Jurnal Sisfotek Global
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v15i2.16008
Vol. 15, No. 2, September 2025, pp. 64-73

Design and Usability Evaluation of a Web-Based Edutainment for Grade VII Science: Task Performance and System Usability Scale Findings

Nurhuda Budi Pamungkas[1] & Sanriomi Sintaro[2*]

[1] SMP Negeri 27, Bandar Lampung, Indonesia, 35223
[2] Sam Ratulangi University, Manado, Indonesia, 95115
E-mail: [1]hudaforlan@gmail.com, [2*]sanriomi@unsrat.ac.id

ARTICLE HISTORY

Received : August 28, 2025
Revised : September 14, 2025
Accepted : September 27, 2025

KEYWORDS

Construct 2
Edutainment
System Usability Scale
Task Performance
Usability Testing

ABSTRACT

This study designed and evaluated a browser-based edutainment prototype for Grade VII science (gravity and Newton's laws). Usability was examined through task-performance metrics and the System Usability Scale (SUS). Thirty students at SMPN 27, Bandar Lampung, completed an eight-task scenario covering core flows: launching the app, entering and navigating the World Map, starting and stopping gameplay, and accessing Learn and Credits. Core navigation tasks showed high effectiveness with short median times (e.g., launch: 96.7% success, 6.53 s; enter World Map: 93.3%, 7.79 s). The in-level physics puzzle was purposely more demanding (83.3% success, 35.74 s; 0.52 errors per participant), indicating a need for clearer in-level cues. Perceived usability was excellent (mean SUS 87.42, SD 5.02). Overall, the prototype appears classroom-ready; targeted refinements inside levels (first-run micro-tutorials, stronger affordances, on-demand hints) should further smooth first-time play. Future work will examine multi-session learnability, broaden device/browser coverage, and link gameplay to learning outcomes via pretest–posttest designs.

1. Introduction

Science education in junior high schools not only targets conceptual mastery but also fosters scientific thinking through science process skills: observing, asking questions, planning, measuring, analyzing data, and drawing conclusions[1], [2]. This emphasis aligns with the Science Learning Outcomes Phase D (*Capaian Pembelajaran Fase D*) in the Independent Curriculum and the Learning and Assessment Guidelines, which position formative assessment as a driver of student-centered learning.[3]. Both will influence learning experiences that provide space for conceptual exploration and ongoing inquiry practices, making the need for interactive media aligned with the curriculum increasingly relevant for Grade VII Science Semester 1 [4].

However, in practice, students are often strong in memorization but less trained in logical reasoning[5], for example when classifying based on scientific attributes, assessing causal relationships, identifying control variables, or interpreting evidence to draw conclusions[6], [7]. Science education literature emphasizes that process skills (observation, measurement, classification, variable control, hypothesis, data interpretation, modeling, etc.) are the foundation of scientific thinking that need to be explicitly trained through authentic tasks[8]. Recent editorials and reviews also emphasize that scientific reasoning competency is a key construct that should be systematically measured and developed in science education[9], [10].

In the last five years, research has shown that game-based learning/edutainment approaches contribute positively to motivation and science learning outcomes when the design is clearly tied to learning objectives [11]. Meta-analyses in STEM domains report moderate–large effects for game-based learning over conventional learning, and similar findings emerge in cross-level gamification studies highlighting increased engagement and academic performance [12]. These findings strengthen the theoretical and empirical basis for adopting edutainment in a measurable manner in science subjects[13].

The context of Grade VII Semester 1 science material such as the classification of living things, substances and their changes, energy, and interactions between living things and the environment, naturally demands evidence-based reasoning, pattern recognition, and simple variable control. Additional evidence in the realm of science shows that interactive/dynamic visualizations can support conceptual understanding compared to static visualizations, so web-based edutainment media is a strong candidate to facilitate concept exploration as well as practice learning logic[14]. However, good edutainment must also be supported by the readiness of the ecosystem where the edutainment will be implemented[15].

In terms of ecosystem readiness, SMP Negeri 27 Bandar Lampung is recorded as having an official website[12] and E-learning [16], This indicates the school's infrastructure and policy support for digital learning. Furthermore, the web-based edutainment project "*Petualangan Ayam*" is now publicly available through modern browsers (cross-device web mode and no installation required), providing a potential example of an interactive activity to support scientific reasoning before being piloted more widely in the classroom. This foundation provides a strong starting point for developing edutainment that is more closely tied to the 7th grade science logic indicators.

Meanwhile, in the context of the curriculum, the need to improve logical reasoning, and the readiness of the school's digital ecosystem, this study aims to (1) design a Grade VII, Semester 1 science edutainment prototype that targets curriculum-aligned logic indicators, (2) evaluate the prototype's usability through task performance metrics and the System Usability Scale (SUS), and (3) describe deployment readiness within the school's digital ecosystem. This framework is expected to provide practical contributions for science teachers and empirical evidence for the development of edutainment media that is aligned with the Independent Curriculum. [17].

Despite curriculum goals that emphasize scientific reasoning, many Grade VII learners still depend on recall and struggle with evidence-based skills such as classification, identifying causal relations, and interpreting data[18], [19]. Game-based "edutainment" can increase engagement, but classroom adoption often stalls when basic usability is uncertain[20], [21]. Few studies report both task-performance indicators (success rates, time on task, errors) and standardized perceived usability using the System Usability Scale (SUS) for K–12 science activities[22], [23], [24]. This gap limits confident deployment of lightweight, browser-based learning tools that align with early physics topics (e.g., gravity and Newton's laws).

