journal of Applied Linguistics* 12, no. 1 (2022): 126–36. <https://doi.org/10.17509/ijal.v12i1.46539>.
- Gall, M.D., J.P. Gall, and W.R. Borg. *Educational Research: An Introduction. Recording for the Blind & Dyslexic*. Boston: Allyn & Bacon, 2003.
- Hamalik, O. *Evaluasi Kurikulum*. Bandung: PT. Remaja Rosdakarya, 1993.
- Hula, I.R.N. *Evaluasi Dan Tes Bahasa Arab: Tinjauan Teori Abstrak*. Language Development Center, 2021.
- Isbah, Faliqul, M. Fairuz Rosyid, and Nilna Uswah Tamama. "Evaluation Implementation of the Arabic Language Intensification Program (PIBA) in Islamic Higher Education." *Arabiyatuna: Jurnal Bahasa Arab* 8, no. 2

- (October 22, 2024): 731–56. <https://doi.org/10.29240/jba.v8i2.11129>.
- Kholis, Nur, and M. Arif Mustofa. “Development of Competency-Based Arabic Language Curriculum in Traditional Islamic Boarding Schools.” *Arabiyatuna: Jurnal Bahasa Arab* 8, no. 2 (November 1, 2024): 827–48. <https://doi.org/10.29240/jba.v8i2.10819>.
- Maulana, Asep, Universitas Islam, Negeri Kiai, Haji Achmad, and Siddiq Jember. “دَدَعْتَلْمَا مَلْعَتَلَا لِنَاسُو رِيوُطْتَ بِلَاظِلَا ءَافَلَكْ ةَيَقْرُ تَلْ لَجُوجْ عَقَاوُمْ بَلْعْ ءَانِبْ تَادِرْقَلْمَا ةَمَجْرَتْ يَفْ مَدَقْتَلَا” هَمْعَدِي يَذَلَا ، 4. 0 ةَعَانِصِلَا رِصْعْ يَفْ يَمَقْرَلَا لُوْحَتَلَا ةَرُورُضْ يَفْ بِلَاظِلَا بَدَلْ سَامَحَلَا ةَلْقُو تَايِدَحَتَلَا . تَنَرْتَنَلَا هَمْعَدِي يَذَلَا ، 4. 0 ةَعَانِصِلَا رِصْعْ يَفْ يَمَقْرَلَا لُوْحَتَلَا ةَرُورُضْ يَفْ بِلَاظِلَا بَدَلْ سَامَحَلَا ةَلْقُو تَايِدَحَتَلَا . تَنَرْتَنَلَا 8 ” no. 1 (2024): 115–46.
- Muhammad. *Memahami Struktur Dan Tata Bahasa Toefl*. Yogyakarta: LINGUA Publishing, 2008.
- Musa, Mofeed Ahmed Abu, and Jihad Ali Al Momani. “University Students’ Attitudes towards Using the Nearpod Application in Distance Learning.” *Journal of Education and E-Learning Research* 9, no. 2 (2022): 110–18. <https://doi.org/10.20448/jeelr.v9i2.4030>.
- Pangestika, Eliana, R. Umi Baroroh, and Salma Lulu Maknuna. “Innovation Discrete Approach Based on Arabic Language Test Assessment.” *Arabiyatuna: Jurnal Bahasa Arab* 8, no. 1 (May 27, 2024): 395. <https://doi.org/10.29240/jba.v8i1.8992>.
- Pateda, M. *Linguistik Terapan*. Flores: Nusa Indah, 1991.
- Pey, Khor Ching, Low Hui Min, and Lee Lay Wah. “Relationship between Oral Reading Fluency and Reading Comprehension among ESL Students.” *GEMA Online Journal of Language Studies* 14, no. 3 (2014): 19–32. <https://doi.org/10.17576/gema-2014-1403-02>.
- Pringgawidagda, S. *Strategi Penguasaan Berbahasa*. Yogyakarta: Adicpta Karya Nusa, 2002.
- Ridlo, Ubaid. “Multiple Intelligences in Evaluation of Arabic Learning at Islamic Boarding School.” *Arabiyatuna : Jurnal Bahasa Arab* 6, no. 1 (April 11, 2022): 125. <https://doi.org/10.29240/jba.v6i1.3798>.
- Sudarmono, E. *The King TOEFL : Tips & Trik Plus Bank Soal*. Wahyumedia, 2015.
- Sulasmianti, Nova. “Pembelajaran Berbasis Web Dengan Google Sites.” *Cendekia Jurnal Pendidikan Dan Pembelajaran*, February 2022, 1–66.
- Tarigan, H.G. *Pengajaran Kompetensi Bahasa*. Bandung: Angkasa, 1990.
- Woan, Lim Hui, and Lee Soon Tat. “Assessing Children’s Native Language in Mandarin Using the Adapted New Reynell Developmental Language Scales-Mandarin (NRDLS-M).” *GEMA Online Journal of Language Studies* 17, no. 2 (2017): 123–45. <https://doi.org/10.17576/gema-2017-1702-08>.
- Zaidi, Shafqat, Fariha Yasmin, Vahid Nimehchisalem, Zalina Mohammad Kasim, and Afida Mohammad Ali. “Grammatical Knowledge of English Cleft Constructions among Pakistani ESL Learners across L2 Proficiency Levels and Learning Styles.” *GEMA Online Journal of Language Studies* 23, no. 3 (2023): 76–97. <https://doi.org/10.17576/gema-2023-2303-05>.