



## Development Of Arabic Language Learning Outcome Assessment Instrument Based On Hot Potatoes Application At Junior High School

Taskia Qalbi<sup>1</sup>, Yusring Sanusi Baso<sup>2</sup>, Yusuf T<sup>3</sup>

<sup>1,2,3</sup>UIN Alauddin Makassar, Indonesia

[taskiaqalbi123@gmail.com](mailto:taskiaqalbi123@gmail.com)

### Abstract

Copyright (c) 2026 Taskia Qalbi, Yusring Sanusi Baso, Yusuf T(Author)

This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).



Doi:

<https://doi.org/10.47435/jpdk.v11i01.4376>

Arabic language education at SMPIT Darul Fikri Makassar faces challenges with conventional assessments that lack interactivity, subjectivity, and fail to motivate Generation Z students, particularly in skills like qira'ah (reading), kitabah (writing), istima' (listening), and kalam (speaking). This study aims to develop a learning outcome assessment instrument based on the Hot Potatoes application using the simplified Borg and Gall Research and Development (R&D) model across eight stages: needs analysis, planning, prototype development, expert validation, revision, field testing, and final product. Primary data sources include Arabic teachers and 8th-grade students, with secondary sources like syllabi and lesson plans (RPP); data collection involved observation, interviews, questionnaires, and tests. Results indicate the instrument has very high content validity (96.43%) from Arabic, educational evaluation, and media experts, plus empirical validity with 26 out of 30 items valid. Reliability reaches 0.87 (Cronbach's Alpha, high category), balanced difficulty levels (63% moderate), and good item discrimination. Practicality is rated very high by teachers (95.83%) and students (87.62%), thanks to automatic scoring, instant feedback, and appealing design. Effectiveness testing shows pre-test (60.20) to post-test (80.20) improvement, raising mastery from 40.60% to 85.67%, with t-value 8.724 ( $p=0.000$ ) and effective gain score. The instrument proves valid, reliable, practical, and effective in enhancing learning outcomes. Overall, this development offers an innovative digital assessment model for Arabic education in Islamic schools, supporting hybrid learning post-pandemic and student motivation.

**Keywords:** Assessment instrument, Hot Potatoes, Arabic language

### 1. Introduction

Arabic language education at SMPIT Darul Fikri Makassar still faces significant challenges in measuring and improving student learning outcomes, particularly in aspects of vocabulary, grammar, and their practical application. Preliminary research shows that conventional assessment instruments such as written tests and manual observations often fail to comprehensively depict student understanding, leading to inaccuracies in evaluation (Arifianto et al., 2021).

The use of these traditional assessment tools also results in a lack of interactivity, making students tend to be passive and reducing their learning motivation (Kahar et al., 2021). In the current digital era, such conventional approaches are no longer relevant to the characteristics of Generation Z students who are more responsive to interactive technology (DEwi et al., 2024).

Assessing learning outcomes is a crucial element in the Arabic language learning process, encompassing four main skills: qira'ah (reading), kitabah (writing), istima' (listening), and kalam (speaking) (Kartika & Arifudin, 2023). Accurate assessment is needed to ensure competency



achievement aligns with curriculum goals (Ahsanuddin et al., 2023), yet field practices are often hindered by teacher subjectivity and time constraints (Baro'ah et al., 2023).

The main challenges include difficulties in evaluating complex aspects such as morphology, pronunciation, and contextual understanding, which are time-consuming and prone to bias (Kurniati et al., 2022). This not only burdens teachers but also negatively impacts student confidence, hindering the holistic development of Arabic language skills (Asngad et al., 2023).

Preliminary observations at SMPIT Darul Fikri reveal that teachers still rely on simple worksheets or manual spreadsheets, which do not provide instant feedback. This situation underscores the urgency of developing innovative digital-based assessment instruments to enhance learning efficiency and appeal.

Post-COVID-19 pandemic, hybrid learning models have become the norm, accelerating the need for adaptive digital tools (Buwono et al., 2022). Generation Z students respond positively to gamification and personalization (Bustang, 2022), which can dynamically measure learning outcomes and bridge the gap between theory and practice in Islamic schools (Destriani, 2022).

