PAPER NAME

6721-19541-1-SM.doc

WORD COUNT

4959 Words

CHARACTER COUNT

29300 Characters

PAGE COUNT

14 Pages

FILE SIZE

628.5KB

SUBMISSION DATE

Oct 10, 2023 8:10 AM GMT+7

REPORT DATE

Oct 10, 2023 8:11 AM GMT+7

● **21% Overall Similarity**

The combined total of all matches, including overlapping sources, for each database.

- 17% Internet database
- Crossref database
- 15% Submitted Works database

- 12% Publications database
- Crossref Posted Content database

● **Excluded from Similarity Report**

- Bibliographic material

- Quoted material

Development of an Arabic Language Receptive Proficiency Test Instrument Based on the Common European Framework of Reference for Languages

Erfan Gazali[1], Hasan Saefulloh[2]
[1]IAIN Syekh Nurjati Cirebon, Indonesia
[2]IAIN Syekh Nurjati Cirebon, Indonesia
E-mail: erfangazali@syekhnurjati.ac.id

Abstract

The purpose of this study is to develop a receptive proficiency test instrument (listening and reading skills) for the Institut Agama Islam Negeri Sheikh Nurjati Cirebon, Indonesia. The preparation of the examination was guided by the Common European Framework of Reference for Languages (CEFR) level B1. Research and development (R&D) based on the ADDIE research model, with five development stages including analysis, design, development, implementation, and evaluation, was employed. The instrument was designed as a 100-question multiple-choice exam with phases for validity and reliability tests, item difficulty tests, and item differentiation indices. Validity of test items is determined with Lawshe's formula, reliability coefficient with the Kuder-Richardson Method (KR20) formula, and differentiation index with the Ferguson formula. On the basis of the trial results, it was determined that (1) there were 94 valid items and the remaining 6 items were revised with a content validity index of 0.97. (3) The reliability coefficient for the high category was 0.94; (2) the level of difficulty of the items was divided into difficult categories as many as 6 questions, moderate 51, easy 41, and very easy 1 item; and (4) the index of the differentiating power of the items, as many as 3 questions were categorized as excellent, 67 as very good, and 11 as sufficient, while the items categorized as bad were 12 and very bad 7 questions. Thus, the devised Arabic language proficiency test instrument is reliable and practicable for measuring Arabic language proficiency at level B1 CEFR.

Keywords: *Arabic proficiency, CEFR, receptive skill, Arabic*

Introduction

Assessment of a language's proficiency has been around as long as teaching itself. (Farhady, 2018) Both Black and Williams(Black & Wiliam, 2010) and Kennedy, Chan, Fok, & Yu, (2008) found that standardised assessment and evaluation boosted student knowledge retention and comprehension. By highlighting areas of strength and improvement, it can aid students in their education. David Boud argues that doing assessments helps students learn and enhance their abilities. (Boud, 1990)

Language testing procedures are fundamentally distinct from those employed in the vast majority of other disciplines. This is due to the fact that teachers of foreign languages have a broad range of assessment tools from which to choose for their students.(Brown & Hudson, 1998) Experts refer to the ability to comprehend, speak, read, and write proficiently in a given language as "language proficiency"(Bachman, 2000; Richards & Schmidt, 2010). These four abilities have been identified by educators in the field of linguistics as essential for language learners. Receptive skills include reading and listening, whereas productive skills include speaking and writing. (Laufer & Goldstein, 2004; Masduqi, 2016; Sreena & Ilankumaran, 2018)

Awamleh (2004) proposes testing grammar, literary style, and cultural sensitivity in addition to the four basic competencies. Language, culture, and education should inform knowledge assessments. Brown & Hudson (1998) identify three main types of language proficiency tests: (a) selected-response tests, which include true-false, matching, and multiple-choice tests; (b) constructed-response tests, which require language learners to write, speak, or do something. (c) Personal response tests, fill-in-the-blank, short-answer, and performance evaluations These exams examine students' abilities during the learning process, depending on their engagement. Self- and peer-evaluation, portfolios, and debates.

