

Development of Multiple-Choice Based on Automatic Assessment through Quizizz Respiratory System Material in Humans

Lavina Azza Khaerina, Aldina Eka Andriani*

Elementary School Teacher Education Study Program, FIPP Semarang State University, Semarang, Indonesia

*e-mail: aldinaekaandriani@mail.unnes.ac.id

Accepted: May 11, 2026. Accepted: June 18, 2026. Published: August 10, 2026

Abstract: Teachers have not been effective in presenting learning evaluations, using a variety of stimuli and question types, resulting in delays in delivering feedback to students. This study aims to describe the development design, test the feasibility, and test the effectiveness of multiple-choice questions based on automatic assessment via Quizizz, using respiratory system material in class V SD N Sekaran 01, Semarang City. The research conducted is R&D (Research and Development) and the ADDIE (analysis, design, development, implementation, Evaluation) development model. The data collection technique uses both tests and non-tests. Data analysis techniques included normality testing, t-tests, and N-Gain. The results of this study include the design of a multiple-choice development based on automatic assessment through Quizizz, with the following question types: standard multiple choice, complex multiple choice, and two-tier multiple choice. Each question is accompanied by a varied, context-specific stimulus in the form of videos, images, tables, and narrative text. The feasibility of multiple-choice automatic assessment via a questionnaire on respiratory system material in humans was demonstrated, with material experts achieving 92.85% and media experts achieving 89.28%, both within very feasible criteria. The effectiveness of multiple-choice questions based on automatic assessment through Quizizz showed an increase in pre-test and post-test by 28.08. The Paired Sample t-test yielded a p-value of 0.00, indicating a significant difference in student learning outcomes, which is supported by the N-Gain test, which yielded 0.5687 with moderate criteria. Thus, a multiple-choice assessment based on Automatic Assessment through Quizizz has been successfully developed, is feasible, and effective in improving students' learning outcomes in grade V at SD N Sekaran 01, Semarang City.

Keywords: Automatic Assessment; Human Respiratory System; Learning Outcomes; Multiple-Choice; Quizizz.

Introduction

Learning is a methodical educational process that involves teacher-student interaction to accomplish the best possible learning objectives through student-centered planning, execution, and assessment [1]. Therefore, aware, meaningful, and joyful learning is an essential principle throughout the learning process. Accordingly, teachers play a strategic role in designing learning experiences that foster students' critical thinking skills sustainably [2]. To create a learning environment that supports students' critical thinking skills, teachers need to stimulate it by fostering an atmosphere where questions, problem-solving, and discussion are actively engaged in, thereby influencing the development of critical thinking [3].

Science learning at the elementary education level emphasizes the development of conceptual understanding and process skills as the main foundation for equipping students with literacy. Understanding concepts helps elementary school students internalize basic scientific and social ideas in a meaningful way and develop process skills, such as observing, conducting simple experiments, and analyzing data to draw conclusions from their learning experiences [4]. By developing understanding of concepts and process skills, science learning in elementary schools encourages students to actively observe, try, and analyze phenomena in simple ways. This process stimulates students'

ability to ask questions, reason cause-and-effect relationships, and draw logical conclusions [5].

The process of methodically gathering, evaluating, and interpreting data to determine the extent to which learning objectives are met is known as learning evaluation [6]. Learning evaluation follows the principle of continuity, which must be sustained to support student development. Comprehensive evaluation is not limited to a single aspect; it encompasses cognitive, affective, and psychomotor domains. A fair and objective evaluation must be based on reality (data and facts) that are not the result of manipulation or engineering [7]. Cooperative educators should work with all parties to optimize students' learning. In practice, the evaluation tool should be easy for teachers and students to use and should pay attention to the language and instructions for working through the questions [8].

To optimize the evaluation of learning-supporting components, namely the types of questions and stimuli [9]. The variety of question types allows the gradual measurement of students' abilities, from basic understanding to analysis and reasoning, so that students do not rely solely on memorization but are also required to understand concepts in their entirety. In science learning, students are required to understand concepts in context through the observation of natural and social phenomena. The presence of these stimulus variations can provide a more concrete and engaging context, helping students understand problems

How to quote:

L. A. Khaerina and A. E. Andriani, "Development of Multiple-Choice Based on Automatic Assessment through Quizizz Respiratory System Material in Humans", *J. Pijar.MIPA*, vol. 21, no. 4, pp. 759–767, Aug. 2026. <https://doi.org/10.29303/jpm.v21i4.12083>

more deeply and encouraging critical and analytical thinking [10].

