

***Wikalah* iBT: How Is An Internet-Based Arabic Language Proficiency Test Instrument Developed?**

Syihabuddin¹, Maman Abdurrahman², Frida Akmalia³, Aam Abdussalam⁴

Universitas Pendidikan Indonesia, Bandung, Indonesia^{1,2,3,4}

syihabuddin@upi.edu¹, mamanabdurrahman@upi.edu²,

fridaakmalia26@upi.edu³, aamabdussalam86@upi.edu⁴

Abstract

This study discusses the development of 140 items of Arabic language proficiency test which covers listening, reading, and grammar mastery. These items were empirically analyzed, tried out, and revised. Results of this development would then be made a basis to design an application that can evaluate internet-based Arabic language proficiency (Internet-based *Taqwim Kafa`ah Maharah Lughah Arabiyah*) abbreviated as *Wikalah* iBT. To achieve this goal, a developmental research method was employed, namely the development of instruments starting from the needs analysis for tests, the making of instruments, validity and reliability testing, the design of iBT program, and the trial of the *Wikalah* iBT application. Results of the study show that (a) several test items still had morphological, morphosyntactic, and technical errors, (b) 81 test items were valid, (c) average reliability level was 0.67, (d) this test was in the medium and easy categories (63.57), and (e) the iBT application design was used successfully and ran smoothly. Based on the conclusions, it is recommended to add the test item package, improve the quality of the questions, and evaluate the iBT application. We believe that these findings provide initial support for making an overview to those preparing for the iBT test or other similar tests of language proficiency. Study outcomes also inform the extent to which the use of technology can evaluate language proficiency.

Keywords: Arabic Proficiency Test; Evaluation; iBT

INTRODUCTION

During the last five years, language proficiency tests have become an important focus of study in several countries (e.g. Culbertson, Andersen, & Christiansen, 2020; Dimova & Kling, 2018; Ginther, & Yan, 2018; Neumann, Padden, & McDonough, 2019; Pen, Yang, & Cheng, 2021; Wang & Treffers-Daller, 2017). For instance in Indonesia (Karjo & Ronaldo, 2019), using language proficiency tests (e.g. TOEFL, IELTS) as entry as well as exit requirement for university education. Kotby, Khairy, Barakah, Rifaie, and Elshobary (1995); Rifaie (1994) explain that in the Arabic field, the Arabic language proficiency test was the first and pioneering test developed for language evaluation of Arabic speaking (Loomis, 2015), writing (Azmi, Al-Jouie, & Hussain, 2019), reading (De Graaf, 2021), and listening (Zubaidi, 2020).

As time passes, the development of technology turns out to change all forms of behavior, learning models, and also the method of testing language proficiency.

The new method is the Internet Based-Test (iBT). Since 2005, the internet has been used for test purposes, especially in English language proficiency (Poorsoti & Davatgari, 2015; Roud & Hidri, 2021; Soureshjani, 2012) by using TOEFL iBT (Papageorgiou & Manna, 2021). Test takers have a positive attitude towards TOEFL iBT (Gu, Davis, Tao, & Zechner, 2021; Kim, 2017). Zahedkazemi (2015) observed 69% of respondents believe that the test will improve their performance in English. They believe that the internet-based TOEFL (iBT) and IELTS can measure their speaking abilities. Test takers' trust in iBT grew out of their habit of using computers in learning and other activities. Bahari (2020) concluded that students have the readiness to learn through computer and the skills to use it at a good level. This means that today a digital generation has been born with the ability to work on various things digitally, including language tests, using computers and assisted by the internet (Fleckenstein, Keller, Krüger, Tannenbaum, & Köller, 2020; Frost, Clothier, Huisman, & Wigglesworth, 2020). Hence, the use of internet-based device is not something strange but is commonplace in various circles and regions.

Besides students' readiness to take the iBT test, this test provides a positive washback for improving teaching and learning. Barnes' (2016) research in Vietnam concluded that iBT TOEFL[®] influenced both the teaching materials delivered by teachers and the teaching method used. Although the use of computers for testing is common, Mousavi's (2014) study emphasizes the importance of iBT test preparation. His research concluded that preparation had an effect on the scores achieved by students. This preparation includes the introduction of the environment, the mechanism of testing, and hardware readiness. The positive response to iBT is also strengthened by the benefits obtained from its effective and efficient implementation. It means that iBT is time- and cost efficient when compared to tests carried out manually or traditionally. Arabic language proficiency tests have to this date been conducted in various parts of the world for various purposes. The tests are prepared by considering international standards and certainly not based on the higher education curriculum in Indonesia. Harsch, Ushioda, & Ladroue (2017) show that students entering the university based on their TOEFL iBT scores feel very prepared and generally find the test an effective way of preparing for their academic studies in the UK. They cope well with linguistic demands, and most of them graduate successfully.