In 1978, Roushdy Ahmad Toiemah (1978) developed a standardised test to measure language proficiency in Arabic for foreign speakers among students studying Arabic in several American universities. This was one of the first studies to measure language proficiency in Arabic for foreign speakers. Consequently, research pertaining to the standardisation of language proficiency examinations emerged and became the primary focus of Arabic linguists' research. (Ben Khiroun et al., 2014; de Graaf, 2021; Masrai & Milton, 2019; Rifaie, Hamza, & Elfiky, 2021; Winke & Aquil, 2014)

In the 2000s, a lot of attention was paid to study on computer and internet-based tests for learning foreign languages.(Alderson, 2009; Douglas & Hegelheimer, 2007; Hosseini, Abidin, & Baghdarnia, 2014; McCauley & Christiansen, 2017; Roever, 2001; Syihabuddin, Abdurrahman, Akmalia, & Abdussalam, 2021) Arabic language proficiency tests have been developed globally by many institutions in several countries with various test names, including the Arabic Language Proficiency Test (ALPT) developed by Arab academics based in Cairo, the Standardised Arabic Test (SAT) developed by the Saudi Arabia Electronic University in Riyadh, and the Testing of Arabic as a Foreign Language (TOAFL) developed by the Al-Arabiyya-Institute in Germany.

In the Indonesian context, Syihabuddin, Abdurrahman, Akmalia, and Abdussalam noted that several universities have administered multiple Arabic language proficiency tests. For example, UIN Sunan Kalijaga Yogyakarta developed *Ikhtibar Kafaah al-Lughah 'al-'Arabiyyah* (IKLA), UIN Jakarta developed the Test of Arabic as a Foreign Language (TOAFL), IAIN Salatiga developed *Ikhtibarat Lughat al-Arabiyyah ka Lughat al-Ajnaby* (ILAiK), IAIN Surakarta developed the Test Of Standard Arabic (TOSA), and other institutions also developed proficiency tests (Syihabuddin et al., 2021). How then is the Arabic exam administered at IAIN Sheikh Nurjati?

Observations regarding the implementation of the Arabic language proficiency exam at the language development centre IAIN Syekh Nurjati Cirebon, Indonesia, indicate that Arabic proficiency is not measured based on language ability level. The

measurement of language proficiency is limited to aspects of listening, reading, and grammar, but does not include aspects of writing or speaking. The measurement of Arabic language proficiency at IAIN Sheikh Nurjati is still conducted using pencil and paper, with only local measurement standards developed by the IAIN Sheikh Nurjati Language Development Centre (LDC) and not on representative language competency standards from international institutions such as the Common European Framework of Reference for Languages (CEFR) or the American Council on the Teaching of Foreign Languages (ACTFL).

The novelty of devising an Arabic language proficiency measurement instrument at IAIN Sheikh Nurjati from a pencil-and-paper test to an internet-based test (iBT) in accordance with CEFR standards makes this research significant. This measurement standard will enhance the Arabic language proficiency of students at IAIN Sheikh Nurjati and ensure the quality of graduates. This study's language proficiency testing model employs the explicit skills and elements approach for language testing outlined by Lado (Wess, Klock, Siller, & Greefrath, 2021) and Carrol (1961) There is a clear distinction in their formulations between skills and abilities on the one hand and test approaches, methods, or types on the other. The skill component includes listening comprehension, oral productivity, reading, and writing, while the elements include pronunciation, grammatical structure, vocabulary, and cultural significance.

Method

The research and development method is used in this study. The ADDIE study model is used for research and development (R&D). The ADDIE model is a way to plan how to build something. It has five stages: analysis, design, development, implementation, and evaluation (Branch, 2009). The goal of development study is to make things and test how well they work. The stages and procedures for development in this investigation are detailed in Table 1.

Table 1 Research Stages and Procedures

ADDIE STAGE	NO.	PROCEDURE
ANALYSIS	1	Arabic language proficiency test needs analysis
	2	Feasibility study of Arabic language proficiency test
DESIGN	3	Designing Arabic language test instruments
	4	Develop test instrument standards
	5	Develop test management guidelines
	6	Develop an outline of the test content
	7	Developing the design (JM/Flowchart/Storyboard) of the Arabic proficiency test
	8	Compile the script of the instrument to be developed
DEVELOPMENT	9	Conducting script development activities
	10	Design/lead/supervise the test instrument production process
	11	Conducting preview activities of test instrument production results

