© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

AL-LISAN: JURNAL BAHASA
Publisher: LPPM IAIN Sultan Amai Gorontalo
ISSN: 2442-8965 E-ISSN: 2442-8973
Volume 11, No. 2 August 2026
Journal Homepage: https://journal.iaingorontalo.ac.id/index.php/al

Deep Learning Approach-Based Gamification to Improve English Vocabulary Mastery among Fifth-Grade Elementary School Students

[1]Ainsha Ainun Habibie
(Corresponding Author)
ainsha.habibie05@gmail.com
Primary Teacher Education Department, Faculty of Education, Universitas Negeri Gorontalo, Indonesia

[2]Karmila Machmud
karmila@ung.ac.id
English Education Department, Faculty of Letters and Culture, Universitas Negeri Gorontalo, Indonesia

[3]Wiwy Triyanty Pulukadang
wiwy_pulukadang@ung.ac.id
Primary Teacher Education Department, Faculty of Education, Universitas Negeri Gorontalo, Indonesia

ABSTRACT

Background: English is now mandatory in Indonesian elementary schools, but vocabulary is still often taught through memorized lists from textbooks. Meanwhile, the curriculum emphasizes deep learning, yet interactive digital media that support both engagement and understanding remain scarce at this level.

Aims: This study developed gamification media grounded in the deep learning approach, and examined their validity, practicality, and effectiveness for improving mastery of Parts of the Body vocabulary among fifth-grade elementary school students.

Methods: Research and Development (R&D) was conducted using the 4D model of Thiagarajan, Semmel, and Semmel (1974). Participants were four experts and 26 fifth-grade students of SDN 5 Limboto, Gorontalo, Indonesia. Effectiveness was examined through a one-group pre-test-post-test design, using expert validation sheets, a student response questionnaire, and a vocabulary achievement test.

Results: The media were rated highly valid, with validation scores of 87.06% (media expert), 87.37% (content expert), and 86.76% (deep learning teaching-module experts), giving an overall mean of 87.06%. Student responses placed the media in the highly practical category, and observed active participation averaged 87.94%. The class mean rose from 42.31 on the pre-test to 78.85 on the post-test, a gain of 36.54 points, with a normalised gain of 0.63 in the medium category; 24 of 26 students (92.31%) reached the minimum criterion of learning goals.

Implications: Six web platforms (Baamboozle, Educaplay, Blooket, BookWidgets, Kahoot, and Canva) arranged across three phases offer primary English teachers a workable model in which game elements are assigned to distinct instructional functions rather than applied uniformly.

Keywords: Deep learning approach; gamification; vocabulary mastery; learning media; primary English; 4D model

Article Info:
Received: 16 April 2026
Accepted: 14 August 2026
Published: 26 August 2026

How to cite:
Habibie, A. A., Machmud, K., & Pulukadang, W. T. (2026). Deep learning approach-based gamification to improve English vocabulary mastery among fifth-grade elementary school students. *Al-Lisan: Jurnal Bahasa (e-Journal), 11*(2), 206-220. https://doi.org/10.30603/al.v11i2.7777

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

1. INTRODUCTION

English has become a compulsory subject in Indonesian elementary schools, yet classroom practice has changed more slowly than the policy framing around it. Instruction at this level still leans heavily on the textbook, and vocabulary is commonly presented as lists to be memorised and rehearsed. The current curriculum direction, which places deep learning (*pembelajaran mendalam*) at the centre of classroom practice, asks for something different: learning experiences that are meaningful and engaging and that leave students with usable understanding rather than temporary recall (Fullan et al., 2018). It is worth stating at the outset that deep learning in this article refers to the qualitative approach to learning described by Marton and Saljo (1976) and later elaborated by Fullan and colleagues, and not to the family of artificial neural network techniques that carries the same name.

Fifth-grade students sit at the concrete operational stage of cognitive development, where understanding is built from concrete, enjoyable, and personally meaningful experience rather than from abstract rehearsal (Bada, 2015). Vocabulary knowledge, in particular, is not a single act of recall but a gradual accumulation of form, meaning, and use, and it grows through repeated meaningful encounters with words across varied tasks (Nation, 2013). Learning media matter here precisely because they shape how many such encounters a young learner gets, and in what form (Arsyad, 2020; Suryani et al., 2020).

Gamification, understood as the use of game design elements in non-game contexts to support engagement and motivation (Deterding et al., 2011; Kapp, 2012), is the response to this situation that has attracted the most sustained attention. Meta-analytic evidence indicates a small to moderate positive effect on cognitive, motivational, and behavioural learning outcomes, with effects varying considerably by design and context (Bai et al., 2020; Huang et al., 2020; Sailer & Homner, 2020). Reviews in language education report similar patterns for vocabulary learning specifically (Dehghanzadeh et al., 2021; Zainuddin et al., 2020; Zhang & Hasim, 2023).

The same body of work also reports a design ceiling. Zhang and Hasim (2023) found that the elements most frequently deployed in EFL and ESL gamification are feedback, points, quizzes, badges, and leaderboards, a cluster oriented almost entirely toward extrinsic incentive. Shortt et al. (2023) concluded from the Duolingo literature that once novelty fades, competition and repetition cannot compensate for design decisions that neglect collaboration and meaningful context. Hanus and Fox (2015) documented the sharper version of the problem, in which scoring structures eroded intrinsic motivation over the course of a semester. Among children the risk is not merely dilution: Li and Chu (2021) found that leaderboards and competition discouraged some primary readers rather than drawing them in.

