

تصميم الاختبار لمهارة الكلام العربي وفقا للإطار المرجعي الأوروبي المشترك
للغات (CEFR) لطلاب الجامعة في إندونيسيا

**The Development of Arabic Speaking Test According to the
Common European Framework of Reference for Languages
(CEFR) for University Students in Indonesia**

Agus Yasin¹, Hidia Tarauni^{2*}

Universitas Darussalam Gontor, Indonesia^{1,2}

elyasien@unida.gontor.ac.id¹, hidiatarauni@gontor.ac.id^{2*}

Cite this article:

Yasin, A., & Tarauni, H. (2023). Tashmim al-Ikhtibar li Maharah al-Kalam al-‘Arabi Wafqhan li al-Ithar al-Marja’i al-Urubi al-Musyarak li al-Lughat li al-Thullab al-Jami’ah fi Indunisiya. *Arabiyatuna: Jurnal Bahasa Arab*, 7(1 May), 111-134. doi:<http://dx.doi.org/10.29240/jba.v7i1.6390>

Received: 10-01-2023

Revised: 29-03-2023

Accepted: 14-04-2023

Abstract

This study aimed to develop a speaking-skill test at the B1 of CEFR, to determine the levels of validity, reliability, and difficulty of each item, and to organize the distinguishing power of each item on the prepared test. This research was motivated by the need of Arabic Language Test as an evaluation activity that required for various purposes. Several universities in Indonesia have also intended the existence of a language test as a requirement for continuing studies to a higher degree or to a place where linguistic proficiency is required or work somewhere. Also, the most important is to upgrade students' language skills after knowing their language abilities. This test was in the form of an aptitude test or an ability test for each skill in the Arabic language. But in the speaking skill, this test still required preparation and development of more test questions to practice. Apart from the test questions, the method of assessment and the procedures for its implementation also needed to be developed and considered. Then, the researchers took account of the need to compile and make a good test product in speaking skills by following the steps for preparing a standardized test. This was the Research and Development (R&D) with Borg and Gall's

version. The researchers used the ANATES V.0.5 2022 application for processing data and analyzing test results. Based on the results of the study, it was found that the test was feasible because it met the requirements of a good test. The trial results showed high validity and reliability values, and the level of difficulty of each item also had good distinguishing power. Out of a total of 17 questions, 13 questions were stated to be very valid, 2 questions were valid, and 2 questions were striped (-), indicating that these two questions needed to be re-analyzed on what aspects made them less valid so that the questions could be revised. These values were obtained from the results of two trials conducted in the sixth semester.

Keywords: CEFR; speaking skill; test

المخلص

تهدف هذه الدراسة إلى تصميم الاختبار لمهارة الكلام في مستوى B1 من CEFR، لتحديد صدق الاختبار وثباته ومستوى صعوبة كل الأسئلة، وتنظيم قوة التمييز لكل أسئلة الاختبار المعد. خلفية هذا البحث هي الحاجة إلى وجود الاختبارات اللغوية كنشاط تقييمي مطلوب لأغراض مختلفة. كما غرض بعض الجامعة في إندونيسيا وجود الاختبار اللغوي كشرط من شروط استمرار الدراسة إلى الدرجة الأعلى أو إلى مكان يتطلب فيه الكفاءة اللغوية، والعمل في مكان ما، وأهمه يعني لترقية لغة الطلاب بعد معرفة قدرتهم اللغوية. وهذا الاختبار على شكل اختبار الكفاءة أو اختبار القدرة على كل مهارة في اللغة العربية. ومع ذلك، كانت مهارة الكلام لا تزال اختبارها يتطلب إعدادا وتطويرا من أسئلة الاختبار التي يتعين ممارستها. من غير أسئلة الاختبار، فإن طريقة التقييم وإجراءات تنفيذها تحتاج أيضًا إلى التطوير. لهذا، رأى الباحثان بالحاجة إلى تطوير الاختبار وتصميمه في مهارة الكلام باتباع خطوات تصميم الاختبار المعياري. المنهج الذي نهجه الباحثان لهذا تصميم الاختبار هو منهج البحث والتطوير (Research and Development Borg and Gall). في معالجة البيانات وتحليل نتائج الاختبار، استخدم الباحثان برنامج ANATES.

2022 V.0.5 بناءً على نتائج هذه الدراسة يعتبر أن الاختبار الذي صممه الباحثان جيد، لأنه يستوفي شروط صحة الاختبار الجيد، تظهر نتائج الاختبار أن الصدق وثباته عال، ومستوى صعوبة كل الأسئلة وقوة تمييز الاختبار جيدة. من مجموع 17 سؤالاً، تم تحديد 13 سؤالاً على أن صدقه عال جداً، وسؤالان عال، وسؤالان ناقص (-) ، مما يعني أن هذين السؤالين يحتاج إلى إعادة تحليل بشأن الجوانب التي تجعلهما لم يكن صادقا بحيث يمكن تحسين الأسئلة. تم الحصول على هذه النتيجة من نتائج تجربتين في الفصل السادس .

الكلمات المفتاحية: الإطار المرجعي الأوروبي؛ مهارة الكلام؛ الاختبار

المقدمة

اختار الباحثان مهارة الكلام لتصميم الاختبار لأنها من أهم أهداف تعليم اللغة العربية، وهو أهم المهارة من جميع المهارات اللغوية الأربعة، لأن الأفراد الذين يتعلمون لغة فيشار إليهم ليكون متحدثي اللغة المستهدفة.¹ كما رأى سيدة مسومة أن الكلام هو أهم المهارة للاهتمام بها ولترقيتها حتى يكون المحادثة بين الناس مفهوما. الكلام هو مهارة جديدة بالاهتمام في كل من اللغة الأولى والثانية، وهو أهم الجانب في تعلم اللغة الثانية أو اللغة الأجنبية، ويتم قياس نجاح اختباره بناءً على القدرة في إجراء المحادثة باللغة المستهدفة.²

ولكن لقد تغيب الكلام عن الاختبار بسبب مشكلة تقييمه بموضوعية والوقت الذي يستغرق إجراء اختبار الكلام عندما نتحدث عن الكلام، فإننا لا نعني مجرد قول الكلمات من خلال الفم. يعني إيصال الرسالة من خلال الكلمات الشفوية. غالبًا ما يتم تجاهل هذه المهارة في بعض فصول المعلمين. ليس لدى المتعلمين فرصة سواء في

¹ Faisal Mubarak, Abdul Sattar, dan Ahmad Fahmy Arief, "Thasmim Mawad Maharah al-Kalam Fi Istimdadiha Min al-Ayat al-Istifhamiyah Fi Surah al-An'am wa al-Mulk Li Thulab al-Ma'had al-Aly Ulum al-Qur'an Amuntai," *Arabiyatuna: Jurnal Bahasa Arab* 4, no. 1 Mei (8 Mei 2020): 74–75, <https://doi.org/10.29240/jba.v4i1.1320>.

² Leong Lai Mei dan Ahmadi Seyedeh Masoumeh, "An Analysis Of Factors Influencing Learners' English Speaking Skill" 2, no. 1 (1 Januari 2017): 34–41.

فصولهم الدراسية أو في الخارج للتحديث باللغة. للأسف، التحدث ليس جزءًا مهمًا من امتحانات المعلمين. يحتاج المتعلمون إلى الكثير من الممارسة لتعلم التحدث.³ من البيان سابقا، وجد أن الاختبار لمهارة الكلام محتاج للأداء بشديد الاحتياج، إما لاستفءاء شرط من شروط شيء ما حتى لترقية لغة الطلاب بعد معرفة قدرتهم. وبنسبة إلى المشكلة المذكورة سابقا، اختبار مهارة الكلام يحتاج إلى أن يكون التقييم موضوعية مهما لا يمكن موضوعيته مائة بالمائة، بل يمكن التقييم أقرب إلى موضوعي بوجو قائمة التقييم الدقيق على شكل التحليلي كما استُخدم هذه القائمة للاختبار العملي.⁴

فانطلاقًا من هذه الخلفية، أراد الباحثان أن تصمم أسئلة الاختبار الجيد للامتحان الشامل بل تخصص هذه الأسئلة الاختبار مهارة الكلام للمستوى المتوسط، لأن هذه المهارة من شكل رئيسي الذي سيوصله تعليم اللغة العربية، ليكون اللغة العربية هي اللغة المفخر لدى الطلبة.⁵ وكان اختبار الكلام قد تغيب في عملية الاختبار بسبب الصعوبة في التقييم مع أنه مهم للأداء. فمن هذا البحث سعي الباحثان التصميم بما يحتاج اختبار الكلام الجيد. كما قال أسامة في كتابه، أن معلم اللغة العربية للناطقين بغيرها يحتاج تطبيقات في تصميم الاختبار اللغوي الجيد الصادق.⁶ ووصم الباحثان في مستوى المتوسط الأدنى لأنه المستوى المثل والوسط،⁷ ولا تصمم الاختبار لكل المستوى حتى تكون أسئلة الاختبار مركز.

