

Development of a Wordwall-Based Test Instrument for Qur'an and Hadith Learning Assessment at Madrasah Al Yusra in Gorontalo City

Muthmainah Putri Bolilio^{1*}, Ibnu Rawandhy N. Hula², and Moh. Zulkifli Paputungan³

UIN Sultan Amai Gorontalo
muthmainahbolilio24@gmail.com*

ARTICLE INFO

Article history

Received June 23, 2026
Revised July 29, 2026
Accepted August 22, 2026

Keywords: Wordwall, Test Instrument, Learning Assessment, Al-Qur'an Hadith, Research and Development.

ABSTRACT

This study aims to develop a Wordwall-based test instrument for assessing Al-Qur'an and Hadith learning at Madrasah Aliyah Al Yusra in Gorontalo City and to determine its validity and initial effectiveness based on a limited trial. This study employed a Research and Development (R&D) approach using the ADDIE model, consisting of Analysis, Design, Development, Implementation, and Evaluation. The research subjects included media experts, subject matter experts, teachers of Al-Qur'an and Hadith, and five tenth-grade students involved in the limited trial. Data were collected through observation, questionnaires, documentation, and learning outcome tests. Data were analyzed using quantitative descriptive analysis and N-Gain analysis. The results showed that the final average validation score was 88.75%, categorized as highly valid, consisting of a media expert validation score of 95% and a final subject matter expert validation score of 82.5%. In the limited implementation trial, the students' average score increased from 80.2 in the pre-test to 91.6 in the post-test, with an average N-Gain score of 0.60, categorized as moderate. These findings indicate that the Wordwall-based test instrument has met the validity criteria and shows promising initial results as a digital assessment instrument for Al-Qur'an and Hadith learning. However, because the implementation involved only five students, the findings should be interpreted as preliminary and cannot yet be generalized to a wider population. The developed instrument integrates interactive features, images, audio, and HOTS-oriented questions to provide a more engaging learning-assessment experience.

I. INTRODUCTION

The purpose of education is to help individuals reach their full potential and acquire the social, emotional, spiritual, and intellectual skills and abilities required to function in society (Ichsan et al, 2021). The transfer of knowledge is no longer the only goal of education in the age of digital transformation; it also needs to keep up with the quick developments in science and technology. The paradigm of learning has changed from teacher-centered to student-centered due to advancements in information technology. Teachers must be able to use digital technology as an essential component of the learning process, including during the learning assessment phase, in order to implement this paradigm change (Riska et al, 2024). The need for an appropriate assessment instrument becomes particularly important in Al-Qur'an and Hadith learning because the subject does not merely require students to recall verses, hadiths, definitions, or concepts (Magdalena et al 2023). Students are also expected to understand the meaning of religious texts, relate Qur'anic and Hadith teachings to real-life situations, analyze the implications of Islamic values, evaluate behaviors based on religious principles, and formulate appropriate solutions to contextual problems. Therefore, assessment in Al-Qur'an and Hadith should be able to measure students' cognitive abilities beyond factual recall while maintaining the characteristics of the subject matter. An assessment instrument that relies

predominantly on routine recall questions may not adequately measure these competencies (Junaidah et al, 2024). In addition, Al-Qur'an and Hadith learning contains distinctive learning materials, including Qur'anic verses, hadith texts, meanings, interpretations, and contextual religious issues, which require appropriate stimuli and assessment formats. This condition creates a specific need for a digital assessment instrument that is not only interactive but also aligned with the content, learning objectives, and cognitive demands of Al-Qur'an and Hadith. In the curriculum for the Qur'an and Hadith, the use of technology in learning assessment is growing in significance. Students must absorb the principles of the Qur'an and Hadith into their everyday lives in addition to mastering cognitive skills like memorization and comprehension of verses and hadiths. As a result, the evaluation procedure must be created in a way that allows for a thorough measurement of these diverse competencies. Learning outcomes that do not accurately reflect students' true abilities may result from the use of boring and traditional evaluation tools, which may discourage students from taking part in the process (A'la et al, 2023). Innovation is therefore required in the creation of evaluation tools that both meet the objectives of 21st-century learning and are consistent with the features of the Qur'an and Hadith curriculum.

The research gap can be seen more systematically by comparing the focus of previous studies. Research on Wordwall has examined its use as an evaluation medium in Indonesian language learning, while other development studies have used Wordwall to develop game-based learning evaluation or interactive assessment products in other subjects and educational contexts (Nur et al 2023). These studies demonstrate the feasibility of Wordwall for digital assessment, but their subject matter, learning objectives, item characteristics, and product specifications differ from those required in Al-Qur'an and Hadith learning. The present study therefore extends the existing research by developing a Wordwall-based test instrument specifically for Al-Qur'an and Hadith, integrating religious content with multimedia stimuli and HOTS-oriented items, and subjecting the product to media and subject-matter expert validation. The novelty is consequently positioned not merely in the use of Wordwall, but in the systematic development and validation of a subject-specific digital assessment instrument. The theoretical foundation of the assessment instrument is also related to Higher Order Thinking Skills (HOTS). In the revised Bloom's Taxonomy, cognitive processes are classified into six levels: remembering (C1), understanding (C2), applying (C3), analyzing (C4), evaluating (C5), and creating (C6). The C4, C5, and C6 levels require students to engage in more complex cognitive processes than merely recalling or reproducing information (Larasati, et al, 2023). Analyzing involves breaking information into relevant components and identifying relationships among them; evaluating involves making judgments based on criteria and standards; while creating involves developing a new idea, solution, or product. In Al-Qur'an and Hadith learning, HOTS should be connected to the substantive characteristics of Islamic religious knowledge. At the C4 level, students can be asked to analyze the meaning, message, or relationship between a Qur'anic verse or hadith and a particular situation. At the C5 level, students can evaluate behaviors, statements, or solutions by applying relevant Qur'anic and Hadith principles as criteria. At the C6 level, students can formulate appropriate solutions, recommendations, or applications of Islamic teachings to contextual problems. Therefore, HOTS-oriented assessment in Al-Qur'an and Hadith is not simply characterized by difficult questions, but by questions that require students to use religious knowledge to analyze information, make reasoned judgments, and formulate appropriate solutions. Wordwall-based learning materials can improve student engagement, learning motivation, and learning results in a variety of courses, according to numerous prior research. Nonetheless, the majority of these research continue to concentrate on using Wordwall as an interactive test platform or learning aid. There is still a dearth of study that uses a research and development (R&D) approach to create Wordwall as a testing tool intended especially for assessing the learning of the Qur'an and Hadith.

