Ceddi Journal of Education
Vol. 5 No. 1 June (2026)
https://doi.org/10.56134/cje.v5i1.206

ISSN: 2830-0017 (print)
ISSN: 2830-0025 (online)

Development of ZooPedia Based on Gamification and Evaluation of Changes in Learning Outcomes of Animal Recognition in Elementary School Students

Bella Nitaliya[1*], Nabila Felisia Ikhtiaranata[2]

[1]*Information Technology Education, Aisyah University Pringsewu, Pringsewu, Indonesia*
[2]*Elementary School Teacher Education, Safin Pati University, Pati, Indonesia*

Abstract

Animal recognition lessons in elementary schools require learning resources that help students identify animals and understand the characteristics used to classify them. This study developed ZooPedia, a gamified learning application, and evaluated its functionality, content relevance, student perceptions, learning activities, and changes in test performance. The development followed the ADDIE framework, and the evaluation involved 16 sixth-grade students using a single-group pretest–posttest design. Data were collected through a 20-item test, three raters' assessments, student questionnaires, and observations across three learning sessions. All 14 reported functional testing scenarios produced the expected outputs. Mean test scores increased from 49.06 to 69.06, a difference of 20.00 points (95% CI: 10.42–29.58; $p < 0.001$). Scores increased for 13 students and decreased for three. Aiken's V coefficients ranged from 0.444 to 1.000, identifying instructional clarity and gamification mechanisms as improvement priorities. Mean observed learning activity did not increase across sessions. The study contributes empirical evidence by distinguishing changes in test performance from student perceptions and observed learning activities, highlighting the value of evaluating gamified media through multiple indicators. However, measurement limitations and the absence of a comparison group prevent causal attribution of the observed changes to ZooPedia.

Keywords: Learning assessment; Gamification; Learning media; Animal recognition; Elementary school

Received: 12 February 2026 Revised: 24 March 2026 Accepted: 23 June 2026

Introduction

Animal recognition in elementary school science learning includes recognizing examples, comparing characteristics, and explaining the basis for grouping. This learning needs to connect visual recognition with an understanding of organism characteristics. (Hooykaas et al., 2022) found that the ability to identify species relates to knowledge of habitat, diet, origin, and behavior, but correct identification does not always reflect in-depth knowledge. Although the study involved adults, the findings provide a conceptual basis for distinguishing the ability to recognize animals from an understanding of their characteristics. In elementary school learning, this distinction matters so practice is not limited to memorizing names or pictures but also guides students to the reasoning behind groupings.

At SD Negeri 1 Pringsewu Barat, animal identification lessons use printed textbooks and teacher explanations as learning resources. Development needs focused on providing supporting media that connect material, identification exercises, and feedback within structured activities. These media give students opportunities to use information about animal characteristics to answer questions and review their understanding. Based on these needs, ZooPedia was developed as a complementary learning resource and a set of teacher-guided learning activities.

Gamification uses game elements in non-game contexts, including learning. Elements such as scores, progress, and badges can organize activities and mark achievements. This use differs from game-based learning, which utilizes games as the primary learning activity or environment. (Zainuddin et al., 2020) demonstrate the diversity of mechanisms, approaches, and outcome indicators in educational gamification research. (Krath et al., 2021) also identify diverse theoretical foundations in research on gamification, serious games, and game-based learning. Therefore, using game elements requires explaining their pedagogical function and how they relate to learning objectives.

This alignment is a concern in primary and secondary education. (Dehghanzadeh et al., 2024) offer a framework for gamified learning that considers individual factors, the learning environment, the learning process, and outcomes. This framework suggests that gamification's benefits depend on how design components connect. In ZooPedia, materials describe animal characteristics, quizzes provide identification practice, and feedback helps students review their answers. Meanwhile, scores, progress, and badges show outcomes and activity completion. For example, an exercise

*Corresponding author.
E-mail address: bellanitaliya16@gmail.com (Bella Nitaliya)

© [2026] The Author(s). This articleis licensed under a
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

on aquatic mammals can draw attention to the biological characteristics underlying classification, so that habitat is not the sole basis for answers.

The presentation of material and exercises should also consider the student's learning process. (Mayer & Moreno, 2003) explain the importance of managing cognitive load in multimedia learning by presenting information appropriate to the learner's processing capacity. Furthermore, (Agarwal et al., 2021) show that information retrieval practice can support learning in a variety of classroom contexts. (Shute, 2008) emphasizes the role of formative feedback in fostering improved thinking or learning behavior. These three studies informed ZooPedia's design, which links animal characteristic presentation, answer practice, and answer review, although it has not tested the contribution of each component separately.

