DESIGN AN ARABIC LANGUAGE PROFICIENCY TEST FOR NON-NATIVE SPEAKERS IN LIGHT OF THE STANDARDS OF THE CEFR USING THE ISPRING SUITE PROGRAM

Halomoan [1]*, Kasmiati [2], Afriza [3], Rini Rahman [4], Indah Muliati [5], Rahmi Wiza [6] and Fakhriyah Annisa Afroo [7]

[1,4,5,6,7] Universitas Negeri Padang, Indonesia.
[2,3] Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia.
Email: [1]halomoan@fis.unp.ac.id (*Corresponding Author), [2]kasmiati@uin-suska.ac.id,
[3]afriza@uin-suska.ac.id, [4]rinirahman@fis.unp.ac.id, [5]indahmuliati@fis.unp.ac.id,
[6]rahmiwiza@fis.unp.ac.id, [7]fakhriyahannisa@fis.unp.ac.id

Abstract

The Arabic language proficiency test is something that needs to be done and continues to be developed following international standards and on a digital basis. One of the frameworks that is currently an international reference is CEFR (Common European Framework of Reference for Languages). The aim of this research is to find a standardized Arabic language proficiency test design. The research method used is the D & D method with four stages, namely: Define, Design, Develop and Disseminate. The results of this research are in the form of a design framework for an Arabic language ability test based on CEFR (Common European Framework of Reference for Languages) using the Ispring Suite application with designs A1, A2, B1 and B2 with the provisions A1: 40 questions, A2: 40 questions, B1: 30 questions , and B2, 30 questions, a total of 140 questions distributed across three language elements, namely: istima' (listening): 50 questions, Qiraah (reading): 50 questions, and tarakib (language structure): 40 questions.

Keywords: Design, Proficiency Test, Arabic, CEFR.

INTRODUCTION

Evaluation in the educational process has become essential to measure learners' competence. To do this, energy, time, and funding are spent efficiently in the learning and teaching process. The current century imposes challenges in various areas of contemporary life, and the new world order requires human potential that possesses distinct competencies and advanced creative functional skills, enabling it to make a positive and effective contribution to human achievements, economic prosperity, and improving the quality of life of individuals (Salahuddin mahmud alam, 2009). Evaluation is an important component of the curriculum, as it consists of educational objectives, content, teaching methods, and evaluation (al Busyairi & Said, 1415).

The importance of evaluation in the process of teaching the Arabic language is many, including determining the level of students before starting the program so that they can be placed at the appropriate linguistic level, helping students choose the appropriate program for them, and enabling them to continue self-direction in light of what they know about their linguistic level and others (Thuaimah, Rusydi ahmad; manna', 2001). The public who has the opportunity to learn about the results of work at an Arabic teaching institute becomes more convinced of it, more willing to participate in its service, and more willing to join it (Thuaimah, Rusydi ahmad; manna', 2001).

Evaluation has begun to play an essential role in providing accurate information to educational leaders regarding improvement, innovation, and development in the educational system (Mahmud rusydi khothir, 1998). One of the tools for evaluating Arabic language teaching is the test (Thuaimah, 1989).

The intended test is a test of Arabic language proficiency for non-native speakers. The Arabic language proficiency test for non-native speakers does not have a specific standard in Indonesia. The researcher chose the Common European Framework of Reference for Languages (CEFR) as a standard for testing Arabic language proficiency for non-native speakers because all languages in Europe are subject to this frame of reference, including the Arabic language, as well as the necessity of formulating Arabic language proficiency tests based on the established classification (Bahruddin, 2016).

This research was limited to four levels because the C1 and C2 levels are specific to those who live in Arab countries (Bahruddin, 2020). The Common European Framework of Reference for Languages has succeeded in establishing an agreed-upon basis among Europeans in the field of second language teaching. The specifications of a good test are (1) validity (al Khuli, 1982), (2) stability (Mudzaffar, 2009), (3) ease of application, Technology is easy to use and can facilitate the learning process (Syahrul, 2020), demikian juga hal serupa diharapkan dapat terjadi pada proses evaluasi (4) representation, (5) discrimination, and (6) time (al Khuli, 1982), (7) economic, (7) objectivity (9) transparency (Lioyd et al., 2008).

Penggunaan fasilitas teknologi seperti komputer diharapkan dapat meningkatkan motivasi siswa dalam proses pembelajaran.

Testing on paper can harm the quality of the test in terms of time (al Khuli, 1982), and testing using the electronic program iSpring Suite, either on the computer or by phone, reduces costs because the examinees do not need the paper and the university does not need to prepare it. Muhbib Abdul Wahab pointed out that the proficiency test must be developed so that the test is based on the Internet (Wahab, 2022) . Just as learning will be more effective if it is computerized, evaluation will also be more effective using technology (Wulansari et al., 2023).