Additionally, prior research typically only assessed gains in learning motivation or learning outcomes without going through a product creation process that was approved by subject matter or media experts. Therefore, there is still room for more research on the development of innovative digital evaluation tools that are appropriate for the features of learning the Qur'an and Hadith.

Based on the identified needs, research gap, and theoretical considerations, this study focuses on the development of a Wordwall-based test instrument for Al-Qur'an and Hadith learning at Madrasah Aliyah Al Yusra in Gorontalo City. Accordingly, the research questions are formulated as follows: (1) What are the needs of teachers and students for a Wordwall-based test instrument for Al-Qur'an and Hadith learning? (2) How is the Wordwall-based test instrument developed using the ADDIE model? (3) How valid is the developed Wordwall-based test instrument based on media expert and subject matter expert validation? and (4) What are the initial results of implementing the developed instrument based on students' learning outcomes and N-Gain scores in the limited trial? The creation of a Wordwall-based assessment tool that incorporates audio recitations of the Qur'an, pictures of Qur'anic verses, hadiths, and a variety of educational activities as a means of assessing learning is what makes this study novel. In addition to being a tool for evaluating learning outcomes, the proposed product is an interactive evaluation medium that can improve student involvement, motivation, and the efficacy of the evaluation process. Additionally, the product was created utilizing the ADDIE paradigm, which includes the phases of Analysis, Design, Development, Implementation, and Evaluation. Before being used with students, it was verified by media and subject matter experts. By offering a paradigm for creating digital evaluation tools that may be modified for religious education classes in madrasah settings, this method adds to the body of scientific research. In light of the aforementioned discussion, this study intends to create a Wordwall-based test instrument as a means of assessing Al-Qur'an and Hadith learning at Madrasah Aliyah Al Yusra in Gorontalo City, as well as to test the product's validity and examine the impact of its application on students' learning outcomes. It is anticipated that the findings of this study will both theoretically advance the development of digital-based learning assessment technologies in the field of Islamic Religious Education and practically assist educators in creating more creative learning assessments.

II. METHODS

In order to create a Wordwall-based exam instrument appropriate for use as an evaluation tool for Al-Qur'an and Hadith education at Madrasah Aliyah Al Yusra in Gorontalo City, this study uses the Research and Development (R&D) process. The research and development approach was selected because it generates educational products through a methodical process of design, development, validation, implementation, and assessment in addition to being focused on testing hypotheses. This project produced a digital exam instrument that uses interactive technology to enhance the quality of learning assessment. The methodical and adaptable steps of the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) model, which is frequently employed in studies on the creation of educational media and learning tools, led to its selection.

Analysis

The analysis was conducted to identify the needs of teachers and students regarding assessment tools for the study of the Qur'an and Hadith. The needs analysis was carried out through field observations, the distribution of questionnaires to teachers and students, and a review of documentation regarding the implementation of assessments already in use at Madrasah Aliyah Al Yusra in Gorontalo City (Isgandi et al, 2024). This stage aims to gather information regarding the actual conditions of the learning evaluation process, student characteristics, and the challenges teachers face in implementing digital-based evaluations. The results of the needs analysis serve as

the basis for determining the specifications of the product to be developed so that it meets user needs. The present study focused primarily on the development and expert validation of the Wordwall-based test instrument. The limited implementation involving five students was intended to obtain preliminary information regarding the use of the instrument and changes in students' learning outcomes. Therefore, the study did not conduct statistical item analysis involving item-level validity, reliability, item difficulty, or discrimination indices. Such analyses require a substantially larger number of respondents to produce stable estimates. Consequently, the 35 items developed in this study should be regarded as expert-validated assessment items at the development stage rather than as statistically calibrated test items. Further research involving a larger sample is required to examine item-level validity, reliability, difficulty, and discrimination.

Design

Design which involves developing a product design that includes determining core competencies, learning outcome indicators, creating a test instrument blueprint, developing test items, selecting a Wordwall template, and designing the layout of the assessment medium (Sulistiyorini et al, 2023). At this stage, the test instruments are developed based on the Qur'an and Hadith curriculum applicable in Madrasah Aliyah, taking into account the principles of content validity, alignment with learning objectives, item difficulty levels, and student characteristics. In addition, the media design also integrates audio recitations (murottal), images of Qur'anic verses and hadiths, and various interactive activities available on the Wordwall platform to make the assessment process more engaging and provide a more meaningful learning experience.

Development

Development involved transforming the product design into a Wordwall-based assessment tool that can be used in the learning evaluation process. The developed product was then validated by media experts and subject matter experts to determine its suitability before being implemented with students. Validation was conducted using an assessment instrument in the form of a questionnaire covering aspects such as media presentation, ease of use, content appropriateness, question quality, language use, and the product's alignment with learning objectives (Gunawan et al, 2023). Feedback and suggestions from the validators served as the basis for revising the product so that the resulting test instrument met the criteria for validity and was suitable for use during the implementation phase.

