Research and Development Journal Of Education
Vol. 12, No. 2, Oktober 2026, Pp: 880 – 891
DOI: https://doi.org/10.30998
Available online at https://journal.lppmunindra.ac.id/index.php/RDJE

p–ISSN 2406-9744
e–ISSN 2657-1056

DEVELOPMENT OF A QUIZIZZ-BASED ASSESSMENT INSTRUMENT FOR HEAT AND SOUND ENERGY IN FOURTH-GRADE SCIENCE

Muhsinin[1], Esty Setyo Utaminingsih[2(*)], FR. Wuriningsih[3], Idammatussilmi[4]
STAI Muhammadiyah Blora, Indonesia[1]
Universitas Negeri Semarang, Indonesia[2]
STPKat Santo Fransiskus Asisi Semarang, Indonesia[3]
Institut Islam Nahdlatul Ulama Temanggung, Indonesia[4]
estys_utami@students.unnes.ac.id[2]

Received: 15 November 2025
Revised: 12 Oktober 2026
Accepted: 12 Oktober 2026

Abstract

This study developed a Quizizz-based science assessment instrument for fourth-grade students, with a focus on heat and sound energy. Research and development followed the four stages of a 4D model: define, design, develop, and disseminate. A needs assessment using teacher interviews identified continued reliance on textbook- and paper-based questions. The product comprised 20 items in four formats: five multiple-choice, five short-answer, five constructed-response, and five matching items. Three independent experts, respectively responsible for media, language, and subject matter, evaluated 20 criteria per domain using a five-point relevance scale. The expert-rating scores were 90/100 (90%) for media, 95/100 (95%) for language, and 95/100 (95%) for content, corresponding to a descriptive overall percentage of 93.33%. These results indicate favorable expert judgments regarding the preliminary suitability of its design and content. However, no student usability assessment, field-test responses, item analysis, reliability estimates, or learning-outcome comparisons were reported. Consequently, the findings support expert-assessed feasibility rather than empirical effectiveness. Classroom piloting is needed before the instrument is used to make consequential decisions about student achievement.

Keywords: Science Assessment; Quizizz; Heat Energy; Sound Energy; Expert Appraisal

(*) Corresponding Author: Utaminingsih, estysetyoutami7@gmail.com

How to Cite: Muhsinin, Utaminingsih, E. S., Wuriningsih, F., & Idammatussilmi. (2026). DEVELOPMENT OF A QUIZIZZ-BASED ASSESSMENT INSTRUMENT FOR HEAT AND SOUND ENERGY IN FOURTH-GRADE SCIENCE. *Research and Development Journal of Education*, 12(2), 880-891.

INTRODUCTION

In elementary science education, assessment should elicit evidence of students' understanding of observable phenomena rather than merely record task completion. Scientific literacy in primary school depends partly on opportunities to explain and reason about science ideas (Irsan, 2021). From a formative-assessment perspective, useful evidence must inform instructional decisions and feedback. Moreover, the interpretation and intended use of scores require supporting evidence about content, responses, and scoring, not simply a visually appealing delivery format (American Educational Research Association et al., 2014; Sondergeld, 2019).

Heat and sound are suitable contexts for everyday scientific reasoning because students can encounter heat transfer and sound sources in their homes and schools. Studies

on technology-enhanced science activities in elementary settings underline the importance of aligning learning tasks with conceptual goals (Ningrum et al., 2022; Nipyrakis et al., 2023). Research on primary-level science resources has also emphasized curriculum alignment and scientific literacy (Khasanah et al., 2022; Parinduri et al., 2022). Within Indonesia's Merdeka Curriculum, such alignment is more consequential than the choice of a particular digital application (Indarta et al., 2022).

Digital resources may expand opportunities to present science content and responses, but evidence about teaching media must be distinguished from evidence about test quality. Science teachers' technology use varies across contexts (Adov & Mäeots, 2021), while studies of digital modules have investigated learner motivation, science literacy, or disciplinary understanding rather than the validity of Quizizz test scores (Diana & Zulherman, 2024; Faridah et al., 2022; Liana et al., 2024). Related development work in STEAM and science instruction similarly illustrates the potential of technology-supported materials without establishing that their assessment scores are interchangeable with conventional tests (Utaminingsih & Ellianawati, 2025; Utaminingsih et al., 2023).