The question this raises is not whether gamification works, but whether game elements can be arranged so that the learning they support is deep rather than merely busy. That is a question about instructional sequence rather than about the elements themselves, and it is what motivates combining gamification with the deep learning approach, which foregrounds conceptual understanding, higher-order engagement, and the capacity to use knowledge in new situations (Biggs et al., 2022; Fullan et al., 2018). The present study therefore developed gamification media in which game elements are deliberately sequenced across three learning phases, and examined how far the resulting product met the criteria of validity, practicality, and effectiveness.

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

1.1 Research Gap and Novelty

Empirical work on gamified vocabulary learning at primary level has established that game elements can raise engagement, but it has approached the question from a narrower angle than the present study. Zou (2020) implemented Kahoot and Edpuzzle in a gamified flipped English classroom with primary students in Hong Kong and reported gains in motivation, confidence, and self-regulated learning, drawing on perceptions rather than on measured depth of understanding. Homer et al. (2018) compared a digital badges-and-points system with a conventional classroom token system among elementary ESL learners, and found effects on classroom behaviour that did not translate uniformly into English learning outcomes. Sandberg et al. (2014) added a gaming context and adaptive sequencing to a mobile vocabulary application and traced the added value to the adaptation rather than to the game framing. Dindar et al. (2021) compared gamified cooperation with gamified competition for English vocabulary learning and found comparable effects on effort and achievement but higher social relatedness under cooperation.

Two patterns recur across this literature. First, most studies at primary level evaluate a single platform or a single configuration, so what is known concerns the effect of a tool rather than the effect of an arrangement of tools. Second, gamification is typically justified through motivational theory alone, with the depth of the resulting learning left unexamined (Koivisto & Hamari, 2019; Majuri et al., 2018). Where the configuration of elements has been varied deliberately, as in Dindar et al. (2021), the differences that emerge suggest that arrangement matters more than presence. Studies that make the link to the deep learning approach explicit, and that do so in Indonesian primary English classrooms, remain scarce.

This study contributes on three points. It sets out an explicit conceptual mapping between gamification elements and deep learning principles at the design stage, expressed as four design principles rather than as a general claim of compatibility. It arranges six web platforms into a single coherent instructional sequence, in which game elements are withdrawn as cognitive demand rises, instead of using one platform in isolation or sustaining the same elements throughout. And it locates the work at elementary level, where the tension between game-like activity and depth of understanding is sharpest and where the empirical base is thinnest.

1.2 Research Questions

This study aims to develop gamification media based on a deep learning approach for the topic *Parts of the Body* using the 4D development model, and to test its feasibility through validation by media experts, content experts, and teaching-module experts. Furthermore, this study also aims to determine the practicality of the media in classroom use and to measure the extent of improvement in fifth-grade students' vocabulary mastery after using it. This study is guided by the following research questions:

1. How is deep learning approach-based gamification media for the topic Parts of the Body developed through the stages of the 4D model?
2. How valid is the developed media according to media, content, and deep learning teaching-module experts?
3. How practical is the media, and what change in fifth-grade students' mastery of Parts of the Body vocabulary accompanies its use?

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

2. METHODS

2.1 Research Design

This study used Research and Development (R&D), which aims to produce an instructional product and to test its quality (Mulyatiningsih, 2021; Sugiyono, 2021). The development followed the 4D model of Thiagarajan, Semmel, and Semmel (1974), comprising the Define, Design, Develop, and Disseminate stages. Data were collected at SDN 5 Limboto, Gorontalo Regency, Gorontalo Province.

The effectiveness component employed a one-group pre-test-post-test design. No comparison class was used, since the purpose at this stage was to establish the feasibility and internal quality of a newly developed product rather than to compare it against an alternative treatment. Change in vocabulary scores is accordingly reported descriptively, and no causal language is used in interpreting it.

2.2 Research Subjects

Expert participants were selected purposively on the basis of field of expertise, academic qualification, and teaching or professional experience: one media expert with a background in instructional media and digital learning design, one content expert with a background in English language teaching for young learners, and two experts in the development of deep learning teaching modules. Student participants were 26 fifth-grade students of SDN 5 Limboto, comprising the whole of one intact class, so that the trial took place under ordinary classroom conditions rather than with a selected subset. A small-group formative trial involving eight students preceded the field trial and served to identify usability problems before full implementation. Because the participants are minors, permission to conduct the study was obtained from the school principal, and written informed consent was obtained from parents or guardians. Students were informed that participation was voluntary, that they could withdraw at any point without consequence, and that their scores would be reported anonymously. No personally identifying information was recorded in the research instruments.

2.3 Research Procedures

Define

This stage comprised front-end analysis of the instructional problem, learner analysis of students' characteristics and needs, task analysis of the vocabulary skills to be mastered, concept analysis organising the Parts of the Body material hierarchically, and the formulation of measurable learning objectives aligned with the curriculum.

Design

Criterion-referenced pre-test and post-test instruments were constructed, the media format was determined, and an initial prototype was designed around three learning phases: Understanding, Applying, and Reflecting. These phases were derived from the deep learning literature, in which understanding is built by moving from acquisition of meaning, through application in varied contexts, to reflection that consolidates and transfers what has been learned (Biggs et al., 2022; Fullan et al., 2018).

