DEVELOPMENT OF *MAHÂRAH AL-ISTIMÂ'* TEST INSTRUMENT FOR ELECTRONIC BASED ARABIC STUDENT USING THE *KAHOOT!* APPLICATION

Ahmad Zubaidi
Universitas Islam Indonesia, Yogyakarta, Indonesia
Jl. Kaliurang KM. 14.5, Kabupaten Sleman, Yogyakarta, 55584, Indonesia
Corresponding E-mail: ahmad.zubaidi@uii.ac.id

Abstract

This study aimed to develop the feasibility of the *Mahârah al-Istimâ'* test instrument for electronic-based Arabic students using the *Kahoot!* application at UIN Sunan Kalijaga Yogyakarta. The method used in this research is the research and development of Borg and Gall model through analysis, design, and testing. Eligibility is based on the expert validator test and the final operation field test. The results of the study can be concluded that: 1) Preparation of the Test Instrument refers to the 5 objectives or indicators of Mahmud Kamil an-Naqah into 50 questions in product design using the *Kahoot!* application, 2) From the results of the expert validator it is known that the results of the material quality are obtained by an average 5.37 and the media quality of the application is obtained by an average of 4.75 (very feasible), 3) After being obtained from the main field test to 20 students there are 6 questions or 12% that are not feasible because they are invalid and have not discrimination index, then the author revise and test the revision results to 70 students with 100% valid results and are suitable for use.

Keywords: *test instrument, mahârah al-Istimâ', Kahoot! application*

Introduction

The activity of assessing the process and learning outcomes is an educational activity that is often under-taken seriously by teachers and lecturers. For example, there are still many teachers and lecturers making assessment test instruments without going through the steps of developing test instruments properly.[1] Some criteria are grouped traditional approaches - This approach although classified as traditional This approach is oriented towards practices aimed at the development of the intellectual aspects of students. The aspects of skills and attitude development received less serious attention. In other words, students are only required to master the subject.

---

[1] Abdul Munip, *Penilaian Pembelajaran Bahasa Arab,* (Yogyakarta: Fakultas Ilmu Tarbiyah dan Keguruan, 2019), iii.

Evaluation activities are also more focused on product components only, while process components tend to be ignored. According to Zainal Arifin in addition to the traditional approach there is such a thing as the Systems Approach In contrast to the traditional approach which only touches the product components, a systematic approach is an approach that is focused on the evaluation component which includes the requirements and feasibility components, input components, process components, product component. These components must be the basis of consideration in the systematic evaluation of learning[2] - to measure the quality of test instruments including validity, reliability, difficulty level of different items, but it is still used to test learning evaluation tools. A modern approach, one of which uses a logistical approach known as IRT - IRT stands for Item Response Theory was first proposed in the field of psychometry for assessment skills, which is increasingly widely used in education to calibrate and evaluate items in tests, questionnaires , instruments and to assess subjects about abilities, attitudes, or other latent traits.[3] Where the software is based on open source, making it easily accessible and does not violate copyright regulations. Modern analysis or better known as the item response theory is more perfect in acquiring the results of the analysis. Weaknesses in classical theory such as the inability to estimate item parameters can be noted further.

Human resources quality is prerequisite of developed country, particularly school. Meanwhile, for attaining the great human resources, there must possess a set of good instrument test for assessing them.[4] *Mahârah al-Istimâ'* (listening skill) is ability of both catching and understanding (receptive) what is heard by other people.[5] As one out of four language skills, listening is skill which enable language user to comprehend orally language used. This competence is highly significant for being owned of every single language user due to the numerous orally daily communication. Without good listening skill, it is going to evoke ample misconception within communication between language users which can cause various obstacles in doing daily work. Therefore, listening skill may not be ignored in teaching language which has goal in comprehending the language.[6] A survey conducted by Hussein shows that Analysis of the responses reveals that Arabic instructors consider listening comprehension an important element in the curriculum. They avail themselves of a wide variety of

---

[2] Zainal Arifin, *Evaluasi Pembelajaran,* (Bandung: PT. Remaja Rosdakarya, 2011), 85.

[3] R. D. Hays, L. S. Morales, and S. P. Reise, "Item Response Theory and Health Outcomes Measurement in the 21st Century", *Medical Care,* Vol. 38, No. 9, 2000: II 28-42.

[4] Wiwiek Tamsyani, "Pengaruh Model Pembelajaran Dan Kesadaran Metakognitif Terhadap Hasil Belajar Peserta Didil SMA dalam Materi Pokok Asam Basa", *Journal of Educational Science and Technology (EST),* Vol. 2, No. 1, 2016, 10-25.

[5] Moh Matsna HS, "Metode dan Strategi Pembelajaran Bahasa Arab", (presented at the Diklat Fasilitator Guru Bidang Studi Bahasa dan Sastra Arab pada Madrasah Tsanawiyyah oleh Departemen Agama RI, Jakarta, 18 Mei 2004), 9.

[6] M. Soenardi Djiwandono, *Tes Bahasa Dalam Pengajaran,* (ITB Bandung, 1996), 54.

imaginative teaching techniques and ancillary resources to improve their students' listening skills.[7]

Listening skill engages understanding the meaning of oral language form. The Ability to understanding -oral language is the target of the assessment and evaluation of listening skills. In other hands, teachers, as one of the primary factors in elevating the quality of education, should be skilled in assessing the proses and outcome of the learning including developing a set of instrument evaluations, processing data, diagnosing the obstacle of the learning, and utilizing the result of the assessment.[8] However, what has happened up to now is that the instrument is not fulfilling the ideal requirements the both those used for daily tests and those used in the selection of prospective new students by the campus for entrance examinations and national exam questions that have long been known in Indonesia.