Synthesized evidence demonstrates the potential of gamification and the variability in its implementation outcomes. (Bai et al., 2020) found that gamification benefits learning outcomes, while (Aldalur & Perez, 2023) distinguished its effects on cognitive, motivational, and behavioral outcomes. (M. Li et al., 2023) also showed that effect sizes varied by participant characteristics, learning domain, design principles, implementation duration, and learning environment. These findings support the development of gamified media, but the success of specific applications still needs to be evaluated based on the conditions of use and the evaluation indicators used.

Evaluations also need to distinguish between test performance, user perceptions, and learning activities. (L. Li et al., 2024) showed that gamification's effects on intrinsic motivation and perceptions of autonomy, connectedness, and competence were not uniform. Ratings of an app's appeal or ease of use reflect the user experience, while test scores indicate performance on the measured tasks. Activity observations capture behaviors observed during learning. Thus, these three indicators can complement each other without being treated as substitutes.

After reviewing 54 empirical studies in primary and secondary education, (Dehghanzadeh et al., 2024) identified limitations in the consistent application and evaluation of gamified learning dimensions. (Yusri & Zainal, 2025) further examined gamification implementation in primary education through a systematic literature review. Their findings highlight gamification's potential to support learning outcomes and student engagement while emphasizing the need for further investigation into its sustained educational benefits. These findings reinforce the importance of evaluating gamified learning applications using multiple indicators rather than relying solely on test-score improvements. This gap underscores the need for research that clarifies relationships among learning objectives, media features, student activities, and collected evidence. The ZooPedia study addresses this need by developing and initially evaluating an animal recognition learning module. The contribution lies in aligning material on animal characteristics, identification exercises, and feedback with evaluation that differentiates changes in test scores, user perceptions, and learning activities to inform media refinement.

This study describes the development of gamification-based ZooPedia and evaluates the suitability of the application's functions, the relevance of instruments and media, user perceptions, learning activities, and changes in animal recognition scores. The research questions include how the application's features align with the learning objectives, which aspects require improvement based on the product evaluation, and how student scores and activity patterns changed during the evaluation. As a developmental study with a single-group pretest–posttest design, the results are interpreted by considering measurement quality, sample size, and the limitations of drawing causal conclusions.

Method

Research Design

This study used a research and development approach with the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) framework, as described by (Spatioti et al., 2022). The learning evaluation used a single-group pretest-posttest design without a comparison group to measure changes in scores before and after ZooPedia use. This design allows examination of within-group changes but limits causal inferences (Reynolds, 2022). The study was conducted at SD Negeri 1 Pringsewu Barat from April to June 2025 and involved all 16 sixth-grade students. The number of participants was based on the scope of the class being evaluated, in accordance with one of the participant size justification approaches discussed by (Lakens, 2022) . All students participated in two measurements and three learning sessions. Data were collected through animal recognition tests, instrument and media relevance assessments, observation of learning activities, application function testing, and user questionnaires.

Media Development and Implementation

The analysis phase covers material needs, student characteristics, learning resources, and device-use conditions. The design phase includes material development, navigation, quizzes, feedback, scoring, progress tracking, and badges.

Materials are organized into pedagogical categories: mammals, birds, reptiles, amphibians, fish, and insects, for feature recognition rather than as fully equivalent taxonomic groups. The development phase produces a learning menu, material search, quizzes, quiz results, and achievements. The alignment of features, learning objectives, and evaluation indicators follows the gamified learning framework (Dehghanzadeh et al., 2024). Functional testing compares actual output with expected output across 14 scenarios. The results are analyzed separately from the educational evaluation.

The pretest was conducted on April 15, 2025, and the posttest on June 21, 2025, each lasting 26 minutes. Learning took place in three sessions, each lasting 37 minutes, covering mammals and birds, reptiles and amphibians, and fish and insects. Students used 16 smartphones individually. Bella Nitaliya acted as learning facilitator, test proctor, and observer. The researchers provided instructions at the beginning of the activity. During the test, assistance was limited to technical instructions, not guidance on answers. The study obtained school permission, and participant identities were disguised using codes.

Measurement Instruments and Quality

The animal recognition test consists of 20 items with dichotomous scoring: correct answers are scored 1 and incorrect answers are scored 0. The 0–100 scale score is obtained by multiplying the number of correct answers by 5. The pretest and posttest use the same test structure. The test grid is presented in Table 1.