The development in this study originated from teacher interviews at SD Negeri 1 Kedung Satriyan, Blora, where assessment was described as relying mainly on textbook- and paper-based questions. This local information motivated the design of an online option. Because these interviews were not a representative survey, they cannot establish how frequently similar assessment practices occur in other schools. A 20-item prototype focusing on fourth-grade heat and sound was consequently prepared for expert review before any student-based evaluation.

Quizizz-based instruments are not new. Previous studies developed assessment instruments for physics knowledge (Elisa et al., 2021) and higher-order-thinking science questions in Grade V (Dermawan et al., 2023). Another elementary science-literacy study used a game-based assessment and examined its reliability through student data (Mukti et al., 2023). A review of Quizizz in elementary science similarly identified potential educational uses (Aqtoina & Rohmani, 2025), and a study of elementary evaluation investigated students' motivation when Quizizz was used (Quispe Ccoa et al., 2023). These publications help position the present product as a specific assessment design rather than a new assessment technology.

Research on gamified response platforms gives a reason to examine engagement but not to assume it. Work with Kahoot! in animal science education described engagement in its own setting (Cameron & Bizo, 2019); a comparison of Kahoot and Quizizz reported differences between the two systems (Orhan Göksün & Gürsoy, 2019). A broader literature review identified both opportunities and practical problems, including time pressure and technology constraints (Wang & Tahir, 2020). Meta-analytic and review evidence likewise suggests that gamification effects depend on implementation and context (Sailer & Homner, 2020; Zainuddin et al., 2020). A review focused on primary education indicates the need to consider infrastructure and learning design (Aji et al., 2025). None of these external findings substitutes for usability or achievement data from the students in the present study.

The contribution of this study is therefore a transparent, four-format blueprint for fourth-grade heat-and-sound assessment, with five multiple-choice, five short-answer, five constructed-response, and five matching items. Prior Indonesian science-instrument research has developed assessment products for other content areas (Prastiawan & Utaminingsih, 2024; Utaminingsih, Ellianawati, et al., 2024; Utaminingsih, Ramadhani, & Puspita, 2024). The current blueprint also includes some alternative-energy items that warrant checking against the specified learning objectives. More generally, expert appraisal

offers an initial basis for development but does not constitute complete validation of score interpretations (Conn et al., 2020; Sondergeld, 2019).

Two research questions guided the report: (1) How was the 20-item, multi-format Quizizz assessment for fourth-grade heat and sound designed through the 4D development process? (2) How did media, language, and subject-matter experts appraise the resulting prototype? The purpose was to report instrument development and expert ratings while clearly identifying the evidence still required to establish classroom practicality, score reliability, and effectiveness.

METHODS

This study used research and development (R&D) organized around the 4D model: Define, Design, Develop, and Disseminate (Thiagarajan et al., 1974). The process was undertaken at Sekolah Dasar Negeri 1 Kedung Satriyan, Ngawen, Blora, Central Java, during November 2023. Teacher interviews supplied contextual information for the needs analysis; the number and selection procedure of interview participants were not specified in the study records reported here. The development sequence is summarized in Figure 1.