Four design principles governed the arrangement of activities within these phases. The first holds the lexical set constant while changing the task at every encounter, since vocabulary knowledge accumulates through repeated encounters that differ in what they ask of the learner (Nation, 2013). The second raises the involvement load of successive tasks, moving from recognition under support to unsupported production, following the

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

account of need, search, and evaluation set out by Laufer and Hulstijn (2001). The third matches game elements to the behaviour each phase is meant to induce and withdraws them when that behaviour changes: on Landers' (2014) theory of gamified learning, game attributes reach learning outcomes only through the behaviours and attitudes they produce, so an attribute that induces speed will not serve a task that rewards care. The fourth closes the sequence with a learner-generated artefact rather than a scored quiz, so that consolidation takes the form of use rather than recall.

Six platforms were selected against four criteria: the fit between the task type each platform affords and the demand of the phase it was assigned to; availability at no cost; operability in a browser on the devices already available at the school; and the absence of any requirement for students to create personal accounts. Table 1 shows how each platform was assigned to a phase.

Table 1 Mapping of Learning Phases, Platforms, Gamification Elements, and Deep Learning Principles

Phase	Platform and activity	Gamification elements	Deep learning principle served
Understanding	Baamboozle (Body Parts Race); Educaplay (Froggy Jumps, Memory Game)	Team-based scoring, points, countdown timer, immediate right or wrong feedback	Acquisition of form-meaning links; visual-verbal association through a high volume of supported encounters
Applying	Blooket (Arcade Challenge); BookWidgets (Jigsaw Puzzle)	Levels, progressively harder challenges, in-game rewards, immediate corrective feedback	Use of vocabulary in progressively demanding tasks, where the learner must retrieve rather than recognise
Reflecting	Kahoot (Quiz Challenge); Canva (My Body Story)	Leaderboard and time pressure in the quiz; no game elements in the creative task	Consolidation, self-monitoring, and creative re-expression of what has been learned

Note. The sequence moves from a low to a high involvement load while the target vocabulary remains constant. Game elements are densest in the Understanding phase, where the behaviour to be induced is participation and exposure, and are absent from the final creative task, where the behaviour to be induced is deliberate, self-paced use of the words.

Develop

The prototype was submitted to expert validation using the three instruments described in Section 2.4. Recommendations arising from validation were incorporated into the prototype before any trial took place. The revised prototype was then tested in a small-group trial (n = 8), refined again on the basis of observed usability problems, and implemented in a field trial with the full class (n = 26). Results of the validation and of the trials are reported in findings.

Disseminate

The media were packaged with a user guide setting out the sequence, the platform links, and the classroom management arrangements each phase requires, and were shared with English teachers at the participating school. Dissemination at this stage was deliberately limited in scope, and the study makes no claim about uptake beyond the site of development.

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

2.4 Instruments

Three validation sheets were used, each employing a five-point Likert scale. The media validation sheet covered technical quality, visual design, navigation, interactivity, and device compatibility. The content validation sheet covered accuracy of material, curricular alignment, language appropriateness, and difficulty level. The deep learning teaching-module validation sheet covered cognitive engagement, depth of understanding, scaffolding, and the six competences of communication, collaboration, critical thinking, creativity, character, and citizenship as formulated by Fullan et al. (2018). Items on all three sheets were written for this study from the aspects listed above and were reviewed for clarity before use.

Practicality was measured with a student response questionnaire administered after the field trial, covering ease of use and navigation, clarity of instructions and content, interest and enjoyment, and perceived helpfulness for learning vocabulary. Items were read aloud to the class before responding, given the age of the participants. Active participation during the trial was recorded through structured classroom observation and reported separately as a behavioral indicator.

Vocabulary mastery was measured with a criterion-referenced achievement test with a maximum score of 100, constructed from a test blueprint covering the Parts of the Body sub-topics taught and the cognitive levels targeted in the learning objectives. Content validity was established through expert judgement, and the same instrument served as pre-test and post-test. Two features of this arrangement bound the claims that can be made from it. Using an identical form twice introduces a testing effect that a one-group design cannot separate from learning, and a test of this type measures receptive and controlled productive knowledge of form and meaning rather than communicative use.

2.5 Data Analysis

Expert validation data were converted to percentages using the formula V = (score obtained / maximum score) x 100%, and interpreted against the criteria of 80-100% (highly valid), 60-79% (valid), 45-59% (moderately valid), and below 45% (less valid) (Akbar, 2013). Student response data were analysed with the same percentage formula and interpreted against the corresponding practicality criteria. Change in vocabulary scores was analysed descriptively. The gain was computed as the difference between the post-test and pre-test class means, and the normalised gain as g = (post-test - pre-test) / (maximum score - pre-test), yielding a decimal value interpreted as high (g > 0.7), medium (0.3 <= g <= 0.7), or low (g < 0.3) (Hake, 1998). The formula is not multiplied by 100%; expressing g as a decimal keeps it consistent with the criteria applied. Mastery was determined against the school's minimum criterion of learning goals. No inferential test of the pre-test-post-test difference was conducted, and the improvement is therefore reported as an observed change rather than as a statistically significant effect.

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

3. FINDINGS AND DISCUSSION

3.1 Findings

This section presents the findings of the study, organized according to the three research questions that guided the investigation. The results are divided into three main parts: first, the developmental process and product resulting from the application of the 4D model; second, the validity scores and feedback provided by the media, content, and teaching-module experts; and third, the practicality of the media as observed during classroom trials, along with the quantitative data on fifth-grade students' vocabulary mastery before and after the intervention.

The Developmental Process and Product Resulting from the Application of the 4D Model

The Define and Design stages produced a set of gamified activities distributed across six web platforms and organised into the three phases set out in Table 1. Three features of the resulting product carry the design argument.