Even so, in Qatar research has been carried out by Shaalan who describes the reliability and validity of four language tests: The Sentence Comprehension Test (SCT), the Expressive Language Test (ELT), the Sentence Repetition Test (SRT), and the Arabic Picture Vocabulary Test (APVT). These tests were administered to two groups of Qatari Arabic-speaking children: A typically developing group (n=81 to 88) aged 4;6-9;4 years old and a chronologically age-matched group with specific language impairment (SLI) (n=26). The results of the four language tests showed high levels of reliability and validity and support the usefulness of these tools to diagnose children with SLI, whose performance on the tests was mostly consistent with findings in other languages.[9] Different from what is in Indonesia, there exists a thesis which analyzes the quality of exam questions of MTs in D.I. Yogyakarta. It reveals that the question developed by Arabic MGMP of MTs possesses poor instrument quality in validity and distractor function that theoretically there are many questions that have to be revised.[10] It is in line with Indonesia, there has not a set of Arabic instrument Internasional standardized up to now, especially for speaking and listening skills.

Listening is totally different from hearing. If someone is hearing something, it has been certain that He or She pays attention to what they heard. Listening is activity of both consciously hearing and paying attention towards what they are hearing.[11] Even, Linguists distinguish three terms related to hearing, which heard (*Simâ'*), listening (*Istimâ'*) and heard with serious or carefully (*inshâth*). *Sima'* is only limited acceptance of the ear to the frequency of the sound from the source specified without the concern that accidental, like hearing the sound of plan or train . *Simâ'* is a simple

---

[7] Hussein Elkhafaifi, "Teaching Listening in the Arabic Classroom: A Survey of Current Practice", *al-'Arabiyya,* Vol. 34, 2001, 55–90.

[8] Komang Ayu Sri Inteni, I Made Candiasa, and Ni Ketut Suarni, "Pengembangan Instrumen Tes Objektif Pilihan Ganda yang Diperluas Berbasis Web untuk Mata Pelajaran TIK Kelas XI SMAN Di Kabupaten Karangasem", *Jurnal Penelitian dan Evaluasi Pendidikan Indonesia,* Vol. 3, No. 1, 2013.

[9] Saleh Shaalan, "Reliability and Validity of Four Arabic Language Tests: A Comparison of Performance of Qatari School-Aged Children with and Without Language Impairment", *Arab Journal of Applied Linguistics,* Vol. 2, No. 1, 2017, 20-48.

[10] Abdul Munip, *Penilaian Pembelajaran Bahasa Arab,* iii.

[11] Abdul Munip, *Penilaian Pembelajaran Bahasa Arab,* 161.

process that does not need to be learned. While *Istimâ'* is a process that requires the attention of a special and deliberate on what is heard, like listening to the explanation of teachers or lecturers. As for *inshâth* is the level at the top of *Istimâ'* which requires both concentration and attention more is kept constantly to realize the goal specific, such as listening to al-Qur'ān and sermons. *Istimâ'* sometimes disconnected or there is a pause in the process of listening, while *inshâth* not be interrupted and must continue to be continued in listening.[12] In conclusion, the good instrument of *Mahârah al-Istimâ'* is giving the instrument including the audio-visual, not only simply the audio, since it is still categorized as hearing not listening.

One of the alternative test instrument that is able to connect the concept of *Mahârah al-Istimâ'* is by giving the audio, animation, or other picture illustration, and being hoped that the students are able to apply in their daily life later on. The new generation requires rapid access and quick rewards, is impatient with linear thinking and displays a novel capacity for multi-tasking.[13] Research on the use of mobile devices (phones, tablets, etc.) in education has gained currency in recent years; varying from improving the foreign language.[14] Games, on the other hand, have been found to be beneficial for academic achievement, motivation and classroom dynamics in K-12[15], Game-based learning can be seen in both primary, and secondary schools, universities, adult education, military training, and medical practice.[16] Games can mainly be integrated into education in different ways, one of which can be to integrate them as part of a traditional classroom lecture to improve learning, motivation and engagement.[17] Digital games are designed to integrate content material with game play; this allows the brain to process information from short- to long-term memory.[18] One of the interactive game-based learning platform that is able to be the alternative is *Kahoot!*.[19]

Kahoot! is a game-based learning platform and CRS aimed at schools and universities. *Kahoot!* was launched in 2013 as a game-based response system (GSRS). Mobitroll has developed this application, which is a collaboration between the

---

[12] Ali Ahmad Madkur, *Tadris Funun al-Lughah al-'Arabiyah,* (Kairo: Dar al-Fikr al-'Arabi li al-Thaba'ah wa al-Nasyr, 1998), 83.

[13] Chris Jones, et.al., "Net Generation or Digital Natives: Is There a Distinct New Generation Entering University?", *Computers & Education,* Vol. 54, No. 3, 2010, 722-732.

[14] Vaida Kadyte, "Learning Can Happen Anywhere: A Mobile System for Language Learning", in *Learning with Mobile Devices: Research and Development*, edited by Jill Attewel and Carol Savill-Smith. (London: The Learning and Skills Development Agendy, 2004), 73–78.

[15] Ricardo Rosas et al., "Beyond Nintendo: Design And Assessment Of Educational Video Games For First And Second Grade Students", *Computers & Education,* Vol. 40, No. 1, 2003, 71-94.

[16] Nicola Whitton, "The Place of Game-Based Learning in an Age of Austerity", *Electronic Journal of e-Learning,* Vol. 10, No. 2, 2012, 249-256.

[17] Alf Inge Wang, "The Wear Out Effect of a Game-Based Student Response System", *Computers & Education,* Vol. 82, 2015, 217-227.

[18] Alison K Banikowski and Teresa a Mehring, "Strategies to Enhance Memory Based on Brain-Research", *Focus on Exceptional Children,* Vol. 32, No. 2, 1999, 1.