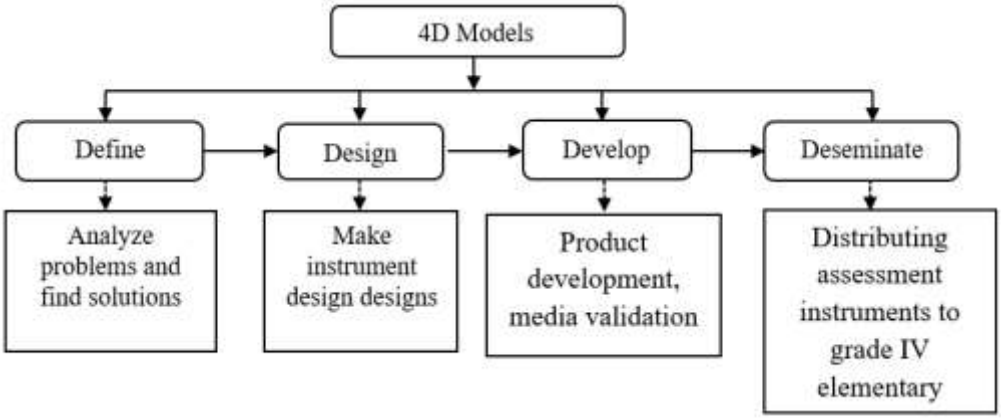


Figure 1.
Four-Stage Development Process of the Assessment Instrument

In the Define phase, the researchers used teacher interviews to identify difficulties associated with conventional paper-based assessment and reviewed the science topics and instructional objectives used at the study school. The local curriculum information guided the initial selection of heat and sound content. These findings informed product design rather than serving as evidence of student performance.

In the Design phase, the researchers established a blueprint relating learning outcomes, observable question indicators, response formats, and item numbers. The plan comprised 20 items: five multiple-choice, five short-answer, five constructed-response, and five matching items. Quizizz was selected as the online delivery platform because it accommodated the intended formats. The assessment was developed within Quizizz; it was not a separately programmed Android application.

In the Develop phase, the blueprint was converted into digital items and reviewed by three experts: one lecturer or practitioner responsible for media, one for language, and one for science content. Each expert rated 20 criteria within the assigned domain. The five ordered response categories were Highly relevant (5), Relevant (4), Moderately relevant

(3), Less relevant (2), and Irrelevant (1). The ratings are reported descriptively by aspect; the study did not estimate agreement among multiple judges rating the same domain. Expert comments were intended to inform revisions, although item-level revision histories were not available in the reported dataset. The interpretation bands used in the original development are reproduced in Table 1.

Scores were calculated as: percentage = (sum of the 20 item ratings / [20 × 5]) × 100. Thus, each expert had a maximum possible raw score of 100. For example, the media expert assigned fifteen ratings of 5 and five ratings of 3, yielding (15 × 5) + (5 × 3) = 90 points, or 90%. The combined percentage across the three distinct domains was (90 + 95 + 95) / 300 × 100 = 93.33%; this is a descriptive average of expert scores, not a statistical validity coefficient. The instrument was classified using the predefined percentage bands, and no inferential tests were performed.

Table 1.
Expert-Rating Interpretation Criteria

Achievement level	Category	Recommended action
85–100%	Excellent	Proceed to classroom pilot
75–<85%	Good	Proceed to classroom pilot
65–<75%	Adequate	Revise and re-evaluate
55–<65%	Limited	Revise and re-evaluate
0–<55%	Very limited	Revise and re-evaluate

In the Disseminate phase, a link to the developed Quizizz instrument was shared through online and social media channels as described in the original project documentation. Distribution did not constitute a classroom field trial. In particular, the dataset presented in this manuscript does not include student response records, usability questionnaires, observations of test-taking, administration-time data, or pretest–posttest comparisons. Consequently, practicality, psychometric reliability, and effectiveness were outside the scope of the present analysis.

RESULTS & DISCUSSION

Results

The needs assessment was conducted at Sekolah Dasar Negeri 1 Kedung Satriyan in Blora Regency, Central Java, in November 2023. Teacher interview data indicated continued use of conventional assessment tasks and a perceived need for more varied digital formats. The school used the Merdeka Curriculum. These interview-based observations provided the rationale for a locally developed question blueprint; no baseline scores or quantified measures of student boredom, engagement, or achievement were collected for the present report.

Following the curriculum and needs review, the researchers organized the questions by intended learning outcome, indicator, response format, and item number. Table 2 presents the original 20-item blueprint.