Computerized or computer-based testing is important because it reduces the burdens of application and procedures represented in monitoring or supervision by making the test available. Also, using the Internet to provide tests is of particular importance to test preparers in terms of the possibility of exploiting technology to improve access to the network (Chapelle & Douglas, 2006). In this research, the researcher used the iSpring Suite program because this program was specifically designed as an evaluation tool and can be used on or offline. Recent research also says that online tools can facilitate evaluation in addition to learning (Zainil et al., 2024).

There are many studies related to this research and it has been found that a good Arabic language proficiency test for the Islamic secondary stage provides the modified Bloom's classification with all its levels, which are (Bahruddin & Qodri, 2020) LOTS, MOTS, HOTS, and that teaching the skill of good listening must take into account the Exercises in listening to Arabic topics or phrases expressed by native speakers (Saputra et al., 2022). The development of educational materials that improved reading skills using the experimental learning model had a significant impact on improving students' ability to read, as the learning outcome for students' reading skills increased.

The research questions are: How to design an Arabic language proficiency test for non-native speakers in light of the Common European Framework of Reference for Languages (CEFR) standard using the iSpring Suite program? How valid is the designed Arabic language proficiency test? This research aims to design a test of Arabic language proficiency for non-native speakers in light of the Common European Framework of Reference for Languages (CEFR) standard using the iSpring Suite program. The researcher wanted to research the development of an Arabic language proficiency test for non-native speakers in light of the Common European Framework of Reference for Languages (CEFR) standard, using the iSpring Suite program.

METHOD

This research used the R&D approach, which is the research method to obtain a specific product and an effective experience (P. D. Sugiyono, 2016). In other words, it is an organized approach to uncover a specific product, model, method, strategy, method, or step, specify, repair, develop, produce, and test its effectiveness to be effective and useful (Borg & Gall, 1984; Putra, 2011). The procedures and steps for research and development are diverse. In this research, the researcher used a model with five steps. This model includes analysis, design, development, application, and evaluation (S. Sugiyono, 2017).

The researcher chose this model to be suitable for developing the test. In the analysis, the researcher analyzed the needs through the interview and analyzed the reality of the Arabic language proficiency test at Sultan Syarif Kasim Riau Islamic State University. In the design, the researcher defined the test objectives, chose the test materials, prepared the test standard, determined the type of test, and designed the Essence Suite, meaning that the researcher sets the question items in light of the Common European Framework of Reference for Languages (CEFR) standard in the iSpring Suite program. In development, the researcher developed the product and tested its validity several times until he obtained a product according to the specified specifications.

The question items were one hundred and forty questions consisting of fifty (50) questions for listening, fifty questions for reading (50), and forty questions for understanding the structures (40). The ratio between easy questions should be 30%, medium questions should be 50%, and difficult questions should be 20%, and the order of the questions should be from easy to difficult. The number of experts for certification is four: an expert in the field of the Arabic language testing subject for non-native speakers, an expert in the field of test design, an expert in the Arabic language, and an expert in educational technology.

The methods of collecting data that the researcher used in this research are the interview, questionnaire, and document, and given that this research is developmental, there is qualitative and quantitative data. The method of analyzing data is according to the doctrine of Miles and Huberman (Robert, 2006), where analysis in this doctrine consists of four activities, which are collecting data, presenting data, classifying data, and concluding(Miles et al., 1992). Quantitative data analysis by questionnaire analysis.

RESULT AND DISCUSSION

The need for standard-based testing and digitization of testing

A good test has certain qualities. The proficiency test in place at Sultan Syarif Kasim Riau State Islamic University needs to be developed, as the director of its language center said.

"The Arabic language proficiency test for non-native speakers at Sultan Syarif Kasim Riau State Islamic University did not have a specific basis, the voice actor in the listening questions was not an Arabic speaker, and the test was not computerized or based on the computer because the university's language laboratory was not usable. I saw that the Arabic language proficiency test at the Language Promotion Center needed to be developed (Syukri, 2021)."

It was clear from the previous data that the proficiency test at the university was not based on a scientific basis and the test is still on paper, hence the test needs to be corrected and changed as necessary. This opinion was supported by Mr. Candoko, one of the authors of the proficiency test questions, and he said:

"The Arabic language proficiency test here is very much focused on questions of daily activities, grammar, and morphology, and it was a standard used in the Western language proficiency test at the Language Promotion Center of Sultan Syarif Kasim Riau State Islamic University. It did not have a specific standard (Candoko, 2021)."

In addition to these previous data, the researcher initially analyzed the proficiency test question items in terms of the test's validity and concluded that one hundred and five of the question items were not characterized by honesty, while only thirty-five of the test items were characterized by validity (Halomoan et al., 2022).