Therefore, it is urgent to develop Arabic language proficiency tests based on the Indonesian curriculum. In the Indonesian context, several Arabic language proficiency tests have been conducted by several universities. For example, in UIN Sunan Kalijaga Yogyakarta developed the *Ikhtibar Kafaah al-Lughah 'al-'Arabiyyah* (IKLA), UIN Jakarta developed the Test of Arabic as a Foreign Language (TOAFL), in IAIN Salatiga developed *Ikhtibarat Lughat al 'Arabiyyah ka Lughat al Ajnaby* (ILAiK), IAIN Surakarta developed *Test Of Standard Arabic* (TOSA), and other institutions also developed proficiency tests. However, the tests developed were still computer-based or done manually, so they were less efficient. The problem is then: how is an internet-based Arabic language proficiency test instrument developed? Has the test at the Arabic Language Education Study

Program, UPI Bandung, met the criteria of a good test, which are valid, reliable, and objective? Has the existing proficiency test met the criteria of practicality, efficiency, and effectiveness in terms of management?

This research is considered important to do, because it has a novelty level in the aspects of application development formatted in the form of iBT. So that it can be used self instructional, self contained, can be operated standalone, user friendly, and adaptive to technological developments. The reality in the field also shows that the iBT program that has upgraded the Arabic language proficiency tests model has not been presented as much as the TOEFL questions in general. In addition, Arabic language proficiency tests which is carried out in several Indonesian Universities, its utilization is more oriented internally to each language institution. Therefore, researchers are encouraged to do the same in order to support language testing services and can be accessed both online and offline.

METHOD

This research uses descriptive method and is explored through quantitative data. As stated by Sudjana and Ibrahim (1989) that descriptive research is research that seeks to describe a symptom, events and events that occur at the present time where the researcher tries to photograph the events and incidents that are the center of attention and then describe them as they are. Ali (1982) emphasized that descriptive research methods are used to solve and answer problems that occur today. This is done by taking the steps of collecting, classifying and analyzing or processing data, making conclusions and reports with the main objective of making an objective description of a situation in a description. Sukmadinata (2002) explains that descriptive research does not manipulate or provide certain treatments to variables or design something that is expected to happen to variables, but all activities, circumstances, events, aspects, components or variables run as they are. Meanwhile, what is meant by a quantitative approach is the approach used in research by measuring the indicators of the research variables in order to obtain an overview of these variables. The purpose of the quantitative approach according to Surakhmad (1998) is to measure the dimensions to be studied.

The use of quantitative descriptive methods is aligned with research variables that focus on actual problems and phenomena that are happening at the moment with the form of research results in the form of numbers that have meaning. As stated by Sudjana (1997) that descriptive research methods with a quantitative approach are used if the aim is to describe or explain events or events that are happening now in the form of meaningful numbers. The purpose of this descriptive study with a quantitative approach is to explain a situation to be studied with the support of a literature study so that it further strengthens the analysis of researchers in making a conclusion. Where the research results were obtained from the calculation of the research variable indicators then presented in writing by the author.

The quantitative approach is used because the results of research on the development of 140 items of Arabic language proficiency test are written in the form of a score or percentage which is then described according to predetermined conditions. This research was conducted at the Department of Arabic Language Education Program at a University in Indonesia. The questions will be answered through research that describing the validity of test items in terms of the principles of test item writing, test empirical validity, test reliability, and the level of difficulty of the test. The test items analyzed were 140 in total, which were prepared specifically for the purposes of this research by the Arabic Education Study Program lecturers. The test consists of three parts, namely *Fahmul Masmu'*, *Fahmul Maqru*, and *Qawa'id*. Empirical test was carried out using the ANATES application. To meet the criteria of practicality, efficiency, and effectiveness of the Arabic language proficiency test, the researchers collaborated with programmers in designing an iBT application called *Wikalah*. It is designed by considering theoretical aspects, objectivity, and efficiency. After this application was designed, 140 items were uploaded into the *Wikalah* iBT application.