Based on the results of the pre-research through observation, interviews, questionnaires, and document data at SDN Sekaran 01 Semarang City, several problems were found in the process of implementing social studies learning in grade V of SDN Sekaran 01 Semarang City, namely, teachers have not optimally utilized technology in the preparation of learning evaluation instruments. This is shown by teachers who still use conventional, paper-based, manual learning evaluations, generally manifested in the form of written tests such as multiple-choice, short-answer, and description questions, administered on printed question sheets, and in an assessment process that remains manual [11]. Conventional learning evaluation is considered less practical because it remains paper-based and relies on manual examination and assessment, which take longer and increase teachers' workload. In addition, conventional evaluations tend to provide delayed and limited feedback, thus being less supportive of ongoing monitoring of student learning progress [12]. Thus, the use of digitally based learning evaluation, which is an automated assessment, can be a more effective and efficient alternative because the system is fast and provides real-time feedback, thereby demonstrating the value of student learning outcomes [11].

The evaluation also did not present a variety of stimuli, as evidenced by the use of only text, while the stimulus for HOTS questions must vary, including images, text, tables, videos, and less complex question types to train students' literacy skills [13]. The lack of stimulus in learning evaluations makes the questions less contextual and unable to measure deep thinking. As a result, students tend to rely on memorization, interest in learning decreases, and competency achievement is not optimal [14]. Thus, a combination of stimulus variations and variations of question types of multiple choice questions, namely ordinary multiple choice, complex multiple choice, and two-tier, combined with the use of stimulus in the form of text narratives, tables, images, and short videos, makes a deeper contribution to improving the quality of learning [15]. The variety of these types of questions allows the measurement of students' abilities gradually, from basic understanding to analysis and reasoning skills, so that students not only rely on memorization, but are also required to understand concepts in their entirety [10].

This is shown by the relatively low student learning outcomes in IPAS subjects, as evidenced by the results of the formative assessment: 64% (18 out of 28 students) have not met the KKTP set by the school, namely 80. while the remaining 36%, namely 10 out of 28 students, have exceeded the set KKTP. Low learning outcomes are due to the lack of variety in stimuli and question types in learning evaluations, which causes students to be uninterested in working through the questions and to rely solely on memorization [16].

The low learning results in the subject matter of the human respiratory system are caused by several factors, one of which is the use of scientific terms that are relatively foreign and rarely used in the daily lives of elementary school students. This condition makes it difficult for students to fully understand concepts [17]. In addition, students often have difficulty distinguishing among parts of the respiratory organs that share similar terms and structures, such as the larynx, pharynx, bronchi, and bronchioles. The similarity in

the names and functions of these organs often leads students to go back and forth when identifying the sequence and roles of each organ in the breathing process.

This is consistent with study findings. related to learning difficulties in the human respiratory organ, IPAS material, and the factors that cause students to have difficulty mastering the material include the physical and psychological condition of the students, the cognitive abilities of the students, the teacher's teaching methods and strategies, as well as the role and support of the school environment and the students' families [18]. As a result, the internal and external factors mentioned affect the low learning outcomes in the human respiratory system material.

Another supporting factor, deemed highly feasible for use as an interactive learning evaluation, as evidenced by validation test results from material experts (91.25%) and media experts (87.5%) [19]. However, the study still uses 2 types of multiple-choice questions: ordinary and complex. This research attempts to close this gap by creating multiple-choice questions for automatic assessment, including 3 types: ordinary, complex, and two-tier.

Another supporting research is the research conducted by [20]. The results of the development research indicate that the Quizizz-based science learning evaluation product meets the validity criteria, with validity ratings of 89% from material experts, 87% from media experts, and 91% from linguists. However, the research focuses only on social studies, has not specifically examined the intended material, and has not examined the human respiratory system in greater depth. The research examines students' motivation to learn. This research aims to fill this gap by developing multiple-choice items based on automated assessment to improve the learning outcomes of grade V students on the human respiratory system.

Based on the descriptions of the problems, it is necessary to develop technology-based learning evaluations. Therefore, the researcher developed Multiple Choice Based on Automatic Assessment Through Quizizz to Improve the Learning Outcomes of Social Science Students in Class V of SDN Sekaran 01 Semarang City with the Objectives: (1) to decipher the development design of Multiple Choice Based on Automatic Assessment Through Quizizz to Improve the Learning Outcomes of Social Science Students in Class V of SDN Sekaran 01 Semarang City; (2) testing the feasibility of Multiple Choice Based on Automatic Assessment Through Quizizz to Improve the Learning Outcomes of Social Science Students in Grade V of SDN Sekaran 01 Semarang City; and (3) testing the effectiveness of Multiple Choice Based on Automatic Assessment Through Quizizz to Improve Social Science Learning Outcomes of Grade V Students of SDN Sekaran 01 Semarang City. This research is expected to assist teachers in conducting assessments more effectively and efficiently, and to provide direct feedback that supports the improvement of the learning process, thereby maximizing student learning outcomes.

Research Methods

This research is R&D (Research and Development). This research uses a mixed-methods approach (qualitative and quantitative). The development model used in the ADDIE model. This model consists of several steps, namely

analysis, design, development, implementation, and evaluation [21].

The subjects of the research were students of V SD N Sekaran 01, totalling 28 students: 14 male and 14 female. This activity involves 6 students in class VA at SDN Sekaran 01, who were selected based on variations in cognitive ability: 2 with excellent, 2 with intermediate, and 2 with low cognitive ability. through purposive sampling techniques. Furthermore, a large-group study was conducted with 22 students.