From the previous data, it is clear that the proficiency test used at Sultan Syarif Kasim Riau University needs to be corrected and its items developed so that it is consistent with the specifications of a good test.

The specific standard is a standard decided upon by the university or a standard decided upon by universities and is considered important in measuring students' Arabic language proficiency. And the Arabic language proficiency test for non-native speakers at Sultan Syarif Kasim Riau State Islamic University, where Adek candoko said:

"The Arabic language proficiency test here is very much focused on questions of daily activities, grammar, and morphology, and it was a standard used in the Western language proficiency test at the Language Promotion Center of Sultan Syarif Kasim Riau State Islamic University. It did not have a specific standard (Candoko, 2021)."

Based on this evidence, it becomes clear that setting language proficiency test questions does not depend on a specific, correct criterion in the test, but rather there is no criterion on which it is based.

In the current era, technology is also a necessary thing and also in education. Education needed educational technology, including the test, as the test at Sultan Syarif Kasim Riau State Islamic University was based on paper, as Mr. Mahyuddin Syukri, the director of the center, said:

"The Arabic language proficiency test in this center was not computerized or computer-based because the university's language laboratory was not fit for use."

From the statement, it is clear that the Arabic language proficiency test at that university is still on the paper. In fact, the use of applications in this era is a necessity as a study found that students believe in the significance of online applications to simplify creating, distributing, and grading assignments (Zulkifli & Rozimela, 2021).

Design

Determine test objectives

The objectives of this test are based on the Common European Framework of Reference for Languages standard (D'Europa, 2018). Therefore, this test standard developed in this research is suitable for application in Indonesia as a unified standard for all Islamic universities in Indonesia. This research confirms what Uril Bahruddin presented about the necessity of having a unified classification of Arabic language proficiency(Bahruddin, 2016), although the existence of a standard An important unifier in Indonesia so that teaching the Arabic language in Islamic universities has a clear goal. The test developed in this research aims to measure students' proficiency in the Arabic language by adding it as a second language for the university stage. This test consists of three groups: listening, reading, and structures.

As for the indicators developed by the researcher in the developed Arabic language proficiency test for non-native speakers, they are ten: 1) mentions, 2) knows, 3) understands, 4) explains, 5) identifies, 6) extracts, 8) deduces, 8) translates, 9) Uses, 10) Apply compositions. These goals and indicators are based on the Common European Framework of Reference for Languages standard. This indicator indicates Bloom's classification which contains three levels, which are LOTS, MOTS, and HOTS (Brookhart, 2010).

Selection of test materials

There are three types of tests in this test: listening, reading, and structures. Listening has nine competencies, reading has nine competencies, and understanding structures has ten competencies (D'Europa, 2018). In the first proficiency, the researcher identified the test materials familiar to the students, including greetings, getting to know each other, daily activities, numbers, and the clock, so that the students would not face difficulty with the first questions. The indicator developed by the researcher for this competency is simple and easy.

Prepare the test standard

Based on the Common European Framework of Reference for Languages standard (D'Europa, 2018), the researcher developed the basic competency and its indicators as follows:

Listening

Basic listening competency

As for the basic listening proficiency in the Arabic language proficiency test for non-native speakers and its indicators in light of the Common European Framework of Reference for Languages standard, they are as follows:

Table 1: Basic competency of listening

Level	Number	Basic Competency
A1	1	That the student understands the meanings and ideas contained in audio-recorded Arabic speech about greetings, getting to know each other, daily activities, numbers, and the hour, in short texts that are easy to contextualize and dictionary.
A2	2	The student understands the meanings and ideas contained in the audio-recorded Arabic speech about family, shopping, and geographical locations in context and dictionary.
	3	That the student understands and recognizes the main idea from an audio text. The texts consist of dialogue in the profession and the surrounding environment, in food, and in the phone call, in context and lexicon.
B1	4	The student understands the meanings and ideas contained in audio-recorded Arabic speech about work, university, and entertainment places, contextually and lexically.
	5	That the student understands the main points of the audio material about politics, public areas, and hopes in context and dictionary
B2	6	The student understands the meanings and ideas contained in the audio-recorded Arabic speech about health, vacation, media, economy, and trade in context and dictionary.
	7	The student can analyze the content of the audio text of contemporary news, Islamic discoveries, and the era of globalization in context and dictionary.
	8	The student can analyze the content of the audio text regarding the Islamization of science and Islamic history in context and dictionary.