Table 2.
Assessment Item Blueprint

Learning Outcomes	Question Indicators	Question Types and Numbers
Describe the heat and sound energy present in the surrounding environment and its properties	Identifying nearby sources of heat energy	Multiple Choice : 1 Multiple Choice : 2 Short Fill :1 Description : 1
	Demonstrate and explain the existence of heat transfer and its types	Multiple Choice : 3 Short Fill :2 Short fill :3 Description : 2 Description : 3
	Make a list of sound sources in the surrounding environment.	Multiple choice : 4 Short fill : 4 Description :4
Explain the different alternative energies and how they can be used	Searching for information Various alternative energy sources	Multiple choice : 5 Short fill : 5 Description : 5
Describe the heat and sound energy present in the surrounding environment and its properties	Identify sources of heat and sound energy	Matchmaking :1 Matchmaking :2 Matchmaking :3 Matchmaking :4 Matchmaking :5

As shown in Table 2, the instrument contains five multiple-choice, five short-answer, five constructed-response, and five matching items. The topics include identification of heat sources, mechanisms of heat transfer, sources of sound, and a small set of questions on alternative energy sources. The latter extend beyond the narrow heat-and-sound focus and should be reviewed for alignment with the specific learning objectives used in subsequent pilots. The item blueprint was implemented in the Quizizz environment. Selected development screenshots appear in Figures 2–5; they document the digital presentation and do not provide evidence of student usability.

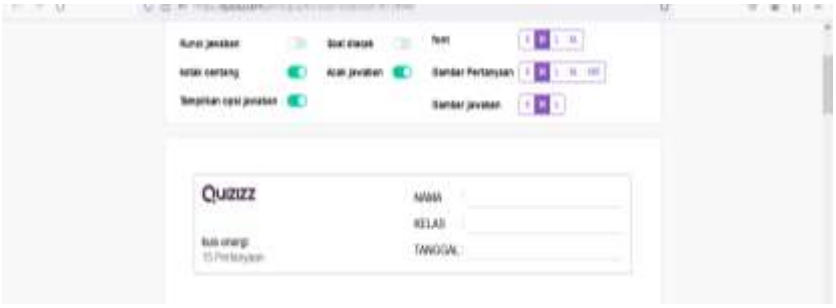


Figure 2.
Quizizz Login Initial View

Figure 2 shows the quiz configuration/interface used during product development. The screenshot documents the setup of the digital assessment rather than verified student registration behavior. Figure 3 illustrates the multiple-choice display, where students select one answer from the available options.

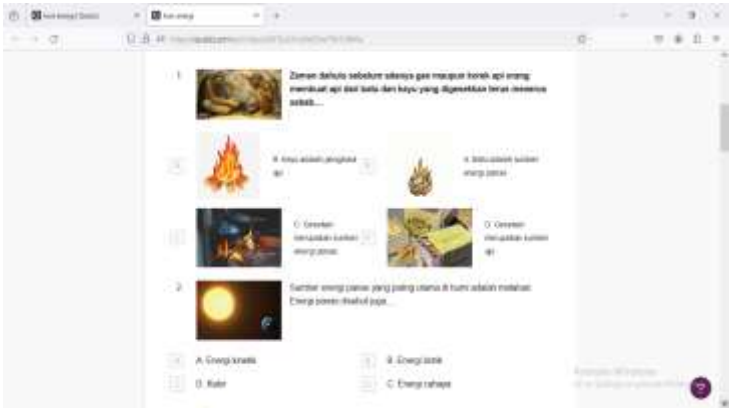


Figure 3.
Multiple Choice Questions

Figure 4 illustrates the short-answer format, which requires students to enter a concise response in a text field. Correct-answer settings and scoring rules should be checked during piloting, particularly where spelling or equivalent scientific expressions could affect automated marking.



Figure 4.
Short Fill Questions

Figure 5 presents a constructed-response item. In contrast to closed-ended questions, this format enables students to explain an answer in their own words; however, a scoring rubric and evidence of scoring consistency would be needed before the resulting scores could support strong interpretive claims. The instrument also includes matching items according to Table 2, although a separate screenshot of the matching format was not included in the available manuscript.