From an educational theory perspective, Hot Potatoes software offers an ideal open-source solution for creating interactive exercises such as multiple-choice, sentence arrangement, matching, and crossword puzzles (Yasa et al., 2020). Its automatic feedback feature has proven to improve the accuracy of Arabic learning outcome measurements (Arridha & Roy, 2025).

The Research and Development (R&D) approach using the Borg and Gall model was chosen because it provides 10 systematic stages, from needs analysis to dissemination (Sugiono, 2009). This model ensures the developed instrument is valid, reliable, and contextual for SMPIT Darul Fikri Makassar.

This research is inductive in nature, starting from local problems toward general contributions to Arabic education. Its goal is to develop Hot Potatoes instruments aligned with school curriculum, enhancing assessment effectiveness and overall learning quality. This development is expected to serve as a replicable model for other Islamic schools in Indonesia, bridging the challenges of conventional assessment instruments with the potential of digital technology for innovative, sustainable, and student-centered religious education.

## **2. Methods**

This research employs a research and development (R&D) approach commonly known in the field of education. In general, research and development or R&D is understood as a systematic process or series of steps aimed at creating new products or improving the quality of existing ones (Fadli, 2021). This is because the study focuses on developing the Hot Potatoes application as an Arabic language learning medium and testing its effectiveness.

### **2.1 Data Sources**

Data sources refer to the subjects from which data can be obtained. Data sources can be objects, movements, humans, places, and so forth. The data sources used by the researcher in this study consist of two types: primary data sources and secondary data sources (Yusanto, 2020). Primary data in this research comprise Arabic subject teachers and 8th-grade students. Teachers serve as expert material validators and primary users of the instrument, while students act as product trial respondents to assess the effectiveness aspects of the developed assessment instrument. Secondary data sources required in this research include written documents and school archives such as: syllabus documents, Lesson Plan Implementation (RPP), and student grade recaps, which are used as references to design assessment indicators aligned with the basic competencies of the curriculum applied at SMPIT Darul Fikri Makassar.



## 2.2 Data Collection Methods

Data collection techniques in this research and development study are adapted to each stage of the Borg and Gall model, encompassing preliminary study, initial product development, limited trial, extensive trial, up to final product revision (Sanjaya, 2013).

- a. Observation was conducted during Arabic language classroom learning processes to gather data on how teachers perform assessments and the level of student engagement in learning. Observation was also carried out during instrument trials to evaluate the feasibility and practicality of using the Hot Potatoes application.
- b. Structured interviews were conducted with Arabic language teachers to gather information regarding the needs for assessment instruments, difficulties in evaluating student learning outcomes, and their views on the use of digital tools in assessment.
- c. Questionnaires were administered to material experts and media experts for content validation, construction, and instrument appearance. Meanwhile, questionnaires were given to teachers and students to assess the practicality and attractiveness of the instrument after trials.
- d. Tests were conducted using the Hot Potatoes application containing Arabic language questions aligned with basic competencies. Test result data were used to analyze item validity, reliability, difficulty level, and discrimination power, as well as to measure the instrument's effectiveness in improving student learning outcomes.

## 3. Results and Discussion

### Development of Hot Potatoes-Based Arabic Learning Assessment Instrument

The initial stage of this research instrument development involved needs analysis to gain a comprehensive understanding of student conditions, teachers, and evaluation media used during Arabic learning. Based on observation and interview results, most students lacked motivation when facing evaluation processes. This was caused by the use of conventional assessment instruments that lacked variety and digital media support to attract attention and enhance student interaction. Additionally, some students admitted difficulty understanding question instructions, quickly feeling bored, and lacking confidence during evaluations.

This situation shows that traditional assessment processes have not fully facilitated 21st-century learning needs demanding interactivity and clear media. Therefore, researchers initiated the development of a Hot Potatoes application-based assessment instrument as a form of modernizing Arabic learning evaluation. This digital evaluation media is expected to create more engaging assessment experiences, ease of use, and provide direct feedback to students.