The types of data are qualitative and quantitative. Quantitative data were collected through media expert assessment sheets, assessment questionnaires for material experts, classroom teacher response questionnaires, student response questionnaires, and student learning test results. Qualitative data can take the form of teacher and student interviews, comments and recommendations from subject matter experts and learning media specialists, and teacher learning tool documentation.

The data collection technique consists of two techniques, namely the non-test data collection technique and the test data collection technique. Non-test data collection techniques include observations, interviews, and questionnaires [21]. A questionnaire is a method of data collection in which participants are given a list of questions or written prompts. for answers [21]. The questionnaire technique is used to obtain data on students' learning processes in the classroom. The test data collection technique is divided into two techniques, namely the initial test (pre-test) and the final test (post-test). This initial test is often referred to as the Pre-Test. This test is administered before the start of the learning process to identify how well the students have mastered the material to be delivered [22]. This test is better known as the Post-Test. This evaluation is conducted at the final stage of learning the material to measure students' understanding of the material and key concepts learned [22].

The validity of the test instrument is assessed using expert ratings. Teacher and student response questionnaires were used to assess the product's practicality. Reliability analysis using Cronbach's Alpha yielded a coefficient of 0.633, which indicates reliability. Pre-test and Post-test questions are used to assess students' understanding of the human respiratory system. The pre-test and post-test consist of 20 multiple-choice questions, including complex and two-tier items, all aligned with the learning objectives. Each correct answer is scored 1 point.

Results and Discussion

Research and development of multiple-choice learning evaluation based on automatic assessment through quizizz in the subject of social science of respiratory system material in humans in class V SD N Sekaran 01 Semarang City, resulting in three research objectives, namely: 1) Design and development of multiple choice learning evaluation based on automatic assessment through quizizz, 2) The feasibility of multiple choice learning evaluation based on automatic assessment through quizizz, 3) The effectiveness of multiple choice learning evaluation based on automatic assessment through quizizz.

Learning evaluation design, Multiple Choice based on Automatic Assessment

Learning evaluation development design: Multiple-choice-based automatic assessment through Quizizz. This study is organised around the ADDIE stages: analysis, design, development, implementation, and evaluation. The analyze stage is to analyze the need for product development, evaluate learning, and analyze the feasibility and requirements for product development[23]. The goal is to identify issues, student traits, and learning needs that arise in the IPAS learning process, as well as the gap between ideal and actual conditions [23]. This stage produces an analysis of students' problems, potential, and needs.

Stages Design (design) is a systematic process that begins with designing the product's concept and content [23]. Planning at this stage is systematic and structured, starting with designing the learning flow, determining the approach, and compiling an initial design for learning evaluation. The design produced at this stage becomes a guideline for the development process [24]. Design that is still a prototype. The products are modified based on the findings of the teacher and student needs questionnaire [21].

The teacher and student requirements questionnaire findings were examined and utilised as a guide for creating Multiple Choice based on Automatic Assessment. The design of multiple choice learning evaluation based on automatic assessment through quizizz, from the preparation of learning evaluation instruments that are tailored to the learning objectives to be implemented, including instruments that have been prepared on the quizizz for teacher platform by utilizing the facilities provided by the quizizz platform namely, the 3 types of multiple choices include regular multiple choices, complex multiple choices, and two tier multiple choices. In addition to using multiple-choice variations, researchers also combine stimuli to stimulate students' critical thinking skills. The following stimuli are used: short narrative text, images, tables, and short videos.

The researcher's goal is to employ a range of stimuli in the form of short narrative texts, pictures, tables, and short videos, because elementary school students, according to Piaget's theory of learning, enter the concrete operation stage, namely in the age range of 7-11 years [25]. At this stage, the student can perform tasks that already exist but cannot yet solve abstract problems. Concrete operation is a reversible mental action that is applied to a real, concrete object. Therefore, the stimulus must use concrete objects, such as pictures or videos, to help students solve a problem more effectively.

Based on Figure 1, students are presented with statements or questions that present difficulties and include potential responses in standard multiple-choice questions; students are required to select the most acceptable or correct response [26]. Based on Figure 2, using this complex multiple choice is almost the same as the regular multiple choice form, but the complex multiple choice form has a combination of statements; From the combination of these statements, there is only one correct answer[27]. Based on Figure 3, using two-tier multiple-choice questions to deepen students' knowledge of the material presented: tier 1 contains the answers to the submitted questions, and tier 2 provides the reasons why students choose those answers. The use of

two-tier multiple-choice question forms is considered more effective for learning evaluation. Because the form of objective questions in the form of two-tier multiple choices can also reduce students' tendency to answer questions by guessing, because this form of question requires students to give reasons that are in accordance with the answers they have chosen [28].