RESULT AND DISCUSSION

Arabic Language Proficiency

Proficiency is defined as perfect linguistic performance in terms of listening, speaking, reading, and writing based on the understanding of a certain language (an-Naqah, 1997). The term “test” or “*taqwim*” and “*ikhtibar*” means testing to know the quality of something or someone. Thuáimah (1989) and Ubaidat (1984) define *ikhtibar* as a collection of questions prepared by the teacher or developer, which must be answered by students. The questions are aimed at measuring the level of students’ language proficiency, knowing the extent of their progress in learning, and comparing their abilities to those of their peers, both quantitatively and qualitatively. These skills include listening, speaking, reading, and writing (Mustafa, 2002; Veronika and Darnita, 2015). In addition to listening, speaking, reading and writing, proficiency tests should measure a person’s abilities in grammatical, stylistic, literary, and cultural aspects. Therefore, proficiency tests must be prepared based on linguistic foundation, cultural foundation, and educational foundation (Hasib, 2004; Madkur, 2000).

Analysis of Test Items in Terms of Linguistic Principles

The minimum standard that needs to be met by a test is that the test must be correct and clear. Correctness is related to the suitability of the concepts tested to scientific principles, while clarity is related to the language used. In terms of language, some errors were found such as the writing of *ya* that was incorrectly written as *ba*, the unnecessary addition of *ya*, inverted *hamzah* writing, lack of *alif* writing, lack of some letters, and *hamzah* writing errors. For example, the word *yaghtasilu* is written *yastaghsilu*, one letter mistakenly added. Such errors can be categorized as morphological errors.

In other items, some adjectival phrases were found to be not aligned between their *sifat* and *maushuf*, use improper conjunctions, and use incorrect definitive markers for subordinate phrases. For example, *al-ahadits* was written *li`ahadits*, with a lack of *alif* on the adjectival phrase. This error can be categorized as morphosyntax error. Additionally, there are errors related to typing, such as not using spaces or using excessive spaces, different fonts, irregular letter positions, and typing formats that are not consistent. This error is categorized as technical ones. The results of the analysis of test items on the three aspects of language proficiency above are presented in table 1.

Table 1. Results of Test Item Analysis Based on Test Item Writing Guidelines

Proficiency Criteria	Number of Items	Incorrect	Correct
<i>Masmu'</i>	50	4	46
<i>Maqru'</i>	50	17	33
<i>Qawa'id</i>	40	4	36
Total	140	25	115

The whole errors found in the 25 items as shown in Table 1 were immediately corrected so that the whole items could be subsequently tested empirically.

Validity and Reliability Tests

The results of the validity and reliability tests of the aspects of *Fahmul Masmu'*, *Fahmul Maqru'*, and *Qawa'id* mastery show that out of 140 items only 81 were valid, while the rest 59 items were invalid. In terms of reliability, the results of the analysis show that the *fahmul masmu'* aspect of the test items was more reliable than the other two aspects. Overall, the reliability of the three aspects was 0.67, which is categorized as quite reliable. The results are displayed in Table 2.

Table 2. Results of Validity and Reliability of the Test Items

No	Proficiency Aspect	Number of Items	Valid	Invalid	Reliability Level
1.	<i>Masmu'</i>	50	32	18	0.83
2.	<i>Maqru'</i>	50	27	23	0.69
3.	<i>Qawa'id</i>	40	22	18	0.50
	Total	140	81	59	Average: 0.67

Table 2 above shows that the reliability coefficient of the test items in terms of *Fahmul Masmu'* was higher than that of other aspects, and the reliability coefficient of the *Qawa'id* group was lower than that of the other two groups of questions. This is partly due to the level of difficulty of *qawa'id* compared to the other two skills.

Difficulty Level

The results for measurement of difficulty level of the 140 items are summarized in Table 3.

Table 3. Difficulty Level of the Test Items

No	Proficiency Aspect	Number of Items	Very Difficult	Difficult	Moderate	Easy	Very Easy
1.	<i>Masmu'</i>	50	10	10	24	3	3
2.	<i>Maqru'</i>	50	1	9	39	1	0
3.	<i>Qawa'id</i>	40	0	8	26	5	1
Total		140	11	27	89	9	4
Percentage			7.86	14.28	63.57	6.43	2.86
Criteria of difficult, moderate, easy			22.14	63.57	9.29		

Table 1.3 above shows that as many as 63.57% of the test items were at the moderate level, 6.43% were easy, and 2.86% were very easy. Meanwhile, as many as 7.86% were categorized into difficult questions and as many as 14.28% were moderately difficult. If the five categories were simplified into three criteria, then the percentages of difficult, moderate, and easy items were 22.14%, 63.57%, and 9.29%, respectively.