³ Mei dan Masoumeh, 34.

⁴ Ihwan Mahmudi, "Rubrik Analitik Penilaian Hasil Belajar Praktik Pendidikan Agama Islam," *Fikrah: Journal of Islamic Education* 2, no. 2 (2018): 19.

⁵ Ahmad Hidayatullah Zarkasyi, "Madza NaPal Li Ajli RaFi al Fakhor fi Istikhdam al Lughah al Arabiyyah" (Al Muhadharah fi Maddati Tathwir al Manahij, Mabna Jami'ati Darissalam Gontor Li Dirasat al Ulya, September 2022).

⁶ Osamah Zakky As Sayyid Ali, *Al Iktibar al Lughawiyah, Muqarabah Manhajiyah Tatbiiyyah* (Riyadh: Markaz al Mulk Abdullah Bin Abdul Aziz al Duwaliy, 2016), 11.

⁷ Anna Siyanova Chanturia dan Stefania Spina, "Multi-Word Expressions in Second Language Writing: A Large-Scale Longitudinal Learner Corpus Study," *Language Learning A Journal of Research in Language Studies* 70, no. 2 (20 Desember 2019), <https://doi.org/10.1111/lang.12383>.

بعد الاختبار سوف ينال الطلاب درجاتهم أو نتيجتهم،⁸ فيحدد بدرجة A1, A2, B1, B2, C1, C2, هذه الدرجة كلها مما أعطاه الإطار المرجعي الأوروبي المشترك للغات (CEFR)،⁹ وهو وثيقة من المجلس المعروف في أوروبا، أعدها لإقامة نظم التحقق من القدرة اللغوية.¹⁰ وهذا المعيار قد وصل إلى العرب في عام 2007 م.¹¹

والجامعات الإندونيسية التي تهتم كثيرا في تعليم اللغة العربية واختباراتها إحداهما جامعة دار السلام كونتور.¹² قد عقد هذه الجامعة الاختبار في اللغة العربية لكل المهارة وأهمها مهارة الكلام.¹³ وترجى من تعليم اللغة العربية في هذه الجامعة القدرة على أن يكون الطلبة معلماً حسناً للغة العربية حتى تنشر اللغة العربية في المجتمع.¹⁴ لتحقيق مدى إتقانهم في اللغة العربية، عقدت جامعة دار السلام كونتور الاختبار اللغوي لجميع الطلاب، ولعقد هذا الاختبار يحتاج إلى مجموعة من الأسئلة المقنن.¹⁵

⁸ Rini Rini dan Partomuan Harahap, "Designing the Holistic Evaluation in Teaching Reading/Tasmim al-Taqwim al-Syumuli Fi Ta'lim al-Muthala'ah," *Arabiyatuna: Jurnal Bahasa Arab* 5, no. 1 Mei (7 Mei 2021): 143, <https://doi.org/10.29240/jba.v5i1.2522>.

⁹ Council of Europe, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (Strasbourg: Language Policy Unit, no year).

¹⁰ Majelis Auruba, *Al Ithar al Marja'i al Aurubiy al Musyarak lil Lughat (Dirasah, Tadris, Taqyim)* (al Qahirah: Dar Ilyas al 'Ishriyyah, 2000), 54.

¹¹ Islam Yasri Ali, "Muraja'atu Kitab al Ithar al Marja'iy al Aurubiy al Musyarak Lil Lughat: Dirasah," *Majallah al Dirasat al Lughawiyah wa al Adabiyah*, Juni 2016, 197.

¹² Nur Hanifatuz Sholeha, Abdul Hafidz Zaid, dan Fairuz Subakir, "Tashmim al-Kitab al-Raqmy Muassasan 'ala Tathbiq Flippdf Professional fi Maharah al Istima' li al-Fashl al-Tahdhiri," *Arabiyatuna: Jurnal Bahasa Arab* 6, no. 2 November (4 November 2022): 674, <https://doi.org/10.29240/jba.v6i2.5247>; Frida Akmalia, Syihabuddin Syihabuddin, dan Yayan Nurbayan, "Intensive Course: the Alternative Program for Teaching Arabic Speaking and Reading Skills at an Islamic Boarding School/al-Daurah al-Mukatsafah: al-Barnamij al-Badil Li Ta'lim Maharah al-Kalam wa al-Qira'ah Fi al-Ma'had," *Arabiyatuna: Jurnal Bahasa Arab* 5, no. 1 Mei (7 Mei 2021): 116, <https://doi.org/10.29240/jba.v5i1.2306>.

¹³ Apri Wardana Ritonga dkk., "Teaching Maharah Kalam on the Basis of Culture Using the Textbook 'al-'Arabiyyah Baina Yadaik' at the Middle School/Ta'lim Maharah al-Kalam 'Ala Asas al-Tsaqafah bi Istikhdam Kitab al-'Arabiyyah Bayna Yadayk fi Marhalah al-Mutawassithah," *Arabiyatuna: Jurnal Bahasa Arab* 5, no. 1 Mei (7 Mei 2021): 2, <https://doi.org/10.29240/jba.v5i1.1942>; Abdul Wahab Rasyidi dan Suci Ramadhanti Febriani, "Uslub Ta'lim Maharah al-Kalam Bi Asas al-Dzaka'at al-Muta'addah Fi al-Marhalah al-Ibtidaiyyah," *Arabiyatuna: Jurnal Bahasa Arab* 4, no. 2 November (17 November 2020): 194, <https://doi.org/10.29240/jba.v4i2.1432>.

¹⁴ Hamid Fahmy Zarkasyi dkk., *Pekan Perkenalan Khatbatu-L-'Ary Universitas Darussalam Gontor* (Ponorogo: UNIDA Gontor Press, 2020).

¹⁵ Muhammad Ismail, *Al Ikhtibar al Lughawiy bi Jam'i'ati Darissalam Gontor*, 15 Juni 2022.

إن هذا البحث يستخدم المنهج التجريبي التطويري (R&D) وفقا لبرغ وغال¹⁶ (Borg and Gall) وهذا هو المنهج لتحقيق والتطوير المستهدف إلى التحسين.¹⁷ والخطوات في إجراء البحث على هذا المنهج يعني: إجراء الدراسة التمهيديّة، ثم التخطيط، ثم تصميم الاختبار، ثم المقابلة مع الخبراء للحصول على البيانات عن الاختبار المصمم ثم تحليلها، ثم تحسين مواد الاختبار متعمدا إلى نتائج الخبراء، ثم التجربة الأولى للاختبار المصمم إلى طالبات جامعة دار السلام كونتور، ثم تحليل نتائج الاختبار وتحسين الاختبار على أسس نتائج التجربة، ثم التجربة الثانية للاختبار، ثم تحليل النتائج وتحسين الاختبار، والإنتاج الأخير.¹⁸

يتألف مجتمع البحث لتطبيق هذا الاختبار من الطالبات في جامعة دار السلام كونتور في الفصل السادس، وتتكون من 20 طالبة الجامعة في كلية البنات جامعة دار السلام كونتور و 30 طالبة من الحرام الثاني للبنات.¹⁹ واختار الباحثان هذه المرحلة لأنها المستوى المناسب بهذا الاختبار من أجل الامتحان الشامل للمستوى المتوسط.²⁰

فقد كانت الأسئلة في الاختبار المبدئي 18 سؤالا بنوع الاختبار اختبار الأداء الشفهي، مما يتكون من الطالبة الإجابة من الأسئلة شفهيًا، والإعادة الشفهية أي الطالبة تكرر مما سمعت، والبيان من الصورة المعينة، ثم الحوار مما يتعلق بالأنشطة اليومية وما أشبه ذلك.