Table 1. Animal recognition test grid

Material	Question number	Number of Items
Mammals	1–3	3
Birds	4–6	3
Reptiles	7–9	3
Amphibians	10–12	3
Fish	13–15	3
Insects	16–18	3
Cross-group implementation	19–20	2
Total	1–20	20

The relevance assessment covers ten aspects of the instrument and media as listed in Table 4, including the suitability of objectives and questions, conceptual accuracy, language clarity, answer key accuracy, and distractor suitability. Three raters, consisting of two assessors of science materials and one assessor of learning media, assigned scores of 1–4 to each aspect. We summarized the assessments using Aiken's V, following the methodological discussion in (Merino-Soto, 2023) in equation 1.

$$V = \frac{\sum_{i=1}^{n}(r_i - l_o)}{n(c-1)} \quad (1)$$

In the equation, (r_i) represents the rating assigned by the (i)-th rater, (l_o) denotes the lowest possible rating on the assessment scale, (n) is the number of raters, and (c) represents the number of rating categories. With three raters and a scale of 1–4, the denominator is 9. The V-value is used descriptively to identify priorities for improvement. Empirical item analysis and test reliability estimates have not been reported. Observations included attention to task, participation, activity completion, and use of feedback. Scores of 1–4 indicate behavior not present despite opportunities (1), present with repeated directions (2), present with little direction (3), and appearing independent on most occasions (4). The observations yielded 48 participant-session recordings with 192 indicator scores. Because only one observer was involved, we did not calculate inter-rater agreement. Furthermore, the questionnaire included five positive statements on a 1–4 scale about ease of use, interface, materials, interactivity and gamification, and perceived usefulness. All students responded to all five statements. User perceptions varied by test performance, learning activities, and intrinsic motivation (L. Li et al., 2024).

Data analysis

Data examination included summing responses, calculating scores, and matching pairs based on codes P01–P16. All complete pairs were analyzed without excluding low scores or negative changes. Descriptive statistics included the mean, sample standard deviation, range, and direction of individual change. Differences were calculated as posttest scores minus pretest scores. The primary analysis used a two-tailed paired t-test at a significance level of 0.05, with 95% confidence intervals for the mean difference and Cohen's (dz). We calculated effect sizes as the mean difference divided by the standard deviation of the difference, following the approach of standardizing difference scores in repeated measures (Langenberg et al., 2023). We examined the distribution of differences using the Shapiro–Wilk test. For sensitivity analysis, we used a two-tailed Wilcoxon signed-rank test with an asymptotic approach, without continuity correction, and accounting for equal absolute differences. There were no zero differences.

Calculations were performed using Python and SciPy. We summarized relevance assessments by aspect using the questionnaire mean, sample standard deviation, and range, and by indicator and session for observations. Repeated observations of the same student were not treated as independent units. Interpretation focused on the pattern of results without assigning an overall media suitability category.

Results and Discussion

Results

Product and Functional Testing

The development of ZooPedia resulted in an application that connects animal introduction materials, quizzes, feedback, and progress and achievement markers. The materials cover mammals, birds, reptiles, amphibians, fish, and insects. Table 2 summarizes the relationship between the application components and learning activities.

Table 2. ZooPedia components and their learning functions

Component	Functions in learning design
Animal material	Presents examples and characteristics of animals as a basis for recognition and comparison.
Quiz	Provides practice using information about animal characteristics to answer questions.
Feedback	Supports review of answers after quizzes.
Scores and progress	Displays the results of work and activities that have been completed.
Badges	Indicates achievements within the application

Table 2 shows that the app design places gamification elements alongside the materials and exercises. These components reflect the product design, while the educational benefits are examined through different evaluation indicators. The functional test report indicates that all 14 scenarios produced expected outputs. These results are limited to the reported scenarios because detailed documentation regarding inputs, actual outputs, implementation, and testing time is not yet available. Therefore, the test results cannot be used to conclude that the app is error-free or effective in enhancing learning. Figure 1 shows the ZooPedia interface design, which links material presentation, identification exercises, concept feedback, and progress and badges.