The proportion of difficulty level based on the three test criteria is determined as follow: (1) 20% difficult questions, (2) 60% moderate questions, and (3) 20% easy questions. Because the number of easy questions was only 9.29%, which is far lower than the 20% criterion, the 140 items then can be said to have not fully met the criteria. In other words, the level of difficulty of this test package was not proportional.

Test Trial

Subsequently, the 140 test items were tested on students. The trial results were compared to the minimum passing score set by the Arabic Language Study Program, which is 425. Based on this criterion, it was found that there were 24 participants or test takers who met the minimum passing score and only 6 test takers who did not.

The criteria used to determine the quality of the test items were: (1) if 27% of the participants fail, then the items are considered easy; (2) if 28-72% of the participants fail, then the questions are considered moderate; and (3) if 73% of the participants fail, then the test items are considered difficult. Because the number of participants who failed was only 6 or below 27%, which is equivalent to 20%, the test package was included in the easy category.

Wikalah Program

Besides having to meet the criteria of a good test, which are valid and reliable, a test must also be practical, namely easy to implement and the results are easily administered. To meet the third criterion, the test needs to be developed with information technology interventions in the form of computers and internet networks. Such a test is known as an internet-based test which is abbreviated as iBT.

The application was developed based on two principles. First, the principle of truth, validity, reliability, and honesty as academic norms that applies to the test. Second, the principle of practicality, convenience, and objectivity that applies to information and communication systems. Based on these two principles, an internet-based Arabic language proficiency test or *Taqwim Kafa'ah Lughah al-Arabiyyah*, abbreviated as *Wikalah* iBT, was developed. The display of this application is shown in Figure 1.

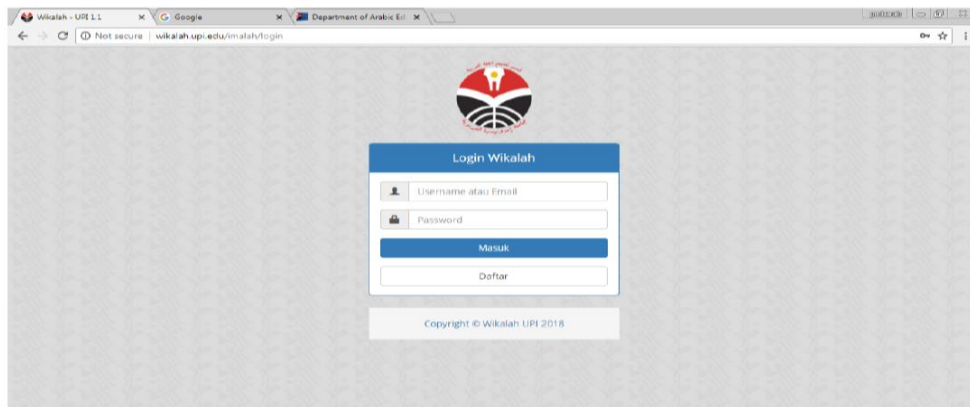


Figure 1: The Display of the Internet-based Arabic Language Proficiency Test

The *Wikalah* iBT application system has three main components, namely admin, test item development team (TPS), and user. Each component has its own authority, obligations, and work procedures, such as illustrated in the workflow chart in Figure 2.