Reading

Basic reading competency

As for the basic reading proficiency in the Arabic language proficiency test for non-native speakers and its indicators in light of the Common European Framework of Reference for Languages standard, they are as follows:

Table 2: Basic reading competency

Level	Number	Basic Competency
A1	1	The student can mention vocabulary related to greetings, getting to know each other, daily activities, hours, and numbers
A2	2	The student understands the meanings and ideas contained about family, shopping, and geographical locations in context and dictionary.
B1	3	The student can identify the main idea from a text about the profession, the surrounding environment, a restaurant, and a phone call, in context and dictionary.
	4	That the student understands the meanings and ideas about work, university, and entertainment places in context and lexicon.
	5	That the student understands the main points of the material about politics, public areas, and hopes in context and dictionary
B2	6	The student understands the meanings and ideas about health, vacation, media, economy, and trade in context and dictionary.
	7	The student can analyze the content of the text from contemporary news, Islamic discoveries, and the era of globalization in context and dictionary.
	8	The student can analyze the content of the text regarding the Islamization of science and Islamic history in context and dictionary.
	9	That the student can infer from the expressions and text the dialogue between two or more people

From the above, it is clear that all of these basic competencies aim to: 1) measure the student's ability to mention vocabulary related to greetings, getting to know each other, daily activities, hours, and numbers, 2) measure the student's ability to understand meanings and ideas about family, shopping, geographical locations, work, university, places of entertainment, health, and vacation. Media, economics and trade, context and dictionary. 3) Measuring the student's ability to know the main idea of a text about the profession, the surrounding environment, the restaurant, the phone call, politics, public areas, and hopes in context and lexicon. 4) Measuring the student's ability to analyze the content of the text from contemporary news, Islamic discoveries, the era of globalization, the Islamization of science, and Islamic history in context, lexicon, and analogy. Their ability to infer from expressions and text dialogue between two or more people.

structures

As for the basic competence to understand the structures in the Arabic language proficiency test for non-native speakers and its indicators in light of the Common European Framework of Reference for Languages standard, they are as follows:

Table 3: Basic structures competency

Level	Number	Basic Competency
A1	1	The student understands the definite article, pronouns, demonstrative nouns, nominal sentence, number, present participle, and noun.
A2	2	For the student to understand the prohibited morphology, the relative noun, prepositions, adverbs, the sound verb - the past, the verbal sentence, the present tense, the genitive, and the genitive.
	3	For the student to understand the present accusative and the present tense, the command, may, negation and prohibition, not, why...won / not... nor / will not... and the like, the original numbers.
B1	4	The student understands the irregular verbs - the past and present tense, if and the subject in the sentence, the irregular verbs - the present tense, ordinal numbers.
	5	For the student to understand the verb more with one letter, with two letters, kan and its sisters, kaad and its sisters, the infinitive, the infinitive, and the negative.
B2	6	The student understands the passive voice form, the verb with three letters, the active participle and the active participle, exaggerated nouns, the instrument noun, the noun of place and time, and the diminutive.
	7	For the student to understand verbs that take two objects, the superlative, the distinction, the incorrect addition, and the conditional letter in conditional sentences.

The basic competency includes three indicators: 1) The student can use verbs that take two objects in the sentence, 2) The student can use the superlative noun, the distinction in the sentence, 3) The student can use the unreal addition, the subjunctive letter in conditional sentences in the sentence.

Development

Writing question items

After the researcher determined the test objectives, test materials, test standard, and test type. The researcher writes the question items, and this stage is an important stage and the test maker must take into account the matters related to it. The researcher divided the question items into four levels, namely Level 1A, A2, B1, and B2, as in the following table:

Table 4: Question posting map

No.	Skill	beginner (A1)	Basic (A2)	Medium (B1)	High Medium (B2)
1.	Listening	15	15	10	10
2.	Reading	15	15	10	10
3.	Structure	10	10	10	10
	Amount	40	40	30	30

Looking at the previous table, it is clear that the number of question items for the beginner level is forty (40) questions, the number of question items for the basic level is forty (40) questions, the number of question items for the intermediate level is thirty (30) questions, and the number of question items for the high level is thirty (30). A question. The question items are divided into three sections: questions for listening, reading, and composition. Available time: 120 minutes. This test, with one hundred and forty questions, took 0.8 minutes. The Arabic language proficiency test requires focus. Therefore, the listening test was developed at the beginning so that the participants could concentrate on answering the questions. He placed understanding structures last, considering that it is a difficult subject. The researcher arranged the basic competency and indicators according to the test preparation theory. Core competency ranges from simple to complex and from easy to difficult. The test was arranged as a listening test, then reading, and then understanding the compositions.

Writing items for the Arabic language proficiency test for non-native speakers in light of the Common European Framework of Reference for Languages standard, in proportion to the subjects taught at the Language Promotion Center at Sultan Syarif Kasim Riau State Islamic University, is one of the possibilities for setting the Arabic language standard in Indonesia, as presented by Tulus Mustafa, that Indonesia must benefit. With the experiences of those countries and making those standards to set standards and a frame of reference for teaching the Arabic language (Musthofa, 2018).