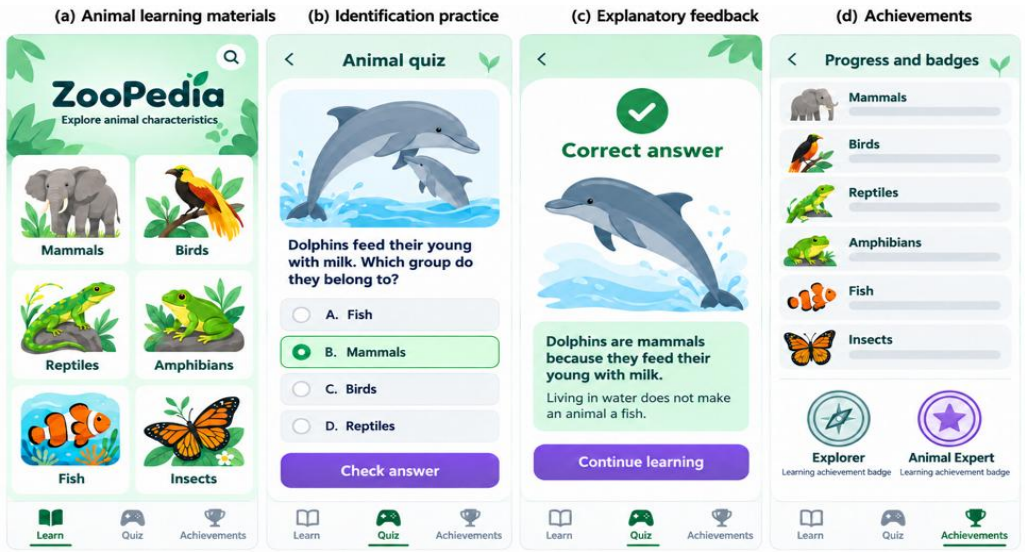


Figure 1. ZooPedia interface

Panel (a) presents the categories of animal material, while panel (b) exemplifies an exercise that uses organism characteristics as the basis for answering. Panel (c) explains that dolphins are mammals because they nurse their young, so classification is not based solely on aquatic habitats. Panel (d) depicts progress and badges as markers of achievement. This view shows ZooPedia's feature design. The following section reports on participants' learning outcomes and activities, evaluated through tests and observations.

User Perception

The questionnaire summary covers five aspects rated by 16 users on a scale of 1–4. The author's report presents the results in Table 3.

Table 3. Summary of user ratings of ZooPedia

Aspect	Average	Standard deviation	Range
Ease of use	3.75	0.447	3–4
Application Interface	3.88	0.342	3–4
Presentation of material	3.38	0.719	2–4
Interactivity and gamification	3.25	0.856	2–4
Perceived Learning Benefits	3.81	0.403	3–4

The application's appearance received the highest average, while interactivity and gamification received the lowest. Interactivity and gamification also had the largest standard deviations, indicating more mixed assessments than the other aspects in the summary. These results reflect user perceptions and should not be considered evidence of changes in knowledge or intrinsic motivation. Questionnaire results are interpreted at the aspect level without assigning overall eligibility categories. Evaluation includes assessing the relevance of 10 aspects of instruments and media. Table 4 presents the three assessors' scores and the Aiken's V coefficient.

Table 4. Relevance assessment scores and Aiken's V coefficient

Aspect	Rater 1	Rater 2	Rater 3	V
Suitability of objectives and questions	4	3	3	0.778
The accuracy of the animal concept	4	3	4	0.889
Clarity of children's language	4	3	3	0.778
Accuracy of answer key	4	3	3	0.778
Conformity of the distractor	4	2	3	0.667
Suitability of application content	4	4	4	1.000
Appropriateness of gamification mechanisms	2	2	4	0.556
Visual readability	4	4	4	1.000
Clarity of instructions	2	3	2	0.444
Conformity of measurement to objectives	3	3	4	0.778

The coefficients in Table 4 ranged between 0.444 and 1.000, with an average of 0.767. Clarity of instructions had the lowest coefficient, followed by the suitability of gamification mechanisms and distractors. This pattern indicates which aspects to prioritize in further analysis. Table 5 presents descriptive statistics for pretest and posttest scores and the difference between the two.

Table 5. Descriptive statistics of test scores on 16 data pairs

Statistic	Pretest	Posttest	Difference
Number of Paired Observations	16	16	16
Mean	49.06	69.06	20.00
Standard deviation	19.08	15.83	17.98
Minimum	5	35	−10
Maximum	80	90	45

Notes: The difference is calculated as the posttest score minus the pretest score. The standard deviation is calculated using the denominator (n-1), with n = 16.

Table 5 shows an average increase of 20.00 points on a 0-100 scale. Changes were not uniform: 13 participants' scores increased, 3 decreased, and no pairs had a difference of zero. All pairs were retained in the analysis. Table 6 presents the distribution of differences, score differences, and effect sizes.

Table 6. Analysis of paired score differences

Analysis	Results
Shapiro–Wilk on the difference	W = 0.946; p = 0.430
Two-tailed paired t-test	t(15) = 4.449; p = 0.000469
95% confidence interval for the mean difference	10.42–29.58 points
Cohen's dz	1.112
Wilcoxon two-way signed-rank	W = 9.000; p = 0.002240

The Shapiro–Wilk test did not indicate a statistically significant departure from normality in the paired score differences (W = 0.946, p = 0.430). However, this result does not establish that the differences follow a normal distribution. Paired t-tests showed significant differences in means, with confidence intervals all above zero. Wilcoxon sensitivity analyses also showed significant differences. Both tests used the same pairs and therefore do not constitute confirmation of independence. These results indicate differences in scores between the two measures but do not explain ZooPedia's specific effect because this study did not use a comparison group. Furthermore, learning activities through observation

included attention to task, participation, activity completion, and use of feedback over three sessions. A summary of the scores is presented in Table 7.