Implementation

Implementation involved piloting the use of Wordwall-based assessment tools with students at Madrasah Aliyah Al Yusra in Gorontalo City. At this stage, teachers used the developed product as an assessment tool for Quran and Hadith instruction. In addition to observing the assessment process, the researcher also collected data on teachers' and students' responses to the product's use through questionnaires. To determine the product's effectiveness on student learning outcomes, measurements were taken using a pre-test and a post-test following the use of the Wordwall-based assessment tool (Idzi',2022). The results of both tests were then analyzed to determine the improvement in student learning outcomes after using the developed product. The implementation stage involved only five tenth-grade students and was therefore designed as a limited trial. The purpose of this stage was to examine the initial usability of the developed instrument and obtain preliminary evidence from students' pre-test and post-test scores. The limited number of participants does not provide sufficient statistical power for generalization or for robust item-level analysis. Accordingly, the results of this trial were not interpreted as evidence of the general effectiveness of the instrument across a wider population.

Evaluation

Evaluation was conducted both formatively at each stage of development and summatively after the product's implementation was completed. The evaluation aimed to assess the generated product's degree of success based on its efficacy, practicality, and validity. The evaluation's findings serve as a foundation for improving the product so that the resulting Wordwall-based test instruments are of high quality and aligned with the needs of Al-Qur'an and Hadith instruction in Madrasah Aliyah (Filahanasari et al, 2023). The evaluation stages also ensure that each revision made has improved the product's quality based on recommendations from validators and the results of field implementation. Research Subjects and Participants. The participants in this study consisted of several groups involved at different stages of the ADDIE development process. The needs analysis involved one Al-Qur'an and Hadith teacher and tenth-grade students at Madrasah Aliyah Al Yusra in Gorontalo City. The product validation stage involved one media expert and one Al-Qur'an and Hadith subject matter expert. The implementation stage was conducted as a limited trial involving five tenth-grade students. The teacher participated in the needs analysis and provided feedback on the practicality and applicability of the developed assessment instrument, while the experts evaluated the validity of the product in terms of media and subject matter. The five students participated only in the limited implementation trial to obtain preliminary evidence regarding students' responses and learning outcomes. Expert Validator Characteristics. Product validation was conducted by two validators with expertise relevant to the developed instrument. The media expert was selected based on academic and/or professional expertise related to educational technology and digital learning media, particularly the development and evaluation of digital learning resources. The subject matter expert was selected based on expertise in Al-Qur'an and Hadith education and familiarity with the curriculum and assessment of the subject. The validators were selected because their academic background, professional experience, and field expertise were considered relevant to assessing the quality of the developed Wordwall-based test instrument. The validation process focused on the appropriateness of the digital media, content accuracy, item construction, relevance to learning objectives, and suitability of the assessment instrument for Al-Qur'an and Hadith learning.

Validation Scale and Criteria. The expert validation questionnaire used a four-point Likert scale to assess each aspect of the developed instrument. The response categories were 4 = very appropriate, 3 = appropriate, 2 = less appropriate, and 1 = inappropriate. The percentage of validity was calculated by comparing the total score obtained with the maximum possible score and multiplying the result by 100. The resulting percentage was then interpreted using predetermined validity criteria. The validity category was classified as very valid (81–100%), valid (61–80%), less valid (41–60%), and invalid (0–40%). The criteria were used to determine whether the developed Wordwall-based test instrument was sufficiently valid for implementation. N-Gain Analysis. The initial learning results obtained from the limited implementation trial were analyzed using the normalized gain (N-Gain) to describe the relative improvement between students' pre-test and post-test scores. First, the pre-test was administered before students used the developed Wordwall-based assessment instrument, and the post-test was administered after the implementation. Second, the mean pre-test and post-test scores were calculated. Third, the N-Gain score was calculated using the following formula:

$$g = \frac{S_{post} - S_{pre}}{S_{max} - S_{pre}}$$

Where S_{post} represents the post-test score, S_{pre} represents the pre-test score, and S_{max} represents the maximum possible score. The resulting N-Gain value was then interpreted using the predetermined categories: high ($g \geq 0.70$), medium ($0.30 \leq g < 0.70$), and low ($g < 0.30$). Because the implementation involved only five students, the N-Gain result was interpreted as preliminary

evidence of learning improvement rather than as evidence of a statistically established effect of the developed instrument.

III. RESULTS AND DISCUSSION

This research and development project resulted in a WordWall-based test instrument for evaluating students' performance in the Qur'an and Hadith course for pupils in the tenth grade at Gorontalo City's MA Al Yusra. The ADDIE paradigm, which consists of five stages analysis, design, development, implementation, and evaluation was used to methodically carry out the development process. The ADDIE model was selected because of its is one of the most comprehensive, structured, and adaptable instructional development models for digital learning needs.

3.1. Analysis

The analysis phase is the cornerstone of instructional media development. During this phase, the actual needs in the field are identified, from the perspectives of both teachers and students. The needs analysis phase was conducted at MA Al Yusra in Gorontalo City by distributing a questionnaire designed to identify the needs of teachers and students.

Table 3.1 Results of the Teacher Needs Analysis

NO	Aspect of Needs	Skor	Skor Maks	Percentage	Category
1	Interactive Evaluation Form	4	4	100%	Very much needed
2	Template Variations	3	4	75%	Very much needed
3	Ease of Manufacturing	4	4	100%	Very much needed
4	Accessibility Link	3	4	75%	Very much needed
5	Contextual Material of the Qur'an and Hadith	4	4	100%	Very much needed
Amount		18	20	90%	Very much needed

Based on Table 3.1, it can be seen that the total score from the teacher needs analysis is 18 out of a maximum score of 20. When expressed as a percentage, this amounts to 90%, which falls into the "Highly Needed" category. Specifically, there are three aspects that received the highest score of 100% in the "Highly Needed" category. These three aspects are the need for game-based or interactive forms of assessment to make students more active; the need for media that is easy to create without requiring coding skills or complex design; and the need for media that allows for the insertion of images and audio into assessment questions on the Qur'an and Hadith. Meanwhile, the other two aspects the need for a variety of templates and accessibility via links received a score of 75%, falling into the "Needed" category. Based on the results of this analysis, it can be concluded that teachers have a strong need for the development of interactive, user-friendly, and contextually relevant assessment tools for the Qur'an and Hadith curriculum. Therefore, the development of Wordwall-based assessment tools is considered appropriate to address this need, as Wordwall offers a variety of game templates, is easy to use without coding, can be shared via a link, and supports the insertion of images and audio.