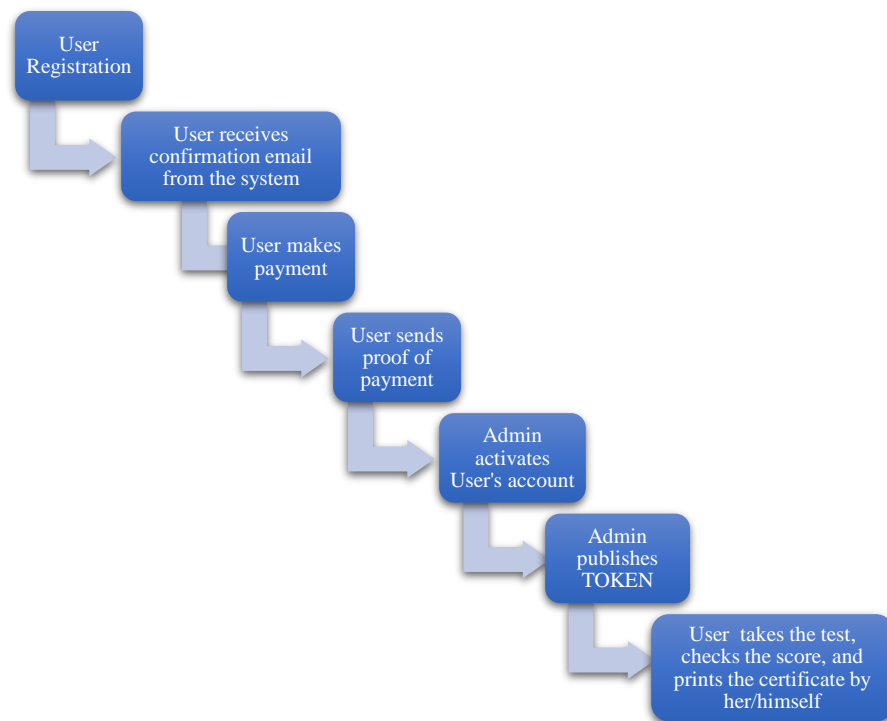
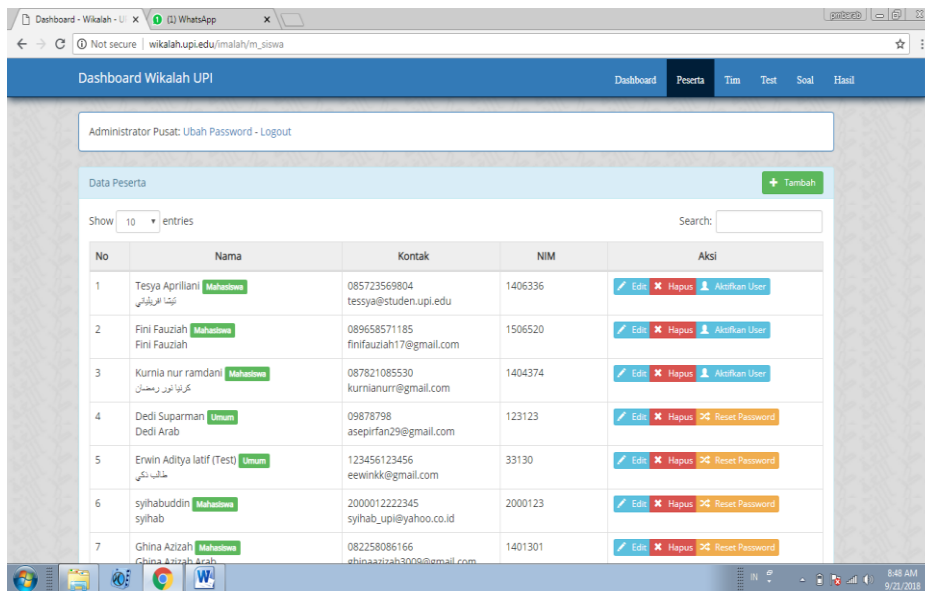


Figure 2: Workflow of the Three Main Components of the Test Application

The first component is admin. An admin has the authority to access all the features found at <http://Wikalah.upi.edu/>, starting from registering and/or deleting accounts, uploading/editing/deleting members of test development team, uploading/editing/deleting test items, managing test categories, and seeing a test taker's test score. However, the admin does not have the authority to conduct a test, as it can only be done by the Test Development Team. In addition, the admin has an obligation to maintain the smooth running of the system. Therefore, s/he must maintain the system periodically, clear caches, debug the system, correct database server down, update the system regularly, and serve several administrative activities.

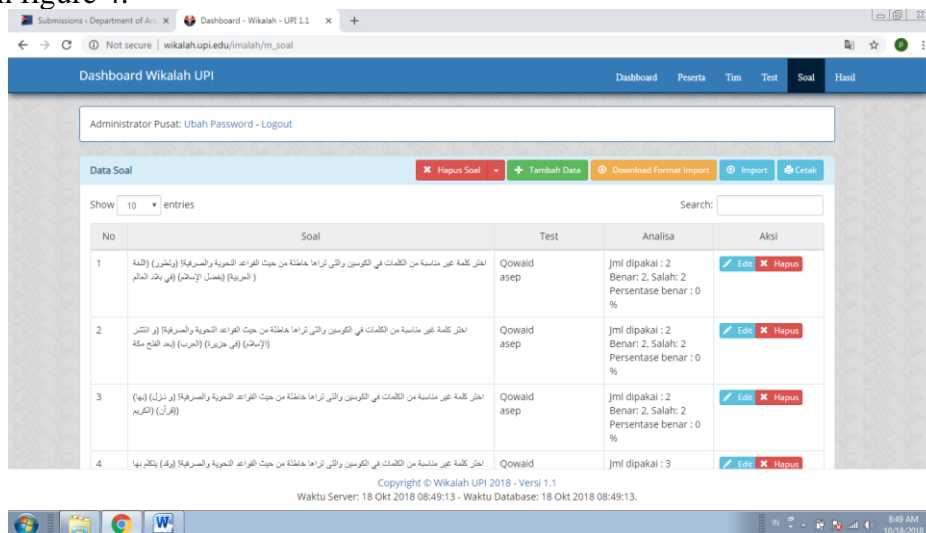
In order for the admin to carry out his/her authority and obligations, an access account should be made for him/her by the application development team. After having an account, the admin can open the *URL* <http://Wikalah.upi.edu> and enter the password and username, so a page that displays information on Dashboards, Teams, Test Takers, Tests, Test Items, and Results will appear as shown in the display in Figure 3.