The lexical set is held constant while the task changes at every encounter. A student meets each target item first as a team-answered recognition prompt, then as a timed visual match, then as a selection under time pressure, then as a fragment to be reassembled into a whole, then as a retrieval item in a shared quiz, and finally as a word to be chosen and used in describing themselves. Six encounters, six task types, one set of words.

Game elements thin out as the sequence proceeds. Team points and countdowns are dense in the Understanding phase, where the behaviour worth inducing is participation and volume of exposure. They are absent from the closing creative task, where the behaviour worth inducing is care rather than speed. The design intent is not that motivation ceases to matter later, but that the kind of behaviour a scoring structure produces stops being the kind of behavior the task rewards.

The sequence ends in production rather than assessment. In My Body Story, students select from the target set and use the words to describe themselves, which is the only point in the product where the vocabulary must be evaluated against a context of use rather than matched to a given prompt. Figures 1 to 6 present the interface of each activity, and Figure 7 shows the flow of the 4D development procedure.

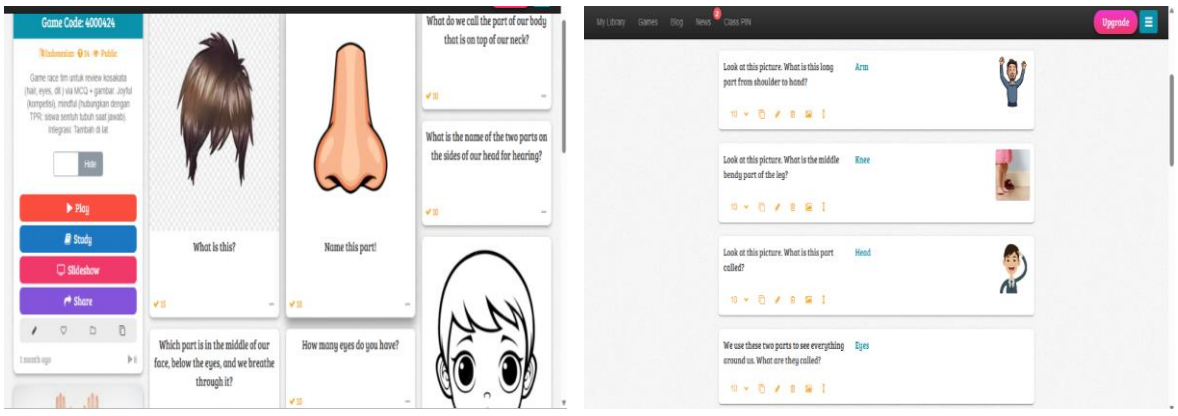

Figure 1. Baamboozle: Body Parts Race (Understanding phase)

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

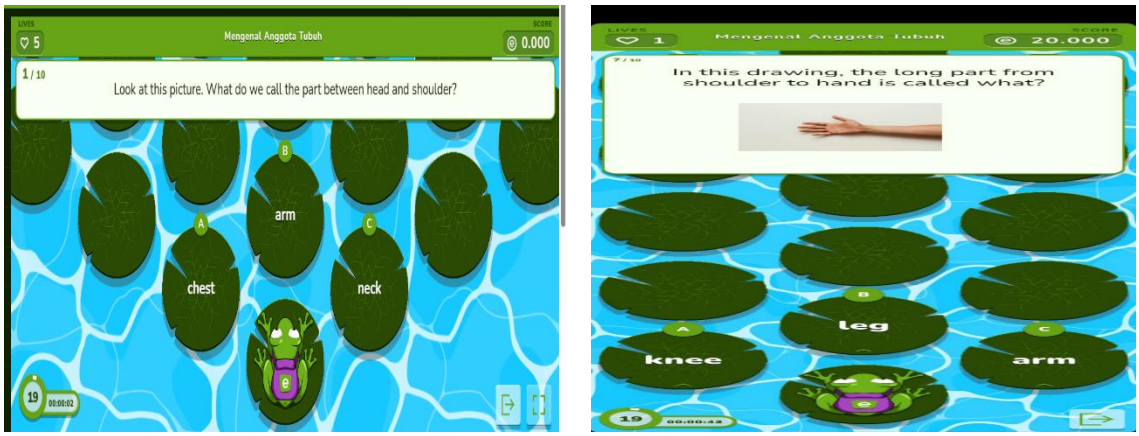

Figure 2. Educaplay: Froggy Jumps and Memory Game (Understanding phase)

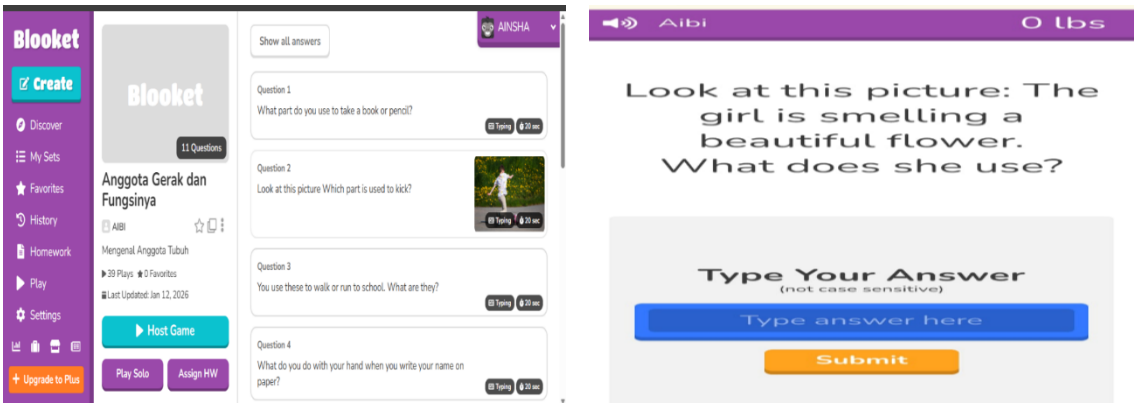

Figure 3. Blooket: Arcade Challenge (Applying phase)