The design stage was the researcher's step in creating a digital assessment format suitable for competencies and student characteristics. At this stage, researchers compiled question grids, formulated achievement indicators, and developed items in three main forms (multiple-choice, short fill-in, and matching). All questions were then integrated into the Hot Potatoes application, utilizing Jquiz, Jcloze, and Jmatch features to produce interactive evaluation instruments. Media display was designed as attractively as possible, considering Arabic font size, user-friendly colors, and easy navigation. At this design stage, researchers also prepared supporting instruments in the form of expert validation sheets and student response questionnaires to measure product validity, practicality, and effectiveness.

The development stage included the process of creating and refining digital assessment instrument prototypes that had undergone evaluation by experts: Arabic language experts, educational evaluation experts, and digital learning media experts. After receiving assessments and recommendations from the experts, researchers conducted revisions resulting in Prototype II, which was improved in content, construction, and media appearance. Prototype II was then tested through trials on students. Trial results showed that the Hot Potatoes-based assessment instrument was not only academically feasible but also appealing to students and helped improve their understanding of tested



materials. Additionally, teachers stated that this media simplified the assessment process due to automatic scoring features and direct feedback generated by the application.

Thus, the development of Hot Potatoes-based assessment instruments proved capable of addressing problems identified in the initial analysis stage and providing more modern, effective assessment alternatives suited to learning needs at SMPIT Darul Fikri Makassar.

### **Validity Level of the Arabic Learning Outcome Assessment Instrument Based on Hot Potatoes Application**

In this research, the validity of the Arabic learning outcome assessment instrument was examined through two approaches: content validity and empirical validity. This dual approach aligns with modern educational evaluation perspectives emphasizing that instrument validity must be supported by theoretical suitability as well as empirical evidence from field data.

Content validation results showed that the developed assessment instrument met alignment with basic competencies, competency achievement indicators, and grade IX Arabic learning materials at SMPIT Darul Fikri Makassar. Experts assessed that the items represented the domains to be measured, both in vocabulary mastery aspects, sentence construction, and imla' skills. These findings indicate that instrument construction was based on clear curriculum foundations and learning objectives.

Very high validity percentages, averaging 96.43% and categorized as very valid, demonstrate optimal integration between learning objectives, indicators, materials, and question forms. According to content validity theory, an instrument is valid if each item representatively reflects the measured competency domain and is relevant to the learning context. Thus, these validation results strengthen that the developed assessment instrument has strong academic legitimacy from the content aspect.

Additionally, media expert validation showed that the Hot Potatoes application-based instrument had attractive visual displays, easy-to-use navigation, and media functions supporting the learning evaluation process. This is important considering digital assessment instruments are not only required to be substantively valid but must also meet display clarity and ease-of-use principles to not disrupt student ability measurement. Some revision suggestions from experts, particularly regarding language wording and distractors, were part of the instrument refinement process and did not reduce overall instrument feasibility.

Empirical validity was analyzed through Pearson Product Moment correlation based on student responses to 30 trialed items. Analysis results showed 26 items declared valid, while 4 items were invalid and recommended for revision or deletion. These findings indicate most items had significant relationships between item scores and total scores, enabling consistent student ability measurement.

Good empirical validity reflects the instrument's discriminative ability to distinguish students with high and low material mastery levels. This distinguishing ability is a key characteristic of quality assessment instruments, as good instruments not only measure learning outcomes but also identify ability variations among students. The presence of some invalid items shows empirical testing plays a crucial role in filtering theoretically good items that don't perform optimally in the field.

Reliability testing using Cronbach's Alpha formula showed a coefficient value of 0.87, falling in the high reliability category. This value indicates the assessment instrument has very good internal consistency and can provide stable measurement results. In educational evaluation contexts, high reliability is an absolute prerequisite for trustworthy assessment data serving as learning decision-making bases.

Reliable instruments ensure score differences obtained by students are caused by ability differences, not instrument inconsistencies. Therefore, the 0.87 reliability value strengthens findings that the Hot Potatoes-based Arabic learning outcome assessment instrument is feasible for use as a dependable measurement tool.

Difficulty level analysis showed most items fell in the moderate category (63%), followed by easy (20%) and difficult (17%). This composition reflects ideal question difficulty balance. Moderate difficulty questions are considered most effective for measuring student abilities because they are neither too easy nor too difficult, providing more accurate ability pictures.