This study asks: How effective and efficient is the prototype for key user flows (e.g., launching, navigating the world map, starting/stopping gameplay, accessing learning content), as measured by task success, completion time, and errors? and How do learners perceive the usability of the prototype, as measured by the System Usability Scale (SUS)?

By combining task-performance analytics with SUS, this study provides a robust, complementary view of usability in K–12 edutainment and offers reproducible reporting conventions (metrics, tables, and thresholds) for future evaluations.

Practical/educational. The findings inform small, high-impact design tweaks (e.g., in-level cues and first-run guidance) that make the prototype classroom-ready for short activities tied to Grade VII physics objectives, supporting teachers who need quick, reliable tools that run in a browser.

2. Research Method

This study employs a structured research methodology following the Multimedia Development Life Cycle (MDLC) framework [25], [26], which consists of six sequential stages: concept, design, material collection, assembly, testing, and distribution. Each stage plays a critical role in ensuring the systematic development of the edutainment system.

2.1 Concept

In this stage, science learning objectives are mapped to "learning logic" indicators (attribute-based classification, simple cause-and-effect, and variable control), and interactive activities that stimulate this reasoning are determined[27], [28]. The design phase then translates ideas into navigation structures, storyboards, and flowcharts to ensure clear and consistent interaction paths.

2.2 Design

At the design stage, the appearance of the application being created is designed. The design stage is carried out with three images, navigation structure, storyboard and flowchart view.

2.2.1 Nagivation Structure

In the first design phase, the navigation structure explains the direction of the buttons when pressed. Figure 1 shows the navigation structure in the science edutainment application.

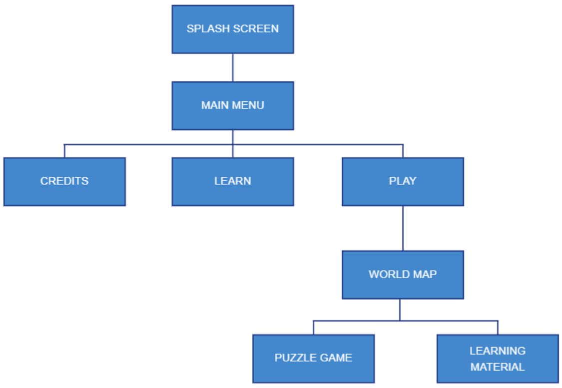


Figure 1. Navigation Structure

Figure 1 explains the navigation structure, where the splash screen displays the main menu. From the main menu, users can select three menus: credits, learn, and play. The Play menu displays a world map and allows users to select puzzle games or learning materials.

2.2.2 Storyboard

When creating a storyboard, the design will be adjusted to the navigation structure in which each scene is created and determined. The following is an example of a Main Menu Storyboard which can be seen in Figure 2.

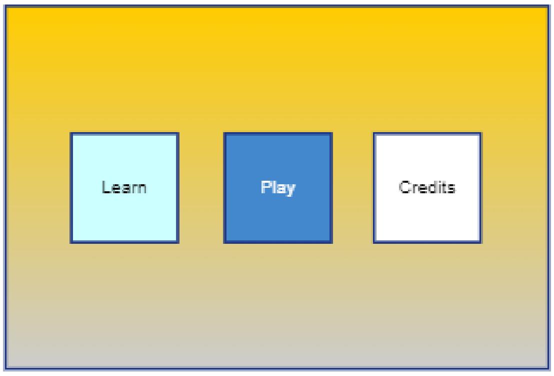


Figure 2. Storyboard Main Menu

From Figure 2, we can see that there are three buttons that will function to move to another scene, for example, when pressing the Play button, the user will be shown the World Map Scene, the following is the storyboard design of the World Map scene which can be seen in Figure 3.

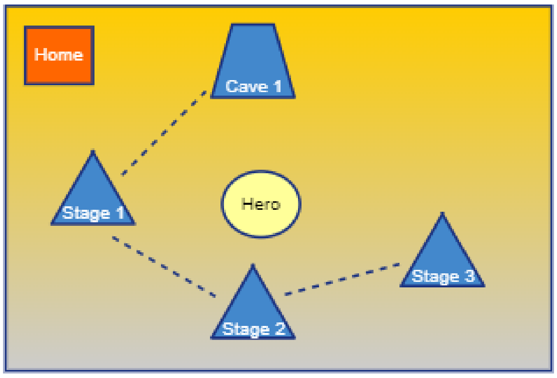


Figure 3. Storyboard World Map

From Figure 3, we can see that there is a return to the Main Menu button in the upper left corner, and the hero will be in the center of the screen. There are also several stages and caves containing learning materials. Upon entering the game, players will engage in a puzzle game, as seen in Figure 4.