[19] Iwan Binanto, *Multimedia Digital - Dasar Teori dan Pengembangannya,* (Yogyakarta: Andi, 2010), 46.

Norwegian University of Science and Technology (NTNU) and the British company We Are Human.[20] *Kahoot!* first introduced in lectures at NTNU in 2012, and was still in beta version at the time. This application utilizes mixed learning, which combines face-to-face classroom methods with computer-mediated activities, creating an environment where students learn partly through online delivery and independent learning.[21] In 2014 *Kahoot!* was announced as the winner of the technological achievements of the year by Teknik Ukeblad[22], and in March 2014 *Kahoot!* has been played by more than 3 million players, and has been growing rapidly with 150,000 new users per week.[23] *Kahoot!* is an application, which is presented in the form of game, and it is aimed to engage the learners in the term of responding to quizzes discussions, and surveys. Learners who are involved in the game do not require an account because it could be accessed directly through a web browser in their own gadget, laptop or computer. Quizzes, discussion forums, or survey forms can becreated after the teacher created a new account on the *Kahoot!* page.[24]

Kahoot! is a game application with two different website addresses, https://kahoot.com/ for teachers and https://kahoot.it/ for learners. *Kahoot!* could be accessed freely (free of charge), including all the features in it. This application only requires an internet connection for the user to be able to play this application. Quizzes features consist of a multiple-choices question, which is available in four selections. Quizzes are not only in the form of written questions but could be inserted by pictures, videos, and songs to support critical thinking in understanding the questions. Questions and answers require a limited number of characters, 80 characters for the questions and 60 characters for the answer choices. In the quizzes features, teachers could set a time setting if needed.

Much research has been done on the development of tests and instruments. as well as *Kahoot!* and *Mahârah al-Istimâ'*. *The First* literature review on this variable was carried out by Jian Cheng et.al researching about "Automatic Assessment of Spoken Modern Standard Arabic" that they present the design and validation of the Versant Arabic Test, a fully automated test of spoken Modern Standard Arabic, that evaluates test-takers' facility in listening and speaking. Experimental data shows the test to be highly reliable (testretest r=0.97) and to strongly predict performance on the ILR OPI

[20] Nafiye Çigdem Aktekin, Hatice Çelebi, and Mustafa Aktekin, "Let's Kahoot! Anatomy", *International Journal of Morphology,* Vol. 36, No. 2, 2018, 716-721.

[21] C.J Bonk and C.R Graham, *The Handbook of Blended Learning: Global Perspective and Local Design*, Eds., Eds (San Fransisco: CA: Pfeiffer Publishing, 2006).

[22] Sigurd Øygarden Flæten, "Sjokkert og glad vinner av Teknologibragden", *Tu.no*, last modified February 6, 2014, accessed January 10, 2020, https://www.tu.no/artikler/sjokkert-og-glad-vinner-av-teknologibragden/226876.

[23] Greg, "This Norwegian EdTech Startup Is Growing 150,000 Users a Week", *ArcticStartup*, March 18, 2014, accessed January 10, 2020, https://arcticstartup.com/this-norwegian-edtech-startup-is-growing-100000-users-a-week/.

[24] Ryan Dellos, "Kahoot! A Digital Game Resource for Learning", *International Journal of Instructional Technology and Distance Learning,* Vol. 12, No. 4, 2015, 49-52.

(r=0.87), a standard interview test that assesses oral proficiency.[25] *Second,* research about "Developing of Listening Subject in Test of Arabic as a Foreign Language based on Test of English as a Foreign Language TOEFL at State Islamic University Maulana Malik Ibrahim of Malang in East Java, Indonesia" The research shows that the developing of listening material of TOAFL based on TOEFL in UIN in the academic year 2010-1011 can improve listening score of the students. This has evidently proved that the result of pre-test and post-test that show significant improvement on the level 5,334. During conducting this research, the researcher feels the importance of university lecturers to follow the development of current software to support the teaching and learning process compared to the students who master using a computer.[26] *Third,* research about "Development of *Mahârah al-Istimâ'* Learning Materials 'Based on Lectora Inspire Media in Ulumuddin Lhokseumawe Private Madrasah Aceh" conclude that The use of *mahârah al-istimâ'* based on Lectora Inspire media is very effective in learning *mahârah al-istimâ'* with the average pre-test score for students of Madrasah Aliyah Ulumuddin is 60.45 and an average post-test value is 85.06 with a significance value of $p = 0,000 < 0.05$.[27] *Fourth,* research about "Arabic Language Learning Development Using *Kahoot!* Application In MTsN 2 Kota Malang" conclude that the development of test using *Kahoot!* is quite valid, reasonable and practically used in learning Arabic.[28] *Fifth,* research about "The Effectiveness of Utilizing Beesmart Software for the *Mahârah al-Istimâ'* and *Qirâ'ah* Test" conclude that the software is effective in improving Beesmart maharah test it ' and qira'ah. This is evidenced by an increase in the average value of which in value then at 72,5 pretes postes 78.2, with the difference between the value of 5.7. There are two factors that influence the effectiveness of software Beesmart i.e., focus student growing and the second due to simple.[29] The literature review above shows that the variables in this study give good results, but no one has developed the *Mahârah al-Istimâ'* test instrument using the *Kahoot!*.

The novelty in this study is to try to elaborate on the development of *Mahârah al-Istimâ'* test instrument for Arabic students electronic-based by using *Kahoot!* application. The asked question is how is the procedure of the development, and how is expert validation towards this instrument, also how is the apropriateness of the *Mahârah al-Istimâ'* test instrument electronic-based by using *Kahoot!* application.

---

[25] Jian Cheng et al., "Automated Assessment of Spoken Modern Standard Arabic", *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications,* (Boulder, Colorado: Association for Computational Linguistics, 2009).