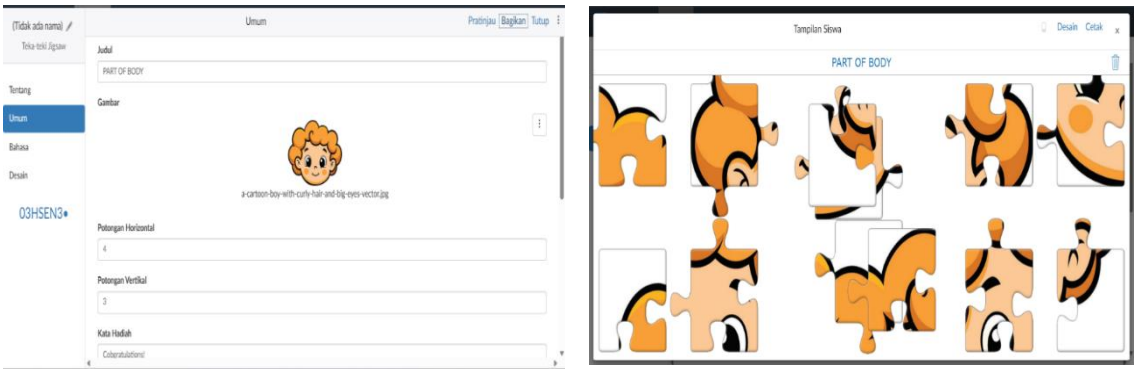

Figure 4. BookWidgets: Jigsaw Puzzle (Applying phase)

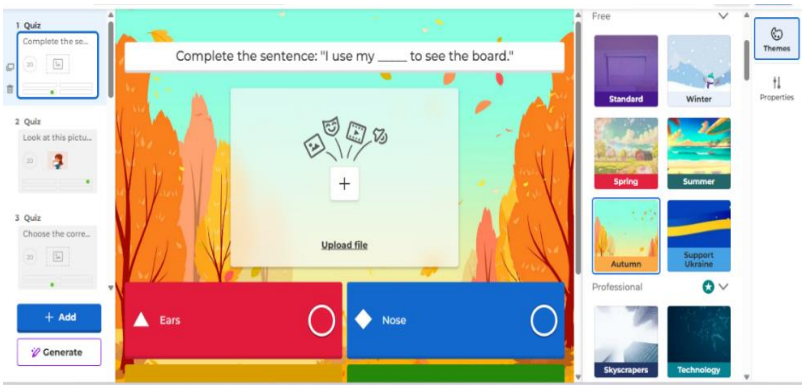

Figure 5. Kahoot: Quiz Challenge (Reflecting phase)