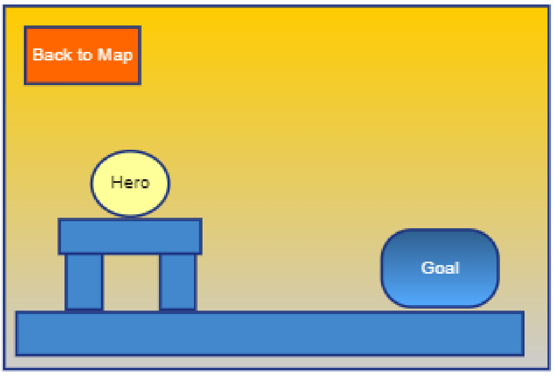


Figure 4. Puzzle Game

The design in Figure 4 represents the initial design of the game. It includes a goal that requires the hero to move toward the goal. The player will then press/touch the wooden blocks to direct the hero toward the goal.

After the storyboard is created, the next step is to create a flowchart view, which will further explain the scenes and the flow of the game between them.

2.2.3 Flowchart View

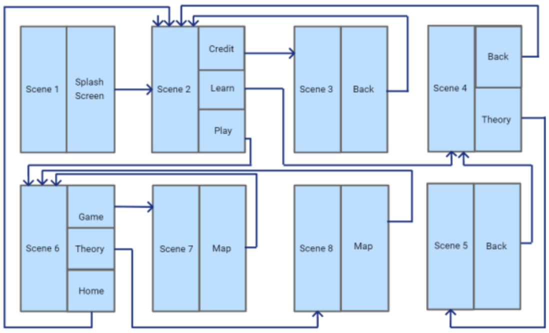


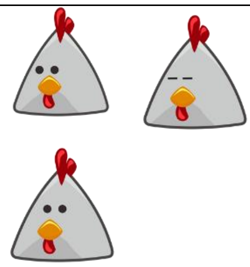
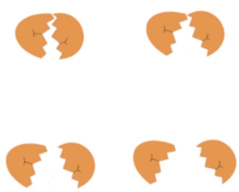
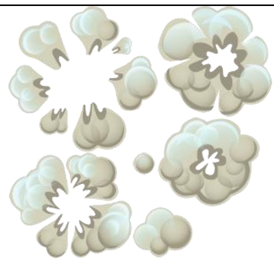
Figure 5. Flowchart View

In Figure 5, we can see that the buttons will navigate to other scenes. After the flowchart view is created, the next step is to gather the materials that will be used to create the edutainment.

2.3 Material Collecting

The next phase in the development process is material collection, during which relevant data is compiled to ensure the accuracy and effectiveness of the system. Some examples of materials collected can be seen in Table 1.

Table 1. Material Collecting

No	Material	Info
1		Main Character
2		One of the Animations Used
3		Smoke Animation that appears when the player touches a block

No	Material	Info
4		Some examples of icons used in the game
5		Main Logo

After collecting the designs that will be used in edutainment, the next step is to carry out the assembly process which combines design, sound and logic to make edutainment that can be played.

2.4 Sampling technique

We used purposive (criterion-based)[29] sampling to recruit participants who match the intended users of the edutainment, Grade VII students enrolled in science at SMPN 27 (State Junior High School 27), Bandar Lampung, Lampung, Indonesia. Participants were drawn from intact Grade VII classes identified by the school's science teachers.

Inclusion criteria. (1) Enrolled in Grade VII at SMPN 27; (2) taking the science subject during the study period; (3) able to operate a school-provided smartphone or laptop with a modern browser.

Exclusion criteria. Prior participation in pilot testing or any condition that would prevent completion of screen-based tasks.

Recruitment and consent. Invitations were distributed via science teachers. Participation was voluntary with parental/guardian consent and student assent, in accordance with school policy.

Rationale. Purposive sampling ensures alignment between the sample and the target classroom context (Grade VII science at SMPN 27), supporting ecological validity for a formative usability evaluation.

2.5 Participants and characteristics

The final analytic sample comprised 30 Grade VII students (N = 30) from SMPN 27 (State Junior High School 27), Bandar Lampung, Lampung, Indonesia. All participants were currently enrolled in the Grade VII science course covering mechanics topics aligned with the prototype content. Ages reflected the typical Grade VII range in Indonesia (approximately 12–13 years). All met the inclusion criteria (ability to operate a school-provided smartphone or laptop with a modern browser). No participants were excluded, and all 30 completed the eight-task scenario.

2.6 Data collection procedure

Sessions were conducted in a classroom/lab setting at SMPN 27 (State Junior High School 27), Bandar Lampung, Lampung, Indonesia during a single class period. After a brief orientation and reminder about voluntary participation, students worked individually on the prototype using school-provided smartphones or laptops with a modern browser and a stable internet connection.

Task scenario. Each participant completed a fixed, eight-task scenario representative of typical use:

TS1 : Launch the application.

TS2 : Enter the World Map from the Main Menu.

TS3 : Navigate within the World Map (locate/choose a stage).

TS4 : Start gameplay for a selected level.

TS5 : Complete the physics puzzle in the level.

TS6 : Return to the Main Menu from within gameplay.

TS7 : Open Learn (learning materials).

TS8 : Open Credits.

Observed metrics. Trained observers recorded for each task: (a) task success/failure (goal achieved without external help), (b) completion time for successful attempts (start = first actionable tap/click on the relevant screen; stop = goal reached), and (c) count of observable errors (slips/mis-taps with recovery). Unsuccessful attempts were coded as failures; their times were not included in time summaries.

Protocol controls. Participants completed tasks without assistance; facilitators only clarified task wording if needed. If a participant could not progress reasonably, the task was marked unsuccessful and the session proceeded to the next task to avoid fatigue.