[26] Muhammad Amiruddin, *Developing of Listening Subject in Test of Arabic as a Foreign Language based on Test of English as a Foreign Language TOEFL at State Islamic University Maulana Malik Ibrahim of Malang in East Java, Indonesia,* (Malang: Universitas Islam Negeri Maulana Malik Ibrahim, 2011).

[27] Muhammad Ediyani, "Pengembangan Materi Pembelajaran Maharah Al-Istima' Berbasis Media Lectora Inspire di Madrasah Aliyah Swasta Ulumuddin Lhokseumawe Aceh", *At-Tajdid: Jurnal Ilmu Tarbiyah,* Vol. 6, No. 2, 2017, 197–226.

[28] Rasydiana, "Pengembangan Tes Pembelajaran Bahasa Arab Menggunakan Aplikasi Kahoot Di MTsN 2 Kota Malang", *Prosiding Konfererensi Nasional Bahasa Arab,* Vol. 5, No. 5, 2019, 437-447.

[29] Miftahul Mufid, "Efektifitas Pemanfaatan Software Beesmart Untuk Tes Mahārah Istimā' Dan Qirā'ah", *Al-Murabbi: Jurnal Studi Kependidikan dan Keislaman,* Vol. 5, No. 1, 2018, 91–100.

Therefore, it is hoped that the *Mahârah al-Istimâ'* test instrument electronic-based by using *Kahoot!* application can be found.

Method

The researcher used Research and Development as the method of this research. The researcher uses the model of development of Borg and Gall, in the implementation and development of this research, the researcher used several methods, which are descriptive, evaluation, and experimental. Borg and Gall development research model include ten activities, which are the identification of problems (potential and problems), the data collection, products design, design validation, revision design, product main field testing, product revision, product operational field testing, final revision product and mass production.[30]

The type of data in this study is quantitative data, which is data that has numbers. The data collection is found by doing unstructured interviews, which is free interviews, documentation, which is used to obtain a collection of *Istimā'*, Questionnaire questions, questionnaire which is used for expert validation, and Tests/Trials , is used to see the ability of the testee to answer the items which the author has designed in the product.
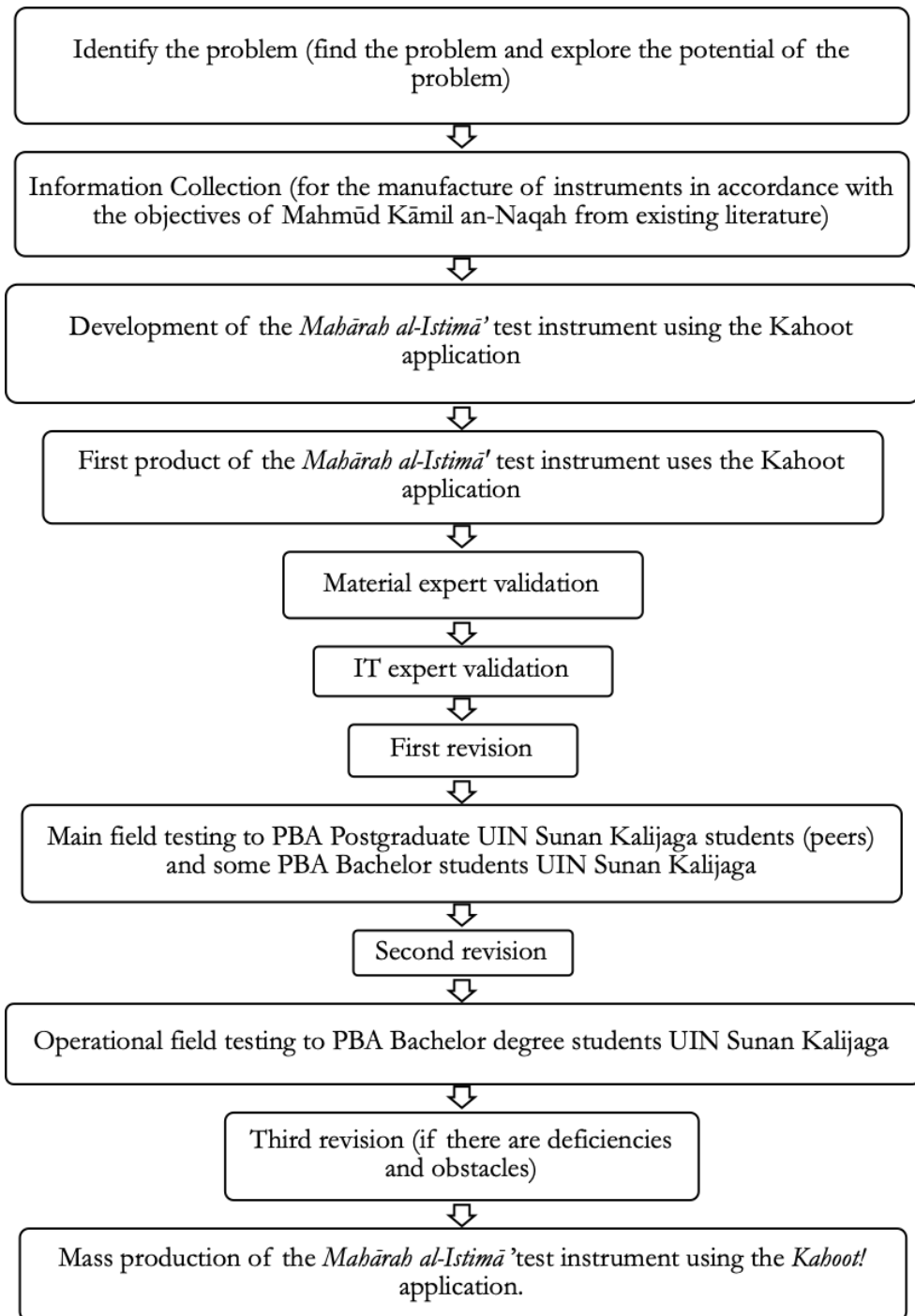
The sample used in this study were students of Arabic Language Education at UIN Sunan Kalijaga Yogyakarta undergraduate students and post-graduate students, 10 students from the graduate program and 10 students from post-graduate programs were used as respondents in the main field test, while 70 undergraduate students were used as respondents for the use or operation field test.