Verification of question items from experts

After the researcher wrote the test question items, the researcher judged the question items prepared by the researcher. The researcher used two arbitrations: expert arbitration and experiment arbitration. The researcher chose four experts to judge the developed test: an expert in the field of test design, an expert in the field of the Arabic language, an expert in the field of educational technology, and an expert in the field of the Arabic language test subject.

After experts in the field of subject matter, the structural field, the linguistic field, and the technological field judged the product of testing the proficiency of the Arabic language for non-native speakers in light of the Common European Framework of Reference for Languages standard using the iSpring Suite method (applied to students of Sultan Syarif Kasim Riau State Islamic University), which had been prepared, the researcher repaired What was modified by the experts, and the results of their observation, modification, comment and suggestion are as follows:

Expert in the construction field

The researcher corrected the test construct after it was approved by the expert. The items for evaluating the test construct consist of seven indicators. The researcher obtained a result of 5 in five items and a result of 4 in two items. The researcher did not correct the items of the questionnaire questions in this field except a little of what the expert modified because they were rated as good and very good.

The expert in the field of the test subject

The researcher corrected the test material after the expert-approved it. The items for evaluating the test material consist of eight items. The researcher obtained a result of 3 on three items and a result of 4 on five items, so the researcher corrected all the items because three items were a result of 3, and five items were a result of 4. After the items were repaired, the researcher submitted those modified items to the expert to evaluate them again, and the researcher obtained a result. 5 in five items and a result of 4 in three items.

Expert in the linguistic field

The researcher fixed the language used in the test and after the test was approved, the items to evaluate the language used in the test consist of six indicators. The researcher obtained a score of 5 on five items and a score of 4 on one item. The researcher corrects the linguistic field only a little of what the expert adjusts because it is to an excellent degree with a result of good and very good.

Expert in the technological field

The researcher reformed the design of the technology that the researcher used in testing the proficiency of the Arabic language for non-native speakers after it was approved by the expert. The educational technology evaluation items in the test consist of nine indicators. The researcher obtained a score of 3 on one item, a score of 4 on five items, and a score of 5 on three items. The researcher repaired the six items whose results did not reach a score of 5. After the items were repaired, the researcher presented the modified question items to the expert to evaluate them again, and the researcher obtained a result of 5 in seven items and a result of 4 in two items.

design of iSpring Suite.

The questions items prepared by the researcher are combined in the iSpring Suite with audio recording and answer choices. The first question slide is as follows:
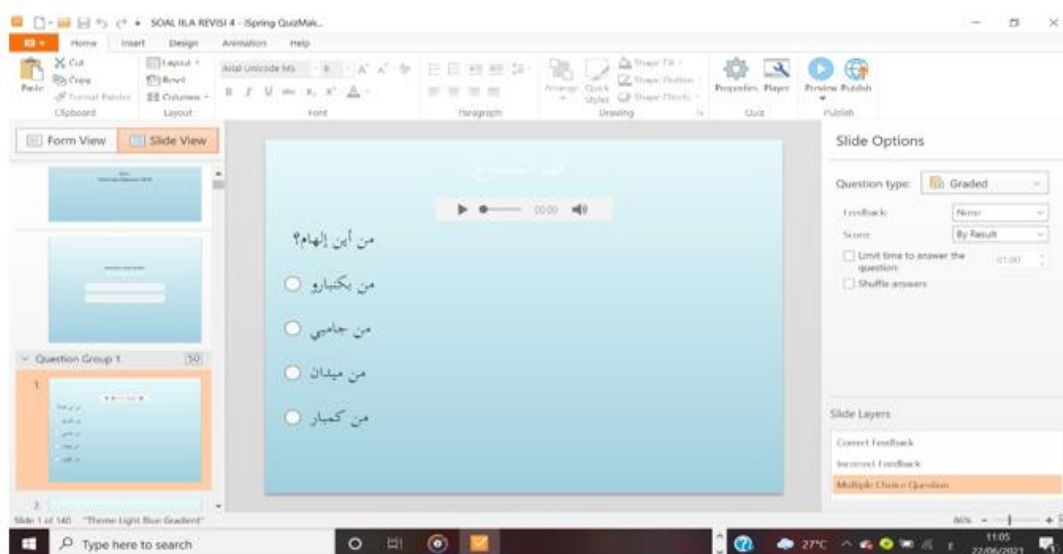

Image 1: Image question 1 for the listening test

This is like the question in Asfaring Sweet with audio recording and answer choices. These are all questions from question one to one hundred and forty.

Description of the content validity, empirical validity, and reliability of an Arabic language proficiency test developed for students of Sultan Syarif Kasim Riau State Islamic University.