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

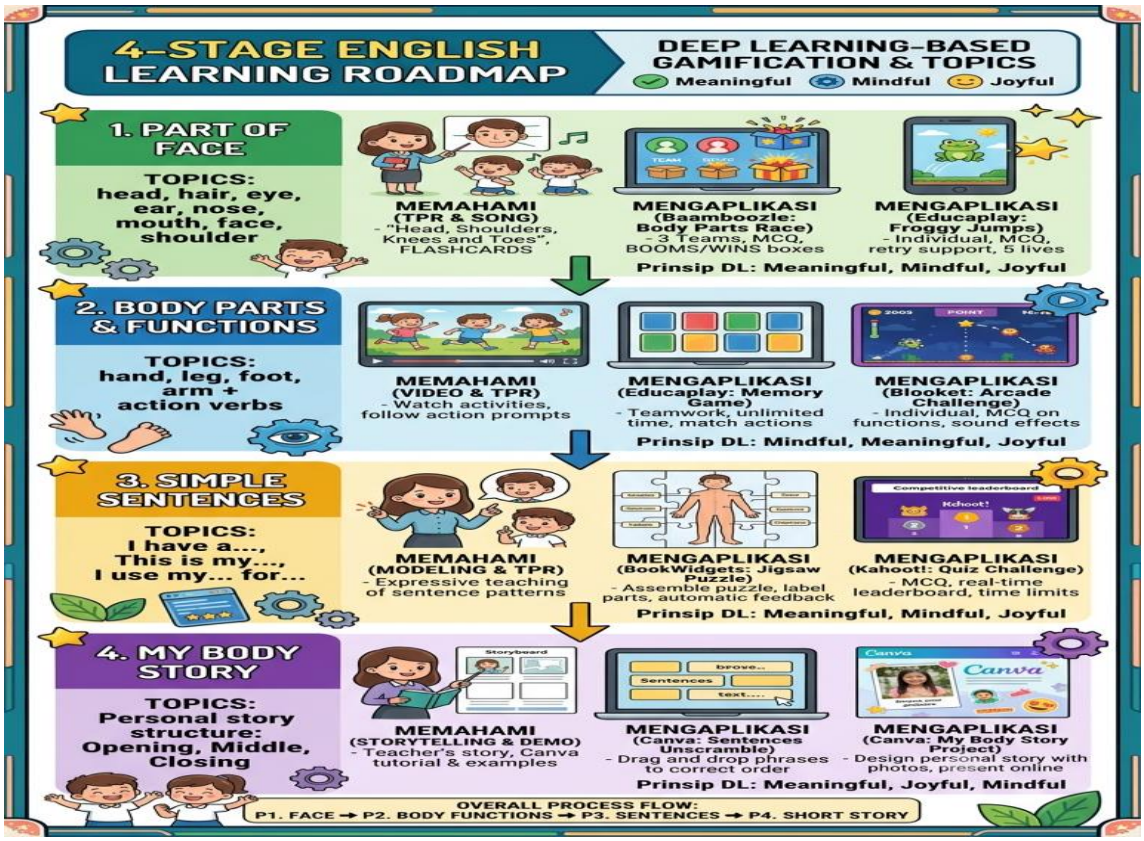

Figure 6. Flow of the 4D development procedure

Validity of the Media

Expert validation placed the media in the highly valid category on all three instruments, as shown in Table 2. Recommendations made during validation concerned refinements to the prototype rather than changes to the phase structure, and all were incorporated before the small-group and field trials.

Table 2 Expert Validation Results by Instrument

Validator	Aspects assessed	Score (%)	Category
Media expert	Technical quality, visual design, navigation, interactivity, device compatibility	87.06	Highly valid
Content expert	Accuracy of material, curricular alignment, language appropriateness, difficulty level	87.37	Highly valid
Deep learning teaching-module experts (mean of two)	Cognitive engagement, depth of understanding, scaffolding, six competences	86.76	Highly valid
Overall mean	**Mean of the three instrument scores**	**87.06**	**Highly valid**

Note. The overall mean is the unweighted mean of the three instrument scores (87.06, 87.37, 86.76), not a mean across the four individual raters; the two figures differ, and the basis is stated here to avoid ambiguity. Because the panel comprised four raters, these percentages are

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

reported as indicators of expert agreement about the product rather than as precise estimates of content validity.

Practicality of the Media

Responses from the 26 students in the field trial placed the media in the highly practical category. Ratings were consistently favourable across ease of use and navigation, clarity of instructions and content, interest and enjoyment, and perceived helpfulness for learning vocabulary, and no indicator fell below the highly practical band. Alongside the questionnaire, classroom observation during the trial recorded an average active participation rate of 87.94%, which is reported here as a separate behavioural indicator rather than as a measure of practicality. Self-reported data of this kind carry a known risk of acquiescence when children respond to adults associated with the intervention, and the observation figure is reported partly for that reason, as an independent behavioural check on the questionnaire.

Effectiveness of the Media

Scores rose substantially between the pre-test and the post-test, as summarised in Table 3. The class mean moved from 42.31 to 78.85, a gain of 36.54 points on a 100-point scale, and the normalised gain of 0.63 places the result in the upper part of the medium category. Twenty-four of the 26 students, or 92.31%, reached the minimum criterion of learning goals at post-test, against a considerably smaller number at pre-test.

Table 3 Pre-test and Post-test Results (n = 26)

Measure	Pre-test	Post-test	Change
Mean score (maximum = 100)	42.31	78.85	+36.54 points
Normalised gain (g)	-	-	0.63 (medium)
Students reaching the minimum criterion	-	24 of 26 (92.31%)	-

Note. Mastery was determined against the school's minimum criterion of learning goals. Because the design includes no comparison condition and no inferential test was conducted, the change reported here is an observed improvement across the implementation period rather than an estimate of the effect attributable to the media.

3.2 Discussion

Arranging Game Elements to Serve Understanding

The pattern in the validation and trial data is consistent with the design premise of the study, namely that game elements are more defensible pedagogically when they are sequenced rather than simply added. Points, timed tasks, and scoreboards mainly serve the motivational function documented in the gamification literature, where satisfaction of competence and autonomy needs is the mechanism, most consistently identified (Ryan & Deci, 2000; Sailer et al., 2017). What the three-phase arrangement adds is a division of labour, and Landers' (2014) framework explains why that division should matter. Game attributes do not act on learning outcomes directly; they act on learner behaviours and attitudes, which then mediate or moderate the effect of instructional quality. Team points and countdowns reliably induce participation and pace, and pace is exactly what a recognition-heavy opening phase converts into learning, because what that phase needs is a high volume of form-meaning encounters. The same attributes attached to a construction task would induce speed where care is wanted. Read through this framework, the recurring finding that gamification effects are heterogeneous and

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

design-dependent (Bai et al., 2020; Sailer & Homner, 2020) is not a puzzle but a prediction: it is what follows when identical attributes are attached to instructional demands they do not fit.

The involvement load analysis supplies the second half of the account. Holding the lexical set constant while raising need, search, and evaluation across encounters is the condition under which Laufer and Hulstijn (2001) predict stronger retention, and it is what distinguishes the present arrangement from a series of games that merely share a topic.

One qualification belongs here rather than in the limitations. The Kahoot step retains a leaderboard and time pressure at a point in the sequence where the involvement load is already high, which sits awkwardly with the third design principle. It was retained because retrieval practice across the full set is valuable and the platform delivers it efficiently, but a stricter application of the principle would either remove the competitive framing from that step or move it earlier in the sequence. This is the clearest candidate for revision in a further development cycle.