Discrimination power analysis results showed most items fell in adequate to good categories. This indicates the instrument adequately distinguishes high and low ability students. Some items with low discrimination power were recommended for revision, particularly in question wording and distractor aspects. These recommendations align with instrument development principles emphasizing discrimination power as a key item quality indicator directly contributing to overall instrument validity.

Based on all these findings, it can be concluded that the Hot Potatoes-based Arabic learning outcome assessment instrument developed at SMPIT Darul Fikri Makassar has good to very good validity levels, both content and empirical validity, supported by high reliability and proportional item characteristics. Thus, this instrument is feasible for use as an Arabic learning outcome assessment tool and can serve as a basis for further development stages.

### **Practicality Level of the Arabic Learning Outcome Assessment Instrument Based on Hot Potatoes Application**

Based on teacher responses, the developed assessment instrument achieved 95.83% practicality percentage and 95.00% attractiveness, both falling in the very practical category. These high percentages indicate teachers experienced no significant obstacles in using the instrument, both in application operation, instruction clarity, and smooth access via available school devices. These findings reinforce the view that assessment instrument practicality is highly determined by ease of use and design suitability with user needs.

Additionally, teachers assessed that Hot Potatoes application use improved time efficiency in assessment implementation, particularly in correction and learning outcome processing. This application's automatic assessment feature enables teachers to obtain evaluation results quickly and accurately, optimizing learning time utilization. This aligns with research findings stating digital technology-based assessments contribute to reducing teacher administrative burdens and improving evaluation activity effectiveness. Thus, the developed instrument is not only technically practical but also provides functional benefits in daily learning practices.

From the student perspective, questionnaire results showed 87.62% practicality and 90.00% attractiveness percentages, falling in very practical and very attractive categories. Students rated the instrument easy to understand, clear question completion steps, and attractive application display. This condition shows Hot Potatoes-based assessment instruments adapt well to junior high student characteristics responsive to digital technology use in learning. Previous research confirms interactively designed digital assessments improve student comfort and engagement in evaluation processes.

High attractiveness aspects also contribute to practicality perceptions. Question form variations and visual displays provided in the Hot Potatoes application make evaluation processes non-monotonous and more enjoyable for students. This positively impacts student motivation in taking tests, so evaluations are no longer viewed as stressful activities. These findings align with opinions that visual attractiveness and interactivity are important components in digital assessments for improving learning interest and student participation.

Additionally, students felt benefits from quickly obtained completion feedback. This feedback helps students know their learning achievements and understand unmastered material parts. In modern learning evaluation contexts, fast feedback is one practicality indicator, supporting assessment functions as learning media (assessment for learning), not merely learning outcome measurement.

Overall, this discussion shows high practicality levels of the Hot Potatoes-based Arabic learning outcome assessment instrument influenced by ease of use combinations, time efficiency, instruction clarity, display attractiveness, and pedagogical benefits for teachers and students. Therefore, the developed instrument is not only technically feasible but also relevant and applicable for use as an assessment tool in Arabic learning at SMPIT Darul Fikri Makassar.

### **Effectiveness Level of the Arabic Learning Outcome Assessment Instrument Based on Hot Potatoes Application**

Instrument effectiveness becomes an important indicator in educational development research, showing how far the instrument impacts student learning outcome achievement. An instrument is



considered effective if its implementation produces better learning outcome improvements compared to pre-use conditions. Therefore, in this research, the Hot Potatoes-based Arabic learning outcome assessment instrument was analyzed through pre-test and post-test score comparisons, gain score calculations, and paired sample t-test. This approach aligns with educational evaluation principles emphasizing instrument effectiveness proven through significant quantitative and statistical learning outcome changes.

Research results showed student average scores experienced meaningful increases, from 60.20 on pre-test to 80.20 on post-test. This score increase indicates that after Hot Potatoes-based assessment instrument use, students had better understanding of tested Arabic materials. In learning evaluation studies, such test score increases reflect that assessment instruments not only function as measurement tools but also serve as media driving concept understanding strengthening and student competition. Thus, the developed instrument positively contributed to student learning outcome achievement.