واستخدم هذا البحث أسئلة الاختبار المصمم في تجارب ميدانية لتحديد صدق الاختبار وثباته بالطرق الإحصائية الكمية باستخدام البرنامج ANATES V.0.5 2022 ، وهو الذي سيكون آلة لقياس ثبات الاختبار وصدقه وما يحصل إلى ذلك من جميع ما

¹⁶ Meredith Gall, Joyce Gall, dan Walter Borg, *Educational Research An Introduction* (New York: Library of Congress Cataloging, 2003).

¹⁷ Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D* (Bandung: Alfabeta, 2013), 311.

¹⁸ Gall, Gall, dan Borg, *Educational Research An Introduction*.

¹⁹ Suharsimi Arikunto, *Prosedur Penelitian Suatu Pendekatan Praktik* (Jakarta: PT Asdi Mahasatya, 2013), 74.

²⁰ Ihwan Ihwan Mahmudi, *Evaluasi Pendidikan* (Yogyakarta: Lintang Books, 2020).

يحتاجه الاختبار به بحيث يكون اختبارًا موضوعيًا جيدًا بقياسه،²¹ وذلك بعد أن قام الباحثان بالتطبيق أو التجربة مع الطالبات، بحيث حصل الباحثان إلى الهدف المنشود بالنتائج بعد التطبيق أو التجربة في هذا البحث.

نتائج البحث ومناقشتها

تصميم الاختبار لمهارة الكلام للمستوى المتوسط الأدنى وفقا بالإطار المرجعي الأوروبي المشترك للغات

مرحلة التحليل

الاختبار اللغوي في جامعة دار السلام كونتور هو الامتحان الشامل الذي عقده ليعرف الطلبة مدى إتقانهم في اللغة العربية. عقد لكل السنة الدراسية لجميع الطلبة تحت مسؤولية مركز اللغة بجامعة دار السلام كونتور. لكل فصل أسئلة مختلفة في الاختبار يهدف إلى أن يعرف كل الطلبة كفاءتهم في اللغة حتى يقيم لغتهم ويستخدم اللغة العربية في أنشطتهم اليومية.²² وفي ناحية أخرى، ألزمت جامعة دار السلام كونتور كتابة رسالة الجامعة باللغة العربية أي اللغة الإنجليزية.²³

بعد الدراسة التمهيديّة في الساحة التربوية، وجد الباحثان الحاجة في تصميم الاختبار لمهارة الكلام لأن الكلام هو أهم الهدف في تعليم اللغة العربية،²⁴ وفي تصميم الاختبار لمهارة الكلام لا يساوي بتصميم الاختبار في مهارة أخرى، لأنه يحتاج إلى قائمة

²¹ Enrique Colino dkk., "Validity and reliability of two standard test devices in assessing mechanical properties of different sport surfaces," *Polymer Testing* 62 (September 2017), <https://doi.org/10.1016/j.polymertesting.2017.06.011>; C. Delgado, A. Araneda, dan M. I. Behrens, "Validation of the Spanish-language version of the Montreal Cognitive Assessment test in adults older than 60 years," *Neurologia (English Edition)* (Elsevier, 2019), <https://www.sciencedirect.com/science/article/pii/S2173580818302074>.

²² Ismail, Al Ikhtibarat al Lughawiy bi Jami'ati Darissalam Gontor.

²³ Amal Fathullah Zarkasyi, *Dauru Jami'ati Darissalam Gontor fi Y'dadi al Mujaddidin* (Ponorogo: Jami'atu Darissalam Gontor Press, 2009).

²⁴ Rusydi Ahmad Thu'aimah, *Al Marja' fi Ta'limi al Lughab al Arabiyyah Li al Nathiqin bi Lughatin 'Ukbro* (Jami'atu Ummul Qura: Ma'had al Lughah al Arabiyyah, 1986), 507; Fathiy Ali Yunus dan Muhammad Abdu Ar Ra'uf, *Al Marja' fi Ta'limi al Lughab al Arabiyyah Lil Ajanib (Min al Nazhariyyah Ila al Tatbiiq)* (Kairo: Maktabah wa Hibah, 2003).

النتيجة الدقيقة حتى يقيس قدرة الطلبة في مهارة الكلام،²⁵ وهذه القائمة يسمى بقائمة تحليلية كانت أدق قائمة، ولم يكن الاختبار بسؤال مكتوبة وقائمة أدق.²⁶ فقد أعطى جامعة دار السلام كونتور الحرية للممتحنين في إعطاء النتيجة وفي إعطاء الأسئلة.

قبل تصميم الاختبار، لاحظ الباحثان نحو المعايير للإطار المرجعي الأوروبي المشترك للغات للمستوى المتوسط الأدنى في مهارة الكلام، وهي كما يلي:²⁷

الجدول 1. المعايير للمستوى المتوسط الأدنى (B1) من CEFRR

المشاركة في الحديث	الحديث المترابط
يمكنه الخوض في أغلب المواقف التي عادة ما يواجهها المرء أثناء رحلة في الدولة المتحدثة للغة، كما يمكنه المشاركة في الحديث عن الموضوعات التي أعرفها أو التي تحزي باهتمامي الشخصي أو التي تتعلق بأمور داخلية في الحياة اليومية كالحديث عن الأسرة والهواية والعمل والسفر والأحداث المعاصرة.	يستطيع الحديث في سياق متصل ويصف الأحداث والتجارب ولعرض آماله وأحلامه وأهدافه الخاصة، كما يمكنه في إيجاز أن يشرح وأن يفسر خطته وآرائه. يمكنه أن يحكي قصة أو أن يعرض أحداث فيلم أو كتاب وأن يوضح رد فعله تجاهه.

B1

مرحلة التصميم

اتبع الباحثان الخطوات في تصميم الاختبار الجيد وهي:

1. تعيين هدف الاختبار، يمكن إجراء اختبارات التحصيل التعليمي لأغراض مختلفة مثل التشخيص والاختيار وما إلى ذلك.

²⁵ Muhammad Ali al Khuliy, *Al Iktibar al Tabsiliyah, 'Daduba wa Ijra'uba wa Tabliluba* (Jordan: Dar al Falah, 1998).

²⁶ Mahmudi, "Rubrik Analitik Penilaian Hasil Belajar Praktik Pendidikan Agama Islam," 12.

²⁷ Ali, "Muraja'atu Kitab al Ithar al Marja'iy al Aurubiy al Musytarak Lil Lughat: Dirasah," 59.

2. تحليل المنهج، يهدف تحليل المناهج إلى تحديد مقدر كل مادة والتي سيتم استخدامها كأساس لتحديد عدد العناصر لكل موضوع.
3. كتابة شبكة الأسئلة، الشبكة تعني عبارة عن جدول مصفوفة يحتوي على مواصفات الأسئلة المطلوبة إجراؤها. هذه الشبكة هي مرجع لكاتب الأسئلة بحيث يقوم أي شخص يكتب أسئلة بإنتاج أسئلة بنفس المحتوى ومستوى الصعوبة نسبيًا. تحتوي الشبكة على ما يلي: معايير الكفاءة، والكفاءات الأساسية، والمؤشرات، وتحديد عدد الأسئلة لكل المؤشر، والآخر تحديد مستوى صعوبة الأسئلة (المجال المعرفي).
4. تعيين نوع الاختبار، أشكال الاختبارات الموضوعية التي تُستخدم غالبًا هي الاختيار من متعدد، والصواب والخطأ. يمكن تصنيف اختبارات المقالات على أنها أوصاف موضوعية وغير موضوعية. غالبًا ما تُستخدم اختبارات المقالات الموضوعية في العلوم والتكنولوجيا أو المجالات الاجتماعية حيث تكون الإجابات على الأسئلة مؤكدة وتكون إجابة واحدة فقط صحيحة. غالبًا ما تستخدم اختبارات المقالات غير الموضوعية في مجال العلوم الاجتماعية، أي الإجابات العامة وليس مجرد إجابة واحدة.
5. تحديد الوقت للاختبار، تعيين طول وقت الاختبار من 90 دقيقة إلى 150 دقيقة للاختبار التحريري، والاختبار الشهي ممكن أكثر من ذلك.
6. كتابة أسئلة الاختبار، هذا الوقت لتطبيق شبكة الأسئلة إلى السؤال المكتوب.
7. تقديم الأسئلة إلى الخبراء، في هذه الخطوة قدم الأسئلة إلى الخبراء في المواد واللغة والمنهج.
8. التجربة الميدانية، بعد أن أصلحنا الأشياء الناقص ما تناوله من الخبراء فلا بد من تجربة الأسئلة إلى العينة المقصودة.
9. تحليل نتائج التجربات، يحلل الباحثان البيانات من التجربة الميدانية ليعرف صلاحية الأسئلة.
10. تحسين أسئلة الاختبار، بعد أن عرف الأشياء الناقصة فتحسن الأسئلة.