	12	Conducting test prototype tests
	13	Drafting the script of test accompanying materials
	14	Conducting a feasibility study on the utilization of the Arabic language proficiency test
IMPLEMENTATION	15	Conducting a pilot utilization of the Arabic language proficiency test
	16	Conducting orientation on the utilization of test instruments
	17	Conducting guidance in the utilization of Arabic proficiency tests
	18	Performing consultation services for the utilization of Arabic proficiency tests
	19	To facilitate the utilization of the Arabic proficiency test
	20	Conducting socialization of the utilization of the Arabic proficiency test
	21	Monitoring the utilization of test instruments
	22	Developing an evaluation design for the utilization of Arabic language proficiency tests
EVALUATION	23	Developing an evaluation instrument for the utilization of the Arabic language proficiency test
	24	Evaluating the utilization of the Arabic language proficiency test

Objectivity, reliability, and validity are the three primary indicators, also known as core quality criteria, that have become the standard for evaluating the quality of a test instrument (Ebel & Frisbie, 1991; Linn, 2011; Miller, Linn, & Gronlund, 2012). Objectivity is required for accurate measurement, and accurate measurement is required for instrument validity. (Wess et al., 2021)

In the development research, the product trial design comprises: 1) test design, 2) test subjects, 3) data collection techniques and instruments, and 4) data analysis techniques. The test design that will be developed through research on the development of Arabic language proficiency test instruments involves expert validation and small-scale trials. The subjects of the expert test consisted of experts or experts, namely Arabic grammarians, Arabic learning evaluation experts, and lecturers teaching Arabic language skills.

As for the data analysis technique on the instrument using test item validity, which uses four types of tests. First, the content validity ratio (CVR) and content validity index (CVI) with the formula proposed by Lawshe (1975) in the form of three rating scales, namely: (1) essential, (2) useful but not essential, and (3) unnecessary; second, the reliability test with reference to the Kuder and Richardson formula number 20. Third, the test measures the level of difficulty of the items using the proportion formula (difficulty index) and the fourth is the item discrimination test with the discrimination index.

Result and Discussion

This study led to the creation of a question for an Arabic language competency test. With a total of 100 items, the tests are made up of multiple-choice questions with three to four possible answers (a, b, c, and d). The steps of growth in this study are about how the ADDIE model was made and how it works. Analysis, design, development, Implementation, and evaluation are the five steps in the ADDIE model.

1. Analysis

    In the analysis, there are two stages: needs assessment and front-end analysis. Needs assessment in the form of analysis of the state of the field and participants as well as the collection of reference tests that will be used as the subject matter in the development of test instruments. Field analysis activities were carried out by collecting information about the needs of Arabic language proficiency assessment at IAIN Sheikh Nurjati Cirebon. The results of the information related to the needs of the Arabic language proficiency test, namely:

    a. Language Development Center (LDC) IAIN Sheikh Nurjati Cirebon does not yet have a comprehensive instrument for measuring Arabic language skills that can be used to assess the four competencies (listening, reading, writing, and speaking skills) of Arabic language students in the IAIN Sheikh Nurjati Cirebon environment. In the meantime, the LDC only possesses a test instrument for Arabic listening and reading skills, but not for writing and speaking skills. At the level of implementation, the test instrument is manual, consisting of a pencil-based answer page and audio media to assist in listening to questions pertaining to aspects of listening skills. The Department of Arabic Language and Literature and the Department of Arabic Language Education, on the other hand, do not yet have test instruments to measure the Arabic language abilities of their students.

    The LDC's instrument is still generic and does not distinguish between levels of language proficiency, such as novice, intermediate, and advanced. The total score is comprised of three assessment components: structure and grammar skills, auditory skills, and text comprehension skills. The scoring figures are based on the TOEFL evaluation. The instrument is a modification of the questions that have been accumulated since 2005 by adjusting a subset of questions as a test variation and not in response to user requirements.

    b. There is a need for instruments founded on generally accepted standards for measuring language proficiency levels. Examples include the American Council on the Teaching of Foreign Languages (ACTFL) proficiency guidelines, the Canadian Language Benchmarks (CLB), the Interagency Language Roundtable (ILR) assessments, and the Common European Framework of Reference for Languages, or CEFR for short. This study's measurement standard is the Common European Framework of Reference for Languages (CEFR), which is a language standard established by the European Council in 2001 and serves as a guide for describing students' foreign language skills in Europe, particularly in academic settings. There

are six levels of language standards: A1, A2, B1, B2, C1, and C2. (Council of Europe, 2001)

c.   The CEFR proficiency levels, which range from basic to professional, require many test instruments that can fulfill each level. Therefore, in this study, the development of test instruments focuses on level B1, or the intermediate level, with aspects of measurement on Arabic receptive skills, namely listening skills and understanding texts, which include mastery of grammar and vocabularies.