Figure 1. Regular multiple options



Figure 2. Multiple Choice Complex



Figure 3. Display of two-tier multiple-choice questions

Based on Figure 4, which shows the stimulus questions as pictures in the learning evaluation, Multiple-choice-based Automatic assessment, the use of images serves to clarify the context of the problem, especially to activate students' visual and imaginative powers [29].



Figure 4. Stimulus image question



Figure 5. Stimulus table questions



Figure 6. Stimulus text questions



Figure 7. Stimulus video questions

Based on Figure 5, the stimuli are displayed in a table for learning evaluation: multiple-choice-based automatic assessment. The table can present data, comparisons, or statistical information, which students then analyze [30]. Based on Figure 6, which shows the stimulus questions in text form for evaluating multiple-choice learning using automatic assessment, the text provides an initial context for developing literacy and logic skills and improving concept understanding through narrative or description [31]. Based on Figure 7, the stimulus for in-video questions in the

Multiple Choice learning evaluation using Automated Assessment can be uploaded directly or retrieved from sources such as YouTube. With videos, students can understand complex concepts more concretely [30].

The development stage is the stage of realizing the product that was previously in the form of a product design in the form of a prototype [24]. The development stage aims to create a complete product, test it, and revise it based on expert validation to ensure the product's feasibility, accuracy, and suitability with the learning objectives. At this stage, the learning evaluation question instrument is included and arranged on the Quizizz platform [23]. Next, the researcher made a validation questionnaire for material experts and media experts. Then the results of the questionnaire of material and media experts were analyzed to find out whether or not Multiple-Choice based on Automatic Assessment was feasible to be used in learning evaluation.

The implementation stage is when feedback is obtained on the developed product [23]. One way to get initial input (assessment) is to ask questions about the goals of product development. This stage aims to apply the products developed in real learning activities. The purpose of the implementation stage is to ascertain how the product will be utilised and how instructors and students will react to it [23]. At this stage, the Multiple-choice product is based on Automatic Assessment through Quizizz, has undergone a feasibility test, and is declared feasible by the validator. At the implementation stage, learning evaluation is tested in small and large groups [19].

Small-group trials were conducted to determine students' initial responses and to assess the product's effectiveness before it was applied to a large group [32]. This activity involved 6 students from the VA class at SDN Sekaran 01, selected based on their cognitive abilities: 2 with high, 2 with moderate, and 2 with low. At this stage, the researcher uses pretest and posttest questions to assess the initial level of understanding and measure improvements in learning outcomes after using the product.

In the large-group trial stage, 22 students completed tests that included pretest and posttest questions. This test was conducted to draw conclusions about the effectiveness of IPAS learning on the human respiratory system through a learning evaluation consisting of multiple-choice questions, implemented via Automatic Assessment using the Quizizz platform developed by the researcher [32]. Evaluation stage. Evaluations are conducted to assess the overall effectiveness and efficiency of learning products [24]. Prior to and during the implementation step, this assessment phase seeks to appraise the calibre of the final product and the educational process [33].

At this stage, the researcher conducts an evaluation by analyzing the results of the pre-test and Post-test that the participants have done [24]. The researcher's analysis includes a normality test conducted in SPSS version 21. Based on the Shapiro-Wilk normality test, the p-value is $.625 > 0.05$; therefore, the data distribution meets the assumption of normality. Based on the normality test results, the post-test showed a p-value of $0.068 > 0.05$, indicating that the data distribution met the assumption of normality [19]. Next, based on the Paired Samples Test, the researcher performed a Paired Samples Test. H_0 is rejected since the significance value (2-tailed) is 0.000, which, according to the criterion, is

less than 0.05. H_0 is rejected while H_a is approved since the significance value (2-tailed) is $0.000 < 0.05$. Students' learning results in the classroom have increased. VA SD N Sekaran 01 after using Multiple-Choice Automatic Assessment via Quizizz.

The last analysis uses the N-Gain test, based on the results of the N-Gain test that has been carried out, showing a Red value of 0.5687. This shows that the increase in pre-test and post-test scores falls within the moderate range. Thus, the development of [19] Multiple Choice-based Automatic Assessment through Quizizz is effective in improving student learning outcomes in the science subject of the Human Respiratory System material in grade V at SD N Sekaran 01.

Feasibility of Multiple Choice learning evaluation based on Automatic Assessment

Learning evaluation feasibility: Multiple Choice-based Automatic Assessment. The material on the human respiratory system is validated by materials and media experts. Assessment by material experts includes 3 aspects: the suitability of the material for learning, appropriate support for the learning content, and stimuli that help students understand the material. Assessment by media experts includes 3 aspects: technical/display aspects, content quality and objectives, and instructional aspects [16].