11. الإنتاج الأخير، بعد أن اتبع كل الخطوة في تصميم أسئلة الاختبار فيمكن كتابة الأسئلة لمرّة عديدة وهذه هي الإنتاج الأخير.²⁸

فاتبع الباحثان الخطوات المذكورة حتى صمم الاختبار متبع إلى شبكة الأسئلة التي كتبها قبل التصميم، وتلك الشبكة كما يلي:
الصورة 1. شبكة الأسئلة

شبكة أسئلة الاختبار لمهارة الكلام في المستوى المتوسط الأدنى (B1)							
الوقت	رمز الأسئلة	علم معرفي	الأسئلة	نوع الأسئلة	المادة	المؤشرات	معايير الكفاءة
ثانية 50	B1.KL.1	C1 (القراءة)	اقرأ ما يلي بالترتيب الصحيح!	القراءة الجهرية	الرياضة	الفترة على قراءة فقرة من المقالة بالمخرجات الصحيحة والترتيب المناسب	1. نطق الجمل بشكل صحيح وبالترتيب الصحيح
ثانية 40	B1.KL.2	C1 (التفاصيل)	أية رياضة أحب إليك؟ ولماذا تحبها أكثر؟	إجابة قصيرة	الرياضة	الفترة على ذكر أحب رياضة إليه	2. نطق الجمل بشكل صحيح وبالترتيب الصحيح
ثانية 40	B1.KL.3	C1 (الذكر)	ما رأيت في هذه الصورة؟	إجابة قصيرة	المدرسة	الفترة على ذكر المفردات عن الأدوات المدرسية	3. نطق المفردات المختلفة
ثانية 20	B1.KL.4	C1 (الذكر)	أكمل هذه الجملة بنطق الكلمة المناسبة للصور!	إجابة قصيرة	ظرف المكان	الفترة على إكمال الجملة بالفرد المناسبة	4. نطق الجمل بشكل صحيح حول الحياة اليومية
ثانية 25	B1.KL.5	C2 (الترجمة)	ترجم هذه الجملة إلى العربية!	الترجمة	الأفعال	الفترة على ترجمة الجملة	5. نطق الجمل بشكل صحيح حول الحياة اليومية
ثانية 45	B1.KL.6	C2 (التفصيل)	تكلم عن الأنشطة المفضلة لديك إلى هذه الصور!	التعبير	الحياة اليومية	الفترة على التكلم باستخدام المفردات الكثيرة	6. نطق الجمل بشكل صحيح حول الحياة اليومية

وأخذ الباحثان كل المؤشرات من الإطار المرجعي الأوروبي المشترك للغات²⁹ والمادة مما تعلمه جامعة دار السلام كونتور للطلبة، وهو كما في الجدول التالي:

الجدول 2. التفاصيل من شبكة الأسئلة

وصف شبكة الأسئلة	التفاصيل
من الإطار المرجعي الأوروبي المشترك للغات	معايير الكفاءة
من الإطار المرجعي الأوروبي المشترك للغات	الكفاءة الأساسية
من الإطار المرجعي الأوروبي المشترك للغات	المؤشرات ³⁰
مما تعلمه جامعة دار السلام كونتور	المادة

²⁸ Mahmudi, *Evaluasi Pendidikan*, 57–59.

²⁹ Claudia Harsch dan André Alexander Rupp, "Designing and Scaling Level-Specific Writing Tasks in Alignment With the CEFR: A Test-Centered Approach," *Language Assessment Quarterly* 8, no. 1 (31 Januari 2011): 1–33, <https://doi.org/10.1080/15434303.2010.535575>; Salwa Mohamed, "The Development of an Arabic Curriculum Framework Based on a Compilation of Salient Features from CEFR Level Descriptors," *The Language Learning Journal*, 20 Mei 2021, 1–15, <https://doi.org/10.1080/09571736.2021.1923781>.

³⁰ Cambridge ESOL, "Using the CEFR: Principles of Good Practice," *Cambridge United Kingdom*, 2011, 48.

نوع الأسئلة	من المعايير في كتاب الاختبارات ³¹
عالم المعرفي	من المعايير في عالم المعرفي بلوم (Bloom)
الوقت	من التجارب قبل التجربة الميدانية

عيّن الإطار المرجعي الأوروبي الجانب المقدر لمهارة الكلام، وهي التنوع في الكلام، والصح والخلو من الخطأ، والطلاقة، والتفاعل اللغوي، والترابط.³² كما عرفه عبد الخالق بأن مهارة الكلام هي أن يصل المتعلم إلى مستوى معقول من الطلاقة أي القدرة على التعبير بصورة مفهومة وسليمة ودون التشكك أو التردد. لأجل اختبار ذلك هناك أنواع شكل أسئلة اختبار الكلام، منها اختبار النطق، والمفردات، والتماسك والارتباط، والقواعد، والطلاقة، ثم الأسلوب، لكل شكل بنود مختلفة كقراءة جهرية أو استخدام الرسوم التوضيحية وما أشبه ذلك مما كتب عبد الخالق في كتابه اختبار اللغة.³³

ومن أجل تعيين النتيجة للمختبر، كتب الباحثان القائمة³⁴ وتلتج فيها النطق، المفردات، القواعد، الأسلوب، والتعبير،³⁵ والقائمة كما يلي:

³¹ Muhammad Abdul Khaliq, *Ikhtibar al Lughab* (Ar Riyadh: Mathabi' Jamiati al Mulk Su'ud, 1997).

³² Auruba, *Al Ithar al Marja'i al Aurubiy al Musytarak lil Lughat (Dirasab, Tadris, Taqyim)*, 39–40.

³³ Khaliq, *Ikhtibar al Lughab*, 167.

³⁴ Dexter L. Manzano, "Examining the Interrater Reliability Between Self- and Teacher Assessment of Students' Oral Performances," *International Journal of Language Testing* 12, no. 2 (1 Oktober 2022): 128–44, <https://doi.org/10.22034/ijlt.2022.157130>.

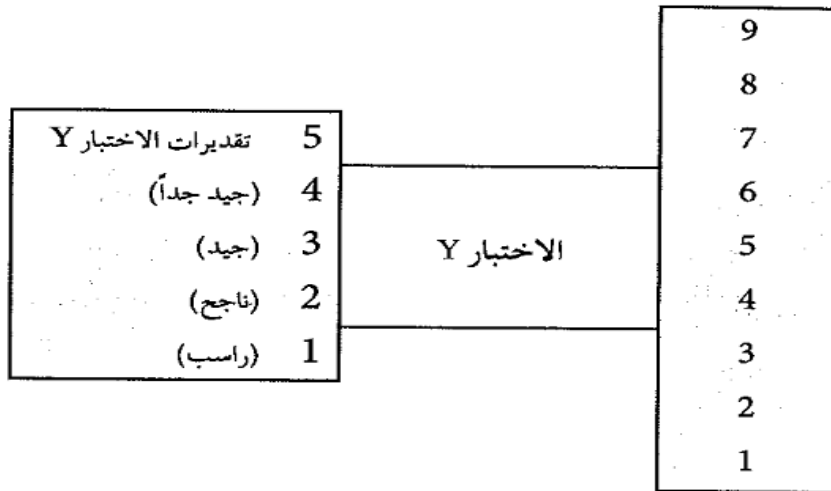
³⁵ Goedele Vandommele dkk., "Assessing the Development of L2 Speaking Skills in Inside-School and Outside-School Settings Through Growth Modelling," *Applied Linguistics* 43, no. 6 (1 Desember 2022): 1094–1115, <https://doi.org/10.1093/applin/amac024>; Kaja Jasińska dkk., "Evaluating Bilingual Children's Native Language Abilities in Côte d'Ivoire: Introducing the Ivorian Children's Language Assessment Toolkit for Attié, Abidji, and Baoulé," *Applied Linguistics* 43, no. 6 (1 Desember 2022): 1116–42, <https://doi.org/10.1093/applin/amac025>.