In this section, the researcher explains the quality of the developed Arabic language proficiency test in terms of validity, reliability, degree of difficulty, discrimination, and dispersion, and explains this as follows:

Content Validity

The researcher presented this test to four experts: an expert in the field of test design and construction, an expert in the field of test material, an expert in the field of linguistics, and an expert in the field of educational technology. The expert in the field of design and construction the test gave a result of 94.3 a very good grade, and the expert in the field of the test subject gave the result of 92.5 a very good grade, and the expert in the field of classical Arabic gave the result of 96.7 as a very good grade, in the field of educational technology, the result was 95.6, indicating a very good grade.

Hence, it has become clear that the result of the developed design of the Arabic language proficiency test for non-native speakers is very good because it reached a result of 94.8, and this result was not obtained by the researcher directly, but rather after making many amendments, suggestions, comments, and reforms by experts.

Experimental validity

The researcher obtained valid data in the specified experiment, stating that the questions that had to be modified were one hundred and twenty-six questions. These questions were modified because their result was less than the r of the table, which is 0.632. Then in the first extended experiment, the researcher obtained the questions that the researcher modified were fifty-five questions because their result was less than the r of the table, which is 0.312. In the second expanded experiment, the researcher obtained that the questions that the researcher modified were twenty-seven questions because their result was less than the r of the table, which is 0.312. In the third extended experiment, there are no modified questions because the result of all question items is more than the r of the table, which is 0.312.

Reliable

The reliability results showed through re-testing that the reliability of the Arabic language proficiency test for non-native speakers in light of the standard of the Common European Framework of Reference for Languages for the students of Sultan Syarif Kasim Riau State Islamic University in the extended experiment was found to be 0.962. This result is greater than the value of 0.700, so the reliability is good. very.

Difficulty level

The steps to know the difficulty scores for question items in SPSS are in the data view, 1) Click "Analysis", then Descriptive Statistics, then Frequency, 2) Fill in the "Variables" box, click "Statistics", click "Average", click "Continue", 3) Then click "OK." Analysis: From the results, the "average" value is interpreted in the statistical table in the difficulty score category, which is:

Table 5: Classification of difficulty indicators

Quality of question	Difficulty index
0,00 – 0,30	Difficult
0,31 – 0,70	Medium
0,71 – 1,00	Easy

The researcher classified the results according to the degree of difficulty of the test, in which the questions were of a high degree of honesty.

Table 6: Difficulty score

difficulty level	Question number
Easy	5,9,11,18,22,23,26,27,31,32,33,36, 43, 46, 47, 48, 53, 55, 58, 65, 67, 70, 76
Medium	2, 3, 4, 6, 7, 8, 12, 15, 16, 17, 19, 20, 21, 24, 25, 27, 29, 30, 34, 35, 37, 38, 39, 40, 41, 42, 44, 45, 49, 50, 51, 52, 54, 56, 57, 59, 60, 61, 62, 63, 64, 66, 69, 71, 72, 74, 75, 77, 78, 79, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140.
Difficult	1, 10, 13, 14, 80, 100.

Based on the previous table, it was found that easy questions are 17%, medium questions are 79%, and difficult questions are 4%.

Discrimination

To determine the degrees of discrimination, the calculation value used is the arithmetic t, which is compared with the following criteria:

Table 6: Discrimination score

Difficulty index	Quality of question
0,40 – 1,00	very good
0,30 – 0,39	Good
0,20 – 0,29	not good
0,00 - 019	Rejected

The computational R can be seen from the Pearson correlation value in determining the validity of the experimental test. The researcher classified the results of the discrimination power of the test in which the questions were highly valid.

Table 8: As a result of the power of discrimination

Degree of discrimination strength	Question
Very good	1, 2, 3, 5, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49, 51, 53, 54, 55, 56, 57, 58, 60, 61, 63, 63, 64, 65, 66, 67, 68, 69, 70, 72, 73, 74, 75, 76, 77, 78, 78, 82, 83, 85, 86, 87, 88, 89, 90, 92, 93, 95, 96, 97, 98, 99, 106, 108, 109, 114, 115, 116, 117, 121, 122, 123, 124, 125, 126, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 140.
good	4, 6, 7, 8, 20, 36, 42, 50, 52, 59, 71, 80, 81, 84, 91, 94, 100, 101, 102, 103, 104, 105, 107, 110, 111, 112, 113, 118, 119, 120, 127, 139.
Rejected	-

Based on the previous table, it is clear that all the developed question items are rated very good and good.

Distraction

To know the quality of dispersion, the researcher used the help of the ANATES program to facilitate the calculation. The steps are to enter the students' answers, then press Record Question, then press Dispersion. We found that the dispersion in the Very Good score is 30 questions (21.4%), in the Good score it is 80 questions (57.1%), and in the Good score it is 18 questions (12, 9%) and 12 questions (8.6%) were graded not good.