Besides average score increases, learning mastery percentages also experienced significant increases, from 40.60% initial conditions to 85.67% after instrument implementation. This increase shows most students achieved Minimum Completeness Criteria (KKM) after following assessments using Hot Potatoes-based instruments. These findings strengthen views that interactively and systematically designed digital assessments improve student engagement and impact more optimal learning outcome achievement. Several Indonesian studies also show technology-based assessment instrument use significantly improves learning mastery compared to conventional paper-based assessments.

Statistical test results using paired sample t-test showed t-count of 8.724 with significance value (p-value) of 0.000 ( $p < 0.05$ ). These results confirm significant differences between pre-test and post-test student scores after assessment instrument use. Statistically, these findings show student learning outcome improvements were not coincidental but real impacts from Hot Potatoes-based Arabic learning outcome assessment instrument use. Paired t-test use as effectiveness analysis tools has been widely used in Indonesian educational research to assess learning intervention or assessment influences.

Gain score analysis further strengthened t-test results by showing student learning outcome improvements in the effective category. Gain scores measure relative student ability improvements considering initial and final abilities. Positive gain values indicate assessment instruments provide real contributions to student competency improvements. In educational evaluation literature, gain score analysis is viewed more representatively for assessing learning intervention effectiveness because it not only compares average scores but also considers achieved improvement levels by students.

Based on all analysis results, it can be concluded that the Hot Potatoes-based Arabic learning outcome assessment instrument proved effective in improving student learning outcomes. Instrument effectiveness is reflected in average score increases, learning mastery percentage increases, positive gain scores, and statistical test results showing significant differences between pre and post-implementation conditions. These findings align with modern educational evaluation paradigms emphasizing effective assessment instruments must provide fast feedback, motivate students, and support optimal learning goal achievement.

#### **4. Conclusion**

The development of the Arabic language learning outcome assessment instrument based on the Hot Potatoes application at SMPIT Darul Fikri Makassar has successfully followed the Borg and Gall model simplified into eight stages. This process encompassed needs analysis, planning, initial product development, expert validation, revision, field trials, and final refinement. This systematic approach ensures the instrument is designed according to teachers' and students' needs, producing an interactive, practical, and digital technology-based assessment tool to support more effective Arabic learning.

The resulting instrument demonstrates very good validity, as evidenced by content validation from Arabic language, educational evaluation, and learning media experts. Aspects of material



suitability, indicators, question construction, as well as media display and functionality, were declared highly valid. Empirical validity testing further confirms that most items are valid and accurately measure student competencies, making this instrument suitable as an evaluation tool for Arabic learning at the junior high school level.

The instrument's reliability reaches a high level with a Cronbach's Alpha coefficient of 0.87, indicating strong internal consistency and stable, trustworthy measurement results. Additionally, feedback from teachers and students categorizes the instrument as very practical, featuring ease of use, automatic scoring systems that expedite processes, and engaging evaluation experiences enjoyable for students.

Overall, this assessment instrument is not only valid and reliable but also practical in implementation, opening opportunities for digital-based assessment innovations in Arabic education. Its use is expected to enhance learning evaluation quality and student motivation at SMPIT Darul Fikri Makassar.