Figure 5.
Description Questions

The media expert evaluated the assessment interface and its visual, device-related, and operational aspects across 20 criteria. Fifteen criteria were rated Highly relevant (5) and five Moderately relevant (3), yielding a raw score of 90/100, equivalent to 90%. Under the reported interpretation bands, this result falls in the Excellent category. The item-level response categories are presented in Table 3.

Table 3.
Media Expert Ratings (n = 20 criteria)

No.	Classification of values
1	Highly relevant
2	Highly relevant
3	Highly relevant
4	Quite relevant
5	Highly relevant
6	Highly relevant
7	Quite relevant
8	Highly relevant
9	Highly relevant
10	Highly relevant
11	Quite relevant
12	Highly relevant
13	Quite relevant
14	Highly relevant
15	Highly relevant
16	Highly relevant
17	Highly relevant
18	Highly relevant
19	Quite relevant
20	Highly relevant

The language expert evaluated the clarity, grammar, and appropriateness of the Indonesian wording used in the items. Fifteen criteria were rated Highly relevant (5) and five Relevant (4). The corresponding raw score was 95/100 (95%), which falls in the Excellent category (Table 4). This percentage summarizes a single expert assessment and should not be interpreted as inter-rater agreement.

Table 4.
Language Expert Ratings (n = 20 criteria)

It	Classification of values
1	Highly relevant
2	Highly relevant
3	Highly relevant
4	Relevant
5	Highly relevant
6	Highly relevant
7	Highly relevant
8	Highly relevant
9	Highly relevant
10	Highly relevant
11	Relevant
12	Relevant

13	Highly relevant
14	Highly relevant
15	Highly relevant
16	Highly relevant
17	Relevant
18	Highly relevant
19	Relevant
20	Highly relevant

The subject-matter expert examined the relevance of the assessment content and its correspondence with the intended science material. As shown in Table 5, fifteen criteria were rated Highly relevant (5) and five Relevant (4), resulting in 95/100, or 95% (Excellent). This finding represents favorable content-focused judgment from one expert, rather than psychometric validation of the individual item scores.

Table 5.
Subject-Matter Expert Ratings (n = 20 criteria)

Item	Value Classification
1	Highly relevant
2	Highly relevant
3	Highly relevant
4	Highly relevant
5	Relevant
6	Highly relevant
7	Highly relevant
8	Relevant
9	Highly relevant
10	Highly relevant
11	Relevant
12	Relevant
13	Highly relevant
14	Highly relevant
15	Highly relevant
16	Highly relevant
17	Highly relevant
18	Highly relevant
19	Highly relevant
20	Relevant

The raw scores and their component rating frequencies are summarized in Table 6. Each domain used the same five-point scale and a maximum of 100 points. The mean percentage of 93.33% is numerically equivalent to dividing the observed total score of 280 by the maximum possible total of 300. Because the three experts assessed different domains, the combined score is an overall descriptive appraisal of design suitability, not evidence of convergent agreement among judges.

Table 6.
Reconstruction of Expert Scores and Percentages

Domain	Rating computation	Expert score
Media	15 × 5 + 5 × 3	90/100 (90%)
Language	15 × 5 + 5 × 4	95/100 (95%)
Content	15 × 5 + 5 × 4	95/100 (95%)

Combined	90 + 95 + 95	280/300 (93.33%)

Ratings 5, 4, and 3 denote Highly relevant, Relevant, and Moderately relevant. No ratings of 1 or 2 were recorded. Overall percentage = 280 ÷ 300 × 100.

Across all three domains, expert scores exceeded the predefined Excellent threshold. The evidence, therefore, supports describing the prototype as favorably assessed by three experts for media presentation, wording, and content. The ratings do not demonstrate that students can use the instrument independently, that items discriminate among achievement levels, that scores are reliable, or that the assessment improves learning. No student-level field-testing evidence was available for analysis.

Discussion

The three expert appraisals yielded 90% for media and 95% each for language and science content, equivalent to a descriptive total of 93.33%. Such ratings are useful for identifying whether the prototype is sufficiently developed for a classroom pilot, but they should not be interpreted as item reliability, student achievement, or a causal effect of using Quizizz. Related Quizizz assessment studies progressed beyond expert judgment to empirical testing (Dermawan et al., 2023; Mukti et al., 2023), making the distinction between product feasibility and measured quality particularly important for interpreting this study.