2. Planning and design of test instruments

The stage of designing receptive Arabic language proficiency test instruments includes setting test objectives, creating test grids, and creating questions.

a. Setting Test Objectives

The test objectives are designed to evaluate auditory and reading comprehension of Arabic. The test contains questions based on indicators derived from the Common European Framework of Reference for Languages (CEFR) at the B1 (intermediate) or independent user proficiency level. The equivalence levels between level B1 and other foreign language proficiency standards are as follows:

Table 2 Degree of equivalence of CEFR Level B1 with similar English language measurement standards (Efset.org, 2021)

Test Type	Score equivalent to B1 Level
IELTS	4.0 – 5.0
TOEIC	550-780
TOEFL iBT	42-71

b. Making question grids

The researcher began compiling a grid of test questions after determining the test's purpose. Arabic receptive skills at level B1 contain the following linguistic competencies:

Table 3 Receptive Proficiency at Level B1-CEFR

level	Aspect	Description of competence
B1	Listening	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken.
	Reading	Able to understand familiar text or work-related vocabulary. Can understand descriptions of feelings or desires.

The measured aspects of linguistic competence are translated into four indicators, which are then reduced to ten topics that will be developed into one hundred questions (see Table 7).

c. Create a question

After identifying the developed indicators and topics for assessing listening and writing abilities, 100 multiple-choice questions with four possible answers are compiled. The questions are separated into three categories. Listening (30

questions), comprehensive reading (40 questions), and grammar and vocabulary (30 questions) are the specific sections of the exam.

3. Development

At this stage, errors or deficiencies are rectified through the validation by experts of the queries that have been developed. The aim is to measure whether or not the question items achieve the objectives set. The validators consisted of eight experts, including Arabic linguists, language learning evaluation professionals, and Arabic language instructors.

a. Test and revise instruments

In order to ascertain the level of validity and reliability of the instrument, eight panelists evaluate the question items that have been developed based on these indicators. Improvements will be made based on the panelists' comments and suggestions, and then the questions will be tested with a small group online to determine the items' degree of difficulty and their ability to differentiate.

The 100 question items were submitted for validation to eight panelists or experts, who evaluated the content's viability. Using the CVI formulation (Lawshe, 1975), the content validity index of the receptive Arabic language proficiency test was calculated after determining the CVR index for each item of the assessment instrument. As shown in Table 6, the result of the CVI calculation is essentially the mean CVR for all query items. Based on the evaluations of eight experts, the item-CVI index of the test instrument is calculated to be 96.8 and the average scale value of the CVI score (S-CVI) of all items is 0.96 (see Table 4).

Table 4 Content validity index scale and index scale results

Sum of I-CVI	96.88	Sum of UA	86
S-CVI/Ave	**0.97**	S-CVI/UA	**0.86**
Category	**accepted**		**accepted**

The CVR value derived from the calculation was compared to the CVR critical value determined by the number of validators enumerated in Table 5. The item is accepted if its value is equal to or higher than the CVR critical value, and it is rejected if its value is less than the CVR critical value.(Ayre & Scally, 2014; Wilson, Pan, & Schumsky, 2012).

Table 5 *Simplified Table of CVR critical Including the Number of Experts Required to Agree an Item*

Panel Size	$N_{critical}$ (Minimum Number of Experts Required to Agree an Item Essential for Inclusion)	Proportion Agreeing Essential	$CVR_{critical}$
5	5	1	1.00
6	6	1	1.00
7	7	1	1.00
8	7	.875	.750
9	8	.889	.778
10	9	.900	.800
11	9	.818	.636
12	10	.833	.667
13	10	.769	.538
14	11	.786	.571
15	12	.800	.600

Table 5 Competencies and performance indicators for listening and reading (adapted and processed from the Council of Europe, 2001).