### Bibliography

- Ahsanuddin, M., Mahliatussikah, H., Asrori, I., Wargadinata, W., Hasanah, M., & Hasan, A. M. (2023). Pelatihan pembuatan tes interaktif HOTS bahasa Arab berbasis hot potatoes bagi guru bahasa Arab MI, MTs, dan MA Kabupaten Jember. *Tifani: Jurnal Penelitian Dan Pengabdian Kepada Masyarakat*, 3(3), 8–14.
- Arifianto, M. L., Ainin, M., Ahsanuddin, M., Irhamni, I., Fitria, N., Nikmah, K., & Anwar, M. S. (2021). *Evaluasi pembelajaran dan pengembangan tes interaktif bahasa Arab*. Tonggak Media.
- Arridha, R., & Roy, A. (2025). Polinef-MI Pelatihan Aplikasi Hot Potatoes dalam Pembuatan Instrumen Soal Inovatif bagi Guru Madrasah Aliyah Negeri Fakkak. *Jurnal Inovasi Dan Pengabdian Kaa Mieera*, 3(1), 40–53.
- Asngad, T., Doloh, M., & Mahdi, A. (2023). Curriculum Development Innovation: Vocational High School Program Center of Excellence: Inovasi Pengembangan Kurikulum: Program Sekolah Menengah Kejuruan Pusat Keunggulan. *Edukasi: Journal of Educational Research*, 3(3), 101–112.
- Baro'ah, S., Trisnawati, S. N. I., Ernawati, A., Supatminingsih, T., Aziz, F., Aziz, M., Astuti, R., Isma, A., & Hasyim, S. H. (2023). Kurikulum Merdeka: Inovasi Kurikulum di Indonesia. *Penerbit Tahta Media*.
- Bustang, B. (2022). PENGGUNAAN MEDIA HOT POTATOES DALAM PENGUASAAN KOSAKATA BAHASA ARAB PADA SISWA KELAS VIII MTs GUPPI SAMATA KABUPATEN GOWA. *Jurnal Sarjana Ilmu Budaya*, 2(2 (Mei)).
- Buwono, S. R., Abubakar, L., & Handayani, T. (2022). Kesiapan Perbankan Menuju Transformasi Digital Pasca Pandemi Covid-19 Melalui Financial Technology (Fintech). *Jurnal Poros Hukum Padjadjaran*, 3(2), 228–241.
- Destriani, D. (2022). Inovasi pengembangan kurikulum pendidikan agama Islam di SMK Negeri 1 Rejang Lebong. *INCARE, International Journal of Educational Resources*, 2(6), 614–630.
- DEwi, D. S., Imanika, M. S., & Hidayat, M. (2024). Dampak Sistem Informasi dalam Manajemen Pendidikan: Studi di SDN 1 Parakanmanggu. *Jurnal Pelita Nusantara*, 1(4), 494–499.
- Fadli, M. R. (2021). Memahami desain metode penelitian kualitatif. *Humanika, Kajian Ilmiah Mata*



---

*Kuliah Umum*, 21(1), 33–54.

Kahar, M. I., Cika, H., Afni, N., & Wahyuningsih, N. E. (2021). Pendidikan Era Revolusi Industri 4.0 Menuju Era Society 5.0 Di Masa Pandemi Covid 19. *Moderasi: Jurnal Studi Ilmu Pengetahuan Sosial*, 2(1), 58–78.

Kartika, I., & Arifudin, O. (2023). Evaluasi Kinerja Guru Pendidikan Agama Islam Dalam Meningkatkan Mutu Pembelajaran Di Madrasah. *Jurnal Tahsinia*, 4(2), 510–523.

Kurniati, P., Kelmaskouw, A. L., Deing, A., Bonin, B., & Haryanto, B. A. (2022). Model proses inovasi kurikulum merdeka implikasinya bagi siswa dan guru abad 21. *Jurnal Citizenship Virtues*, 2(2), 408–423.

Salsabiillah, A. G., Astuti, N., & Mytra, P. (2025). Penerapan Model Direct Instruction untuk Meningkatkan Hasil Belajar Matematika Siswa. *Hybrid: Jurnal Pendidikan dan Pembelajaran Sains*, 4(1), 14-19.

Sanjaya, W. (2013). *Penelitian Pendidikan Jenis, Metode dan Prosedur*. Kencana Prenada Grup.

Sugiono. (2009). *Metode Penelitian Pendidikan Pendekatan Kualitatif, Kuantitatif, dan R&D*. Alfabeta.

Warahma, S. M., Wahdania, A., & Mytra, P. (2025). Analisis Implementasi Pembelajaran Berbasis Mindfulness Untuk Meningkatkan Ketenangan Siswa Sekolah Dasar. *Hybrid: Jurnal Pendidikan dan Pembelajaran Sains*, 3(3), 12-21.

Yasa, A. D., Suastika, K. K., & Zubaidah, R. S. A. N. (2020). Pengembangan e-evaluation berbasis aplikasi hot potatoes untuk siswa kelas V sekolah dasar. *Jurnal Ilmiah Sekolah Dasar*, 4(1), 26–32.

Yusanto, Y. (2020). Ragam Pendekatan Penelitian Kualitatif. *Journal of Scientific Communication (Jsc)*, 1(1).