Identification of problems (potential and problems); The researchers first discovered the application, which can be used to support the process of making a test instrument based on the theory the researcher, picked which is *Kahoot!.* Data collection; On this the researcher collects the information and the references which would be used as the material for designing the product. Product design; At this stage, the researcher begins the author begins to make questions, the form of the questions is in line with the indicators which is wanted to be achieved. Design validation; The researcher on the validation and assessment steps were contemplate whether the question has been in line with the indicator or not by presenting expert instrument tests and expert e-learning/ IT. Design revision, Product testing and Revision; at this stage, the researcher revises the product, which has been validated through discussion with some experts by giving critics and advices for the product. Product usage testing, Final revision product, and Mass production; At this stage, the researcher does product operational field testing by asking the testee to bring their smartphone. In the end, the researcher showed and evaluated the score, which they got using Microsoft Excell 2016 instead of SPSS. After that, the score is used to be the material of questions analysis. This stage of revision would be done, if obstacles or deficiencies were still be found in the usage of the application in a wider scope. On the last step

---

[30] Sugiyono, *Metode Penelitian Pendidikan:(Pendekatan Kuantitatif, Kualitatif Dan R & D)* (Bandung: Alfabeta, 2008), 409.

the researcher does a mass production after getting approved on the effectiveness, efficiency and feasibility to be produced of the product after testing it.

Diagram 1. Process of research design

Identify the problem (find the problem and explore the potential of the problem)

⇩

Information Collection (for the manufacture of instruments in accordance with the objectives of Mahmūd Kāmil an-Naqah from existing literature)

⇩

Development of the *Mahārah al-Istimā'* test instrument using the Kahoot application

⇩

First product of the *Mahārah al-Istimā'* test instrument uses the Kahoot application

⇩

Material expert validation

⇩

IT expert validation

⇩

First revision

⇩

Main field testing to PBA Postgraduate UIN Sunan Kalijaga students (peers) and some PBA Bachelor students UIN Sunan Kalijaga

⇩

Second revision

⇩

Operational field testing to PBA Bachelor degree students UIN Sunan Kalijaga

⇩

Third revision (if there are deficiencies and obstacles)

⇩

Mass production of the *Mahārah al-Istimā'* test instrument using the *Kahoot!* application.

Result

The researcher start to making questions, the form of the questions is in accordance with the indicators to be achieved. researchers make 10 questions on each one indicator.

Appendix 1. Test instrument indicator

No.	Indicator	Description	Question Form	Number of Questions
1.	Students are able to distinguish Arabic sounds, intonation patterns, and compare their equivalents in mother tongue (Indonesian).	Mention the correct sentence of an Arabic utterance from native speaker	Multiple choice	7 items
		Mention the equivalent of sentence patterns spoken by native speakers	Multiple choice	3 items
2.	Students are able to find the global meaning of the message conveyed by the speaker.	Mention sentences that are not in the place mentioned by native speaker	Multiple choice	5 items
		Determine the number of an Arabic text that has been read by a native speaker.	Multiple choice	5 items
3.	Students are able to recall messages heard.	Question and answer material that is listened to based on memories that are remembered by students in accordance with Arabic conversation	Multiple choice	5 items
		Question and answer material that is listened to based on memories that are remembered by students in accordance with Arabic texts read by native speakers	Multiple choice	5 items
4.	Students are able to understand messages and interact with these messages.	Answering from the clue that has been spoken by native speakers.	Multiple choice	5 items
		Answering questions and filling in hiatus sentences spoken by native speakers	Multiple choice	5 items
5.	Students are able to discuss and apply hidden messages	Find the meaning implicit in a sentence spoken by native speaker	Multiple choice	10 items

The table above is a matrix of *Istimâ's* test grid as a whole. From the description contained in the matrix here, the author takes 50 questions in the form of multiple choice to be tested. The lattice matrix of the *Mahârah al-Istimâ'* test instrument was arranged according to the purpose of Istimā' according to Mahmud Kamil an-Naqah in the book *Ta'lîm al-Lugah al-'Arabiyyah Lighairi Nathiqîna Bihâ* which the writer formulated as a test indicator.

The test instrument produced gets through the validation stage by a team of experts. The expert team consists of material experts and e-learning experts. Both of

them gave an assessment through a questionnaire. Material eligibility by material experts is 5.37 (Very Good).

Appendix 2. Average score of the material expert validator

No.	Assessment Aspects	Average Score	Criteria
1.	Content	4,8	Very Good
2.	Construction	4,9	Very Good
3.	Language	6,25	Very Good
Amount		**15,95**	
Cumulative average score		**5,37**	**Very Good**

The feasibility of the media by e-learning experts is 4.75 (Very Good). After getting advice from experts the authors revised the product according to the expert's recommendations.