No.	competence	Indicator	Question theme	Number of questions
1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken.	1. Can understand the main points of clear standardized input on matters commonly encountered in work, at school, at recreation	1. Personal and professional hopes and dreams for the future	10
2	Able to understand familiar text or work-related vocabulary. Can understand descriptions of feelings or desires.	2. Can handle most situations that are most likely to occur while traveling in the area where the language is spoken.	2. Watching television and favorite shows.	10
		3. Capable of producing straightforward, connected texts on familiar or personal topics.	3. Education and future plans	
		4. Possesses the ability to describe experiences and events, as well as dreams, hopes, and aspirations, and to concisely explain opinions and plans.	4. Favorite music, music or entertainment trends and plans to go to Arabic drama performances.	10
			5. Healthy practices, diet, and lifestyle, as well as giving and receiving advice in these areas.	10
			6. Meet people on social networks.	10
			7. Going to a restaurant, ordering food, having supper conversation, and paying for dinner.	10
			8. Participate in negotiations	10
			9. Safety concerns, accident reporting, and the explanation of regulations.	10
			10. Polite behavior and respond appropriately to impolite behavior.	10

This result can be interpreted as indicating that the devised test instrument has a very high content validity index (CVI) based on the 7-statement assessment instrument used by experts.

The next stage is to revise the questions that need improvement based on the notes and feedback from the experts. Then, it goes to the stage of assembling questions for limited trials in a learning management system based on MOODLE (Modular Object-Oriented Dynamic Learning Environment) version 4.1 by dividing the questions into three groups of questions: grammar, listening, and reading comprehension. The following is a screenshot of the uploaded Arabic exam on the Moodle platform:
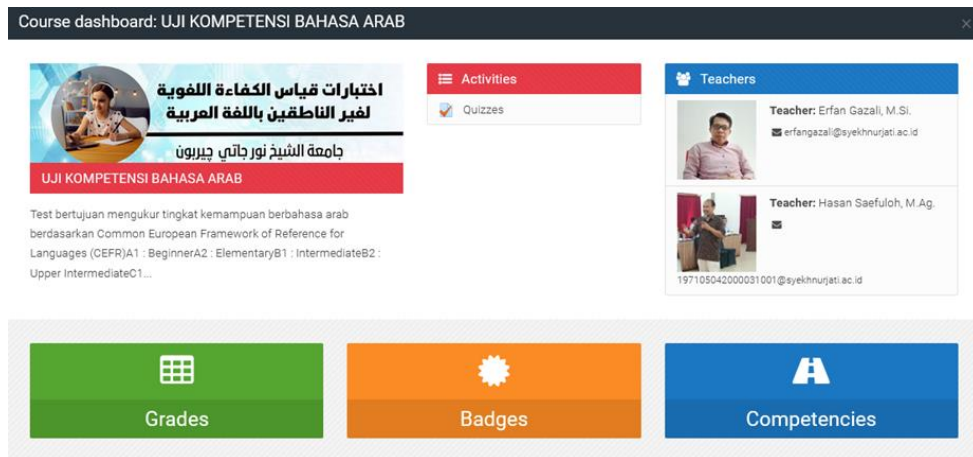
Figure 1 Screenshot of the dashboard for the Moodle-based online Arabic Proficiency Test
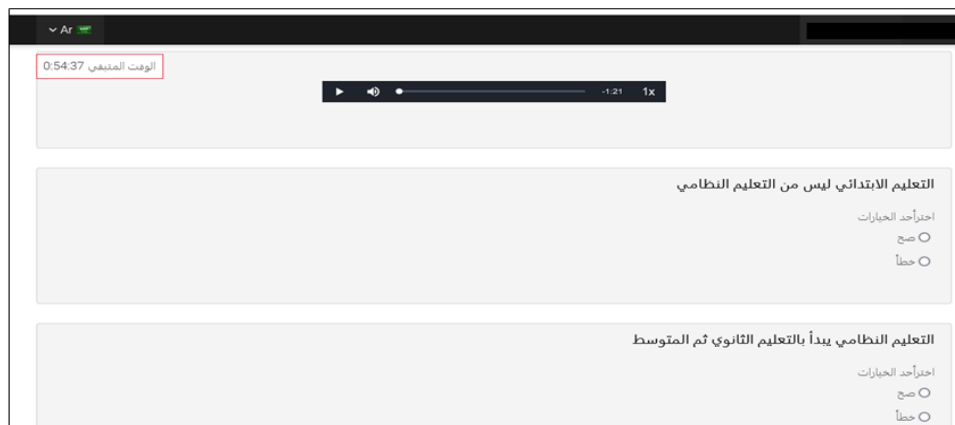
Figure 2 Screenshot of the reading comprehension session question (*fahmu al-maqru*)

Figure 3 Screenshot of Arabic vocabulary and reading comprehension questions

4. Implementation

   a. Small-Group Testing

The test prototype that has been integrated into the Moodle software was then tested online with a small group of 15 students from the Department of Arabic Language and Literature, IAIN Sheikh Nurjati Cirebon. The data from this trial are used to calculate the level of question item reliability, the level of question difficulty and differentiation, and to assess the product's usability. Online assessments allow students to conduct product evaluations. After the student completed the online test, the researcher distributed a questionnaire to the respondents to evaluate the receptive Arabic language proficiency measurement test instrument based on the application's usability, sound and image quality, and question presentation.