الصورة. 2 قائمة تحليلية لتقييم الاختبار لمهارة الكلام

قائمة تحليلية لتقييم مستوى مهارة كلام B1							
الإسم:							
جانب المقدر	رقم الأسئلة	فشل	راسب	ناقص	كاف	جيد	جيد جدا
جانب المقدر	رقم الأسئلة	0	النتيجة (3-1)	النتيجة (4)	النتيجة (5)	النتيجة (6-7)	النتيجة (8-9)
النطق	1	لا تفهم السؤال فلا تجب على الإطلاق	قرأ بالكامل ، دون التتغيم المناسب، وهناك العديد من المخرجات الخاطئة	قرأ كاملة والتتغيم غير مناسب وأخطأ كثيرا في المخرج	قرأ كاملة ، وتتغيم صحيح ، بـ 3 أو 4 مخرجات خاطئة	قرأ كاملة ، والتتغيم صحيح بـ 1 أو 2 مخرج خاطئ	قرأ كاملة ، والتتغيم صحيح ، ومخرجات صحيحة

وكل النتيجة أخذها من قائمة التقييم للإطار المرجعي الأوروبي المشترك للغات³⁶، وهي كما يلي:

الصورة.3. مقياس الكفاءة



³⁶ Katrin Wisniewski, "Empirical Learner Language and the Levels of the Common European Framework of Reference," *Language Learning: A Journal of Research in Language Studies* 66, no. 1 (12 Januari 2017), <https://doi.org/10.1111/lang.12223>.

مرحلة التطوير

طلب الباحثان التقييم للاختبار من الخبراء في اللغة العربية والاختبارات ليعطي الاقتراحات في لغة أسئلة الاختبار وكذلك مواد الاختبار والاختبار نفسه.³⁷ فوجد الباحثان أن الاختبار الذي صمما الباحثان فيها الأشياء وجوب تحسينها وإصلاحها، وكل تلك الإصلاحات أخذها الباحثان لتحسين الاختبار، والنقطة الأخيرة بعد التقييم يعني كل السؤال صالح للاستخدام في التجربة الميدانية بعد أن حسب الباحثان نتائج التقييم من الخبراء بالرموز الآتي:

$$V = \sum s / (n (c - 1), S = r - l_o$$

$$\text{صلاحية المحتوى} = V$$

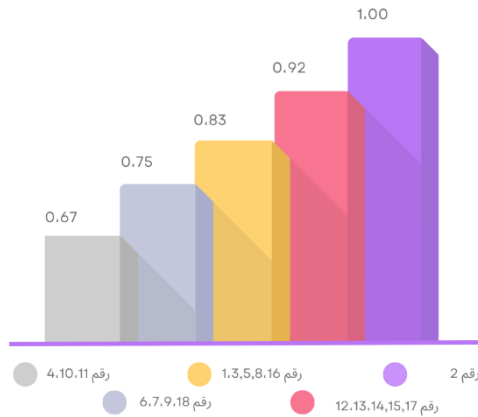
$$\text{النتيجة من الخبير} = r$$

$$l_o = \text{أدنى النتيجة للتقييم}$$

$$n = \text{عدد الخبير}$$

$$c = \text{عدد اختيار النتيجة}^{38}$$

الرسم البياني 1. نتيجة التصديق من الخبراء



³⁷ Ihwan Mahmudi, *Pengembangan Instrumen Penelitian Sosial (Konsep, Contoh, dan Analisis Data Menggunakan SPSS)* (Yogyakarta: Lintang Books, 2020), 70.

³⁸ Iwan Suhardi dan Hendriyati Eko Budiyaningsih, "Validasi Pakar Konten pada Pengembangan Paket Soal Bahasa Indonesia untuk Basis Data Computer Based Testing," 2022, 5.

الصورة 4. أسئلة الاختبار

شاهد الفيديو الآتي ثم أجب عن الأسئلة لرقم ١١-١٤ (مدة الفيديو دقيقتان)

١١. اسرد مختصرًا عن محتويات الفيديو!


١٢. أعد بعض الحوارات في الفيديو!

١٣. كيف أكلت ليل الطعام؟ ولماذا تتأذى أسنانها؟

١٤. أخبر عن العبرة أو الدروس الذي يمكن استخلاصه من محتويات الفيديو القصير!

١٥. عرِّ عن مشاعرك عندما كنت في هذا المكان!

الأسلوب



١٦. أحك عن أيام صفارك! الطلاقة

الأسئلة في الصورة 4 هي الأسئلة التي أصلح الباحثان بعد تقديمها إلى الخبير في اللغة العربية، والخبير في الاختبار اللغوي والخبير في التقييم التربوي. فمن الإصلاحات والمدخلات من الخبراء هي:

1. أن تكون أسئلة اختبار الكلام تقيس كفاءة الطلاب في الكلام بعناصره الكامل، وهي النطق بمخارج الحروف الصحيحة، ثم المفردات، ثم القواعد، ثم التماسك والارتباط، ثم الطلاقة، والآخر الأسلوب.
2. يجب أن يتوافقت المؤشرات على مهارة الكلام مع المستوى التصنيفي المحدد. يمكن أن تبدأ من القدرة على الذكر، والشرح، ثم التحليل في النهاية.
3. الوقت لم يكن مثاليًا، يرجى التكيف مع متطلبات الإجابات والصعوبات المحددة.
4. لم تكن الأسئلة متنوعًا، وتثري أنواع الأسئلة من خلال استكشاف قدرات التحدث لدى الطلاب، مثل: الإشارة، والإجابة، والرد على الأسئلة، والوصف، وإعطاء أمثلة، وبناء جمل جديدة، وتغيير الجمل، وتغيير القواعد إلى جمل، وجمل قصيرة، جمل طويلة، إلخ.

كل المدخلات من الخبراء تطابق المعايير في تصميم اختبار الكلام كما بين الباحثان سابقًا.

مرحلة التنفيذ

وبعد التصميم، عقد الباحثان التجربة الميدانية للفصل السادس في جامعة دار السلام كونتور. تقصد من التجربة الميدانية معرفة مدى صلاحية أسئلة الاختبار،³⁹ فبعد إتمام تحسين الاختبار قام الباحثان بالتجربة الأولى في جامعة دار السلام كونتور للبنات للفصل السادس العام 2022-2023 بعدد طلبة الجامعة المشتركات 20 طالبة، عقدت هذه التجربة الأولى في التاريخ الثالث من نوفمبر 2022. وبعد إجراء هذه التجربة فحصلت إلى نتائج الاختبار وقام الباحثان بحسابها باستخدام برنامج ANATES V.0.5 .2022

الصورة 5. البيانات من نتيجة الاختبار

```

SKOR DATA
=====
Rata2= 100,27
Standar Deviasi= 15,86
Nama berkas: E:\2. M.PD\4. THESIS\EVERYTHING\TEST 2.AUR

Nomor  Nomor  No. Butir Baru ---->  Skor  1  2  3  4  5  6  7  8
Urut  Subyek  No. Butir Asli --->  1  2  3  4  5  6  7  8
      Nama|Skr Ideal ->  9  9  9  9  9  9  9  9

1      1      1      Sonali Karela Khilmiya
126    8      8      9  7  9  5  9  9
2      2      2      Azizah Qurrota A'yun ...  109  8  7  8  5  9  5  4  7
3      3      3      Khoirunnida Salsabila  135  9  8  9  7  9  9  9  9
4      4      4      Aliefia Salsabila Put...  104  7  6  6  5  4  5  4  5
5      5      5      Alfi Khilmi Khusnia    89  5  6  6  5  5  5  4  7
6      6      6      Margareth Aliyatul Ma...  99  5  6  6  5  4  7  4  5
7      7      7      Syifaurohmah Putri A...  96  4  5  7  5  5  5  4  5
8      8      8      Annisaa Diva Febriyanti  116  7  8  6  7  7  7  4  9
9      9      9      Dera Febrianti         81  5  4  7  4  7  4  4  4
10     10     10     Fazati Afida El-Aqwami  103  7  5  7  4  6  9  4  4
11     11     11     Syifa Auliya Hanifah    126  9  8  5  5  5  8  8  8
12     12     12     Navila Salsabila       107  6  7  8  7  6  6  4  4
13     13     13     Sya'na Sekar Izaty      93  6  5  4  4  6  5  4  5
14     14     14     Risma Nada Azkiya      114  9  8  7  5  5  8  4  7
15     15     15     A'isyatu Makarimil Ch...  97  7  7  7  5  3  7  7  7
16     16     16     Zakiah Ainun Jariyah    95  5  6  6  4  6  8  4  5
17     17     17     Salsa Adella Dyani H...  94  5  5  6  6  7  8  4  4
18     18     18     Sukma Angraini         81  5  5  4  4  6  5  4  5
19     19     19     Layyinatuzzahro        110  5  8  8  8  7  9  4  6
20     20     20     Shaliha Annisa         117  9  6  6  9  9  9  0  9
21     21     21     Mutiara Rahmah Riawan  109  6  9  7  9  7  9  0  4
22     22     22     Tiara Eki Wahyu Rinawati  120  9  5  5  7  4  9  7  8
23     23     23     Wahyuningtyas Islami    103  9  9  5  4  7  7  5  8
24     24     24     Atika Cahyaning Singgih  91  9  7  5  5  7  5  0  7
25     25     25     Vina Veronika Maya Ta...  89  6  4  7  4  9  5  4  6
26     26     26     Bayyinatun Mazidah     86  4  5  7  4  7  6  4  7
27     27     27     Fauzia Rozifatul Haq    86  7  5  4  4  4  6  3  6
    
```