Post-task questionnaire. Immediately after the eight tasks, participants completed the 10-item System Usability Scale (SUS). Responses were anonymous and linked to task data by a non-identifying code.

Ethics and consent. Invitations were distributed via science teachers; participation was voluntary with parental/guardian consent and student assent in line with school policy. No personally identifying information was stored in the dataset.

2.7 Instruments

A structured sheet was used to record, for each task (TS1–TS8): (a) task success/failure (goal achieved without facilitator intervention), (b) completion time for successful attempts only (start = first actionable tap/click on the relevant screen; stop = goal reached), (c) count of observable errors (mis-taps, unintended actions, recoveries), and (d) error-free completion (binary). These definitions ensured consistent coding across participants.

Per Brooke's 10-item, technology-agnostic questionnaire, participants rated each item on a 5-point Likert scale (1 = Strongly Disagree … 5 = Strongly Agree). Items alternate in polarity (odd = positive, even = negative). We scored SUS using the standard transformation to a 0–100 scale: for odd-numbered items, $(x - 1) \times 2.5$; for even-numbered items, $(5 - x) \times 2.5$; then sum across all 10 items. Higher scores indicate greater perceived usability. For interpretability, we report descriptive statistics (mean, median, SD, range) and an adjectival/letter-grade band commonly used in SUS reporting.

To match Grade VII reading levels, the facilitator read each item aloud and provided neutral vocabulary clarifications when asked (without coaching on answers). The questionnaire was completed individually immediately after the eight tasks.

2.8 Data analysis

For each task, we computed:

1. Success rate (SR) = (number of successful completions ÷ number of attempts) × 100%.
2. Time on task for successful attempts only; reported as mean (M), median (Md), and interquartile range (IQR). Times from unsuccessful attempts were excluded from time summaries.
3. Errors per participant = mean count of observable errors during the task.
4. Error-free rate = (participants with zero errors ÷ number of attempts) × 100%.

System Usability Scale responses were transformed to a 0–100 score using the standard method (odd items: $(x - 1) \times 2.5$; even items: $(5 - x) \times 2.5$; then sum across items). We report mean, median, standard deviation, and range. For interpretability, we also provide an adjectival/letter-grade band commonly used in SUS reporting.

Task times are typically right-skewed; therefore medians and IQRs are emphasized. We flagged extreme times using the 1.5 × IQR rule. Data points attributable to clear procedural/technical interruptions (e.g., connection drop, bell/teacher interjection) were excluded from time summaries and noted as protocol deviations; the corresponding success status remained recorded. Otherwise, flagged values were retained. If any SUS item response was missing, that questionnaire was excluded from SUS analysis (listwise). Descriptive analyses and tables were prepared using Microsoft Excel 365.

3. Result and Discussion

The next step is the assembly phase, where we transform the outputs from earlier stages into a functional edutainment experience in Construct 2. We organize assets, align content with learning goals, and wire up interactions and core game logic so the flow feels cohesive and engaging. We then integrate visual and audio elements, refine navigation, and run quick checks to ensure each scene behaves as intended. The figures that follow show representative screens from the edutainment that already been made, illustrating how the initial design evolves into a playable, learning-focused product.

3.1 Assembly

In the assembly phase, the edutainment creation process combines the previous stages, such as structure, flow, and design, to create an edutainment prototype. The assembly process was carried out using the Construct 2 application. The following is an example of the developed display, namely the Main Menu, which can be seen in Figure 6.

Figure 6. Main Menu

Figure 6 displays buttons adapted from the design created in the previous phase. These buttons will take you to Play, Learn, and Credits. Another example of this edutainment is the world map, which displays the in-game map. Here, players can move by tapping the screen to determine the direction the main hero is moving, as seen in Figure 7.

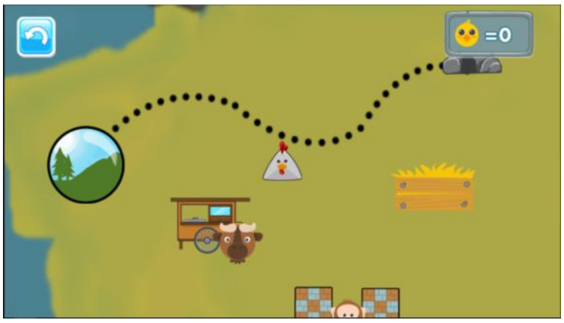


Figure 7. World Map

In Figure 7, the top right corner displays the number of chicks found. To find chicks, players must complete each level. Each successful level earns one chick. The number of chicks unlocks the next stage. One example of a game can be seen in Figure 8.

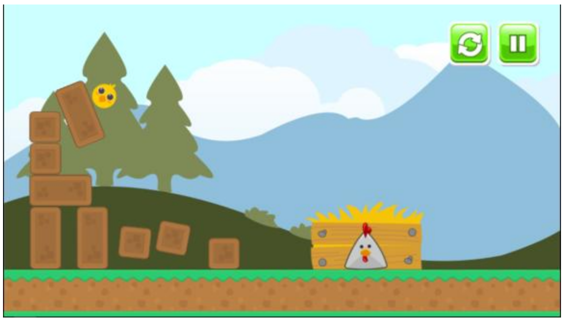


Figure 8. Gameplay.