Table 7. Average activity observation scores on a scale of 1–4

Session	Attention	Participation	Completion of activities	Use of feedback	Overall
1	2.375	2.625	2.625	2.563	2.547
2	2.375	2.625	2.438	2.500	2.484
3	2.313	2.375	2.563	2.688	2.484

Table 7 does not show an overall average increase between sessions. Changes in certain indicators were also not uniform. Feedback usage had the highest average in the third session, while attention and participation were lower than in the first session. This summary is descriptive and does not show changes compared to learning before using the app because no baseline observations were available for comparison.

Discussion

The Meaning of Score Changes and Differences in Evaluation Indicators

The main finding of this study was a mean increase of 20.00 points in animal recognition scores after the ZooPedia use period, with a 95% confidence interval of 10.42–29.58 points. This change indicates improved performance on the test, but it was uneven: 13 students increased their scores, while 3 decreased. Thus, interpret group-level results alongside individual variation. Educationally, this pattern points to two needs: maintaining learning-supportive activities and identifying struggling students who have not yet improved.

The direction of change aligns with (Langenberg et al., 2023; M. Li et al., 2023) meta-analysis, which found benefits of gamification on learning outcomes, with effects varying by participant characteristics, learning area, design, duration, and implementation environment. (Zeng et al., 2024) also reported a positive influence on academic achievement through a synthesis of 22 experimental studies with Hedges's g of 0.782. Both studies place ZooPedia's results in the context of evidence supporting gamification's potential. However, the similar direction of the results does not indicate equal effect sizes because this study measured changes within one group, whereas Zeng et al. compared gamified and non-gamified learning.

These design differences are also important in interpreting Cohen's dz of 1.112. This measure indicates the magnitude of the average change relative to the standard deviation of individual changes. This within-group effect size cannot be directly compared with between-group effect sizes to determine the relative effectiveness of ZooPedia. As (Langenberg et al., 2023) explain, standardizing effect sizes across repeated measures depends on the data structure and variability used. Therefore, consider differences in points, confidence intervals, and individual changes alongside the effect size.

As an empirical comparison, (Kaya & Ercag, 2023) found improvements in academic achievement and motivation through a challenge-based gamification program involving 60 students in experimental and control groups over a semester. These findings demonstrate the potential of structured gamified activities, but differences in educational level, content, duration, and evaluation design limit their application in ZooPedia. This research contributes preliminary evidence for animal recognition learning in a single elementary school classroom.

Score increases may also be related to opportunities to practice information retrieval. (Agarwal et al., 2021) show that information retrieval practice and quizzes can support classroom learning. In ZooPedia, students not only learn about animal characteristics but also use that information to answer questions. These activities may help explain changes in performance. However, practice cannot be separated from materials, feedback, mentoring, and rewards, as these components work together as a single learning component. The educational significance of score increases still needs to be distinguished from mastery. Confidence intervals indicate the accuracy of the estimated mean change, while mastery requires criteria established before measurement. Without that basis, the study results support conclusions about changes in test performance, not claims that all students have mastered the material.

Activity Test Performance and User Perception

The increase in average scores was not accompanied by an increase in average observed activity between sessions. This discrepancy indicates that ZooPedia's evaluation results do not show a uniform pattern of improvement across all indicators. The test records performance on measurement items, while the observation records attention, participation, activity completion, and use of feedback during learning. These two measures provide different information, so a lack of increased activity does not negate score changes, but it does limit the interpretation that app use improves all aspects of learning.

(Kaya & Ercag, 2023) also reported a different pattern of outcomes. In their study, improvements in academic achievement and motivation were not accompanied by significant increases in flow. While the path differs from the ZooPedia observational indicators, these findings suggest that the benefits of gamification may depend on the construct being evaluated. Consequently, media evaluations should maintain indicators that do not increase so that assessments of benefits do not rely solely on favorable outcomes. A similar distinction applies to user perceptions. (L. Li et al., 2024) found that gamification's effects on intrinsic motivation and perceptions of autonomy, connectedness, and competence were not uniform. (Howard et al., 2021) also showed that different motivation types related differently to educational outcomes. In this study, the questionnaire assessed ease of use, appearance, content, interactivity, and perceived usefulness. Therefore, positive responses are better interpreted as assessments of the user experience than as evidence of increased intrinsic motivation or conceptual mastery.