No	Nama	Kontak	NIM	Aksi
1	Tesyra Aprillani تيسرا ابريلاني	085723569804 tesyaa@studen.upi.edu	1406336	Edit Hapus Aktifkan User
2	Fini Fauziah فينا فوزيا	089658571185 finifauziah17@gmail.com	1506520	Edit Hapus Aktifkan User
3	Kurnia nur ramdan كورتيا نور رحمان	087821085530 kurnianur@gmail.com	1404374	Edit Hapus Aktifkan User
4	Dedi Suparman Dedi Arab	09878798 asepirfan29@gmail.com	123123	Edit Hapus Reset Password
5	Erwin Aditya latif (Test) عرب اديتيا لاتف	123456123456 eewinnk@gmail.com	33130	Edit Hapus Reset Password
6	syihabuddin syihab	2000012222345 syihab_upi@yahoo.co.id	2000123	Edit Hapus Reset Password
7	Ghina Azizah غينا اذيا	082258086166 ghinaazizah3006@gmail.com	1401301	Edit Hapus Reset Password

Figure 3: Admin Staff Display

The features are described in the book of guidelines for operating *Wikalah iBT*. The second component is the Test Development Team. The team has the authority to access the page <http://Wikalah.upi.edu/> after having been given a username and password by the admin. The test development team can access the test item feature (Input-edit-delete questions), hold tests, and see the test results of test takers. The team has the obligation to make test items, update the items, evaluate the items, hold tests, and publish *TOKEN* for users. The feature is shown in figure 4.



No	Soal	Test	Analisa	Aksi
1	اختر كلمة غير مناسبة من الكلمات في الكوسين والتي تراها خاطئة من حيث القواعد النحوية والصرفية (إشطار) (الفة) (الحوية) (لخيل) (إشطار) (إي بلاد الشام)	Qowaid asep	Jml dipakai : 2 Benar: 2. Salah: 2 Persentase benar : 0 %	Edit Hapus
2	اختر كلمة غير مناسبة من الكلمات في الكوسين والتي تراها خاطئة من حيث القواعد النحوية والصرفية (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار)	Qowaid asep	Jml dipakai : 2 Benar: 2. Salah: 2 Persentase benar : 0 %	Edit Hapus
3	اختر كلمة غير مناسبة من الكلمات في الكوسين والتي تراها خاطئة من حيث القواعد النحوية والصرفية (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار)	Qowaid asep	Jml dipakai : 2 Benar: 2. Salah: 2 Persentase benar : 0 %	Edit Hapus
4	اختر كلمة غير مناسبة من الكلمات في الكوسين والتي تراها خاطئة من حيث القواعد النحوية والصرفية (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار) (إشطار)	Qowaid	Jml dipakai : 3	Edit Hapus

Figure 4: Test Item Feature

The third component is the user. A user has the authority to access the page <http://Wikalah.upi.edu/> after s/he registers and gets a confirmation email from the system. However, users will not be able to take the test if the Test Development

Team has not provided *TOKEN*. Users can access the test results after completing the test, and they can print their own certificates after fulfilling their obligations, namely registering on the registration page, paying the test fees, and working on the test.

In the next step, test takers/users are ready to work on the test. On this page, the test takers can access a number of features, namely, item numbers that can be accessed randomly, test items/questions, audio questions, answer choices, time allocation, “next question” button, “hesitation” button, and “complete” button. This feature is displayed in Figure 5.