Table 3.2 Results of the Analysis of Students' Needs

NO	Aspect of Needs	Total Score	Skor Maks	Percentage	Category
1	Interest in Learning	17	20	85%	Very much needed
2	Reducing Boredom	18	20	90%	Very much needed
3	Interactivity	18	20	90%	Very much needed
4	Technology Readiness	20	20	100%	Very much needed
5	Understanding the Material	20	20	100%	Very much needed

Amount	93	100	93%	Very much needed
--------	----	-----	-----	------------------

Based on Table 3.2, it can be seen that the results of the student needs analysis yielded a total score of 93 out of a maximum score of 100. Expressed as a percentage, this amounts to 93%, which falls into the “Highly Needed” category. Specifically, the aspects with the highest percentages were technological readiness and understanding of the material, both scoring 100%. This indicates that the school has provided computers for accessing online assessments and that students strongly need assessments that include visual and audio elements to help them better understand the material on the Qur'an and Hadith.

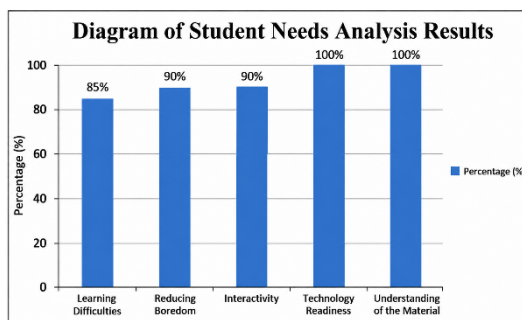


Figure 3.1. Percentage Diagram of Student Needs Analysis Results

Figure 3.1 shows that the highest bar in the diagram is in the aspects of technological readiness and understanding of the material, at 100%. This is supported by the survey results, which indicate that all students agreed that the school has provided the necessary devices and that they need media containing visual and audio elements. Meanwhile, the lowest bar is in the aspect of interest in learning, at 85%. Although this is the lowest percentage, the result still indicates a very high need among students for engaging assessment media. Thus, it can be concluded that students have a strong need for learning assessment tools that are engaging, interactive, technology-based, and incorporate visual and audio elements. The results of this analysis of student needs reinforce the urgency of developing Wordwall-based learning tools, as Wordwall is capable of presenting interactive games, incorporating images and audio, reducing boredom during learning, and being accessible via computers already available at the school.

3.2. Design

The design phase included the creation of a test item matrix based on HOTS (Higher Order Thinking Skills), the design of the Wordwall interface, and the creation of research tools, such as pre-test and post-test items, teacher and student response questionnaires, and an expert validation form. This phase involves designing a test instrument blueprint consisting of 35 items, all of which are designed to measure higher-order thinking skills (HOTS) at levels C4 (Analysis), C5 (Evaluation), and C6 (Creation). This design was carried out by shifting the focus of Qur'an and Hadith material from the domains of memorization and comprehension to the domains of application, analysis, evaluation, and creation through the context of real-life problems in daily life. Thorough planning at this stage is key to the quality of the product to be developed. Based on the results of expert validation and revisions, 35 final test items were selected from the initial draft of 60 items. All 35 items are designed to measure higher-order thinking skills (HOTS) at levels C4–C6. The mapping of each item to indicators and cognitive levels is presented in the following table :

Table 3.3 Mapping of 35 Test Items to HOTS Cognitive Levels

No	Chapter/Material	HOTS Indicator	Cognitive Level	No. Butter	Amount
1	Chapter 1: The Qur'an is My Book	Analyzing the definition of the Qur'an according to scholars	C4	1	1

2	Chapter 1: The Qur'an is My Book	Analyzing the relationship between the name and meaning of the Qur'an	C4	2	1
3	Chapter 1: The Qur'an is My Book	Comparing scholars' opinions on the origins of the Qur'an	C4	3	1
4	Chapter 1: The Qur'an is My Book	Evaluating behavior steadfast in adhering to the Qur'an	C5	4	1
5	Chapter 1: The Qur'an is My Book	Designing solutions based on the values of the Qur'an	C6	5	1
6	Chapter 2: How Authentic My Book Is	Analyzing the conditions of miracles	C4	6	1
7	Chapter 2: How Authentic My Book Is	Classifying physical and spiritual miracles	C4	7	1
8	Chapter 2: How Authentic My Book Is	Evaluating the miraculous nature of the Qur'an from a linguistic perspective	C5	8	1
9	Chapter 2: How Authentic My Book Is	Evaluating the authenticity of the Quran	C5	9	1
10	Chapter 2: How Authentic My Book Is	Constructing an argument for the authenticity of the Quran	C6	10	1
11	Chapter 3: The Purpose and Function of the Qur'an	Analyzing the position of the Qur'an as a source of law	C4	11	1
12	Chapter 3: The Purpose and Function of the Qur'an	Analyzing the function of the Qur'an	C4	12	1
13	Chapter 3: The Purpose and Function of the Qur'an	Evaluating the application of the functions of the Qur'an	C5	13	1
14	Chapter 3: The Purpose and Function of the Qur'an	Designing behavior according to the function of the Qur'an	C6	14	1
15	Chapter 3: The Purpose and Function of the Qur'an	Evaluating the function of the Qur'an in life	C5	15	1
16	Chapter 4: Main Points of My Book	Analyzing the main themes of the Quran	C4	16	1
17	Chapter 4: Main Points of My Book	Analyzing the content of verses on faith and worship	C4	17	1
18	Chapter 4: Main Points of My Book	Evaluating the relevance of science and technology in the Qur'an	C5	18	1
19	Chapter 4: Main Points of My Book	Designing behavior in accordance with the core teachings of the Quran	C6	19	1
20	Chapter 4: Main Points of My Book	Analyzing the interconnections of the main themes in the Quran	C4	20	1
21	Chapter 1: The Qur'an is My Book	Analyzing the definition of the Qur'an (Essay)	C4	21	1
22	Chapter 1: The Qur'an is My Book	Evaluating the names of the Qur'an	C5	22	1