Table 1. Validation Results of Material Experts and Media Experts

Members	Introduce yourself	Categories
Material	92.85%	Highly Worth It
Media	89.28%	Highly Worth It

Based on Table 1. The results of the assessment by the subject matter expert in the first aspect get a score of 4 on 5 indicators and 3 scores on 1 indicator, in the second aspect get 4 scores for 3 indicators and 3 scores for 2 indicators, and in the third aspect get 4 scores for 2 indicators and 3 scores for 1 indicator. The total score obtained is 52 out of a maximum score of 56. It received an assessment score of 92.85% and met the criteria very well according to material experts. The stimulus of questions in each question can train the student's higher-order thinking skills (HOTS) [34]. This is the consideration of material experts when assessing the quality of the prepared RPPM.

The assessment of media experts in the first aspect receives a score of 4 on 5 indicators and 3 on 1 indicator; in the second aspect, 4 on 1 indicator and 3 on 3 indicators; in the third aspect, 4 on 2 indicators and 3 on 2 indicators. The number of scores obtained is 50, with a maximum score of 56. So it received 89.28% and met the criteria very well.

Table 2. Results of Teacher and Student Responses

Response	Introduce yourself	Categories
Teacher	90.9%	Highly Worth It
Pupils	89%	Highly Worth It

Based on Table 2. The results of the Teacher and Student Responses showed that the teacher's response to the use of Very Feasible Multiple Choice based on Automatic

Assessment through Quizizz in the trial received a score percentage of 90.9% with very feasible criteria. This is indicated by the use of multiple-choice-based automatic assessment, which has been proven to help teachers carry out assessments more effectively and efficiently, as the system is fast and real-time and will show the value of student learning outcomes [11]. Usage: Multiple-choice-based automatic assessment. In learning evaluation, it helps teachers see results for per-item and overall questions, per-student and overall students [19].



Figure 8. Results of the Automatic Assessment of all questions and all students

Based on Figure 8, which shows performance across all students and questions, the teacher can easily identify the most difficult questions and which students need further help [30].

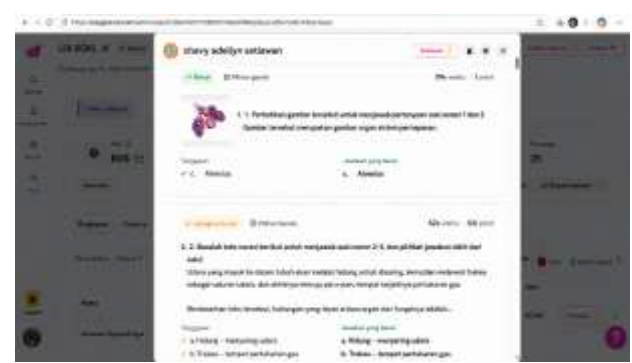


Figure 9. Results of Automatic Assessment per student and per question item

Figure 9 shows the performance of each student and each question. The teacher can easily see each student's results and identify the problem the student is struggling with and needs further action on.[19]. Responses from 28 students showed an average rating of 89% for the criteria as very feasible, indicating that the use of game elements and attractive displays in learning evaluations can increase students' motivation to work on problems [35]. Fun and interactive presentations keep students more enthusiastic and less overwhelmed, while attractive visual designs help keep focus and reduce boredom. Even though it is packaged as a game, the material presented maintains the clarity of the concept, so students' understanding is not reduced and may even be easier to understand [19]. Thus, the multiple-choice quiz based on Automatic Assessment through Quizizz received excellent feedback from instructors and students during the trial, making it highly feasible for use in learning.

The effectiveness of Multiple Choice learning evaluation based on Automatic Assessment

The effectiveness of multiple-choice learning assessment using automatic assessment in this study was evaluated through analyses of student learning outcomes using pre- and post-tests, including normality tests, paired-samples t-tests, and N-Gain tests.

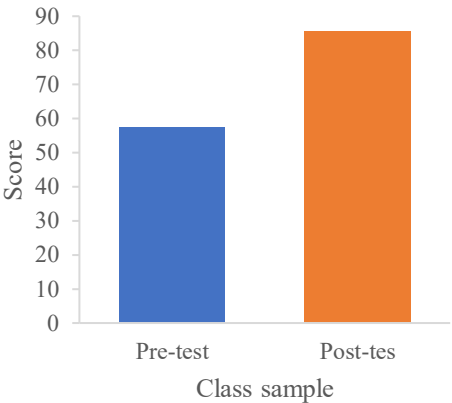


Figure 10. Pre-test and Post-test results

Based on the pre-test and post-test results, the post-test score is higher than the pre-test score. The average pre-test score is 57.5, while the average post-test score increases to 85.58. Thus, there was an increase of 28.08 in the average pre-test and post-test results. This increase indicates that evaluating multiple-choice learning through automated assessment positively impacts students.

The normality test assesses whether the collected data are normally distributed and is used in parametric statistics (inverse statistics) [36]. The normality test for pretest and posttest data was conducted using the One-Sample Shapiro–Wilk test to assess normality and as a prerequisite for further analysis. A Sig. A value greater than 0.05 indicates that the data is normally distributed. In this study, normality testing was performed using the Shapiro–Wilk test [27]. The test was selected from a relatively small sample, namely fewer than 50 students in the research class. The decision-making criteria for this normality test are that if the significance value (Sig) > 0.05, the data is considered normally distributed, whereas if the p-value (Sig) < 0.05, the data is not normally distributed.