الطريقة التي استخدم الباحثان لحساب معامل ثبات الاختبار هي طريقة حساب معامل الترابط بين نصفي الاختبار هما النصف للبنود الفردية وهو يرسم ب (x) ثم النصف للبنود الجوزية يرسم ب (y)، ولكل البنود 9 درجة، ويمكن حسابه

³⁹ Fatih Ozden, Gokhan Coskun, dan Serkan Bakirhan, "The test-retest reliability, concurrent validity, and minimal detectable change of the L test in patients with total hip arthroplasty," *Bulletin of Faculty of Physical Therapy*, 25 Agustus 2021, <https://bfpt.springeropen.com/articles/10.1186/s43161-021-00038-8>.

باستخدام برنامج ANATES V.0.5 2022. فيحسب فيه معامل الترابط بين درجات النصفين وسوف يستخرج الباحثان معامل الثبات الداخلي من البرنامج، ونتيجته كما يلي:

المعدل : 19,50

علاقة بين (x) و (y) : 0,62

ثبات الاختبار : 0,76

من النتيجة السابقة ظهرت لباحثان العلاقة بين قائمة (x) وقائمة (y) تعني (0,62)، فهذا يدل على أن النتيجة في المعامل حصلت إلى درجة "عالية"، وثبات الاختبار يساوي (0,76) كذلك في درجة عالية. كما عرفه غيلفورد (Guilford) بأن نتيجة 00,1-91,0 في درجة عال جدا، و 90,0-71,0 عال، و 70,0-41,0 متوسط، و 40,0-21,0 ضعيف، و 20,0-00,0 ضعيف جدا.

عرفه أريكونطا (Arikunto) إن مدى صعوبة الأسئلة في الاختبار تعرف من إجابة الطالبات، إذا كانت أغلبهن يستطعن على الإجابة بإجابة صحيحة وحصلت إلى النتيجة الكبيرة فهذه تدل على أن الأسئلة في مستوى السهلة، وأما إذا كانت الطالبات لم يقدرن على إجابتها بالإجابة الصحيحة ثم حصلت إلى النتيجة الصغيرة فتعتبر الأسئلة صعبة، وهذه ترى من مجموعة النتائج لكل الأسئلة، والمعايير لتحديد مستوى صعوبة الأسئلة إذا حصلت النتيجة 0 فهي صعبة جدا، و 0,29-0,1 صعبة، و 0,30-0,69 متوسطة، و 0,99-0,70 سهلة، و 1 سهلة جدا. رأى الباحثان أن مستوى صعوبة الأسئلة أغلبها متوسطة لدى الطالبات، يعنى 12 سؤالاً من 18 سؤالاً، وأما درجة غيرها مختلفة العدد في كل المستوى.

بعد حساب نتيجة صدق الاختبار وثباته، ثم معرفة مستوى صعوبة الأسئلة، فبعد ذلك لأبد من حساب مدى قوة التمييز للأسئلة لأن تحليل كل الأسئلة من أمر ضروري، ومن المؤكد أن السؤال يجب أن يكون قادراً على التمييز بين المشتركات المتميزات والضعيفات، فحسب الباحثان قوة تمييز الأسئلة كذلك بنفس البرنامج، والحاصل أن أغلب الأسئلة قوة تمييزه في درجة متوسط وضعيف، فعرفت كذلك عدد

الأسئلة يمكن إصلاحها أو تجديدها، تعني بعض الأسئلة في درجة ضعيف التي تسبب إلى عدم هام الأسئلة.

ففي آخر عرض البيانات وتحليلها، حصل الباحثان إلى النتيجة بأن ثبات الاختبار وصدقه حصل إلى الدرجة العالية، ومستوى صعوبته متوسطة، وقوة التمييز لكل الأسئلة أغلبها متوسط وضعيف. من هذه النتيجة قام الباحثان بالتطوير لمضمون الأسئلة في بعضها ليكون أحسن وأنسب، ولم يكن هذا التطوير لكل واحد من الأسئلة إلا ما يحتاج إلى إصلاحها.

من هذه النتيجة وجد الباحثان 7 أسئلة وجب على الباحثان التحليل لمرة عديدة لتعرف منه الباحثان الأشياء لابد تحسينها.

بعد أن قام الباحثان بإجراء التجربة الثانية في الفصل السادس في جامعة دار السلام كونتور للبنات وحصل الباحثان إلى نتائجهم في الاختبار، قام الباحثان بحسابها في برنامج ANATES V.0.5 2022، والبيانات التي حصل عليها الباحثان تشرحها فيما يلي.

المعدل : 100,27

علاقة بين (x) و (y) : 0,74

ثبات الاختبار : 0,85

من النتيجة السابقة ظهرت لباحثان العلاقة بين قائمة (x) و قائمة (y) تعني (0,74)، فهذه النتيجة تدل على أن نتيجة المعامل حصلت إلى درجة "عالية"، وثبات الاختبار يساوي (0,85) كذلك في درجة عالية، هذا بالإضافة إلى درجة العلاقة عند جون بيس كما في المعيار التي وضعها الباحثان في التجربة الأولى.

استنبط الباحثان من هذا الجدول على أن درجة صعوبة الأسئلة في هذا الاختبار أغلبها المتوسطة، والأخرى سهلة من غير الأسئلة الصعبة، فلا توجد الأسئلة الصعبة في الاختبار للتجربة الثانية.

بعد حساب نتيجة صدق الاختبار وثباته، ثم معرفة مستوى صعوبة الأسئلة، فبعد ذلك لابد من حساب مدى قوة التمييز للأسئلة لأن تحليل كل الأسئلة من أمر ضروري، ومن المؤكد أن السؤال يجب أن يكون قادرًا على التمييز بين المشتركات المتميزات والضعيفات. حسب الباحثان قوة تمييز الأسئلة كذلك ببرنامج ANATES V.0.5 2022. من نتيجة التجليل وجد الباحثان سؤالان الذين وجب على الباحثان التحليل لمرة عديدة ليعرف الباحثان الأشياء يجب تحسينها.

مرحلة التقويم

صدق الاختبار هو أول ما يجب أن تحسنه الباحثان بعد تصميم الاختبار وتجربه في التجربة الميدانية، فهو الشرط الأول لابد أن يحصل إليه الاختبار الجيد، أي بمعنى أن يقيس الاختبار ما وضع من أجله لا لغيره. ولقد تم تصميم الاختبار الذي صممه الباحثان بشروطه الكثيرة التي يجب توافرها في الاختبار الجيد. فشروط الاختبار الجيد يعني أن يحصل الاختبار إلى درجة الصدق العالي في صدق المحتوى وصدق البناء. وقد أتمم الباحثان كل ذلك حتى حصلت إلى درجة الصدق المرجوة. فبعد التجربة الميدانية لمرة الأولى والثانية، وجد الباحثان من نتيجة برنامج "ANATES V.0.5 2022" أن عدد الأسئلة الصادقة في التجربة الأولى يعني 11 وغير صادق 7. وفي التجربة الثانية أن سؤالين غير صادق والباقي 15 أسئلة صادق.

وشروط الاختبار الجيد بعد الصدق يعني ثبات الاختبار، فالاختبار بنتيجة الثبات الجيد هو الذي كان نتائجه متقاربة من وقت إلى وقت آخر، أو كان الحاصل من الاختبار لا يختلف كثيرا إذا طبق أكثر من مرة في ظروف متماثلة.