In Figure 8, players can touch the wooden board supporting the baby chick on the screen to destroy it. The goal of this game is to teach about the laws of gravity and Newton's laws. The goal is to deliver the chick to the Main Hero (Chicken). Players can also tilt the phone screen to activate the gyro to move the Main Hero to catch the baby chicken.

3.2 Testing

After creating the edutainment, the next step is testing. In the testing conducted in this study, the test targets were 30 users who would later be tested for user satisfaction using the SUS questionnaire to produce ratings and average SUS scores[30]. To see the rating of the SUS score, it can be seen in Figure 9.

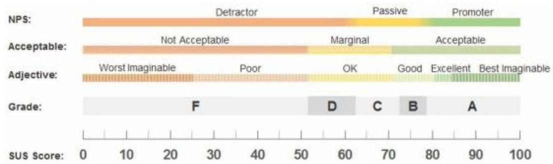


Figure 9. SUS Rating Scale

3.2.1 Testing Performance (Results for RQ1)

Once the target has been identified, the system testing task scenario is defined to evaluate the functionality and effectiveness of the system[31]. A detailed overview of this scenario is provided in Table 2.

Table 2. System Testing Task Scenario

Task	Scenario description	Success criteria
TS1	Display the Main Menu	App launches and the Main Menu shows Play, Learn, and Credits, all tappable.
TS2	Enter the World Map	From the Main Menu, tap Play and the World Map loads with the hero and chick counter visible.
TS3	Navigate the World Map	Tap a destination node and the hero moves; a level entry prompt appears.
TS4	Start a Puzzle Level	Enter a level; the objective is shown and physics elements load correctly.
TS5	Complete a Puzzle Level	Remove wooden boards and tilt the device at least once; the chick reaches the main hero; the chick counter increases and any newly unlocked stage becomes accessible.
TS6	Return to Main Menu mid-level	Open Pause during play and choose Return to Main Menu; the Main Menu appears and prior progress is preserved.
TS7	Open and review Learn Menu	From the Main Menu, open Learn, view one topic, then return to the Main Menu without layout or navigation errors.
TS8	Open Credits	From the Main Menu, open Credits and return to the Main Menu; credits display fully and navigation back works.

With 30 participants (N) that given task (t), the first formula that we need to count is to counting the success rate that can be seen below[32].

$$SuccessRate_t = \frac{\#Successful\ participants\ on\ t}{N} \times 100$$

After we found the success rate, we also counting the average time on task (s, successes only) [33].

$$\bar{T}_t = \frac{1}{S_t} \sum_{i=1}^{S_t} T_{i,t}$$

Where S_t is the number of successful completions for task ttt and $T_{i,t}$ is the completion time in seconds for successful participant *i*. It's standard to compute time from successful attempts only, with failed attempts treated as censored. Report median too for robustness.

After that we count average errors (count per participant).

$$\bar{E}_t = \frac{1}{N} \sum_{i=1}^{N} E_{i,t}$$

Where $E_{i,t}$ is the number of observable slips, mistakes, or recoveries logged during task *t*. This indexes efficiency and helps diagnose UI friction. Task-performance overview. Table 3 summarizes outcomes for the eight tasks (TS1–TS8), success rate, time on task (medians from successful attempts only), errors per participant, and error-free rate. The task scenarios are described in 2.6.

Table 3. Task Performance

Task	Success rate %	Avg time (success)	Median time (success)	Avg errors	Error-free rate %
TS1	96.7	6.58	6.53	0.03	96.7
TS2	93.3	7.94	7.79	0.13	86.7
TS3	96.7	11.97	11.73	0.20	83.3
TS4	90.0	7.49	7.52	0.20	86.7
TS5	83.3	35.86	35.74	0.52	63.3
TS6	93.3	8.14	8.24	0.23	83.3
TS7	93.3	9.04	9.03	0.10	90.0
TS8	100.0	6.47	6.55	0.07	93.3

Across the eight tasks, participants completed core flows with high effectiveness and short times. Launching the app (TS1) showed 96.7% success (Md = 6.53 s). Entering the World Map (TS2) achieved 93.3% (Md = 7.79 s). World Map navigation (TS3) was readily understood (96.7%; Md = 11.73 s). Starting a level (TS4) worked for most users (90.0%; Md = 7.52 s). The physics puzzle (TS5) was intentionally more demanding—83.3% success with a longer Md = 35.74 s, 0.52 errors on average, and 63.3% error-free—suggesting an opportunity to clarify in-level affordances without reducing challenge. Returning to the Main Menu mid-level (TS6) was generally successful (93.3%; Md = 8.24 s). Accessing Learn (TS7) and Credits (TS8) was straightforward (93.3%; Md = 9.03 s) and 100.0%; Md = 6.55 s, respectively.

These task-performance results address RQ1; perceptions of usability (RQ2) are reported next (Table 4).

The physics puzzle demanded more effort, which fits its learning goals. On TS5, the success rate was 83.3 percent and the median completion time was 35.74 seconds. Participants made more observable errors during this task than in the navigation tasks, with an average of 0.52 errors and an error-free rate of 63.3 percent, suggesting that small affordance cues could help guide first-time play without reducing the intended challenge. Exiting a puzzle mid-level and returning to the Main Menu was generally successful. On TS6, the success rate was 93.3 percent with a median time of 8.24 seconds. Accessing learning content and credits was uncomplicated. On TS7, the success rate was 93.3 percent with a median time of 9.03 seconds, and on TS8 every participant completed the task, yielding a success rate of 100.0 percent with a median time of 6.55 seconds. These results indicate that the frame of the experience is robust and that the primary opportunity for improvement lies in clarifying puzzle interactions.