Table 3. Normality Test Results

Categories	Sig	Remarks
Pre-test	0.625	Normal
Post-test	0.068	Normal

The table above shows the results of the pre-test normality test and the post-test of the initial trial, using SPSS 21. From the results of the Shapiro-Wilk normality test, the significance of the pre-test is 0.625, and that of the post-test is 0.068. The data distribution satisfied the assumption of normality because the normality test for the pre-test significance value yielded $p = 0.625 > 0.05$. The data distribution satisfied the assumption of normality, as the post-test normality test indicated a significance level of $0.068 > 0.05$. Therefore, it can be concluded that the pre-test and first post-test data are normally distributed because both are greater than 0.05.

Table 4. Paired Sample t-test results

Actions	Red	Improvement	Sig.
Pre-test	57.5	28.08	0.000
Post-test	85.58		

Based on the results of the paired-samples t-test, the average pre-test score in large-scale trials was 57.5, and the average post-test score was 85.58. Thus, there was an average increase of 28.08 after learning using multiple-choice learning evaluation based on automatic assessment. The significance value of Sig. (2-tailed) is $0.000 < 0.05$, thus indicating a significant difference between the two variables. According to the decision-making standards, if the p-value < 0.05 , then H_0 is rejected and H_a is accepted. So, it can be concluded that there is a significant difference in the learning outcomes of students in class VA SD N Sekaran 01 before and after using Multiple-Choice-based Automatic Assessment through Quizizz.

The N-Gain exam is used to measure how much student learning outcomes have improved before and after therapy or utilising development. Multiple-choice-based Automatic Assessment uses the N-Gain value, calculated as the difference between pretest and posttest scores, and then compares it to the maximum score achieved. The results of the N-Gain test are presented in the following table [19].

Table 5. N-Gain Test Results

Aspects	Red	Remarks
N-Gain Score	0.5687	Medium

Based on the table above, the N-Gain test yielded an average value of 0.5687, which falls within the medium category. This indicates that the pre-test and post-test results fall within the moderate range. The N-Gain test yielded reasonable results, the material and media experts' feasibility test yielded very respectable results, and the instructors' and students' replies were highly feasible. Thus, the acquisition of N-gain values in the medium category is suspected to be influenced by several technical factors that occur during the implementation of learning. One of the main factors is students' readiness and learning styles [16].

The development of multiple-choice questions for automatic assessment in Quizizz has been designed to accommodate a variety of learning styles, including visual and audio-visual. This product still has limitations regarding students' readiness to learn. Not all students have the same level of readiness to face digitally based evaluations, in terms of understanding the material, skills in using technology, and psychological conditions when working on the questions [37]. Therefore, the success of implementing learning evaluation based on automatic assessment remains highly dependent on students' readiness, independence, and responsibility in following the evaluation process optimally [35]. Thus, even though the improvement obtained is in the medium category, Multiple Choice based on Automatic Assessment through Quizizz is still proven to be effective in improving student learning outcomes.

Conclusion

This study aims to develop and test the feasibility and effectiveness of a multiple-choice learning evaluation based on automatic assessment via Quizizz for the respiratory

system material in social science. The learning evaluation was developed using the ADDIE model, with multiple-choice and stimulus-based questions delivered through Quizizz. The results of the study showed that the learning evaluation received very feasible criteria, with percentages of material experts scoring 92.85% and media experts 89.28%, as well as positive teacher responses of 90.9% and students of 89%. In addition, the results of the Paired Samples Test showed an increase in learning outcomes, with an average difference of 28.08 and an N-Gain of 0.5687, meeting the moderate criteria. It can be concluded that multiple-choice questions based on automatic assessment through Quizizz have been successfully developed, are highly feasible, and are effective in improving understanding of respiratory system material among grade V students at SD N Sekaran 01. Practically, this study helps teachers identify levels of understanding and learning difficulties and determine appropriate follow-up for learning science subjects on the human respiratory system. Increasing teachers' knowledge of technology-based learning evaluation makes it easier for teachers to assess students' understanding of the material taught, especially the human respiratory system, and makes the assessment process more time-efficient. However, this study has limitations; therefore, researchers are advised to use other types of multiple-choice questions and more comprehensive features on the Quizizz website to make learning evaluations more interactive.

Author's Contributions

L.A Khaerina: contributes to research, product development, data analysis, and research writing. A.E. Andriani: contributing to compiling, guiding, and revising research.

Acknowledgments

Thank you to the Elementary School Teacher Education Study Program, Faculty of Education and Psychology, Semarang State University. Thank you to Mrs. Aldina Eka Andriani, S.Pd., M.Pd., as a supervisor in the direction, motivation, and support provided. Thank you to parents for their endless prayers and support. Thank you to SD N Sekaran 01 for providing permission and support in the implementation of this research. As well as all friends and colleagues who contribute to every research process, from the entire series of lecture processes to carrying out the final project. Last but not least, thank you to Lavina Azza Khaerina for all the great struggles in going through the entire process until the end of this research.