The results also illustrate the distinction between expert approval of an instrument and the validity of interpretations from its scores. Standards-based assessment reasoning considers evidence from test content, response processes, scoring, and intended use (American Educational Research Association et al., 2014). Studies on multimethod assessment validation show why combining evidence sources is preferable to relying on a single percentage (Conn et al., 2020; Sondergeld, 2019). A useful formative instrument must ultimately support decisions about what students understand and what teachers should teach next (Black & Wiliam, 2009). Consequently, the expert percentages are appropriately described as early, domain-specific appraisals.

The four response formats provide opportunities to elicit different forms of answers. Multiple-choice and matching items can ask students to identify or classify sources of heat and sound; short-answer and constructed-response tasks can invite explanations about mechanisms. Whether these formats actually yield interpretable and consistent evidence is an empirical question. Development studies of science assessment instruments in different elementary topics illustrate the importance of specifying constructs and item content (Prastiawan & Utaminingsih, 2024; Utaminingsih, Ellianawati, et al., 2024; Utaminingsih, Ramadhani, & Puspita, 2024). For the present prototype, scoring rubrics, equivalent-answer rules, item difficulty, and discrimination need further examination before scores are used for instructional decisions.

The platform may make tasks visually engaging and enable prompt feedback, but those affordances have not been evaluated with the participating fourth-grade students. Reports on Quizizz and related quiz systems have examined motivation, engagement, and potential barriers in other classrooms (Orhan Göksün & Gürsoy, 2019; Quispe Ccoa et al., 2023; Wang & Tahir, 2020). Reviews of gamified education note that effects vary by design, population, and learning outcomes (Aji et al., 2025; Sailer & Homner, 2020; Zainuddin et al., 2020). The present manuscript, therefore, does not claim that students learned more, enjoyed the test, or became more motivated. Those hypotheses require direct evidence from students and appropriate comparison or observation.

Pedagogically, the content blueprint also requires attention to local learning goals. Elementary science-literacy literature has emphasized meaningful contexts and conceptual reasoning (Irsan, 2021; Khasanah et al., 2022), while studies of digitally supported science materials point to various ways of representing phenomena (Diana & Zulherman, 2024; Ningrum et al., 2022; Parinduri et al., 2022). Work on technology-supported science experiments offers a related rationale for linking representations to instructional purposes (Nipyrakis et al., 2023). In the current product, three heat-and-sound strands sit alongside some broader alternative-energy questions. These items should be reviewed against the selected learning outcomes rather than presumed to be aligned, because they concern energy in a general sense (Indarta et al., 2022).

A practical next phase could combine a small-scale classroom pilot with teacher observations, student think-alouds, a review of response records, and an analysis of usability constraints. Research on digital science resources shows that technological content and classroom use vary considerably {Formatting Citation}. However, outcomes observed in instructional-media studies are not interchangeable with evidence supporting a test. The follow-up study should record completion rates and technical issues, assess students' interpretations of science vocabulary, and evaluate scoring rubrics before making any claims about wider dissemination or effectiveness.

Finally, the original development used one expert for each of the media, language, and content domains, so inter-rater agreement within a domain could not be estimated. Item-level revisions were incompletely documented, and no student field-test responses or comparison data were available. These limitations are substantive rather than merely procedural. The 93.33% score is an average of expert ratings, not a psychometric validity coefficient or a measure of impact; similarly, disseminating a Quizizz link does not establish the instrument's effectiveness. Subsequent research should explicitly separate content review, response-process evidence, internal structure, reliability, and classroom consequences (American Educational Research Association et al., 2014; Conn et al., 2020; Mukti et al., 2023).