Figure 5: Test Features Accessible by Test Takers

CONCLUSION

The results of the analysis of the 140 test items from the linguistic aspect show that there were 25 items that were incorrect and as many as 115 test items were correct. The errors include morphological errors, morphosyntactic errors, and technical errors related to typing. The test items were subsequently tested for validity and reliability. The results show that of the 140 items only 81 were valid and 59 items were invalid. Overall, the average level of reliability of the test items for proficiency aspects of listening, reading, and mastery of language was 0.67. This average score indicates that the test was quite reliable. In terms of the level of difficulty of the test items or questions, the difficult questions made up 22.14%, the moderate ones 63.57%, and the easy questions 9.29%. If these percentages are compared to the proportion of 20% of difficult, 60% of moderate, and 20% of easy questions, then the test items did not meet the ideal proportion because the proportion for easy questions was only 9.29%.

After being tested, there were 24 participants as test takers who met the minimum passing score, namely 425 points, and only 6 participants who did not meet the minimum score. The figures indicate that the test items were included in the easy category because only 20% of the test takers failed because according to the criteria, a test is said to be easy if only 27% of the test takers fail. All valid items

were then uploaded into the internet-based application and named *Wikalah* iBT which stands for internet-based *Taqwim Kafa`ah Lughah `Arabiyah* test.

Some of the conclusions above show some problems that need to be addressed. Therefore, the researchers suggest that the test package be added and improved so that it is not easy, added with *kitabah* and *tarekat* proficiency, and *Wikalah* iBT be evaluated empirically so that reliability can be objectively proven.

REFERENCES

- Ali, M. (1982). *Penelitian Kependidikan Prosedur dan Strategi*. (Bandung: Angkasa Bandung).
- An-Naqoh, M.K. (1985). *Ta'limul Lughah al-Arabiyah Linnathiqina Bilughatiin Ukhra*. Al-Mamlakah al-Arabiyah as-Saúdiyah: Jamiáh Ummul Qura.
- Azmi, A. M., Al-Jouie, M. F., & Hussain, M. (2019). AAEE–Automated evaluation of students' essays in Arabic language. *Information Processing & Management*, 56(5), 1736-1752. <https://doi.org/10.1016/j.ipm.2019.05.008>
- Bahari, A. (2020). Computer-assisted language proficiency assessment tools and strategies. *Open Learning: The Journal of Open, Distance and e-Learning*, 1–27. <https://doi.org/10.1080/02680513.2020.1726738>
- Barnes, M. (2016). The washback of the TOEFL iBT in Vietnam. *Australian Journal of Teacher Education*, 41(7), 158-174. <https://doi.org/10.14221/ajte.2016v41n7.10>
- Culbertson, G., Andersen, E., & Christiansen, M. H. (2020). Using Utterance Recall to Assess Second Language Proficiency. *Language Learning*, 70(S2), 104-132. <https://doi.org/10.1111/lang.12399>
- De Graaf, A. (2021). *Challenges in Developing Standardized Tests for Arabic Reading Comprehension for Secondary Education in the Netherlands*. *Challenges in Language Testing Around the World: Insights for language test users*, 181.
- Dimova, S., & Kling, J. (2018). Assessing English-medium instruction lecturer language proficiency across disciplines. *TESOL Quarterly*, 52(3), 634-656. <https://doi.org/10.1002/tesq.454>
- Fleckenstein, J., Keller, S., Krüger, M., Tannenbaum, R. J., & Köller, O. (2020). Linking TOEFL iBT® writing rubrics to CEFR levels: Cut scores and validity evidence from a standard setting study. *Assessing Writing*, 43, 1-15. <https://doi.org/10.1016/j.asw.2019.100420>
- Frost, K., Clothier, J., Huisman, A., & Wigglesworth, G. (2020). Responding to a TOEFL iBT integrated speaking task: Mapping task demands and test takers' use of stimulus content. *Language Testing*, 37(1), 133-155. <https://doi.org/10.1177/0265532219860750>
- Ginther, A., & Yan, X. (2018). Interpreting the relationships between TOEFL iBT scores and GPA: Language proficiency, policy, and profiles. *Language Testing*, 35(2), 271-295. <https://doi.org/10.1177/0265532217704010>