Reporting three-score declines also broadens the assessment of variation in media-use outcomes. (Almeida et al., 2023) identified unintended impacts of implementing gamification in education, including motivational issues, decreased performance, and irrelevant game elements. This study supports the importance of examining negative outcomes but does not establish the cause of the score declines. To explain this, future evaluations should link individual changes to incorrect answers, need for assistance, and students' experiences using the app.

Implications for Improving Learning Design

Relevance and user perception assessments yielded similar improvement trends. Clarity of instructions and gamification mechanisms scored relatively low in relevance assessments, while interactivity and gamification scored the lowest on average in the questionnaire. This similarity in direction reinforces the rationale for examining how students understand rules, perform activities, and interpret their achievements. However, because the two instruments assess different aspects, this pattern indicates development priorities, not evidence that both measures confirm a single weakness.

(Dehghanzadeh et al., 2024) emphasize the interrelationship between individual, environmental, process, and outcome factors in gamified learning. Based on this framework, ZooPedia needs to link instructions for use to expected learning actions. Instructions can explain the sequence of activities, how to answer, the meaning of the score, and the steps to take after receiving feedback. Badging rules also need to distinguish completion of activities from achievement of mastery criteria so that the meaning of the award is consistent with the evidence generated by the application. Refining feedback is the next priority. (Mertens et al., 2022) showed that elaborative feedback better supports lower-level learning outcomes and transfer than other feedback types. In ZooPedia, this principle can be applied by briefly explaining the reasons for answers and the distinguishing characteristics between animal groups. For example, after students group dolphins, feedback can highlight relevant mammalian characteristics rather than simply stating whether the answer is correct or incorrect.

The relationship between features and the learning process also needs exploration. In a lecture context, (Sailer & Sailer, 2021) showed that gamification can affect application-oriented knowledge and performance during learning. For ZooPedia, these findings provide a basis for recording whether students read explanations, review errors, and utilize feedback before continuing with an activity. This recording can help clarify how features are used and supplement information about their existence and function. Variations in score changes also highlight the importance of examining user needs. (Oliveira et al., 2023) note the potential to tailor gamification to student characteristics, while also highlighting limited evidence on the benefits of personalization. Therefore, ZooPedia developers could first test incremental assistance, navigation clarity, and exercise difficulty based on observed challenges. Adjustments should reflect identified learning needs, not the assumption that adding game features will improve outcomes.

Understanding the Concept and Contribution of Education

In animal recognition learning, successful identification must be linked to an understanding of the organism's characteristics. (Hooykaas et al., 2022) found that identification skills are related to species knowledge, but correct identification does not always reflect in-depth knowledge. Although involving adults, the study helps clarify the distinction between recognizing specimens and understanding the basis for their classification. This suggests that ZooPedia's exercise development should give students space to explain their answers and compare examples. Future learning activities could incorporate animals with similar habitats to assess whether students classify organisms based on biological characteristics rather than environmental similarities alone. New images and examples are also needed to assess whether students can apply concepts beyond the exercises already completed. Further evaluation should target students' ability to explain their reasoning and apply these concepts, as this study did not specifically measure these skills. The sustainability of learning outcomes also needs examination. (Carpenter et al., 2022) described the role of spaced learning and information retrieval practice in supporting long-term learning. Based on this study, future developments could combine spaced practice with delayed testing. Such an approach would extend the evaluation

beyond immediate post-intervention performance by examining knowledge retention and the application of learned concepts.

This research contributes ZooPedia, which links animal characteristics, identification exercises, feedback, and achievement markers to multidimensional evaluation. The increase in average scores without a corresponding increase in activity suggests that the results should be interpreted through the measured indicators. Meanwhile, the relevance assessment and user perceptions serve as a basis for prioritizing improvements to the instructions and gamification mechanisms. This research provides contextual evidence for using concurrent product and instructional evaluation to enhance animal recognition media in elementary schools. For teachers, ZooPedia can serve as a supporting tool, paired with oral explanations, comparison examples, and checks of students' reasoning. Using the application requires careful consideration of students' reading skills, device experience, and the support they need. The teacher's role remains crucial in linking digital activity completion to understanding the material.