CONCLUSION

This study produced a 20-item, four-format Quizizz assessment prototype for fourth-grade science topics centered on heat and sound energy. The media, language, and subject-matter experts respectively awarded 90/100 (90%), 95/100 (95%), and 95/100 (95%), corresponding to a descriptive combined percentage of 93.33% and the Excellent category in the study's rubric. These results support preliminary expert-assessed suitability. They do not establish student usability, test reliability, learning improvement, or comparative effectiveness. A classroom pilot and subsequent psychometric evaluation, alongside a check of the broader alternative-energy items against the intended learning outcomes, are necessary before stronger claims or high-stakes uses are warranted.

REFERENCES

Adov, L., & Mäeots, M. (2021). What can we learn about science teachers' technology use during the COVID-19 pandemic? *Education Sciences, 11*(6), Article 255. https://doi.org/10.3390/educsci11060255

Aji, W. D., Haryani, S., Subali, B., & Widiarti, N. (2025). Efektivitas penggunaan gamifikasi dalam pembelajaran di sekolah dasar: Analisis tren penelitian 2020–2025.

EDUPROXIMA: Jurnal Ilmiah Pendidikan IPA, 7(4), 2270–2283. https://doi.org/10.29100/.v7i4.7873

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://www.testingstandards.net/open-access-files.html

Aqtoina, A., & Rohmani. (2025). Media analysis of the Quizizz application in elementary school science learning: A systematic literature review. *Indonesian Journal of Educational Research and Review, 8*(1), 55–69. https://doi.org/10.23887/ijerr.v8i1.86468

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Cameron, K. E., & Bizo, L. A. (2019). Use of the game-based learning platform KAHOOT! to facilitate learner engagement in animal science students. *Research in Learning Technology, 27*, Article 2225. https://doi.org/10.25304/rlt.v27.2225

Conn, C. A., Bohan, K. J., Pieper, S. L., & Musumeci, M. (2020). Validity inquiry process: Practical guidance for examining performance assessments and building a validity argument. *Studies in Educational Evaluation, 65*, Article 100843. https://doi.org/10.1016/j.stueduc.2020.100843

Dermawan, D. D., Wardani, S., & Pranoto, Y. K. S. (2023). Development using higher order thinking skills assisted by Quizizz applications in science learning. *Proceedings of International Conference on Science, Education, and Technology, 7*(1), 89–97. https://proceeding.unnes.ac.id/ISET/article/view/1940

Diana, P. A., & Zulherman. (2024). Kvisoft Flipbook Maker-assisted e-modules to foster learning motivation in elementary school students on the topic of changes in the shape of the Earth. *Jurnal Penelitian Pendidikan IPA, 10*(8), 5741–5748. https://doi.org/10.29303/jppipa.v10i8.7935

Elisa, Rambe, A., Mardiyah, A., Siregar, T. A., Roipalah, R., & Zunastri, F. (2021). Development of Quizzizz-based assessment instruments to measure student's competency in physics knowledge. *Journal of Natural Sciences, 2*(2), 72–78. https://doi.org/10.34007/jonas.v2i2.125

Faridah, U., Rahayu, Y. S., & Dewi, S. K. (2022). Pengembangan e-modul interaktif untuk melatihkan keterampilan literasi sains siswa materi transpor membran. *Berkala Ilmiah Pendidikan Biologi (BioEdu), 11*(2), 394–404. https://doi.org/10.26740/bioedu.v11n2.p394-404

Indarta, Y., Jalinus, N., Waskito, W., Samala, A. D., Riyanda, A. R., & Adi, N. H. (2022). Relevansi Kurikulum Merdeka Belajar dengan model pembelajaran abad 21 dalam perkembangan era Society 5.0. *Edukatif: Jurnal Ilmu Pendidikan, 4*(2), 3011–3024. https://doi.org/10.31004/edukatif.v4i2.2589

Irsan, I. (2021). Implemensi literasi sains dalam pembelajaran IPA di sekolah dasar. *Jurnal Basicedu, 5*(6), 5631–5639. https://doi.org/10.31004/basicedu.v5i6.1682

Khasanah, Q., Prasasti, P. A. T., & Kusumawati, N. (2022). Kelayakan e-modul berbasis T-PACK dalam memberdayakan literasi sains pada siswa kelas IV SDN Pilangbango. *Prosiding Konferensi Ilmiah Dasar, 3*. https://prosiding.unipma.ac.id/index.php/KID/article/view/3162