23	Chapter 2: How Authentic My Book Is	Analyzing the aspect of miracles	C4	23	1
24	Chapter 2: How Authentic My Book Is	Analyzing the balance of words and letters in the Qur'an	C4	24	1
25	Chapter 2: How Authentic My Book Is	Mengevaluasi perbedaan mu'jizat para nabi	C5	25	1
26	Chapter 3: The Purpose and Function of the Qur'an	Analyzing the purpose of the revelation of the Qur'an	C4	26	1
27	Chapter 3: The Purpose and Function of the Qur'an	Evaluating the function of the Qur'an as a healing remedy	C5	27	1
28	Chapter 4: Main Points of My Book	Analyzing the main themes of the Quran	C5	28	1
29	Chapter 4: Main Points of My Book	Designing the implementation of the core content of the Quran	C6	29	1
30	Chapter 4: Main Points of My Book	Analyzing the relationship between faith and worship	C4	30	1
31	Chapter 1: The Qur'an is My Book	Preparing a comprehensive description of the Qur'an	C6	31	1
32	Chapter 2: How Authentic My Book Is	Constructing a comprehensive argument for the authenticity of the Quran	C6	32	1
33	Chapter 3: The Purpose and Function of the Qur'an	Designing the implementation of the functions of the Qur'an	C6	33	1
34	Chapter 3: The Purpose and Function of the Qur'an	Evaluating the impact of the application of the Qur'an	C5	34	1
35	Chapter 4: Main Points of My Book	Designing a thematic study of the core content of the Qur'an	C6	35	1

Table 3.4 Summary of HOTS Cognitive Levels

Cognitive Level	Item Number	Amount
C4 (Analyzing)	1, 2, 3, 6, 7, 11, 12, 16, 17, 20, 21, 23, 24, 26, 30	15
C5 (Evaluating)	4, 8, 9, 13, 15, 18, 22, 25, 27, 28, 34	11
C6 (Creating)	5, 10, 14, 19, 29, 31, 32, 33, 35	9
Total	1–35	35

Although Qur'an and Hadith material is often associated with memorization, in this study all test items were designed to assess Higher-Order Thinking Skills (HOTS) using the following strategies: 1) The questions did not directly test textual memorization but instead assessed the analysis of the meaning, wisdom, and implications of verses and hadiths within the context of specific cases; 2) Question stimuli consist of images, audio recitations, and case studies to encourage students to think critically; 3) Answers involve evaluating attitudes and devising solutions, rather than selecting memorized options. Thus, aspects C1–C3 are used only as implicit prerequisites, not as measurement objectives.

3.3 Development

Table 3.3 Results of the Media Expert Validation

NO	Aspect	Skor	Skor Maks	Percentage	Category
1	Function of Media	7	8	87.50%	Very Valid
2	Technical Quality	8	8	100%	Very Valid
3	Design and Appearance	8	8	100%	Very Valid
4	Interactivity	8	8	100%	Very Valid
5	Content Relevance	7	8	87.50%	Very Valid
Amount		38	40	95%	Very Valid

Based on Table 3.3, it can be seen that the media expert validation results achieved a total score of 38 out of a maximum score of 40. When expressed as a percentage, this corresponds to 95%, which falls into the “Highly Valid” category. In detail, out of the 10 criteria evaluated, 8 criteria received a score of 4 and 2 criteria received a score of 3. The criteria that received a score of 4 include the media’s alignment with learning objectives, text readability and visual presentation, feedback provision, visual appeal, ease of navigation, interactive activities, difficulty level adjustment, and alignment of content with the curriculum. Meanwhile, the two criteria that received a score of 3 were the media’s support for effective learning processes and the media’s support for concept reinforcement. In response to the feedback provided by media experts, the researchers took two actions. First, during the midterm exam at MA Al Yusra, the Wordwall-based test instrument was used with the answer discussion feature disabled and real-time scoring turned off while the exam was in progress, thereby ensuring its function as an assessment tool remained intact. Second, in the product user guide, the researchers recommended that this instrument serve a dual purpose: it can be used as a formative practice tool before the midterm exam and as a summative assessment tool for the midterm exam with specific settings. Thus, the results of this media expert validation serve as the basis for enhancing the product, ensuring that it is not only technically valid but also flexible enough to meet teachers’ needs.

Table 3.4 Results of Subject Matter Expert Validation for Cycle I

NO	Aspek	Skor	Skor Maks	Persentase	Kategori
1	Alignment with Learning Outcomes	2	4	50%	Less Valid
2	Update of Materials	2	4	50%	Less Valid
3	Concept Accuracy	2	4	50%	Less Valid
4	Clarity of the Problem Statement	3	4	75%	Valid
5	Language Use	3	4	75%	Valid
6	Level of Difficulty of the Question	2	4	50%	Less Valid
7	Compliance with Indicators	2	4	50%	Less Valid
8	Clarity of Instructions	3	4	75%	Valid
9	Quality of Choice Answer	2	4	50%	Less Valid
10	Ability to Measure Understanding	2	4	50%	Less Valid
Amount		23	40	57.5%	Less Valid

Based on Table 4.4, the results of the subject matter expert validation for Cycle I yielded a total score of 23 out of a maximum score of 40, or 57.5%, falling into the “Insufficiently Valid” category. Of the 10 aspects evaluated, 7 received a score of 2 and 3 received a score of 3.