References

- [1] Y. D. Mara, M. J. Hitipeuw, A. Balisa, and W. Gunawan, "Analisis butir soal *multiple choice* mata pelajaran matematika materi menghitung bangun datar untuk kelas 3 SD Negeri Inpres Kotabaru di Nabire," 2026.
- [2] J. Juliani, Y. Yusrizal, and I. Huda, "Development of four-tier multiple-choice diagnostic tests to know students' misconceptions in science learning," *J. Penelit. Pendidik. IPA*, vol. 7, no. 4, pp. 763–769, 2021, doi: 10.29303/jppipa.v7i4.854.
- [3] T. Nasution, D. Afrianti, T. Tukiyo, S. Sulistyani, and H. Herman, "Critical discourse analysis in the

- classroom: A critical language awareness on early children's critical thinking," *J. Obsesi: J. Pendidik. Anak Usia Dini*, vol. 6, no. 5, pp. 4992–5002, 2022, doi: 10.31004/obsesi.v6i5.2951.
- [4] Donna Meylovia, Julianto, and Alfin, "Inovasi pembelajaran IPAS pada Kurikulum Merdeka Belajar di SDN 25 Bengkulu Selatan," *J. Pendidik. Islam Al-Affan*, vol. 4, no. 1, pp. 84–91, 2023, doi: 10.69775/jpia.v4i1.128.
- [5] S. Halimah, H. Usman, and S. Maryam, "Peningkatan kemampuan berpikir kritis dalam pembelajaran IPA melalui penerapan model *problem-based learning* (PBL) di sekolah dasar," *J. Syntax Imperatif*, vol. 3, no. 6, pp. 403–413, 2023, doi: 10.36418/syntax-imperatif.v3i6.207.
- [6] I. Magdalena, *Desain Evaluasi Pembelajaran SD*. Sukabumi, Indonesia: CV Jejak, 2021.
- [7] B. Pariama, D. Elfariyah, K. Harahap, M. I. Prabudi, and M. A. Fiqri, "Bentuk dan teknik evaluasi hasil belajar," *J. Mudabbir*, vol. 5, 2025, doi: 10.56832/mudabbir.v5i1.656.
- [8] D. N. Rahmawati, A. F. Nisa, and D. Astuti, "Pemanfaatan aplikasi Quizizz sebagai media penilaian pembelajaran ilmu pengetahuan alam," *J. Pendidik. MI/SD*, vol. 2, no. 1, 2022, doi: 10.35878/guru.v2i1.335.
- [9] M. Dewati and H. Widyasari, "Pengembangan instrumen evaluasi pembelajaran Kurikulum 2013 berbasis multimedia pada materi fisika SMA kelas X," in *Proc. Semin. Nas. Fis. dan Apl. (SNFA)*, 2021.
- [10] M. Nilwan and S. L. S. N., "Pengembangan instrumen evaluasi pembelajaran menggunakan aplikasi Quizizz untuk mengukur kemampuan berpikir kritis peserta didik kelas V SDN Inpres Paradowane," *J. Ilm. Pendidik. Dasar*, vol. 10, 2025.
- [11] S. Asral and R. Chandra, "Perbedaan motivasi belajar siswa yang menggunakan media evaluasi pembelajaran Quizizz dengan media evaluasi pembelajaran konvensional," 2021, doi: 10.46574/mted.v2i2.53.
- [12] F. D. S. Purba and D. F. Dwi, "Pengembangan evaluasi hasil belajar berbasis digital dengan aplikasi Educandy pada tema panas dan perpindahannya siswa kelas V di SD Negeri 107423 Tanjung Garbus," *ARMADA J. Penelit. Multidisiplin*, vol. 1, no. 7, pp. 731–739, 2023, doi: 10.55681/armada.v1i7.706.
- [13] Y. Hasnah, P. Ginting, and S. H. Hasibuan, "Analisis evaluasi pembelajaran berbasis HOTS bagi guru SMP," vol. 4, no. 1, pp. 134–143, 2021.
- [14] Wahidin, "Pelaksanaan evaluasi pembelajaran berbasis *higher order thinking skills* (HOTS)," 2023, doi: 10.71382/sinova.v1i3.32.
- [15] Zulrafla, Kamarudin, and Y. Erawati, "Peningkatan kompetensi dan kreativitas guru melalui pelatihan pembuatan soal-soal berbasis HOTS pada kelompok kerja guru," *J. Hum. Educ.*, 2023.
- [16] M. Andra, A. Wd, M. Sondang, L. Hanifah, and L. D. Kusumayati, "Penggunaan aplikasi Quizizz sebagai media untuk melaksanakan evaluasi pembelajaran," 2022.
- [17] N. Haliza and D. F. Dwi, "Analisis faktor kesulitan belajar siswa pada pembelajaran IPAS materi organ pernapasan manusia kelas V SD Negeri 106815 Marindal tahun pembelajaran 2024–2025," *Pendas J. Ilm. Pendidik. Dasar*, vol. 10, 2025.