كما عرفه عبد الخالق أن صدق الاختبار شرط من شروط الاختبار الجيد، فكذلك ثبات الاختبار جزء من شروطه، وله خمسة طرق التحديد ليكون الاختبار ثابتا وهي إعادة الاختبار نفسه والصور المتكافئة والثبات النصفى واتفاق المقيمين والخطأ المعياري. ولقد أجر الباحثان إحدى من هذه الطرق وحسبه ببرنامج ANATES V.0.5 2022 حتى حصل الباحثان إلى نتيجة عالية، بأن نتيجة ثباته بلغ 0,76 في التجربة الأولى و0,85 في التجربة الثانية.

وشروط الاختبار الجيد من غير الصدق والثبات هو التنبيه إلى مدى صعوبة أسئلة الاختبار، وهذا ليكون أسئلته مناسبة بمستوى المجتمع المختبر، فلا يمكن أن يختبر الطالبة في المستوى المبتدئ بالاختبار في مستوى متقدم أو عكسه. والتنبيه إلى حساب مستوى صعوبة الأسئلة كذلك ليسهل وضع الأسئلة لترتيب مضمون الأسئلة من السهلة إلى الصعبة. يمكن معرفة مستوى صعوبة الاختبار إلى مجموعة النتائج من إجابة المشتركات لكل رقم الأسئلة ومجموعة النتائج في الاختبار. وقد حسب الباحثان هذه الأشياء باستخدام برنامج ANATES V.0.5 وحصلت إلى النتيجة عن مدى صعوبة الاسئلة وسهولته.

عدد الأسئلة في التجربة الأولى يعني 18 سؤالاً، سؤال واحد صعبة، 12 أسئلة متوسطة، 3 سؤالاً سهلة، وسؤالان سهل جداً. في التجربة الثانية عدد الأسئلة يعني 17 سؤالاً، 12 أسئلة متوسطة، 5 سؤالاً سهلة ولم توجد الأسئلة الصعبة أو صعبة جداً في التجربة الثانية.

من التجربة الأولى والتجربة الثانية، لم يبلغ الاختبار إلى مائوية مستوى صعوبة الاختبار، بأن يكون عدد الأسئلة الصعبة 25%، والمتوسطة 50%، والسهلة 25%. أو 20% الأسئلة الصعبة، ثم 60% متوسطة، و20% سهلة. أو 15% الأسئلة الصعبة، ثم 70% المتوسطة، و15% السهلة.⁴⁰

استناداً إلى الغرض من تحليل كل الأسئلة، يمكن أن تؤثر القوة التمييزية على نتيجة ثبات الاختبار، عالية كانت أو سافلة. ومن نتائج حساب قوة التمييز، يمكن أيضاً معرفة الأسئلة التي يجب قبولها أو تصحيحها وتجديدها أو رفضها بحيث يمكن استخدام الاختبار كأداة قياس جيدة. ونتيجة حساب قوة تمييز الأسئلة لكل التجربة لا يختلف كثيراً، في التجربة الأولى سؤالان جيد في التمييز و8 أسئلة متوسط و8 أسئلة ضعيف. في التجربة الثانية سؤالان جيد التمييز و10 أسئلة متوسط و5 أسئلة ضعيف.

⁴⁰ Zainal Arifin, *Evaluasi Pembelajaran, Prinsip, Teknik, Prosedur* (Bandung: Remaja Rosdakarya, 2009), 270.

والنتيجة المعنوية لكل الأسئلة في التجربة الثانية أقل من التجربة الأولى. في كان عدد الأسئلة الناقصة في النتيجة المعنوية 7 سؤال، والتجربة الثانية بعدد سؤالين ناقص النتيجة المعنوية، فهذا يدل على أن تحسين الأسئلة التي فعله الباحثان تؤثر في النتيجة المعنوية حتى تنقص عدد الأسئلة التي ليست لها النتيجة المعنوية. فتكون هذه النتيجة هي نتيجة آخرة لمناقشة النتائج بعد التجربة الميدانية. وعموماً أن النتيجة المستنبطة من كل عناصر التي أجر الباحثان تحليلها تعني "جيد".

الخاتمة

بعد أن قام الباحثان بعرض البيانات وتحليلها ومناقشتها، استخلص منها النتائج بأن أسئلة الاختبار الذي صممه الباحثان يعتبر أنه جيد ويمكن استخدامه في الاختبار اللغوي في الجامعة، بأن هذا الاختبار قد وافى الشروط للاختبار الجيد، وهي درجة صدق أسئلة الاختبار في التجربة الثانية أكثره عال غير سؤالين من 17 سؤالاً، والثبات للاختبار كذلك عال، مستوى صعوبة كل الأسئلة وقوة التمييز للاختبار كذلك جيد، ودليل ذلك أن نتيجة صدقه بلغ 0,61 وثباته 0,76 في التجربة الأولى، ثم 0,74 لصدقه و0,85 لثباته في التجربة الثانية.

رغم بذلك، هناك بعض الجوانب لم يحصل عليها الباحثان بأن بعض الرقم من الأسئلة ناقص النتيجة المعنوية وهذا نعني صدق كل أسئلة الاختبار، من مجموع 17 سؤالاً في التجربة الثانية أو الآخرة، تم تحديد 13 سؤالاً على أن صدقها عال جداً، وسؤالان عال، وسؤالان ناقص (-)، مما يعني أن هذين سؤالين يحتاج إلى إعادة تحليل بشأن الجوانب التي تجعلها غير صادق بحيث يمكن تحسين الأسئلة.

لا يزال هذا البحث بعيداً من الكمال بسبب قدرة الباحثان وكفاءتهما على إجراء هذا البحث. تمنى الباحثان من الباحث الأخرى تبحث بعدهما في مجال الاختبارات اللغوية في المهارات اللغوية خاصة لمهارة الكلام، حتى يكون أسئلة اختبار الكلام معياري ومثلي وواسع الانتشار، وفي التصميم لابد أن يتبع الباحث الأخرى المعايير في التصميم كما كتبه الباحثان في هذا البحث أو معايير ما الذي وفق أبحاثها.

References

- Akmalia, Frida, Syihabuddin Syihabuddin, dan Yayan Nurbayan. 2021. "Intensive Course: the Alternative Program for Teaching Arabic Speaking and Reading Skills at an Islamic Boarding School/al-Daurah al-Mukatsafah: al-Barnamij al-Badil Li Ta'lim Maharah al-Kalam wa al-Qira'ah Fi al-Ma'had." *Arabiyatuna : Jurnal Bahasa Arab* 5, no. 1 Mei (7 Mei): 113–26. <https://doi.org/10.29240/jba.v5i1.2306>.
- Ali al Khuliy, Muhammad. 1998. *Al Ikhtibarat al Tabshiliyyah, 'Idaduha wa Ijra'uba wa Tabliluha*. Yordan: Dar al Falah.
- Ali, Islam Yasri. 2016. "Muraja'atu Kitab al Ithar al Marja'iy al Aurubiy al Musytarak Lil Lughat: Dirasah,." *Majallah al Dirasat al Lughawiyah wa al Adabiyah*.
- Arifin, Zainal. 2009. *Evaluasi Pembelajaran, Prinsip, Teknik, Prosedur*. Bandung: Remaja Rosdakarya.
- Arikunto, Suharsimi. 2013. *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: PT Asdi Mahasatya.
- As Sayyid Ali, Osamah Zakky. , 2016. *Al Ikhtibarat al Lughawiyah, Muqarabah Manhajiyah Tatbiiyyah*. Riyadh: Markaz al Mulk Abdullah Bin Abdul Aziz al Duwaliy.
- Auruba, Majlis. 2000. *Al Ithar al Marja'i al Aurubiy al Musytarak lil Lughat (Dirasah, Tadris, Taqyim)*. al Qahirah: Dar Ilyas al 'Ishriyyah.
- Chanturia, Anna Siyanova, dan Stefania Spina. 2019 "Multi-Word Expressions in Second Language Writing: A Large-Scale Longitudinal Learner Corpus Study." *Language Learning A Journal of Research in Language Studies* 70, no. 2 (20 Desember). <https://doi.org/10.1111/lang.12383>.
- Colino, Enrique, Javier Sanchez, Jorge Garcia Unanue, Ester Ubago Guisado, Pascal Haxaire, Aurelien Le Blan, dan Leonor Gallardo. 2017. "Validity and reliability of two standard test devices in assessing mechanical properties of different sport surfaces." *Polymer Testing* 62 (September). <https://doi.org/10.1016/j.polymertesting.2017.06.011>.
- Council of Europe. no year. *Common European Framework of Referennce for Languages: Learning, Teaching, Assessment*. Strsbourg: Language Policy Unit.
- Delgado, C., A. Araneda, dan M. I. Behrens. 2019. "Validation of the Spanish-language version of the Montreal Cognitive Assessment test in adults older than 60 years." *Neurologia (English Edition)*. Elsevier, <https://www.sciencedirect.com/science/article/pii/S2173580818302074>.