Table 3.5 Results of Subject Matter Expert Validation, Cycle II

NO	Aspect	Skor	Skor Maks	Persentase	Kategori
1	Alignment with Learning Outcomes	3	4	75%	Valid
2	Update of Materials	3	4	75%	Valid
3	Concept Accuracy	3	4	75%	Valid
4	Clarity of the Problem Statement	4	4	100%	Very Valid
5	Language Use	4	4	100%	Very Valid
6	Level of Difficulty of the Question	3	4	75%	Valid
7	Compliance with Indicators	3	4	75%	Valid
8	Clarity of Instructions	4	4	100%	Very Valid
9	Quality of Choice Answer	3	4	75%	Valid
10	Ability to Measure Understanding	3	4	75%	Valid
Amount		33	40	82.5%	Very Valid

Based on the validation results, Cycle I yielded a score of 57.5%, meeting the “Sufficiently Valid” criterion. After revisions were made in accordance with expert recommendations, the score in Cycle II increased to 82.5%, meeting the “Highly Valid” criterion. Thus, the score used as the basis for assessing the validity of the final product is the Cycle II score of 82.5%. This 25% increase indicates that the revisions made were in line with the subject matter experts’ input. The Cycle I score was used only as a basis for improvement, not as a determinant of the final product’s validity.

Table 3.6 Summary of Expert Validation

Validador	Cycle I	Criteria	Cycle II/Final	Criteria	Improvement
Material Expert	57,5%	Experto en Medios	82,5%	Very Valid	25%
Media Expert	95%	Very Valid	95%	Very Valid	
Final Average			88,75%	Very Valid	

Based on the table above, the average final validation score was 88.75%, meeting the “Highly Valid” criterion. Thus, it can be concluded that the Wordwall-based test instrument for the Qur’an and Hadith subject is valid for pilot testing without further revisions. The final validation score of 88.75% indicates that the developed Wordwall-based test instrument met the predetermined validity criteria. However, this finding should be interpreted as evidence of expert-based appropriateness rather than as evidence of complete psychometric quality. The high validation score suggests that the instrument was considered appropriate in terms of content alignment, item construction, visual presentation, and technical functionality. More importantly, the result indicates that the development process was able to integrate the characteristics of Al-Qur’an and Hadith learning into an interactive digital assessment format. Nevertheless, expert validation alone cannot demonstrate how individual items function statistically when administered to students. Therefore, further empirical testing with a larger sample is required to examine item-level validity, reliability, difficulty, and discrimination. The findings can also be interpreted from the perspective of HOTS-oriented assessment. The developed instrument consists of items designed at the C4–C6 cognitive levels, which require students to analyze information, evaluate situations, and formulate appropriate responses. This is consistent with the revised Bloom’s Taxonomy, in which analyzing, evaluating, and creating represent increasingly complex cognitive processes. In Al-Qur’an and Hadith learning, these cognitive processes can be applied through questions that require students to analyze the meaning and implications of verses or hadiths, evaluate behaviors based on Islamic principles, and formulate

solutions to contextual problems. Therefore, the 88.75% validation result suggests that the developed instrument was considered appropriate for integrating subject-specific content with higher-order cognitive demands. However, the presence of C4–C6 indicators alone does not automatically demonstrate that every item successfully measures HOTS; empirical item analysis is still required to verify the actual cognitive functioning of the items.

3.4 Implementation

The product was implemented on a limited basis with five 10th-grade students at MA Al Yusra in Gorontalo City over four sessions, each lasting 2×60 minutes. This pilot study took the form of one-on-one and small-group trials aimed at assessing the students' usability and initial responses to the medium, rather than conducting a quantitative analysis of item quality. The implementation stages included administering a pre-test, instruction using the Wordwall platform as the treatment, and concluding with a post-test. Given the limited number of participants, the results of this pilot study cannot yet be generalized and should not be used to test item validity, reliability, or discriminative power. During the implementation process, students appeared enthusiastic and actively engaged in completing the assessment through the Wordwall platform. Additionally, the enjoyable and stress-free assessment environment also contributed to the students' optimal performance.

3.5 Evaluation

The evaluation phase was conducted to measure the effectiveness of the WordWall-based test instrument that had been developed. The evaluation was conducted formatively through teacher and student response questionnaires, as well as summatively through an analysis of the N-Gain Score from the pre-test and post-test results. Formula:

$$N\text{ Gain} = \frac{\text{Skor Posttest} - \text{Skor Pretest}}{\text{Skor Ideal} - \text{Skor Pretest}}$$

Table 3.7 Results of the Effectiveness Test Using the N-Gain Score

NO	Student Name	Pre-test	Post-test	N-Gain	Criteria
1	Moh. Iqbal Nento	78	89	0,50	Currently
2	Fatgan Dunggio	80	91	0,55	Currently
3	Sitti Fatimah Yunus	85	97	0,80	Currently
4	Furqon Syiu	76	87	0,46	Currently
5	Agung P. Nurdin	82	94	0,67	Currently
Average		80,2	91,6	0,60	Currently

The limited implementation trial showed an increase in the students' mean score from 80.2 in the pre-test to 91.6 in the post-test, resulting in an average N-Gain score of 0.60, which falls into the moderate category. This finding indicates an improvement in students' scores following the limited implementation of the developed Wordwall-based assessment instrument. However, the finding should be interpreted cautiously. Because the trial involved only five students and did not include a control or comparison group, the observed increase cannot be interpreted as evidence that Wordwall caused the improvement in learning outcomes. Rather, the result provides preliminary evidence that the developed instrument can be implemented in an Al-Qur'an and Hadith learning-assessment context and was accompanied by improved scores in this limited trial. Further research involving a larger sample and an appropriate comparative design is required to determine the effectiveness of the instrument more rigorously.