Limitations and Directions for Further Research

The single-group pretest–posttest design limits conclusions regarding the specific effects of ZooPedia. Additional learning during April–June 2025, mentoring, and repeated testing may have contributed to the score changes. (Reynolds, 2022) explains that factors other than the intervention could influence these design changes. The use of the same test structure and the lack of reported empirical item and reliability analyses also limit the strength of the measurement's interpretation. Including all 16 students provided comprehensive coverage of the evaluated class but did not guarantee broader representation of elementary school students. As (Lakens, 2022) discussed, defined population coverage should be differentiated based on the need for estimation accuracy and the goal of generalization. Furthermore, a single observer who also served as a test supervisor and proctor raised concerns about potential bias. This study also did not include retention tests or tasks that applied concepts to new examples.

Future research should use comparison groups with equivalent materials, learning time, and practice opportunities. Involving multiple schools, strengthening instruments, adding observers, and measuring retention and concept application would broaden the evaluation framework. Comparisons across feature designs are also needed to distinguish the contribution of feedback and rewards to the benefits of materials and practice. With such designs, further research can more robustly explain not only whether scores change, but also under what conditions and through what learning processes those changes occur.

Conclusions and Suggestions

Conclusions

ZooPedia is an animal recognition tool that integrates animal characteristics, identification exercises, feedback, progress, and badges. Evaluation of 16 students showed a statistically significant average score increase of 20.00 points, but individual changes varied, and average activity between sessions did not increase. Relevance and user perceptions guided improvement priorities for instructional clarity and gamification mechanisms. This study contributes by linking instructional design to evaluation of test performance, user perceptions, and learning activities. Differentiating these indicators provides a basis for assessing the results of media use and determining improvement priorities. Findings support improved test performance following ZooPedia use in the evaluated classes. However, measurement limitations, the scope of a single class, and the lack of a comparison group limit generalization and attribution of changes to the application or gamification elements.

Suggestions

ZooPedia improvements should prioritize clear instructions, feedback that explains the rationale for answers, and badge rules that distinguish between activity completion and mastery of the material. Teachers can use the app as a support tool by asking students to explain their grouping rationale and apply that understanding to examples of animals different from those in the exercises. Future research should strengthen the evidence base by using comparison groups with equivalent materials and learning time, adding retention measures, and including tasks that apply the concepts. Involving multiple schools and additional observers could broaden the scope and strengthen the evaluation's consistency. The number of participants should be justified based on the analysis objectives and the accuracy of the estimates, while comparisons across feature designs are needed to examine the contribution of feedback and rewards to learning outcomes.

Acknowledgements

The author thanks SD Negeri 1 Pringsewu Barat for permission to carry out the activity and the assessors of learning materials and media for their input in the ZooPedia evaluation.

Funding Statement

This research did not receive external funding.

Author Contributions

Bella Nitaliya contributed to methodology development, application development, instrument development, learning assistance, test supervision, activity observation, data collection and analysis, and manuscript drafting. Nabila Felisia Ikhtiaranata contributed to reviewing the basic education content, interpreting the research results, and reviewing and editing the manuscript.

Research Ethics and Informed Consent

The activity was implemented with permission from SD Negeri 1 Pringsewu Barat. Participation was voluntary. Participants' identities were disguised in the analysis files using a code to maintain confidentiality.

Conflict of Interest

The author declares no conflict of interest.

Data Availability Statement

The data used in the analysis are available in Microsoft Excel format (.xlsx) and include test item responses, paired pretest-posttest scores, a relevance assessment matrix, and coded observations. Access requests may be submitted to the corresponding author for consideration with the school. Access is limited to data that can be shared in accordance with participant consent and privacy protection provisions.

Declaration of Generative AI Use

The author used Google Translate, Gemini, and ChatGPT to assist with language editing, narrative organization, and drafting revisions. All output has been reviewed and edited by the author, including for appropriate citations and references. The author is responsible for the accuracy, originality, and integrity of the entire manuscript.

References

Agarwal, P. K., Nunes, L. D., & Blunt, J. R. (2021). Retrieval Practice Consistently Benefits Student Learning: a Systematic Review of Applied Research in Schools and Classrooms. In *Educational Psychology Review* (Vol. 33, Number 4). https://doi.org/10.1007/s10648-021-09595-9

Aldalur, I., & Perez, A. (2023). Gamification and discovery learning: Motivating and involving students in the learning process. *Heliyon*, *9*(1), e13135. https://doi.org/https://doi.org/10.1016/j.heliyon.2023.e13135

Almeida, C., Kalinowski, M., Uchôa, A., & Feijó, B. (2023). Negative effects of gamification in education software: Systematic mapping and practitioner perceptions. *Information and Software Technology*, *156*. https://doi.org/10.1016/j.infsof.2022.107142

Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review*, *30*(December 2019), 100322. https://doi.org/10.1016/j.edurev.2020.100322