3.2.2 Perceived Usability (SUS) (Results for RQ2)

Immediately after completing the eight task scenarios, all thirty participants filled out the ten-item System Usability Scale[34]. The overall SUS mean was 87.42, the median was 87.50, the standard deviation was 5.02, the minimum was 77.50, and the maximum was 95.00. Using the common letter-grade interpretation, twenty-eight participants fell into grade A and two participants into grade B, indicating a consistently positive perception of usability across the sample. These results align with the task performance findings, where navigation and menu flows reached high completion rates, and suggest that users not only could complete the primary actions but also felt confident and satisfied while doing so.

The System Usability Scale produces a single 0–100 score from ten Likert items. Each response is on a five-point scale from one to five. To compute a participant's score, first recode the odd-numbered items by subtracting one from each response, then recode the even-numbered items by subtracting each response from five. Sum these ten adjusted values to obtain S, then multiply by 2.5 to place the result on a 0–100 scale [35].

In formula form:

$$SUS = 2.5 \times [(Q1-1)+(5-Q2)+(Q3-1)+(5-Q4)+(Q5-1)+(5-Q6)+(Q7-1)+(5-Q8)+(Q9-1)+(5-Q10)]$$

Using this scoring, we administered the following ten questions to all participants.

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

The ten items cover complementary aspects of perceived usability so the final score reflects a balanced attitude rather than a single impression. The first item asks about the desire to use the system frequently, which captures overall appeal. The second probes whether the system feels needlessly complex, offering an early check on cognitive load. The third asks whether the system seems easy to use, which anchors general ease. The fourth explores whether users feel they need technical support, highlighting perceived self-sufficiency. The fifth looks at how well the system's functions fit together, tapping perceived integration. The sixth asks about inconsistency, which can signal problems with patterns and standards. The seventh considers how quickly most people would learn the system, reflecting perceived learnability. The eighth asks whether the system feels cumbersome, pointing to friction in interaction. The ninth invites users to report confidence while using the system, which is often built through clear feedback and predictable behavior. The tenth asks whether a lot of learning was required before getting started, which helps reveal barriers at first contact. Together, the odd items tend to be affirming while the even items are deliberately critical, and the scoring balances these voices into a single, interpretable index of perceived usability.

The result of the SUS can be seen from table 4.

Table 4. SUS results summary

Statistic	Value
Mean SUS	87.42
Median SUS	87.50
Standard deviation	5.02
Minimum	77.50
Maximum	95.00
Grade distribution	A: 28, B: 2

Table 4 summarizes perceived usability for thirty participants. The mean score of 87.42 and the median of 87.50 indicate consistently high perceived usability with only modest spread, as reflected in the standard deviation of 5.02. Scores ranged from 77.50 to 95.00, showing that even the lowest individual rating remained in the acceptable band while many participants approached the upper end of the scale. The grade distribution reinforces this picture, with twenty-eight participants in grade A and two in grade B, suggesting that users generally found the prototype easy to operate, coherent in its functions, and confidence-building during use. Read together with the task performance findings, these results imply that the interface not only supports successful completion of key actions but also feels intuitive and reliable to most users.

3.3 Distribution

The edutainment application is distributed through the school's official website at https://edu.smpn-27bdl.sch.id/ipa (accessed August 26, 2025).

3.4 Discussion

Main findings (answers to RQ1 & RQ2). Core navigation and access flows (launching the app, entering and navigating the World Map, returning to the Main Menu, and opening Learn/Credits) showed high effectiveness (generally ≥90% success) with short median times, indicating that the information architecture and labelling are clear for Grade VII learners. The physics-puzzle task was intentionally more demanding and produced lower success with longer times and some errors, which is expected for a challenge-focused activity. Despite this, perceived usability was very high (SUS ≈ 87), suggesting learners found the experience coherent and easy to use overall.

Interpretation. The strong performance on menu and navigation tasks points to effective structuring of entry points (Main Menu → World Map → Level) and consistent UI labels. The relative difficulty in-level likely stems from increased cognitive/ motor demands (timing, coordination, discovery of mechanics). Because SUS remained high, the added challenge did not undermine overall satisfaction—learners appear to distinguish between navigating to content (frictionless) and mastering the puzzle mechanics (appropriately effortful).

Design implications. Keep the current Main Menu/World Map structure. Focus refinements inside levels: (a) add first-run micro-tutorials or staged hints, (b) strengthen affordances (e.g., subtle animations on tappable/tiltable elements), (c) provide on-demand hints instead of global help, and (d) ensure clear feedback on success/failure states. These changes target the specific task with lower success while preserving desirable challenge.

Educational implications. For classroom use, the prototype fits short (≈15–20 minute) activities tied to mechanics topics. A brief teacher-led pre-brief (goal + controls) followed by individual play can minimize early slips. Because out-of-level navigation is already robust, teachers can reliably move students between Learn content and gameplay within one session.