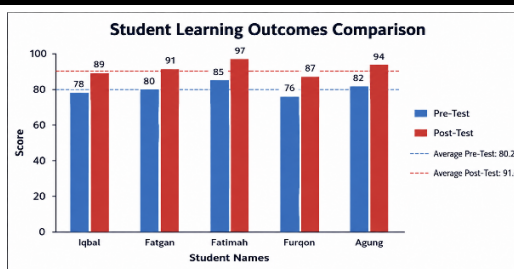


Figure 3.3 Comparison Diagram

Based on Table 3.7 and Figure 3.3, it can be seen that there was an increase in students' learning achievement scores after using Wordwall-based learning media. The average pre-test score of 80.2 increased to 91.6 on the post-test. Based on the N-Gain calculation, the average was 0.60, which falls into the "Moderate" category. In detail, out of 5 students, 4 met the "Moderate" N-Gain criterion and 1 met the "High" criterion. The greatest improvement was observed in a student named Sitti Fatimah Yunus, who achieved an N-Gain of 0.80, falling into the "High" category. The findings of the present study can be further understood by comparing them with previous Wordwall-based studies. Previous research has generally examined Wordwall as an interactive learning or evaluation medium in different subject areas and educational contexts. In comparison, the present study focuses specifically on developing a Wordwall-based test instrument for Al-Qur'an and Hadith learning and integrates HOTS-oriented items with multimedia stimuli. Methodologically, the present study uses a Research and Development approach based on the ADDIE model, followed by expert validation and a limited implementation trial involving five students. These characteristics distinguish the present study from previous studies that may have used different research designs, larger or different participant groups, or focused primarily on the use of Wordwall as a learning medium rather than on the development of a subject-specific assessment instrument. Therefore, the 88.75% validation score and the N-Gain of 0.60 obtained in this study should be interpreted within the scope of its development-oriented design and limited sample. The comparison indicates that the contribution of the present study lies primarily in the development and expert validation of a subject-specific digital assessment instrument rather than in establishing a generalized causal effect of Wordwall on learning outcomes.

Discuccion

The results of this study indicate that the Wordwall-based test instrument developed using the ADDIE model is valid and effective for evaluating learning in the Qur'an and Hadith subject. The findings demonstrate that integrating interactive digital technology into learning assessment can enhance both the quality of the evaluation process and students' engagement during assessment activities. The systematic development process, beginning with needs analysis and continuing through design, expert validation, implementation, and evaluation, ensured that the resulting product met educational and technological standards. The needs analysis showed that both teachers (90%) and students (93%) expressed a very high need for interactive digital assessment media. These findings suggest that conventional paper-based assessments are no longer sufficient to meet the demands of 21st-century learning. Teachers require assessment tools that are easy to develop, practical to implement, and capable of integrating multimedia elements such as images and audio. Meanwhile, students expect assessments that are more engaging, reduce boredom, and provide meaningful learning experiences. These findings support the concept of technology-enhanced assessment, which emphasizes that digital assessment should not only measure learning outcomes but also improve students' motivation and participation during the evaluation process. The design phase successfully produced a HOTS-oriented test instrument consisting of 35 final items covering

cognitive levels C4 (analysis), C5 (evaluation), and C6 (creation). This result demonstrates that assessment in Qur'an and Hadith learning can move beyond memorization toward higher-order thinking skills. Rather than asking students merely to recall verses or hadiths, the developed instrument encourages them to analyze meanings, evaluate real-life situations based on Islamic values, and propose appropriate solutions. This finding aligns with contemporary educational perspectives that emphasize critical thinking, problem-solving, and creativity as essential competencies for students in the digital era. The validation results further strengthen the quality of the developed product. Media expert validation reached 95%, while subject matter expert validation improved from 57.5% in Cycle I to 82.5% after revisions, resulting in an average validation score of 88.75%. This improvement demonstrates the importance of formative evaluation during product development. Feedback provided by validators contributed significantly to refining the content, language, item construction, and technical features of the assessment instrument. Consequently, the final product achieved the "Highly Valid" category and was considered suitable for implementation in classroom learning.

The implementation results also revealed positive responses from students during the learning assessment process. Students were more enthusiastic and actively participated in completing the assessment through the Wordwall platform. The game-based environment reduced students' anxiety and created a more enjoyable assessment experience. This finding indicates that assessment can simultaneously function as a learning activity rather than merely as a measurement tool. Interactive features, multimedia integration, and immediate digital responses appear to increase students' willingness to engage actively throughout the evaluation process. The effectiveness test also supports the usefulness of the developed instrument. The average pre-test score increased from 80.2 to 91.6 after implementation, with an average N-Gain score of 0.60, categorized as moderate. These findings indicate that the Wordwall-based assessment instrument contributed positively to improving students' learning outcomes. Although the trial involved only five students and therefore cannot yet be generalized to a larger population, the observed improvement suggests that interactive digital assessment has considerable potential to enhance students' conceptual understanding of Qur'an and Hadith materials. The findings of this study are consistent with previous studies reporting that Wordwall improves student motivation, engagement, and learning achievement. However, this research extends previous work by focusing not on Wordwall as instructional media, but as a comprehensive digital assessment instrument developed through a systematic Research and Development process using the ADDIE model. Furthermore, unlike many previous studies, this research integrates expert validation, multimedia elements, and HOTS-oriented assessment into one product specifically designed for Qur'an and Hadith learning. Therefore, this study contributes to expanding the application of Wordwall beyond classroom activities into the domain of learning evaluation. Nevertheless, several limitations should be acknowledged. The implementation stage involved only a limited number of participants, so broader field testing is still required to evaluate the practicality and effectiveness of the instrument across different educational contexts. Future studies are recommended to involve larger samples, conduct reliability and item analysis, and compare Wordwall-based assessment with other digital assessment platforms to obtain more comprehensive evidence regarding its effectiveness. Overall, this study demonstrates that the development of a Wordwall-based test instrument provides an innovative solution for digital assessment in Islamic education. The product not only fulfills validity requirements but also offers practical benefits for teachers and creates a more engaging, meaningful, and student-centered assessment experience. These findings contribute to the growing body of knowledge regarding technology-integrated assessment and support the implementation of digital transformation in learning evaluation within Madrasah Aliyah.

