GamifyPlay: A Personalized Gamified System for Assessing Reflective Abstraction through the Tower of Hanoi

Yuri Mercedes Bermúdez Mazuera[1,*]

[1]Unidad Central del Valle del Cauca, Tuluá, Colombia

Abstract

This system-description paper presents GamifyPlay, a personalized gamified ecosystem for supporting and assessing reflective abstraction through the Tower of Hanoi in secondary education. The ecosystem integrates GamifyTest for learner profiling, GamifyPlay for interactive problem solving, Genially for visual scaffolding, Mentimeter for classroom interaction, and Moodle for storing assessment evidence. Reflective abstraction was operationalized as interiorization, coordination, encapsulation, and generalization through four questionnaires containing 9, 7, 7, and 4 items. A one-group pretest–posttest evaluation used complete paired records from 30 sixth-grade students. Mean posttest scores increased in all four forms. Wilcoxon signed-rank tests identified statistically significant differences in interiorization, coordination, and encapsulation; generalization improved descriptively but was not significant. The evaluation shows how the ecosystem connects learner profiles, interactive resources, student perceptions, and traceable assessment records. It also identifies transfer tasks, prediction activities, and richer interaction logs as priorities for redesign.

Keywords

reflective abstraction, gamified assessment, computational thinking, Tower of Hanoi, educational technology

1. Introduction

Assessing computational thinking in schools is challenging because complex cognitive competences cannot be represented adequately by a single final product or correct answer [1, 2]. Computational thinking involves processes that may become visible at different moments of a task. Assessment environments should therefore capture not only whether a learner reaches a solution, but also the representations, decisions, and transformations used while solving the problem.

Abstraction is particularly difficult to observe because learners must select relevant information, construct representations, coordinate actions, and transfer solution structures to new situations [3, 4]. These operations are related but not equivalent. A student may understand the rules without coordinating them into a valid sequence, or may solve a familiar instance without recognizing the underlying pattern. An assessment design should consequently specify the evidence expected at different levels of reasoning.

Gamification can support this process through progression, immediate feedback, interactive representations, repeated attempts, and opportunities to revise strategies [5, 6, 7]. Nevertheless, many educational implementations emphasize motivation or participation without explaining how the system produces evidence of a specific competence. An engaging activity becomes an assessment system only when the target construct, tasks, generated records, and analytical procedures are explicitly aligned.

This paper describes GamifyPlay as an integrated ecosystem for assessing reflective abstraction through the Tower of Hanoi. The problem requires students to understand constraints, coordinate legal moves, recognize intermediate states, and infer patterns. The contribution comprises: (1) a system that connects profiling, interactive problem solving, visual scaffolding, classroom interaction, and assessment records; (2) an operational model that separates interiorization, coordination, encapsulation, and

NALEF 2026, Cartagena, Colombia

*Corresponding author.

ybermudez@uceva.edu.co (Y. M. B. Mazuera)

https://orcid.org/0000-0002-6474-7424 (Y. M. B. Mazuera)

© 2026 Copyright for this paper by its authors. Use permitted under CC BY 4.0. NALEF 2026, Cartagena, Colombia.

Table 1
Components of the GamifyPlay assessment ecosystem

Component	Role	Evidence or output
GamifyTest	Characterizes learning preferences and HEXAD player types.	Profiles used to select interaction, feedback, and exploration strategies.
GamifyPlay	Presents the Tower of Hanoi as an interactive gamified problem.	Interactive exploration of legal moves, strategies, and problem states.
Genially	Provides rules, examples, hints, and preparatory activities.	Visual scaffolds for understanding constraints and validating moves.
Mentimeter	Supports anonymous collective responses and discussion.	Perceptions, reported difficulties, and qualitative comments.
Moodle	Administers the four pretest and posttest questionnaires.	Item responses and normalized scores for quantitative analysis.

generalization; and (3) an empirical evaluation with 30 students that informs technical and pedagogical redesign.

2. Related Work

Gamification uses game design elements in non-game contexts [5]. In education, challenges, feedback, progression, and interactive resources can support sustained effort in structured problem solving [6, 8]. However, points, badges, or rankings do not by themselves demonstrate conceptual understanding. Their educational value depends on how the selected mechanics are connected to the intended learning processes.

Personalized gameful design relates interaction and feedback decisions to user characteristics rather than applying the same incentive structure to all learners [7]. This is relevant in computing education, where students differ in how they explore examples, collaborate, respond to feedback, and revise strategies. Learner profiles can inform visual representations, progression, social interaction, and opportunities for autonomous exploration, but they should be treated as contextual design information rather than fixed categories of ability.