The Role of Multi-Platform Variation

Distributing activities across six platforms was intended to vary task type and modality while keeping the vocabulary constant, so that students encounter the same items through recognition, selection, reconstruction, retrieval, and production. Prior work suggests that varied and appropriately challenging task designs sustain engagement more reliably than repeated exposure to a single format (Hamari et al., 2016), and meta-analytic findings indicate that the effects of gamification depend strongly on design features rather than on the presence of game elements as such (Huang et al., 2020; Sailer & Homner, 2020). What this study cannot claim should be stated without hedging. No single-platform comparison condition was included, so the multi-platform arrangement cannot be shown to be superior to a well-designed single-platform one. Dindar et al. (2021) illustrate what the relevant test looks like, comparing two gamified configurations directly and locating the difference in social relatedness rather than in achievement, and a comparison of that kind is the natural successor to the present work. Variation is justified here in terms of task type and involvement load, not in terms of matching learning styles, since the learning styles hypothesis has not survived empirical scrutiny.

Interpreting the Size of the Improvement

The improvement of 36.54 points, corresponding to a normalised gain of 0.63, places the result at the upper end of the medium category. Read against the meta-analytic picture, in which gamification effects on cognitive outcomes are typically small to moderate and highly heterogeneous (Bai et al., 2020; Huang et al., 2020; Sailer & Homner, 2020), the present result is consistent with rather than exceptional among existing findings. Presenting it as exceptional would invite exactly the scrutiny that a one-group design cannot withstand.

Several factors plausibly account for the gain stopping short of the high category. The intervention covered a single topic within a limited number of meetings, which caps the number of encounters with each item, and vocabulary growth is cumulative rather than immediate (Nation, 2013). Novelty is the other candidate, and not a minor one: unfamiliar digital platforms attract attention independently of their instructional design, and the positive effect of gamification has repeatedly been found to fade once that novelty is exhausted (Shortt et al., 2023; Zhang & Hasim, 2023). A single implementation cycle cannot separate the design from its novelty; only a longer or repeated deployment can. Design factors identified in the findings, in particular

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

structured repetition across varied task types and immediate corrective feedback, remain the most likely contributors to the gain that was obtained.

Implications

For practice, the study offers primary English teachers a concrete arrangement rather than a general endorsement of gamification: a phased sequence in which freely available platforms are assigned to specific instructional functions, in which competitive elements are concentrated where they help and withdrawn where they do not, and in which the final phase asks students to produce something with the vocabulary rather than merely to recognise it. None of this requires funding, specialised software, or student accounts. For theory, the findings are consistent with the proposition that game elements and depth of learning are not in tension when the elements are subordinated to an instructional sequence. The four design principles are stated at a level that allows them to be tested rather than merely adopted: two products can be built that differ only in whether game elements are withdrawn at the production step, and compared directly. Claims about knowledge transfer and long-term attitude change, however, cannot be made from these data, since neither transfer tasks nor delayed measures were administered.

Limitations

Six limitations should be borne in mind when reading these results. The study involved a single class of 26 students in one school, so the findings are not statistically generalisable. No control group was used, which means that the improvement cannot be attributed to the media alone; maturation, teacher effects, and the testing effect of taking the same instrument twice cannot be ruled out. No inferential test of the pre-test-post-test difference was conducted, so the change is reported descriptively and should not be read as a statistically established effect. The intervention was short, and the novelty of unfamiliar digital platforms may itself have contributed to engagement. Practicality data were self-reported by students, with the attendant risk of acquiescence. Finally, no delayed post-test was administered, so nothing can be said about retention beyond the immediate post-test, and the My Body Story artefacts produced in the Reflecting phase were collected but not analysed as data, so the study measured vocabulary mastery rather than communicative use.

4. CONCLUSION

This study developed and evaluated gamification media grounded in the deep learning approach for teaching Parts of the Body vocabulary at fifth grade. The product was judged highly valid by media, content, and deep learning teaching-module experts, with an overall mean of 87.06%, was rated highly practical by students, and was associated with an improvement from a pre-test mean of 42.31 to a post-test mean of 78.85, a normalised gain of 0.63 in the medium category, with 24 of 26 students reaching the minimum criterion of learning goals.

The contribution of the study lies less in these figures than in the design principle behind them. Gamification is often adopted for its motivational appeal and evaluated on that basis alone; here, game elements were assigned to distinct instructional functions across a sequence of understanding, application, and reflection, so that the motivational load sits early, the cognitive demand rises later, and scoring is withdrawn at the point where deliberate construction begins. The results indicate that such an arrangement is feasible with freely available platforms in an ordinary primary classroom, and that it does not require sacrificing depth for engagement.

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

Four directions follow for further work. A quasi-experimental study with a control class and inferential analysis would establish whether the gains reported here exceed those from conventional instruction. A delayed post-test would address retention, which the present design cannot speak to. Analysis of student-produced artefacts would allow claims about communicative use to be tested rather than assumed. And a comparison of two configurations differing only in where game elements are withdrawn would turn the third design principle from a stated commitment into a finding.

Acknowledgements

The authors thank SDN 5 Limboto, its English teachers and participating students, and the experts who validated the media. Appreciation is also extended to the Department of Primary Teacher Education, Faculty of Education, Universitas Negeri Gorontalo, for institutional support.

Authors' Contribution

Ainsha Ainun Habibie was responsible for conceptualisation, methodology, data collection and analysis, and drafting the manuscript. Karmila Machmud and Wiwy Triyanty Pulukadang supervised the research, contributed to the interpretation of results, and reviewed and edited the manuscript. This article is derived from the first author's undergraduate thesis, supervised by the second and third authors. All authors have read and approved the final version and agreed to be accountable for the integrity of the work.