The data analysis technique used to estimate the reliability of the instrument used the internal consistency estimation technique with Kuder and Richardson formula number 20, and the reliability score for multiple-choice questions obtained was 0.95, which indicates the level of reliability is in the very high range.

Table 6 Reliability of multiple-choice items on Arabic receptive proficiency

n/n-1	(St2-∑PQ)St2	r
1.01	0.94	0.95

According to Table 6, the difficulty level of the question items does not have a balanced proportion of questions, as the number of medium category questions has a higher proportion of 51 questions (51%). The lower difficult category amounted to 6 questions (6%), and for the easy category as many as 42 questions (42%), while very easy was 1 question (1%) and questions with a very difficult category did not exist.

The distinguishing power of an item depends on the size of the discrimination index value. According to Table 7, the average distractor of 100 questions is classified into five groups. There are 3 questions (3% in the excellent category), 67 (67% in the good category), 11 (11% in the sufficient category), 12 (11% in the terrible category), and 7 (7% in the very bad category).

Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya Vol.X, No. X |X-X
Al-Ta'rib | p-ISSN 2354-5887 | e-ISSN 2655-5867

2

The average distraction of 100 queries falls into five categories. Specifically, the class of excellent queries There are 3 questions (3% of the total) with an average difficulty index of 1, the good category has 67 questions (67% of the total) with an average difficulty index of 0.78, the fair category has 11 questions (11%) with an average difficulty index of 0.60, the bad category has 12 questions (11%) with an average difficulty index of 0, and the very bad category has 7 questions (7% of the total) with an average difficulty index of -0.22.

Table 7 Item difficulty of multiple-choice questions on Arabic receptive proficiency

N-of Item	Item Number	Average ($\overline{x}$) difficulty index	Category	Percentage (%)
1	37	1	Very easy	1
42	2,4,5,6,7,15,19,20,21,23,25,29,31,32,33,36,41,44,50,51,52,54,58,59,61,63,65,67,69,72,76,78,80,81,82,83,87,90,91,92,95,96	0,78	Easy	42
51	1,3,10,12,13,14,16,17,18,21,24,26,27,28,30,34,35,38,39,40,42,43,45,46,47,48,49,53,55,57,60,62,64,66,70.71,73,74,75,77,798 5,86,88,89,93,94,97,98,99,100	0,60	moderate	51
6	8,9,11,56,68,84	0.28	Difficult	6
0	-	0	Very difficult	0

According to the results of the evaluation of students' responses to the test instrument product, 74.4% (11) of respondents rated the level of ease in operating the test application as "Good," while 86.7% (13) of respondents rated the aspect of sound and images presented as "Very Good," and the aspect of appearance or layout in the presentation of questions and answers received the category "Very Good" with 12 responses, or 80% ot the total.

5. Evaluation

Product evaluation follows validation and testing. As a small-group trial revision, the product is reviewed. Field evaluations must be addressed if shortcomings are detected. Product evaluation results are approved, changed, or rejected. Accepted question items have an essential or valid correlation, high dependability, moderate difficulty, and are good or very good. If the correlation is valid, reliability is low, difficulty is low, and discriminatory power is low The item can be approved with corrections. If the correlation is invalid or non-essential, reliability is high, difficulty level is low, and discriminating power is poor or extremely poor, the question must be eliminated due to its poor quality. From the summary of the questions above, it can be concluded that 97 questions are used without revision, 3 questions (numbers 15, 20, and 23) must be revised, and questions that must be discarded due to poor quality cannot be found.es.

Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya Vol.X, No. X |X-X
Al-Ta'rib | p-ISSN 2354-5887 | e-ISSN 2655-5867

3

Conclusion

The final result of this development research is an instrument for assessing Arabic language proficiency based on the Common European Framework of Reference for Languages (CEFR). The proficiency test instrument must satisfy a limited trial before becoming the final product. However, prior to testing, the product has been validated by eight experts in the disciplines of Arabic grammar, Arabic language skills, and evaluation of Arabic language learning. Expert input was then used to refine the product based on the results of the validation and limited trial. In terms of validity, reliability, level of difficulty, and differentiating power, the final product meets the criteria for quality items.