- [18] N. Haliza and D. F. Dwi, "Analisis faktor kesulitan belajar siswa pada pembelajaran IPAS materi organ pernapasan manusia kelas V SD Negeri 106815 Marindal tahun pembelajaran 2024–2025," *Pendas J. Ilm. Pendidik. Dasar*, vol. 10, 2025.
- [19] A. S. Balqis and A. E. Andriani, "Development of learning evaluation based on automatic assessment through Quizizz paper mode to improve students' natural and social sciences learning outcomes," *J. Penelit. Pendidik. IPA*, vol. 10, no. 6, pp. 3357–3366, 2024, doi: 10.29303/jppipa.v10i6.7251.
- [20] S. R. Zidni, S. R. Ubaidillah, Sahlan, and A. Suhardi, "Pengembangan evaluasi pembelajaran IPAS berbasis Quizizz untuk meningkatkan motivasi dan prestasi belajar kelas V SDN Karanganyar 01 Jember," *Indones. J. Islam. Teach.*, vol. 7, pp. 249–254, 2024, doi: 10.35719/ijit.v7i1.2063.
- [21] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung, Indonesia: Alfabeta, 2023.
- [22] I. Magdalena, M. N. Annisa, G. Ragin, and A. R. Ishaq, "Analisis penggunaan teknik *pre-test* dan *post-test* pada mata pelajaran matematika dalam keberhasilan evaluasi pembelajaran di SDN Bojong 04," 2021.
- [23] W. Dick, L. Carey, and J. O. Carey, *The Systematic Design of Instruction*. Hoboken, NJ, USA: Pearson, 2022.
- [24] T. Siregar, "Implementasi pengembangan model ADDIE pada dunia pendidikan," 2025.
- [25] L. Marinda, "Teori perkembangan kognitif Jean Piaget dan problematikanya pada anak usia sekolah dasar," 2020, doi: 10.35719/annisa.v13i1.26.
- [26] I. Masidjo, *Penilaian Pencapaian Hasil Belajar Siswa di Sekolah*. Berlin, Germany: Walter de Gruyter, 1995.
- [27] R. S. Budi and A. E. Andriani, "Development of multiple-choice learning evaluation based on automatic assessment of hydrosphere material," *J. Dimens. Pendidik. Pembelajaran*, vol. 13, 2025, doi: 10.24269/dpp.v13i2.11865.
- [28] Afifah and Wardhatus, "Pengembangan alat evaluasi pembelajaran matematika berbasis *two-tier multiple choice* berlevel HOTS menggunakan iSpring Suite 9," 2021, doi: 10.30587/postulat.v2i2.3835.
- [29] S. Nurfadhillah, A. P. Cahyani, A. F. Haya, P. S. Ananda, T. Widyastuti, and U. M. Tangerang, "Penerapan media audio visual berbasis video pembelajaran pada siswa kelas IV di SDN Cengklong 3," 2021.
- [30] R. S. Budi and A. E. Andriani, "Development of multiple-choice learning evaluation based on automatic assessment of hydrosphere material," *J. Dimens. Pendidik. Pembelajaran*, vol. 13, 2025, doi: 10.24269/dpp.v13i2.11865.
- [31] A. I. Susanti, *Media Pembelajaran Berbasis Teknologi Informasi dan Komunikasi (TIK)*. Ponorogo, Indonesia: Penerbit NEM, 2021.
- [32] Okpatrioka, "Research and Development (R&D): Penelitian yang inovatif dalam pendidikan," *Dharma Acariya Nusantara J. Pendidik. Bahasa dan Budaya*, vol. 1, 2023, doi: 10.47861/jdan.v1i1.154.
- [33] F. Hidayat and M. Nizar, "Model ADDIE (*Analysis, Design, Development, Implementation and Evaluation*) dalam pembelajaran pendidikan agama Islam," 2021, doi: 10.15575/jipai.v1i1.11042.

- [34] A. M. Ahmadi *et al.*, “Problematika guru dalam melaksanakan evaluasi pembelajaran siswa di kelas IV sekolah dasar,” *J. Binagogik*, vol. 12, no. 1, pp. 33–40, 2025.
- [35] A. Jong and Y. T. B. Tacoh, “Pemanfaatan aplikasi Quizizz untuk meningkatkan motivasi belajar siswa,” *J. Dimens. Pendidik. Pembelajaran*, vol. 12, 2024, doi: 10.24269/dpp.v12i1.7344.
- [36] N. Haniah, “Uji normalitas dengan metode Liliefors,” 2013.
- [37] M. Andra, A. Wd, M. Sondang, L. Hanifah, and L. D. Kusumayati, “Penggunaan aplikasi Quizizz sebagai media untuk melaksanakan evaluasi pembelajaran,” 2022.