Threats to validity. Findings are based on a single school context, one session, and a limited device set; results may vary with different hardware or repeated exposure. Timing data are susceptible to classroom interruptions; although controlled, some variability is inevitable. SUS reflects immediate perceptions; multi-session use could change ratings through learnability effects.

4. Conclusion

This study designed and evaluated a browser-based edutainment prototype for Grade VII science and found that learners achieved high success with short median times on navigation and access flows (launching the app, entering and navigating the World Map, returning to the Main Menu, and opening Learn/Credits), while the in-level physics puzzle was intentionally more demanding and produced longer times and some errors—an acceptable tradeoff for challenge-focused gameplay. Perceived usability was very high (SUS ≈ 87), indicating the experience is coherent and easy to use in the target classroom context. Practically, the prototype is classroom-ready for short activities aligned with mechanics topics; the most valuable refinements are inside levels (first-run micro-tutorials, clearer affordances, and on-demand hints). Limitations include a single school context, one session, and a limited device/browser set, with timing susceptible to classroom interruptions. Future work should test multi-session deployments to observe learnability, broaden device/browser coverage, and link gameplay to learning outcomes using pretest–posttest designs.

5. Acknowledgement

The authors thank SMP Negeri 27 Bandar Lampung for access to facilities and participants and appreciate the support from the school's ICT team

during deployment and testing.

References

[1] I. W. Suja, *Keterampilan Proses Sains dan Instrumen*. PT. RajaGrafindo Persada-RajaGrafindo Persada, 2023.

[2] H. Elselia, "Penggunaan Inkuiri Terbimbing untuk Meningkatkan Keterampilan Sains Siswa di Sekolah Menengah Pertama," *J. Didakt. Pendidik. Dasar*, vol. 7, no. 2, pp. 639–660, 2023, doi: 10.26811/didaktika.v7i2.1025.

[3] A. Mulkani, "CP & ATP - Ilmu Pengetahuan Alam Ipa Fase d," Simpul IPA. Accessed: Aug. 25, 2025. [Online]. Available: https://guru.kemdikbud.go.id/kurikulum/referensi-penerapan/capaian-pembelajaran/sd-sma/ilmu-pengetahuan-alam-ipa/fase-d/%0Ahttps://static.perangkat-ajar.belajar.id/USR.zkWNxRAomR.1690519544-25JkvbkWGv.pdf

[4] Sufyadi *et al.*, *Panduan Pembelajaran dan Asesmen Jenjang Pendidikan Dasar dan Menengah (SD/MI, SMP/MTs, SMA/SMK/MA)*. 2021. [Online]. Available: https://repositori.kemendikdasmen.go.id/24921/1/Panduan Pembelajaran dan Asesmen.pdf

[5] M. S. Hidayatullah, R. S. Dewi, H. Thaibah, I. Larasaty, R. Simanjuntak, and M. Anggraini, "Mengukur dan Memahami Penalaran," 2024, *Komojoyo Press*.

[6] N. Bhaw, J. Kriek, and M. Lemmer, "Insights from coherence in students' scientific reasoning skills.," *Heliyon*, vol. 9, no. 7, p. e17349, Jul. 2023, doi: 10.1016/j.heliyon.2023.e17349.

[7] H. R. Martín, *How do we learn?: A scientific approach to learning and teaching (evidence-based education)*. John Wiley & Sons, 2024.

[8] M. Krell, A. Vorholzer, and A. Nehring, "Scientific Reasoning in Science Education: From Global Measures to Fine-Grained Descriptions of Students' Competencies," 2022. doi: 10.3390/educsci12020097.

[9] R. Tejero, "The Effectiveness of Science Investigatory Project on Students' Science Process Skills," *Int. J. Res. Innov. Soc. Sci.*, vol. IX, pp. 3632–3649, Feb. 2025, doi: 10.47772/IJRISS.2025.9010290.

[10] G. Gizaw and S. Sorsa, "Improving Science Process Skills of Students: A Review of Literature," *Sci. Educ. Int.*, vol. 34, pp. 216–224, Sep. 2023, doi: 10.33828/sei.v34.i3.5.

[11] M. Kalogiannakis, S. Papadakis, and A.-I. Zourmpakis, "Gamification in Science Education. A Systematic Review of the Literature," 2021. doi: 10.3390/educsci11010022.

[12] M. Arztmann, L. Hornstra, J. Jeuring, and L. Kester, "Effects of games in STEM education: a meta-analysis on the moderating role of student background characteristics," *Stud. Sci. Educ.*, vol. 59, no. 1, pp. 109–145, Jan. 2023, doi: 10.1080/03057267.2022.2057732.

[13] M. Sailer, R. Maier, S. Berger, T. Kastorff, and K. Stegmann, "Learning activities in technology-enhanced learning: A systematic review of meta-analyses and second-order meta-analysis in higher education," *Learn. Individ. Differ.*, vol. 112, p. 102446, 2024, doi: https://doi.org/10.1016/j.lindif.2024.102446.

[14] J. Schoenherr, A. R. Strohmaier, and S. Schukajlow, "Learning with visualizations helps: A meta-analysis of visualization interventions in mathematics education," *Educ. Res. Rev.*, vol. 45, p. 100639, 2024, doi: https://doi.org/10.1016/j.edurev.2024.100639.