- Gu, L., Davis, L., Tao, J., & Zechner, K. (2021). Using spoken language technology for generating feedback to prepare for the TOEFL iBT® test: a user perception study. *Assessment in Education: Principles, Policy & Practice*, 28(1), 58-76. <https://doi.org/10.1080/0969594X.2020.1735995>
- Harsch, C., Ushioda, E., & Ladroue, C. (2017). Investigating the predictive validity of TOEFL iBT® test scores and their use in informing policy in a United Kingdom university setting. *ETS Research Report Series*, 2017(1), 1-80. <https://doi.org/10.1002/ets2.12167>
- Hasib, A. (2004). *Maharatut Ta'limil Qira'ah wal Kitabah Lil athfal*. Amman: Dar Wail Liththibaah Wannasyr.
- Karjo, C. H., & Ronaldo, D. (2019). The validity of TOEFL as entry and exit college requirements: Students' perception. In *Eleventh Conference on Applied Linguistics (CONAPLIN 2018)* (pp. 326-330). Atlantis Press. <https://doi.org/10.2991/conaplin-18.2019.277>
- Kim, E. Y. J. (2017). The TOEFL iBT writing: Korean students' perceptions of the TOEFL iBT writing test. *Assessing Writing*, 33, 1-11. <https://doi.org/10.1016/j.asw.2017.02.001>
- Kotby MN, Khairy A, Barakah M, Rifaie N, Elshobary A (1995) Language testing of Arabic speaking children: proceedings of XXIII world congress of the international Association of Logopedics and Phoniatics. Ain Shams University, Cairo.
- Loomis, S. (2015). Using speaking test data to define the advanced proficiency level for L2 Arabic speakers. *Foreign Language Annals*, 48(4), 604-617. <https://doi.org/10.1111/flan.12167>
- Madkur, A.A. (2000). *Tadris Fununul Lughah al-Arabiyyah*. Cairo: Dar El Fikr al-Arabi.
- Mousavi, S.A., Saleh Arizavi, & Namdar, Namdari. (2014). The effect of test preparation on the test performance: The case of the IELTS and TOEFL iBT reading tests. *International Journal of English and Education*, 3(2), 10-22.
- Neumann, H., Padden, N., & McDonough, K. (2019). Beyond English language proficiency scores: Understanding the academic performance of international undergraduate students during the first year of study. *Higher Education Research & Development*, 38(2), 324-338. <https://doi.org/10.1080/07294360.2018.1522621>
- Papageorgiou, S., & Manna, V. F. (2021). Maintaining access to a large-scale test of academic language proficiency during the pandemic: The launch of TOEFL iBT Home Edition. *Language Assessment Quarterly*, 18(1), 36-41. <https://doi.org/10.1080/15434303.2020.1864376>
- Peng Y, Yan W, Cheng L. (2021). Hanyu Shuiping Kaoshi (HSK): A multi-level, multi-purpose proficiency test. *Language Testing*, 38(2), 326-337. <https://doi.org/10.1177/0265532220957298>.
- Rifaie N (1994) The construction of an Arabic test to evaluate child language: MD thesis submitted to phoniatic unit. Ain Shams University, Cairo.

- Roud, L. F. P., & Hidri, S. (2021). Toward a sociocultural approach to computerized dynamic assessment of the TOEFL iBT listening comprehension test. *Education and Information Technologies*, 1-26. <https://doi.org/10.1007/s10639-021-10498-z>
- Sudjana, N. (1997). *Tuntunan Penyusunan Karya Ilmiah, Makalah, Skripsi, Tesis, Desertasi*. (Bandung : Sinar Baru Algensindo).
- Sudjana, N. and Ibrahim. (1989). *Penelitian dan Penelitian Pendidikan*, (Bandung : Sinar Baru).
- Sukmadinata, N. S. (2002). *Metode Penelitian Pendidikan*. (Bandung: Rosda Karya).
- Surakhmad, W. (1998). *Pengantar Penelitian Ilmiah Dasar Metode Tehnik*, (Bandung: Transito).
- Thu'aimah, R.A. (1985). *Dalilu 'Amalin fi I'dadil Mawad at-Ta'limiyyah Libaramiji Ta'limil Lughah*. Makkah: Jami'ah Ummul Qura.
- Thuáimah, R.A. (1989). *Ta'limul Lughah Al-Arabiyyah Lighairin Nathiqina Biha wa Manahijih, wa Asalibihi*. Egypt: Mansyurat al-Islamiyyah Littarbiyah wal Úlum Watstsaqafah.
- Ubaidat, D. (1984). *Al-Bahts al-'Ilmi*. Amman: Darul Fikr.
- Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System*, 65, 139-150. <https://doi.org/10.1016/j.system.2016.12.013>
- Veronika, N., & Darnia, Y. (2015). Rancang bangun aplikasi tes TOEFL menggunakan algoritma quick sort berbasis computer (Designing TOEFL test application using computer-based algorithms quick sort). *Jurnal Pseudocode*, 2(2) 89-97.
- Zubaidi, A. (2020). Development Of Mahârah Al-Istimâ'Test Instrument For Electronic Based Arabic Student Using The Kahoot! Application. *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 7(1), 61-77. <https://doi.org/10.15408/a.v7i1.13002>