Analyze the question items

In analyzing the question items, it was found that all question items are honest because the result of all question items is more than the r of the table, which is 0.312, and the reliability of the Arabic language proficiency test for speakers of other languages in light of the Common European Framework of Reference for languages standard for students of Sultan Syarif Kasim Riau State Islamic University in the extended experiment It is 0.749, so this result is greater than the price of the table, which is 0.361. In terms of the degree of difficulty, it was found that the easy questions were 17%, the medium questions were 79%, and the difficult questions were 4%, and all items of the developed questions were rated very good and good. These question items are valid for use. This research confirms what Nur Komari and friends presented in developing the achievement test for listening skills (Saputra et al., 2022). The items of the questions that were not all honest must be developed, as was found by Partomuan Harahap in analyzing the items of the Arabic language proficiency test for non-native speakers at the Curup State Islamic Institute (Harahap, 2018). However, the test developed does not yet contain a speaking skills test because measuring speaking skills requires a special and authentic assessment and thus requires further development. This was also found in the assessment of English speaking skills for junior high school level, that an authentic assessment model that is by teachers' needs is still needed to measure students' speaking abilities (Zaim & Arsyad, 2020).

CONCLUSION

As for the results of the research that the researcher arrived at, the Arabic language proficiency test for non-native speakers was developed according to the needs required by the test users at Sultan Syarif Kasim Riau State Islamic University. The developed test is characterized by using the Common European Framework of Reference for Languages standard, with four levels specified: Level A1 (40 questions), A2 (40 questions), B1 (30 questions), and B2 (30 questions). It has also been included in the Esvering Suite program to form the test. Digitally. The test consists of 140 items of questions, which are divided into three components: listening (50 questions), reading (50 questions), and understanding structures (40 questions). The test developed in light of the Common European Framework of Reference for Languages (CEFR) standard using the iSpring Suite program is valid for application because the research results indicated a very good measure of Arabic language proficiency for non-native speaking students and given that the results of the approval of the four experts, including an expert in the field of test design and an expert in the field of language Arabic and an expert in the field of educational technology and an expert in the field of the Arabic language test subject obtained 94.8. All question items are honest and the reliability is 0.749 a high degree. As for the quality of the developed Arabic language proficiency test for non-native speakers in terms of

degree of difficulty, it was found that the easy question items were 17%, the medium question items were 79%, and the difficult question items were 4%. From the distinction, it is clear that all the items of the developed questions are rated very good and good, and in terms of dispersion, 30 questions (21.4%) were rated very good, 80 questions (57.1%) were rated good, and 18 questions (12) were rated as adequate. 9%) and 12 questions (8.6%) received a grade of not good. From these results, it can be concluded that this test can be applied to measure the Arabic language proficiency of university students.

References

1) al Busyairi, M. M., & Said, M. malik. (1415). *madkhal Ila al Manahji wa Thuruq at Tadris*. Dar al Liwa'.

2) al Khuli, M. A. (1982). *Asalib Tadris al Lughah al Arabiyah li Ghair an Natiqhina Biha*. Jami' Huquq Mahfuzho lil Muallif.

3) Bahruddin, U. (2020). *Professor of Arabic at Maulana Malik Ibrahim State Islamic University*.

4) Bahruddin, U. (2016). *Tajarrubatu jamiati Maulana Malik Ibrahim Islamiyah hukumiyah Indonesia fi I'dad ikhtibarat lughah arabiyah muwahhadah linnatiqhina bigairiha, A research paper presented at the Second Scientific Conference on Tests for Measuring Arabic Language Skills for*. 16.

5) Bahruddin, U., & Qodri, M. (2020). Tahlîl Mulâimah Bunûd Ikhtibâr Al-Lughah Al-'Arabiyyah Al-Nihâî Al-Wathanî Li Manhaj Mustawâ Al-Wahdah Al-Ta'lîmiyyah (Ktsp) Wa Tashnîf Bloom Al-Mu'addal. *Arabiyat : Jurnal Pendidikan Bahasa Arab Dan Kebahasaaraban*, *7*(1), 115–130. https://doi.org/10.15408/a.v7i1.12806

6) Borg, W. R., & Gall, M. D. (1984). Educational research: An introduction. *British Journal of Educational Studies*, *32*(3).

7) Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Ascd.

8) Candoko, A. (2021). *Dosen Pusat Bahasa Universitas Islam Negeri Sultan Syarif Kasim Riau*.

9) Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge University Press.

10) D'Europa, C. (2018). Common European framework of reference for languages (CEFR). *Learning, Teaching*, 39.

11) Halomoan, H., Ibrahim, F. M. A., & Bahruddin, U. (2022). Taḥlīl Ikhtibār Kafāah al-Lugah al-'Arabiyyah li an-Nāṭiqīna bigairihā fī Jāmi'ah Sulṭān Syarīf Qāsim al-Islāmiyyah al-Ḥukūmiyyah Riau. *LISANIA: Journal of Arabic Education and Literature*, *6*(1), 74–87.