The development of this Arabic language proficiency test instrument has limitations, including: (1) the development of multiple-choice test instruments is limited to receptive competencies of Arabic language skills, namely listening and reading skills, as well as grammar and vocabulary. (2) Product development trials have not reached the stage of large-scale field trials and product testing, and (3) the number of test items has not been altered to interpret the meaning of each indicator theme.

Reference

Alderson, J. C. (2009). Test of English as a Foreign Language™: Internet-based Test (TOEFL iBT®). *Language Testing*, *26*(4), 621–631. https://doi.org/10.1177/0265532209346371

Awamleh, H. (2004). *Mahārāt Taʼlīm al-Qirāʼaẗ wa al-Kitābaẗ Lilāṭfāl*. Amman: Dār Wāʼil lilṭbāʼaẗ Wa al-Nšar Wa-alTaūzīʼ.

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, *47*(1), 79–86. https://doi.org/10.1177/0748175613513808

Bachman, L. F. (2000). Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing*, *17*(1), 1–42. https://doi.org/10.1191/026553200675041464

Ben Khiroun, O., Ayed, R., Elayeb, B., Bounhas, I., Ben Saoud, N. B., & Evrard, F. (2014). Towards a new standard arabic test collection for mono- and cross-language information retrieval. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8455 LNCS*, 168–171. RIADI Research Laboratory, ENSI Manouba University, 2010, Tunisia: Springer Verlag. https://doi.org/10.1007/978-3-319-07983-7_23

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *92*(1), 81–90. https://doi.org/10.1177/003172171009200119

Boud, D. (1990). Studies in Higher Education academic values Assessment and the Promotion of Academic Values. *Studies in Higher Education*, *15*(1), 101–111. https://doi.org/http://dx.doi.org/10.1080/03075079012331377621

Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya Vol.X, No. X |X-X
Al-Taʼrib | p-ISSN 2354-5887 | e-ISSN 2655-5867

4

Branch, R. M. (2009). *Instructional Design: The ADDIE Approach*. New York-Dordrecht-Heidelberg-London: Springer.

Brown, J. D., & Hudson, T. (1998). Alternatives in Language assessment. *TESOL Quarterly*, *32*(4), 653–675.

Carrol, J. B. (1961). Fundamental considerations in testing English proficiency of foreign students. In *Testing the English Proficiency of Foreign Students* (pp. 30–40). Washington D.C: Center for Applied Linguistics.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR)*. Strasbourg: Language Policy Programme Education Policy Division Education Department Council of Europe. Retrieved from www.coe.int/lang-cefr

de Graaf, A. (2021). Challenges in Developing Standardized Tests for Arabic Reading Comprehension for Secondary Education in the Netherlands. In *Fairness in College Entrance Exams in Japan and the Planned Use of External Tests in English* (pp. 181–189). educational assessment at Cito, Netherlands: Springer Singapore. https://doi.org/10.1007/978-981-33-4232-3_13

Douglas, D., & Hegelheimer, V. (2007). Assessing language using computer technology. *Annual Review of Applied Linguistics*, *27*(2007), 115–132. https://doi.org/10.1017/S0267190508070062

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs ,NJ: Prentice Hall.

Farhady, H. (2018). History of Language Testing and Assessment. *The TESOL Encyclopedia of English Language Teaching*, 1–7. https://doi.org/10.1002/9781118784235.eelt0343

Hosseini, M., Abidin, M. J. Z., & Baghdarnia, M. (2014). Comparability of Test Results of Computer based Tests (CBT) and Paper and Pencil Tests (PPT) among English Language Learners in Iran. *Procedia - Social and Behavioral Sciences*, *98*, 659–667. https://doi.org/10.1016/j.sbspro.2014.03.465

Kennedy, K. J., Chan, J. K. S., Fok, P. K., & Yu, W. M. (2008). Forms of assessment and their potential for enhancing learning: Conceptual and cultural issues. *Educational Research for Policy and Practice*, *7*(3), 197–207. https://doi.org/10.1007/s10671-008-9052-3

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399–436. https://doi.org/10.1111/j.0023-8333.2004.00260.x

Linn, R. L. (2011). The Standards for Educational and Psychological Testing: Guidance in Test Development. In T. M. Haladyna & S. M. Downing (Eds.), *Guidance in Test Development* (pp. 27–38). Mahwah, N.J: L. Erlbaum.