Appendix 3. Average score of the e-learning expert validator

No.	Assessment Aspect	Score	Criteria
1.	Attractive design	5	Very Good
2.	Layout	5	Very Good
3.	Font size	5	Very Good
4.	Legible writing	5	Very Good
5.	Ease of instructions	4	Good
6.	Attractive color composition	5	Very Good
7.	Ease of use	5	Very Good
8.	Useful score display	4	Good
Amount		**38**	
Average score		**4,75**	

Then, the product passes the simulation stage (Main Field Test) to determine the product's suitability before being used in the field. The author tested the Test Instrument on 20 Arabic Language Education students with 11 in Undergraduate Program and 9 in Postgraduate Program. Based on the main field test results, there were 3 students who had "Very Good" scores, 4 students had "Good" grades, 9 students got "Enough", 2 students had "Less", 2 people had "Very Less". Then, from the main field test, there are 44 valid questions and 6 questions in the category which are invalid because the correlation results are less than 0.444. However, the test instrument has a high reliability with a value of 0.928. The discrimination index of the instruments tested by the author was found to be 2 items which did not have a discrimination index (NAN), but in the test instrument there were no questions categorized as "Very Difficult" or "Very Easy".

Appendix 4. Recap analysis of main field test items

No.	Item of Question	Discrimination index	Difficulty level	Correlation	Significance of correlation
1.	10	Fair	Medium	0,234	Not Significant
2.	27	No Data	Medium	-0,546	NAN
3.	29	No Data	Medium	0,159	Not Significant
4.	30	Fair	Medium	0,192	Not Significant
5.	38	Fair	Medium	0,205	Not Significant
6.	44	Fair	Medium	0,039	Not Significant

Based on the recapitulation of the test items as a whole main field test, the authors concludes that out of the 50 questions, there were 6 questions that had to be revised/replaced or as much as 12% of the total questions, because of the 6 questions that were insignificant/invalid and not valid has a discrimination index. Therefore, the author revises/replaces the 6 questions to be tested on wider testee (trial of use/eligibility).

After the author revised the product and tested it (Operational Field Testing) to 70 bachelor degree students of Arabic Language Education at UIN Sunan Kalijaga Yogyakarta, the results were 3 students got "Very Good" scores, 16 students got "Good" grades, 36 students got "Fair" scores, 14 students scored "Less", and 1 student got a score of "Very Less". Then the item analysis found that out of the 50 questions tested by the author, all of them have distinguishing features, there are no item, which got difficulty in the "Very Difficult", and "Very Easy" category, all questions are significant/valid, and all items are reliable.

Appendix 5. Recap analysis of operational field test items

Item of Question	Discrimination index	Difficulty level	Correlation	Significance of correlation
1.	Fair	Medium	0,377	Very Significant
2.	Fair	Medium	0,304	Very Significant
3.	Poor	Medium	0,333	Very Significant
4.	Poor	Medium	0,386	Very Significant
5.	Fair	Medium	0,323	Very Significant
6.	Poor	Medium	0,261	Significant
7.	Fair	Difficult	0,349	Very Significant
8.	Fair	Medium	0,266	Significant
9.	Fair	Difficult	0,322	Very Significant
10.	Fair	Medium	0,429	Very Significant
11.	Fair	Medium	0,389	Very Significant
12.	Fair	Difficult	0,270	Significant
13.	Good	Medium	0,364	Very Significant
14.	Fair	Medium	0,399	Very Significant
15.	Fair	Medium	0,404	Very Significant
16.	Fair	Medium	0,321	Very Significant
17.	Fair	Medium	0,392	Very Significant
18.	Fair	Medium	0,270	Significant
19.	Fair	Medium	0,277	Significant
20.	Fair	Medium	0,269	Significant
21.	Fair	Medium	0,333	Very Significant

22.	Fair	Medium	0,258	Significant
23.	Poor	Medium	0,292	Significant
24.	Fair	Medium	0,245	Significant
25.	Fair	Medium	0,252	Significant
26.	Fair	Medium	0,281	Significant
27.	Fair	Medium	0,276	Significant
28.	Fair	Medium	0,271	Significant
29.	Poor	Medium	0,322	Very Significant
30.	Fair	Medium	0,289	Significant
31.	Good	Medium	0,383	Very Significant
32.	Fair	Medium	0,256	Significant
33.	Fair	Medium	0,306	Very Significant
34.	Fair	Medium	0,309	Very Significant
35.	Fair	Medium	0,349	Very Significant
36.	Fair	Medium	0,250	Significant
37.	Poor	Medium	0,259	Significant
38.	Good	Medium	0,353	Very Significant
39.	Fair	Medium	0,275	Significant
40.	Fair	Medium	0,300	Very Significant
41.	Fair	Medium	0,271	Significant
42.	Fair	Medium	0,262	Significant
43.	Fair	Medium	0,242	Significant
44.	Fair	Medium	0,284	Significant
45.	Fair	Medium	0,287	Significant
46.	Fair	Medium	0,267	Significant
47.	Fair	Medium	0,304	Very Significant
48.	Fair	Medium	0,241	Significant
49.	Good	Medium	0,283	Significant
50.	Poor	Medium	0,236	Significant

Discussion

First, answering how the procedure for developing the *Mahârah al-Istimâ'* Test Instrument for Electronic-Based Arabic Language Students Using the *Kahoot!* Application. Based on interviews and problem findings, which are not in line with the theory, the author takes action by developing electronic based test instruments, one of them is using the *Kahoot!* application. Followed by the design phase with the *Kahoot!* application on the website www.kahoot.com, which students would access it on their use on the website www.kahoot.it.

Second, answering the results of the experts' validation. The validation results are determined from the validator evaluation results. Material and e-learning experts aim to evaluate how feasible the material, the media and provide suggestions or revisions if necessary. The value of material worthiness by material experts was 5.37 (Very Good) and the value of media worthiness by e-learning experts was 4.75 (Very Good). This is included in the criteria for the product to be used. Material experts suggest the need to replace question items from credible sources, *mutakallim* should be taken directly from the native speaker and modeled on the English test. Furthermore, media experts suggest the need to add an explanation to the media used. Based on the

validation from the experts, the author revised 1 item. The following display before and after revision.