AI Generative Statement

The authors acknowledge the use of generative artificial intelligence tools solely for language refinement and clarity of expression. The conceptualisation and design of the study, the methodology, the data analysis, the interpretation of results, and the conclusions are entirely the intellectual work and responsibility of the authors. All references were verified against the original sources by the authors. No content was included without critical review, modification, and final approval by the authors.

REFERENCES

Akbar, S. (2013). *Instrumen perangkat pembelajaran*. Remaja Rosdakarya.

Arsyad, A. (2020). *Media pembelajaran* (2nd ed.). Rajawali Pers.

Bada, S. O. (2015). Constructivism learning theory: A paradigm for teaching and learning. *IOSR Journal of Research & Method in Education, 5*(6), 66-70.

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*, Article 100322. https://doi.org/10.1016/j.edurev.2020.100322

Biggs, J., Tang, C., & Kennedy, G. (2022). *Teaching for quality learning at university* (5th ed.). Open University Press.

Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaee, E., & Noroozi, O. (2021). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning, 34*(7), 934-957. https://doi.org/10.1080/09588221.2019.1648298

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference* (pp. 9-15). https://doi.org/10.1145/2181037.2181040

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

Dindar, M., Ren, L., & Jarvenoja, H. (2021). An experimental study on the effects of gamified cooperation and competition on English vocabulary learning. *British Journal of Educational Technology, 52*(1), 142-159. https://doi.org/10.1111/bjet.12977

Fullan, M., Quinn, J., & McEachen, J. (2018). *Deep learning: Engage the world change the world*. Corwin Press.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64-74. https://doi.org/10.1119/1.18809

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*, 170-179. https://doi.org/10.1016/j.chb.2015.07.045

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education, 80*, 152-161. https://doi.org/10.1016/j.compedu.2014.08.019

Homer, R., Hew, K. F., & Tan, C. Y. (2018). Comparing digital badges-and-points with classroom token systems: Effects on elementary school ESL students' classroom behavior and English learning. *Journal of Educational Technology & Society, 21*(1), 137-151.

Huang, R., Ritzhaupt, A. D., Sommer, M., Zhu, J., Stephen, A., Valle, N., Hampton, J., & Li, J. (2020). The impact of gamification in educational settings on student learning outcomes: A meta-analysis. *Educational Technology Research and Development, 68*(4), 1875-1901. https://doi.org/10.1007/s11423-020-09807-z

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. Pfeiffer.

Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management, 45*, 191-210. https://doi.org/10.1016/j.ijinfomgt.2018.10.013

Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & Gaming, 45*(6), 752-768. https://doi.org/10.1177/1046878114563660

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26. https://doi.org/10.1093/applin/22.1.1

Li, X., & Chu, S. K. W. (2021). Exploring the effects of gamification pedagogy on children's reading: A mixed-method study on academic performance, reading-related mentality and behaviors, and sustainability. *British Journal of Educational Technology, 52*(1), 160-178. https://doi.org/10.1111/bjet.13057

Majuri, J., Koivisto, J., & Hamari, J. (2018). Gamification of education and learning: A review of empirical literature. In *Proceedings of the 2nd International GamiFIN Conference* (pp. 11-19).

Marton, F., & Saljo, R. (1976). On qualitative differences in learning: I-Outcome and process. *British Journal of Educational Psychology, 46*(1), 4-11. https://doi.org/10.1111/j.2044-8279.1976.tb02980.x

Mulyatiningsih, E. (2021). *Metode penelitian terapan bidang pendidikan*. Alfabeta.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78. https://doi.org/10.1037/0003-066X.55.1.68

© 2026 Ainsha Ainun Habibie, Karmila Machmud, & Wiwy Triyanty Pulukadang.
This is an open-access article under the CC-BY-SA license.

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*(1), 77-112. https://doi.org/10.1007/s10648-019-09498-w

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371-380. https://doi.org/10.1016/j.chb.2016.12.033

Sandberg, J., Maris, M., & Hoogendoorn, P. (2014). The added value of a gaming context and intelligent adaptation for a mobile learning application for vocabulary learning. *Computers & Education, 76*, 119-130. https://doi.org/10.1016/j.compedu.2014.03.006

Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., & Akinkuolie, B. (2023). Gamification in mobile-assisted language learning: A systematic review of Duolingo literature from public release of 2012 to early 2020. *Computer Assisted Language Learning, 36*(3), 517-554. https://doi.org/10.1080/09588221.2021.1933540

Sugiyono. (2021). *Metode penelitian pendidikan: Pendekatan kuantitatif, kualitatif, dan R&D*. Alfabeta.

Suryani, N., Setiawan, A., & Putria, A. (2020). *Media pembelajaran inovatif dan pengembangannya*. Remaja Rosdakarya.

Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. Indiana University.

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. https://doi.org/10.1016/j.edurev.2020.100326

Zhang, S., & Hasim, Z. (2023). Gamification in EFL/ESL instruction: A systematic review of empirical research. *Frontiers in Psychology, 13*, Article 1030790. https://doi.org/10.3389/fpsyg.2022.1030790

Zou, D. (2020). Gamified flipped EFL classroom for primary education: Student and teacher perceptions. *Journal of Computers in Education, 7*(2), 213-228. https://doi.org/10.1007/s40692-020-00153-w