[15] S. Sintaro, A. Surahman, and N. Khairandi, "APLIKASI PEMBELAJARAN TEKNIK DASAR FUTSAL MENGGUNAKAN AUGMENTED REALITY BERBASIS ANDROID," *TELEFORTECH J. Telemat. Inf. Technol.*, vol. 1, no. 1, pp. 22–31, 2020.

[16] "Fasilitas Sekolah – SMPN 27 BANDAR LAMPUNG." Accessed: Aug. 25, 2025. [Online]. Available: https://smpn-27bdl.sch.id/fasilitas-sekolah/

[17] Kemendikbud, *Capaian Pembelajaran Mata Pelajaran Ilmu Ilmu Pengetahuan Sosial Fase D: Untuk Kelas VII – IX SMP/MTs/Program Paket B*. 2022. [Online]. Available: https://kurikulum.kemdikbud.go.id/file/cp/dasmen/19. CP IPS.pdf

[18] H. Alberida *et al.*, *Evaluasi Proses Dan Hasil Pembelajaran Biologi Pendekatan Teoritis Dan Aplikatif*. Deepublish, 2024.

[19] A. Ridwan, "ANALISIS KEMAMPUAN PENALARAN ILMIAH SISWA DALAM PEMBELAJARAN BERBASIS PROYEK MATA PELAJARAN IPA DI SMPN 2 SAROLANGUN," *EDU Res.*, vol. 5, no. 2, pp. 85–95, 2024.

[20] H. Sindhi, "Pedagogies of Reform: The Role of Game-Based Learning in Recontextualizing Education," 2023, *City University of New York*. [Online]. Available: https://academicworks.cuny.edu/gc_etds/5551/

[21] F. Strada, "Augmented reality and serious games for learning : exploring potentialities , assessing effectiveness , and investigating user

experience .," *Thesis*, 2021.

[22] D. W. Wilson, J. S. Valacich, and J. L. Jenkins, "Behaviorally Measuring Usability By Analyzing Users' Mouse Movement Efficiency1," *MIS Q. Manag. Inf. Syst.*, vol. 48, no. 3, pp. 979–1008, 2024, doi: 10.25300/MISQ/2023/17900.

[23] T. P. de Campos, S. Delabrida, and N. M. C. Valentim, "Usability and UX Evaluation in a Mixed Reality Puzzle Game Using Questionnaires," *J. Interact. Syst.*, vol. 16, no. 1, pp. 589–605, 2025, doi: 10.5753/jis.2025.6089.

[24] M. G. Apolinário, "JellyfishGO!: Exploring Citizen Science Users' Engagement with Gamification Strategies," 2025.

[25] M. R. Siregar and N. Nelmiawati, "Game 3D 'Lawan Narkoba' Menggunakan Metode Game Development Life Cycle (GDLC)," *J. Appl. Multimed. Netw.*, vol. 4, no. 1, pp. 24–31, 2020.

[26] N. F. B. Ridzuan, S. K. B. Jali, A. B. M. Tazuddin, N. B. Jali, M. I. Bin Bandan, and L. P. Chin, "ARventure: Edutainment Meets Science Mobile Application," in *2022 IEEE International Conference on Computing, ICOCO 2022*, IEEE, 2022, pp. 130–135. doi: 10.1109/ICOCO56118.2022.10031868.

[27] S. Sintaro, "Accounting Edutainment an Alternative Accounting Learning (Dictionary and Accounting terms)," *J. Ilm. Sist. Inf. Akunt.*, vol. 3, no. 1, pp. 1–12, 2023.

[28] S. Sintaro, "RANCANG BANGUN GAME EDUKASI TEMPAT BERSEJARAH DI INDONESIA," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 1, no. 1, pp. 51–57, 2020.

[29] Z. Draissi, Y. Rong, and M. A. Ebrahim Suliman, "Employability and Education Mismatch in the Moroccan Urban Workplace," *SAGE Open*, vol. 13, no. 4, p. 21582440231217890, 2023, doi: 10.1177/21582440231217890.

[30] J. Lewis and J. Sauro, "USABILITY AND USER EXPERIENCE: DESIGN AND EVALUATION," 2021, pp. 972–1015. doi: 10.1002/9781119636113.ch38.

[31] I. Kirana, I. Nugraha, and R. Novasani, "Comparative Usability Analysis of University Department Websites: A Case Study of Industrial and Chemical Engineering Programs," *Techno.Com*, vol. 24, pp. 1025–1034, Aug. 2025, doi: 10.62411/tc.v24i3.13778.

[32] N. A. Nik Ahmad and N. Hasni, *ISO 9241-11 and SUS Measurement for Usability Assessment of Dropshipping Sales Management Application*. 2021. doi: 10.1145/3457784.3457794.

[33] P. Weichbroth, "Usability Testing of Mobile Applications: A Methodological Framework," *Appl. Sci.*, vol. 14, pp. 1–27, Feb. 2024, doi: 10.3390/app14051792.

[34] S. Kurniawan, A. I, and A. Ependi, "Analisis Usability Aplikasi C-Access Commuterline Menggunakan System Usability Scale (Sus)," *J. Syntax Admiration*, vol. 4, pp. 894–911, Jul. 2023, doi: 10.46799/jsa.v4i7.671.

[35] M. Fiandhika and B. Kelana, "System Usability Scale Validation from The Expert Perspective," *E3S Web Conf.*, vol. 483, Jan. 2024, doi: 10.1051/e3sconf/202448303008.