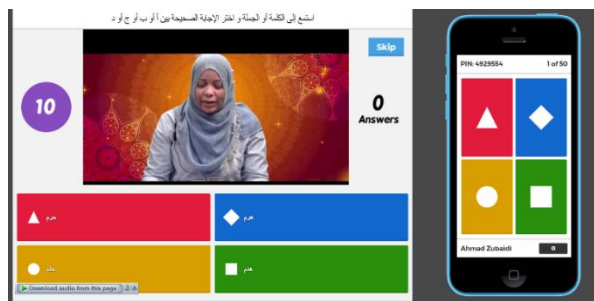**Figure 1** : views before being revised

Figure 2 : views after being revised

Third, answering the product eligibility. The eligibility of this product is determined from the product operational field test by looking at the results and analysis of the items. From the results of the operational field test, it was found that 3 students received "Very Good" scores, 16 students received "Good" scores, 36 students received "Fair" scores, 14 students received "Less" scores, and 1 students received "Very Less" scores. Then, from the item analysis, it was found that out of the 50 questions tested by the author, all of them have distinguishing features, there is no any items which have difficulty in the "Very Difficult" and "Very Easy" category, all questions are significant/valid, and all items are reliable. It can be concluded from the results of the author's last test that the *Mahârah al-Istimâ '* test instrument for electronic-based Arabic students using the *Kahoot!* application was appropriate to be used to test the ability of *Mahârah al-Istimâ'* Arabic students. This test is very suitable to test the ability to hear Arabic undergraduate and master students but this has not been tested on doctoral students so it is unknown whether this test can be used or not for doctoral students. The instrument's validity is very valid and reliable. Following are some of the displays in the *Mahârah al-Istimâ'* Test Instrument developed by the author:
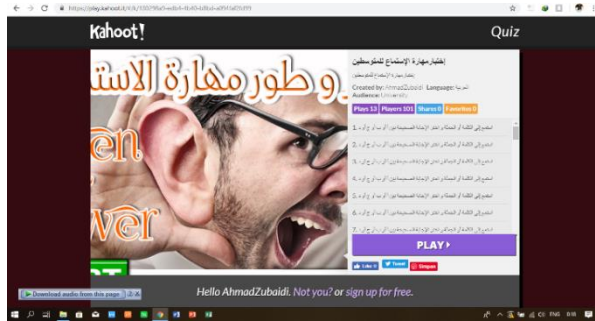
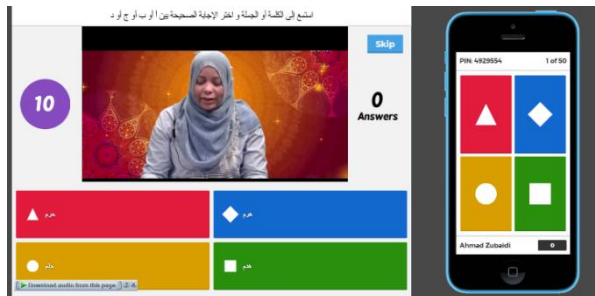**Figure 3** : the *opening* view before *starting*
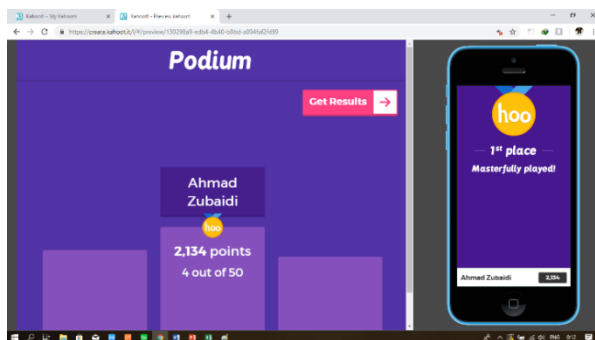

Figure 4 : question display

Figure 5 : the display of final winning

Figur 6 : the result of *final score* in the form of Ms. Excell

Conclusion

Preparation of the Instrument This test refers to the 5 objectives/indicators of Mahmūd Kāmil an-Naqoh. The author compiled 50 questions that were tested with 10 stages based on the development of the Borg and Gall model, the authors analyzed the problem and found the potential, collected data related to the *Mahārah al-Istimā'* test, product design using the *Kahoot!* application, validation to material experts and e-learning, product revisions based on expert advice, trialling limited products before the feasibility test to 20 students, product revisions from main field test results, product usage trials that were tested on 70 students, and mass production.

From the results of expert validators it can be seen that the material quality results are obtained with an average of 5.37 and the media quality of the application is obtained with an average of 4.75. Thus, this product according to expert validator is feasible to use. After being revised according to the experts' recommendations, of the 50 questions that were tested were main field test to 20 Arabic Department as known as PBA UIN Sunan Kalijaga Yogyakarta students, there were 6 questions or 12% which had to be revised because they were invalid and have not discrmination index. After being revised, the authors tested the use of the product to 70 students of PBA UIN Sunan Kalijaga Yogyakarta and obtained from 50 questions overall or 100% feasible to use because all of them have a validation level/validity, have a discrimination index, and nothing is categorized as "Very Easy" and "Very difficult".[]

REFERENCES

Aktekin, Nafiye Çigdem., Hatice Çelebi, & Mustafa Aktekin. "Let's *Kahoot!* Anatomy", *International Journal of Morphology,* Vol. 36, No. 2, 2018.

Amiruddin, Muhammad. *Developing of Listening Subject in Test of Arabic as a Foreign Language based on Test of English as a Foreign Language TOEFL at State Islamic University Maulana Malik Ibrahim of Malang in East Java, Indonesia.* Malang: Universitas Islam Negeri Maulana Malik Ibrahim, 2011.