IV. CONCLUSION

This study developed a Wordwall-based test instrument for Al-Qur'an and Hadith learning at Madrasah Aliyah Al Yusra in Gorontalo City using the ADDIE development model. The final expert validation results reached 88.75%, indicating that the developed instrument was highly valid in terms of its media and subject-matter aspects. The instrument consisted of 35 HOTS-oriented items designed at the C4–C6 cognitive levels and incorporated interactive and multimedia features to support Al-Qur'an and Hadith learning assessment. The limited implementation trial involving five students showed an increase in the mean score from 80.2 in the pre-test to 91.6 in the post-test, with an average N-Gain score of 0.60, categorized as moderate. These findings indicate that the developed instrument demonstrated promising initial potential in supporting the assessment of Al-Qur'an and Hadith learning. However, the result should be interpreted cautiously because the limited trial involved only five students and did not include a control or comparison group. Therefore, the findings cannot yet be interpreted as evidence of the generalized effectiveness or causal impact of the Wordwall-based instrument. Further research is recommended to involve a larger and more representative sample and to conduct empirical item analysis, including item-level validity, reliability, difficulty, and discrimination analysis. Such studies are needed to establish the broader effectiveness and psychometric quality of the developed instrument.

V. REFERENCES

- A'la, N. I., Bilqist, P. S. I., Anisa, T. N., & Aini, V. Z. (2023). Pemanfaatan teknologi dalam evaluasi pembelajaran tingkat sekolah dasar. *Al-DYAS*, 2(3), 604–611. <https://doi.org/10.58578/alldyas.v2i3.1480>
- Filahanasari, E., Wahyuni, R. J., & Efendi, R. (2023). Pengembangan game interaktif menggunakan platform Wordwall.net sebagai alat evaluasi hasil belajar materi pecahan kelas IV. *JP2M (Jurnal Pendidikan dan Pembelajaran Matematika)*, 9(2), 200–210. <https://doi.org/10.29100/jp2m.v9i2.4160>
- Gunawan, S., Susandi, S., & Susilawati, L. (2023). Pengembangan media evaluasi pembelajaran bahasa Indonesia dengan memanfaatkan aplikasi daring Wordwall pada siswa kelas VIII A MTs Annur Al Huda Tajinan Malang. *Alfabeta: Jurnal Bahasa, Sastra, dan Pembelajarannya*, 6(2), 67–73. <https://doi.org/10.33503/alfabeta.v6i2.3688>
- Ichsan, F. N., & Hadiyanto. (2021). Implementasi perencanaan pendidikan dalam meningkatkan karakter bangsa melalui penguatan pelaksanaan kurikulum. *Al-Riwayah: Jurnal Kependidikan*, 13. Artikel jurnal
- Isgandi, Y., & Darmawan, A. A. (2024). Analisis kebutuhan bahan ajar ulumul Qur'an. *Al-I'jaz: Jurnal Studi Al-Qur'an, Falsafah dan Keislaman*, 6(2), 171–196. <https://doi.org/10.53563/ai.v6i2.222>
- Junaidah, Ambiyar, & Rizal, F. (2024). Mobile-based digital assessment transforming the learning support efficiency. *International Journal of Interactive Mobile Technologies*, 18(23), 44–56. <https://doi.org/10.3991/ijim.v18i23.51339>
- Larasati, P., Putrayasa, I. B., & Martha, I. N. (2023). Pemanfaatan media Wordwall.net sebagai media evaluasi dalam pembelajaran bahasa Indonesia. *Nusantara: Jurnal Pendidikan Indonesia*, 3(3), 395–412. <https://doi.org/10.14421/njpi.2023.v3i3-3>
- Layyinnati, I. (2022). Pengembangan media pembelajaran berbasis website (Wordwall) mata pelajaran fiqh kelas VII di Madrasah Tsanawiyah Muhammadiyah 12 Palirangan untuk meningkatkan hasil belajar siswa. *Jurnal Pendidikan Islam*, 8(1), 1–22. <https://doi.org/10.37286/ojs.v8i1.128>
- Magdalena, I., Nurchayati, A., & Apriliyani, D. (2023). Pentingnya peran evaluasi dalam pembelajaran di sekolah dasar. *MASALIQ*, 3(5), 833–839. <https://doi.org/10.58578/masaliq.v3i5.1381>
- Ramadhani, N. H., Prayudha, R., Bone, N. R., Zahroddar, & Hasibuan, S. (2023). Peran desain evaluasi pembelajaran dalam meningkatkan kualitas pembelajaran. *Jurnal Sadewa: Publikasi Ilmu Pendidikan, Pembelajaran dan Ilmu Sosial*, 2(1), 223–233. <https://doi.org/10.61132/sadewa.v2i1.501>
- Sulistiyorini, Y., Napfiah, S., & Mufidah, K. (2023). Pengembangan evaluasi pembelajaran berbasis game menggunakan platform Wordwall. *Prismatika: Jurnal Pendidikan dan Riset Matematika*, 5(2), 162–175. <https://doi.org/10.33503/prismatika.v5i2.2712>
- Tanjung, R. R., Ritonga, A. A., Abdullah, B. M., Siregar, N. A., & Armilah. (2024). Transformasi digital dalam pendidikan: Meningkatkan kualitas pembelajaran melalui teknologi. *Sinar Dunia: Jurnal Riset Sosial Humaniora dan Ilmu Pendidikan*, 3(2), 211–217. <https://doi.org/10.58192/sidu.v3i2.2195>