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. In *Nature Reviews Psychology* (Vol. 1, Number 9). https://doi.org/10.1038/s44159-022-00089-1

Dehghanzadeh, H., Farrokhnia, M., Dehghanzadeh, H., Taghipour, K., & Noroozi, O. (2024). Using gamification to support learning in K-12 education: A systematic literature review. In *British Journal of Educational Technology* (Vol. 55, Number 1). https://doi.org/10.1111/bjet.13335

Hooykaas, M. J. D., Schilthuizen, M., Albers, C. J., & Smeets, I. (2022). Species identification skills predict in-depth knowledge about species. *PLoS ONE*, *17*(4 April). https://doi.org/10.1371/journal.pone.0266972

Howard, J. L., Bureau, J., Guay, F., Chong, J. X. Y., & Ryan, R. M. (2021). Student Motivation and Associated Outcomes: A Meta-Analysis From Self-Determination Theory. *Perspectives on Psychological Science*, *16*(6). https://doi.org/10.1177/1745691620966789

Kaya, O. S., & Ercag, E. (2023). The impact of applying challenge-based gamification program on students' learning outcomes: Academic achievement, motivation and flow. *Education and Information Technologies*, 1–26. https://doi.org/https://doi.org/10.1007/s10639-023-11585-z

Krath, J., Schürmann, L., & von Korflesch, H. F. O. (2021). Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Computers in Human Behavior*, *125*. https://doi.org/10.1016/j.chb.2021.106963

Lakens, D. (2022). Sample Size Justification. In *Collabra: Psychology* (Vol. 8, Number 1). https://doi.org/10.1525/collabra.33267

Langenberg, B., Janczyk, M., Koob, V., Kliegl, R., & Mayer, A. (2023). A tutorial on using the paired t test for power calculations in repeated measures ANOVA with interactions. *Behavior Research Methods*, *55*(5). https://doi.org/10.3758/s13428-022-01902-8

Li, L., Hew, K. F., & Du, J. (2024). Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: a meta-analysis and systematic review. *Educational Technology Research and Development*, *72*(2). https://doi.org/10.1007/s11423-023-10337-7

Li, M., Ma, S., & Shi, Y. (2023). Examining the effectiveness of gamification as a tool promoting teaching and learning in educational settings: a meta-analysis. *Frontiers in Psychology*, *14*, 1253549. https://doi.org/10.3389/fpsyg.2023.1253549

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1). https://doi.org/10.1207/S15326985EP3801_6

Merino-Soto, C. (2023). Coeficientes V de Aiken: diferencias en los juicios de validez de contenido. *MHSalud: Revista En Ciencias Del Movimiento Humano y Salud*, *20*(1). https://doi.org/10.15359/mhs.20-1.3

Mertens, U., Finn, B., & Lindner, M. A. (2022). Effects of Computer-Based Feedback on Lower and Higher-Order Learning Outcomes: A Network Meta-Analysis. *Journal of Educational Psychology*, *114*(8). https://doi.org/10.1037/edu0000764

Oliveira, W., Hamari, J., Shi, L., Toda, A. M., Rodrigues, L., Palomino, P. T., & Isotani, S. (2023). Tailored gamification in education: A literature review and future agenda. *Education and Information Technologies*, *28*(1). https://doi.org/10.1007/s10639-022-11122-4

Reynolds, J. J. (2022). Improving the Assessment of Teaching Effectiveness With the Nonequivalent Dependent Variables Approach. *Teaching of Psychology*, *49*(4). https://doi.org/10.1177/00986283211022390

Sailer, M., & Sailer, M. (2021). Gamification of in-class activities in flipped classroom lectures. *British Journal of Educational Technology*, *52*(1). https://doi.org/10.1111/bjet.12948

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1). https://doi.org/10.3102/0034654307313795

Spatioti, A. G., Kazanidis, I., & Pange, J. (2022). A Comparative Study of the ADDIE Instructional Design Model in Distance Education. In *Information* (Vol. 13, Number 9, p. 402). https://doi.org/10.3390/info13090402

Yusri, A. A., & Zainal, M. Z. (2025). Unleashing gamification: a systematic review in primary schools. *Journal of Education and Learning*, *19*(4). https://doi.org/10.11591/edulearn.v19i4.22009

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*(March 2019). https://doi.org/10.1016/j.edurev.2020.100326

Zeng, J., Sun, D., Looi, C. K., & Fan, A. C. W. (2024). Exploring the impact of gamification on students' academic performance: A comprehensive meta-analysis of studies from the year 2008 to 2023. In *British Journal of Educational Technology* (Vol. 55, Number 6). https://doi.org/10.1111/bjet.13471