Arifin, Zainal. *Evaluasi Pembelajaran.* Bandung: PT. Remaja Rosdakarya, 2011.

Banikowski, Alison K., & Teresa a Mehring. "Strategies to Enhance Memory Based on Brain-Research", *Focus on Exceptional Children,* Vol. 32, No. 2, 1999.

Binanto, Iwan. *Multimedia Digital-Dasar Teori dan Pengembangannya.* Yogyakarta: Andi, 2010.

Cheng, Jian., Jared Bernstein, Ulrike Pado, & Masanori Suzuki. "Automated Assessment of Spoken Modern Standard Arabic", *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications*, Colorado: Association for Computational Linguistics, 2009.

Dellos, Ryan. "*Kahoot!* A Digital Game Resource for Learning", *International Journal of Instructional Technology and Distance Learning,* Vol. 12, No. 4, 2015.

Djiwandono, M. Soenardi. *Tes Bahasa dalam Pengajaran.* Bandung: ITB Bandung, 1996.

Ediyani, Muhammad. "Pengembangan Materi Pembelajaran *Maharah al-Istima'* Berbasis Media Lectora Inspire Di Madrasah Aliyah Swasta Ulumuddin Lhokseumawe Aceh", *At-Tajdid: Jurnal Ilmu Tarbiyah,* Vol. 6, No. 2, 2017.

Elkhafaifi, Hussein. "Teaching Listening in the Arabic Classroom: A Survey of Current Practice", *al-'Arabiyya,* Vol. 34, 2001.

Flæten, Sigurd Øygarden. "Sjokkert og glad vinner av Teknologibragden", *Tu.no.* Last modified February 6, 2014. Accessed January 10, 2020. https://www.tu.no/artikler/sjokkert-og-glad-vinner-av-teknologibragden/226876.

Greg. "This Norwegian EdTech Startup Is Growing 150,000 Users a Week", *ArcticStartup*, March 18, 2014. Accessed January 10, 2020. https://arcticstartup.com/this-norwegian-edtech-startup-is-growing-100000-users-a-week/.

Hays, R. D., L. S. Morales, & S. P. Reise. "Item Response Theory and Health Outcomes Measurement in The 21st Century", *Medical Care,* Vol. 38, No. 9, 2000.

HS, Moh Matsna. "Metode dan Strategi Pembelajaran Bahasa Arab" presented at the Diklat Fasilitator Guru Bidang Studi Bahasa dan Sastra Arab pada Madrasah Tsanawiyyah oleh Departemen Agama RI, Jakarta, 18 Mei 2004.

Inteni, Komang Ayu Sri., I Made Candiasa, & Ni Ketut Suarni. "Pengembangan Instrumen Tes Objektif Pilihan Ganda Yang Diperluas Berbasis Web Untuk Mata Pelajaran TIK Kelas XI SMAN Di Kabupaten Karangasem", *Jurnal Penelitian dan Evaluasi Pendidikan Indonesia,* Vol. 3, No. 1, 2013.

Jones, Chris., Ruslan Ramanau, Simon Cross, & Graham Healing. "Net Generation or Digital Natives: Is There a Distinct New Generation Entering University?", *Computers & Education,* Vol. 54, No. 3, 2010.

Kadyte, Vaida. "Learning Can Happen Anywhere: A Mobile System for Language Learning", In *Learning with Mobile Devices: Research and Development*, Edited by Jill Attewel and Carol Savill-Smith. London: The Learning and Skills Development Agendy, 2004.

Madkur, Ali Ahmad. *Tadris Funun al-Lughah al-'Arabiyah.* Kairo: Dar al-Fikr al-'Arabi li al-Thaba'ah wa al-Nasyr, 1998.

Mufid, Miftahul. "Efektivitas Pemanfaatan *Software Beesmart* Untuk Tes *Mahārah Istimā'* dan *Qirā'ah*", *Al-Murabbi: Jurnal Studi Kependidikan dan Keislaman,* Vol. 5, No. 1, 2018.

Munip, Abdul. *Penilaian Pembelajaran Bahasa Arab.* Yogyakarta: Fakultas Ilmu Tarbiyah dan Keguruan, 2019.

Rasydiana. "Pengembangan Tes Pembelajaran Bahasa Arab Menggunakan Aplikasi *Kahoot!* Di MTsN 2 Kota Malang", *Prosiding Konfererensi Nasional Bahasa Arab,* Vol. 5, No. 5, 2019.

Rosas, Ricardo., Miguel Nussbaum, Patricio Cumsille, Vladimir Marianov, Mónica Correa, Patricia Flores, & Valeska Grau. "Beyond Nintendo: Design and Assessment of Educational Video Games for First and Second Grade Students", *Computers & Education,* Vol. 40, No. 1, 2003, 71–94.

Shaalan, Saleh. "Reliability and Validity of Four Arabic Language Tests: A Comparison of Performance of Qatari School-Aged Children With And Without Language Impairment", *Arab Journal of Applied Linguistics,* Vol. 2, No. 1, 2017.

Sugiyono. *Metode Penelitian Pendidikan: (Pendekatan Kuantitatif, Kualitatif Dan R & D).* Bandung: Alfabeta, 2008.

Tamsyani, Wiwiek. "Pengaruh Model Pembelajaran dan Kesadaran Metakognitif Terhadap Hasil Belajar Peserta Didil SMA dalam Materi Pokok Asam Basa", *Journal of Educational Science and Technology (EST),* Vol. 2, No. 1, 2016.

Wang, Alf Inge. "The Wear Out Effect of a Game-Based Student Response System", *Computers & Education,* Vol. 82, 2015.

Whitton, Nicola. "The Place of Game-Based Learning in an Age of Austerity", *Electronic Journal of e-Learning,* Vol. 10, No. 2, 2012.