- ESOL, Cambridge. 2011, "Using the CEFR: Principles of Good Practice." *Cambridge United Kingdom*.
- Gall, Meredith, Joyce Gall, dan Walter Borg. 2003. *Educational Research An Introduction*. New York: Library of Congress Cataloging,
- Harsch, Claudia, dan André Alexander Rupp. 2011. "Designing and Scaling Level-Specific Writing Tasks in Alignment With the CEFR: A Test-Centered Approach." *Language Assessment Quarterly* 8, no. 1 (31 Januari): 1–33. <https://doi.org/10.1080/15434303.2010.535575>.
- Ismail, Muhammad. 2022. *Al Ikhtibarat al Lughawiy bi Jami'ati Darissalam Gontor*.
- Jasińska, Kaja, Yapo Hermann Akpe, Blahoua Axel Debaye Seri, Benjamin Zinszer, Rodrigue Yoffo Agui-Kouadio, Kelsey Mulford, Erin Curran, Mary-Claire Ball, dan Fabrice Tanoh. 2022. "Evaluating Bilingual Children's Native Language Abilities in Côte d'Ivoire: Introducing the Ivorian Children's Language Assessment Toolkit for Attié, Abidji, and Baoulé." *Applied Linguistics* 43, no. 6 (1 Desember): 1116–42. <https://doi.org/10.1093/applin/amac025>.
- Khaliq, Muhammad Abdul. 1997. *Ikhtibarat al Lughab*. Ar Riyadh: Mathabi' Jamiati al Mulk Su'ud.
- Mahmudi, Ihwan. 2020. *Evaluasi Pendidikan*. Yogyakarta: Lintang Books.
- . 2020. *Pengembangan Instrumen Penelitian Sosial (Konsep, Contoh, dan Analisis Data Menggunakan SPSS)*. Yogyakarta: Lintang Books.
- . 2018 "Rubrik Analitik Penilaian Hasil Belajar Praktik Pendidikan Agama Islam." *Fikrah: Journal of Islamic Education* 2, no. 2: 19.
- Manzano, Dexter L. 2022 "Examining the Interrater Reliability Between Self- and Teacher Assessment of Students' Oral Performances." *International Journal of Language Testing* 12, no. 2 (1 Oktober): 128–44. <https://doi.org/10.22034/ijlt.2022.157130>.
- Mei, Leong Lai, dan Ahmadi Seyedeh Masoumeh. 2017. "AN ANALYSIS OF FACTORS INFLUENCING LEARNERS' ENGLISH SPEAKING SKILL" 2, no. 1 (1 Januari): 34–41.
- Mohamed, Salwa. 2021, "The Development of an Arabic Curriculum Framework Based on a Compilation of Salient Features from CEFR Level Descriptors." *The Language Learning Journal*, 20 Mei1–15. <https://doi.org/10.1080/09571736.2021.1923781>.
- Mubarak, Faisal, Abdul Sattar, dan Ahmad Fahmy Arief. 2020. "Thasmim Mawad Maharah al-Kalam Fi Istimdadiha Min al-Ayat al-Istifhamiyah Fi Surah al-An'am wa al-Mulk Li Thulab al-Ma'had al-'Aly Ulum al-Qur'an

- Amuntai.” *Arabiyatuna: Jurnal Bahasa Arab* 4, no. 1 Mei (8 Mei): 69–90. <https://doi.org/10.29240/jba.v4i1.1320>.
- Ozden, Fatih, Gokhan Coskun, dan Serkan Bakirhan. 2021. “The test-retest reliability, concurrent validity, and minimal detectable change of the L test in patients with total hip arthroplasty.” *Bulletin of Faculty of Physical Therapy*, 25 Agustus <https://bfpt.springeropen.com/articles/10.1186/s43161-021-00038-8>.
- Rasyidi, Abdul Wahab, dan Suci Ramadhanti Febriani. 2020. “Uslub Ta’lim Maharah al-Kalam Bi Asas al-Dzaka’at al-Muta’addah Fi al-Marhalah al-Ibtidaiyyah.” *Arabiyatuna: Jurnal Bahasa Arab* 4, no. 2 November (17 November): 193–210. <https://doi.org/10.29240/jba.v4i2.1432>.
- Rini, Rini, dan Partomuan Harahap. 2021. “Designing the Holistic Evaluation in Teaching Reading/Tasmim al-Taqwim al-Syumuli Fi Ta’lim al-Muthala’ah.” *Arabiyatuna: Jurnal Bahasa Arab* 5, no. 1 Mei (7 Mei): 141–60. <https://doi.org/10.29240/jba.v5i1.2522>.
- Ritonga, Apri Wardana, Wildana Wargadinata, Mahyudin Ritonga, dan Suci Ramadhanti Febriani. 2021. “Teaching Maharah Kalam on the Basis of Culture Using the Textbook ‘al-‘Arabiyyah Baina Yadaik’ at the Middle School/Ta’lim Maharah al-Kalam ‘Ala Asas al-Tsaqafah bi Istikhdam Kitab al-‘Arabiyyah Bayna Yadayk fi Marhalah al-Mutawassithah.” *Arabiyatuna: Jurnal Bahasa Arab* 5, no. 1 Mei (7 Mei): 1–22. <https://doi.org/10.29240/jba.v5i1.1942>.
- Sholeha, Nur Hanifatus, Abdul Hafidz Zaid, dan Fairuz Subakir. 2022. “Tashmim al-Kitab al-Raqmy Muassasan ‘ala Tathbiq Flippdf Professional fi Maharah al Istima’ li al-Fashl al-Tahdhiri.” *Arabiyatuna: Jurnal Bahasa Arab* 6, no. 2 November (4 November): 671–92. <https://doi.org/10.29240/jba.v6i2.5247>.
- Sugiyono. 2013. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Suhardi, Iwan, dan Hendriyati Eko Budiyaningsih. 2022, “Validasi Pakar Konten pada Pengembangan Paket Soal Bahasa Indonesia untuk Basis Data Computer Based Testing.”
- Thu’aimah, Rusydi Ahmad. 1986. *Al Marja’ fi Ta’limi al Lughah al Arabiyyah Li al Nathiqin bi Lughatin ‘Ukbro*. Jami’atu Ummul Qura: Ma’had al Lughah al Arabiyyah.
- Vandommele, Goedele, Koen Van Gorp, Kris Van den Branden, dan Sven De Maeyer. 2022. “Assessing the Development of L2 Speaking Skills in Inside-School and Outside-School Settings Through Growth

- Modelling.” *Applied Linguistics* 43, no. 6 (1 Desember): 1094–1115. <https://doi.org/10.1093/applin/amac024>.
- Wisniewski, Katrin. 2017. “Empirical Learner Language and the Levels of the Common European Framework of Reference.” *Language Learning A Journal of Research in Language Studies* 66, no. 1 (12 Januari). <https://doi.org/10.1111/lang.12223>.
- Yunus, Fathiy Ali, dan Muhammad Abdu Ar Ra’uf. 2003. *Al Marja’ fi Ta’limi al Lughah al Arabiyyah Lil Ajanib (Min al Nazhariyyah Ila al Tathbiq)*. Kairo: Maktabah wa Hibah.
- Zarkasyi, Ahmad Hidayatullah. 2022. “Madza Nafal Li Ajli Rafi al Fakhor fi Istikhdami al Lughah al Arabiyyah.” Dipresentasikan pada Al Muhadharah fi Maddati Tathwir al Manahij, Mabna Jami’ati Darissalam Gontor Li Dirasat al Ulya.
- Zarkasyi, Amal Fathullah. 2009. *Dauru Jami’ati Darissalam Gontor fi I’dadi al Mujaddidin*. Ponorogo: Jami’atu Darissalam Gontor Press.
- Zarkasyi, Hamid Fahmy, Kholid Muslih, Khoirul Umam, dan Yuangga Kurnia Yahya. 2020. *Pekan Perkenalan Khutbatu-L-’Ary Universitas Darussalam Gontor*. Ponorogo: UNIDA Gontor Press.