Masduqi, H. (2016). Integrating Receptive Skills and Productive Skills into a Reading Lesson. *The 2nd International Conference on Teacher Training and Education*, *2*(1), 507–511. Semarang: Sebelas Maret University. Retrieved

Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya Vol.X, No. X |X-X
Al-Ta'rib | p-ISSN 2354-5887 | e-ISSN 2655-5867

5

from https://jurnal.uns.ac.id/ictte/article/view/7476

Masrai, A., & Milton, J. (2019). How many words do you need to speak Arabic? An Arabic vocabulary size test. *Language Learning Journal*, *47*(5), 519–536. https://doi.org/10.1080/09571736.2016.1258720

McCauley, S. M., & Christiansen, M. H. (2017). Computational Investigations of Multiword Chunks in Language Learning. *Topics in Cognitive Science*, *9*(3), 637–652. https://doi.org/10.1111/tops.12258

Miller, D., Linn, R., & Gronlund, N. (2012). *Measurement and Assessment in Teaching* (11th ed.). Upper Saddle River, NJ: Pearson Education Inc.

Richards, J. S., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics* (4th ed.). London: Longman (Pearson Education).

Rifaie, N., Hamza, T. M. A. W., & Elfiky, Y. H. (2021). Standardization of the Revised Arabic Language test for 4-8-year old children. *Egyptian Journal of Ear, Nose, Throat and Allied Sciences*, *22*(22). https://doi.org/10.21608/EJENTAS.2021.92770.1414

Roever, C. (2001). Web-based language testing. *Language Learning and Technology*, *5*(2), 84–94.

Sreena, S., & Ilankumaran, M. (2018). Developing Productive Skills Through Receptive Skills – A Cognitive Approach. *International Journal of Engineering & Technology*, *7*(4.36), 669. https://doi.org/10.14419/ijet.v7i4.36.24220

Syihabuddin, Abdurrahman, M., Akmalia, F., & Abdussalam, A. (2021). Wikalah iBT : How Is An Internet-Based Arabic Language Proficiency Test Instrument Developed ? *Ijaz Arabi*, *4*(3), 535–547. https://doi.org/10.18860/ijazarabi.v4i3.12198

Toiemah, R. A. (1978). *The Use of Cloze to Measure the Proficiency of Students of Arabic: As a Second Language in Some Universities in the United States*. University of Minnesota, Minneapolis.

Wess, R., Klock, H., Siller, H.-S., & Greefrath, G. (2021). *Measuring Professional Competence for the Teaching of Mathematical Modelling*. Switzerland AG: Springer Nature.

Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the Critical Values for Lawshe's Content Validity Ratio. *Measurement and Evaluation in Counseling and Development*, *45*(3), 197–210.

Winke, P. M., & Aquil, R. (2014). Issues in developing standardized tests of arabic language proficiency. In *Handbook for Arabic Language Teaching Professionals in the 21st Century* (pp. 221–233). Department of Linguistics and Germanic Slavic, Asian, and African Languages, Michigan State University, United States: Taylor and Francis. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85118374254&partnerID=40&md5=473574df6fcdf60bd5b605dc60ee2be4

Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya Vol.X, No. X |X-X
Al-Ta'rib | p-ISSN 2354-5887 | e-ISSN 2655-5867

6

● **21% Overall Similarity**

Top sources found in the following databases:

- 17% Internet database
- Crossref database
- 15% Submitted Works database

- 12% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	**ejournal.uin-malang.ac.id** Internet	2%
2	**UIN Maulana Malik Ibrahim Malang on 2020-09-30** Submitted works	2%
3	**ejournal.umm.ac.id** Internet	<1%
4	**Li Wang, Jason Fan. "Working Towards a Proficiency Scale of Business...** Crossref	<1%
5	**educationnz.org.nz** Internet	<1%
6	**Universitas Negeri Surabaya The State University of Surabaya on 2023-...** Submitted works	<1%
7	**Florida Institute of Technology on 2022-04-19** Submitted works	<1%
8	**yumpu.com** Internet	<1%

Sources overview

57 **University of Birmingham on 2020-09-22**

<1%