Liana, T., Sarong, A., & Sofyan, A. (2024). Development of Android-based e-modules in biology learning on water environmental pollution to enhance science literacy. *Jurnal Penelitian Pendidikan IPA, 10*(9), 644–654. https://doi.org/10.29303/jppipa.v10i9.8097

Mukti, T. S., Elvira, M., & Hussin, Z. B. (2023). Development of the game-based HOTS assessment instrument for measuring science literacy skills of Islamic elementary school students. *Al Ibtida: Jurnal Pendidikan Guru MI, 10*(1), 63–80.

Ningrum, K. D., Utomo, E., Marini, A., & Setiawan, B. (2022). Media komik elektronik terintegrasi augmented reality dalam pembelajaran sistem peredaran darah manusia di sekolah dasar. *Jurnal Basicedu, 6*(1), 1297–1310.

Nipyrakis, A., Stavrou, D., & Avraamidou, L. (2023). Designing technology-enhanced science experiments in elementary teacher preparation: The role of learning communities. *Research in Science & Technological Education*.

Orhan Göksün, D., & Gürsoy, G. (2019). Comparing success and engagement in gamified learning experiences via Kahoot and Quizizz. *Computers & Education, 135*, 15–29. https://doi.org/10.1016/j.compedu.2019.02.015

Parinduri, W. M., Rambe, T. R., Kesumawati, D., & Franklin, T. N. D. (2022). The development of digital module for natural sciences to improve Islamic elementary school students' learning outcomes. *MUDARRISA: Jurnal Kajian Pendidikan Islam, 14*(2), 183–204. https://doi.org/10.18326/mdr.v14i2.183-204

Prastiawan, M. D., & Utaminingsih, E. S. (2024). Development of natural science instruments on the topics of green plant assessment. *Research and Development Journal of Education, 10*(2), 1208–1220. https://doi.org/10.30998/rdje.v10i2.21941

Quispe Ccoa, N. M., Farfan Choquehuanca, M. E., & Rucano Paucar, F. H. (2023). An application of the Quizizz gamification tool to improve motivation in the evaluation of elementary school students. *International Journal of Information and Education Technology, 13*(3), 544–550. https://doi.org/10.18178/ijiet.2023.13.3.1837

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*, 77–112. https://doi.org/10.1007/s10648-019-09498-w

Sondergeld, T. A. (2019). Development and validation of a 21st century skills assessment: Using an iterative multimethod approach. *School Science and Mathematics, 119*(6), 312–326. https://doi.org/10.1111/ssm.12355

Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. Leadership Training Institute/Special Education, University of Minnesota.

Utaminingsih, E. S., & Ellianawati, E. (2025). Development of STEAM-based e-modules on human circulatory topics containing critical reasoning and independent characters. *Turkish Online Journal of Distance Education, 26*(1), 48–84.

Utaminingsih, E. S., Ellianawati, E., Sumartiningsih, S., & Puspita, M. A. (2023). STEAM education. *Jurnal Ilmiah Profesi Pendidikan, 8*(3), 1605–1612.

Utaminingsih, E. S., Ellianawati, Sumartiningsih, S., Wuriningsih, F. R., & Puspita, M. A. (2024). Development of science literacy competency evaluation instruments on human circulatory topic. *Thabiea: Journal of Natural Science Teaching, 7*(1), 56–78. https://doi.org/10.21043/thabiea.v7i1.24091

Utaminingsih, E. S., Ramadhani, M. H., & Puspita, M. A. (2024). Fostering superior characters: Development of innovative instruments for critical reasoning and independent character in the realm of science topic. *Jurnal Penelitian Pendidikan IPA, 10*(8), 4650–4661. https://doi.org/10.29303/jppipa.v10i8.7053

Wang, A. I., & Tahir, R. (2020). The effect of using Kahoot! for learning: A literature review. *Computers & Education, 149*, Article 103818.

Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review, 30*, Article 100326.