12) Harahap, P. (2018). ANALISIS SOAL TOAFL IAIN CURUP. *Jurnal Pendidikan Bahasa Dan Sastra Arab*, *4*(2).

13) Lioyd, D., Davidson, P., Coombe, C. A., & Addamigh, K. ben abdel aziz. (2008). *asasiyat taqyim fi ta'limil lughawiy*. King Saud University Press. https://books.google.co.id/books?id=jB0RnQAACAAJ

14) Mahmud rusydi khothir. (1998). *thuruq tadrisil lughoh arabiy wat tarbiyah diniyah fi dhauil ittijahat tarbawiyah haditsah*. muassassah kutub jami'iyah.

15) Miles, B. M., Huberman, A. michael, & Terj. Rohedi, T. R. (1992). *Analisis Data Kualitatif*. UI Press.

16) Mudzaffar, A. L. W. (2009). *Thuruq Tadris wa Asalib al Imtihan*. Syabkah Mudarris Islamiyah.

17) Musthofa, T. (2018). Al-Ithaar Al-Marja'iy Al-Induniisiy Li Ta'liimi Al-Lughah Al-'Arabiyyah Fi Dhaui Al-Ithar Al-Marja'Iy Al-Eropa Al-Musytarak Li Ta'limi Al-Lughaat. *Prosiding Pertemuan Ilmiah Internasional Bahasa Arab*, 1179.

18) Putra, N. (2011). Research & Development-Research and Development: An introduction. *Raja Grafindo Persada*.

19) Robert, K. Y. M. M. D. (2006). *Terj. Studi Kasus Desain & metode*. RajaGrafindo Perkasa.

20) Salahuddin mahmud alam. (2009). *at taqwimu tarbawiyyul badil, ususuhunnazriyah wal manhajiyah wa tathbiqotuhu al maidaniyah*. Dar al Fikri.

21) Saputra, H., Fuadah, F., Yanti, S. A., & Qomari, N. (2022). Penerapan butir soal unsur bahasa Arab (bunyi, kosakata, struktur kalimat). *Muhadasah: Jurnal Pendidikan Bahasa Arab*, *4*(2), 118–138.

22) Sugiyono, P. D. (2016). metode penelitian kuantitatif, kualitatif,dan R&D. In *Alfabeta, cv.*

23) Sugiyono, S. (2017). Metode penelitian dan pengembangan untuk bidang pendidikan, manajemen, sosial, teknik: Research and development. *R&D. Alfabeta*.

24) Syahrul, R. (2020). The Development Of E-Modules Based on Adobe Flash For Indonesian Subjects At IAIN Bukittinggi. *Journal of Physics: Conference Series*, *1471*(1), 12002. https://doi.org/10.1088/1742-6596/1471/1/012002

25) Syukri, M. (2021). *Kepala Pusat Bahasa Universitas Islam Negeri Sultan Syarif Kasim Riau*.

26) Thuaimah, Rusydi ahmad; manna', muhammad sayyid. (2001). *ta'limul 'arabiyah wad din bainal i'ilmi wal fanni*. Dar al Fikri al Arabi.

27) Thuaimah, R. A. (1989). *Ta'lim al Lughah al Arabiyah li Ghairi an Nathiqina Biha: Manahijuhu wa Asalibuhu*. Mansyurat Islamiyah lit Tarbiyah wal Ulum wa Tsaqafah.

28) Wahab, M. A. (2022, January 15). *Perkembangan Tes Bahasa Arab Standar di Indonesia,*. Http://Repository.Uinjkt.Ac.Id.

29) Wulansari, R. E., Marta, R., Sakti, R. H., Dewi, S. M., Safitri, D., Kassymova, G. K., Folkourng, F., & Kumar, V. (2023). Computer Assisted Instruction (CAI) Integrated Case Method-Flipped Classroom: Innovative Instructional Model to Improve Problem-Solving Skill and Learning Outcome of TVET Students. *Journal of Technical Education and Training*, *15*(4), 100–113.

30) Zaim, M., & Arsyad, S. (2020). Authentic Assessment for Speaking Skills: Problem and Solution for English Secondary School Teachers in Indonesia. *International Journal of Instruction*, *13*(3), 587–604.

31) Zainil, M., Helsa, Y., Sutarsih, C., & Nisa, S. (2024). A needs analysis on the utilization of learning management systems as blended learning media in elementary school. *Journal of Education and E-Learning Research*, *11*(1), 56–65. https://doi.org/10.20448/jeelr.v11i1.5310

32) Zulkifli, N. A., & Rozimela, Y. (2021). Online applications to support remote classroom dialogue and assessment. *Journal of Physics: Conference Series*, *1779*(1), 12039. https://doi.org/10.1088/1742-6596/1779/1/012039