Computational thinking has been assessed through tests, rubrics, observations, interviews, and product analysis [9, 10, 11]. For abstraction, the central difficulty is defining observable evidence because it is not a homogeneous skill [3, 4]. Reflective abstraction explains how actions are internalized, coordinated, transformed into stable structures, and projected toward broader formulations [12]. The present work addresses the gap between gamified interaction and competence assessment by integrating these processes within a traceable workflow.

3. System Description

3.1. Design Objectives and Components

The system was designed to represent the Tower of Hanoi visually and interactively; adapt selected interactions to predominant learner profiles; support assessment without reducing the experience to rewards; collect separate evidence for four forms of abstraction; and preserve traceability from classroom activity to Moodle records. It is an ecosystem rather than a single application because specialized and general-purpose tools perform complementary functions (Table 1).

Each component was assigned a distinct function. GamifyTest informed the design decisions; Genially introduced and reinforced the rules; GamifyPlay provided the main problem-solving experience; Mentimeter opened a channel for rapid classroom reflection; and Moodle organized comparable pretest and posttest records. This modular organization allows individual tools to be replaced or extended without modifying the conceptual assessment model.

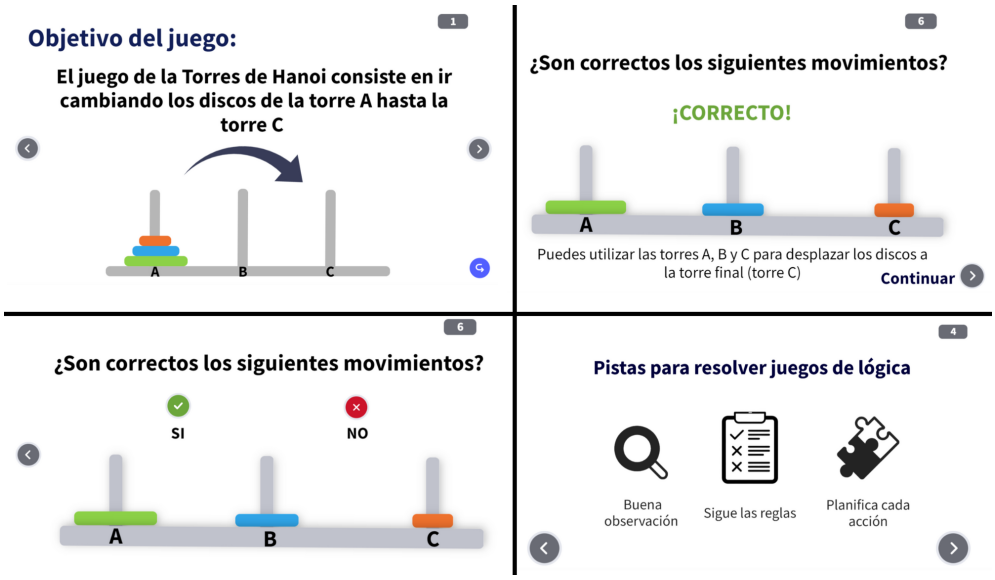

Figure 1: Genially screens used for visual scaffolding, rule presentation, and move validation.

Table 2
Operationalization of reflective abstraction

Form	Assessment focus	Items
Interiorization	Essential elements, rules, disk relations, and symbolic representations.	9
Coordination	Integration of legal moves and intermediate states into coherent sequences.	7
Encapsulation	Interpretation of dynamic sequences as stable states that can be judged.	7
Generalization	Transfer of solution patterns, including disks and minimum moves.	4

3.2. Personalization and Gamification Design

The design was guided by Marczewski's Gamification Design Framework [13]. GamifyTest combines Felder–Silverman learning-style dimensions with HEXAD player types. The predominant characteristics were Active, Sensing, Visual, and Sequential preferences, together with Socializer and Disruptor player types. These results were used as design criteria rather than fixed labels or predictors of student performance.

Visual and sequential preferences informed diagrams and progressive explanations. Active and sensing preferences supported manipulation and concrete problem states. Socializer characteristics informed collective interaction, whereas Disruptor characteristics motivated autonomous exploration and alternative strategies. The goal was not to produce a separate application for every student, but to align the principal interaction pattern with the characteristics identified in the group.

The objective was to assess reflective abstraction, not merely reward task completion. Consequently, progression, feedback, interaction, and strategy revision were prioritized over isolated points or leaderboards. Students encountered the rules visually, judged proposed moves, interacted with problem states, discussed solutions, and completed structured items. Figure 1 shows representative Genially screens that move from basic rules toward move validation and strategic planning.

3.3. Assessment Model and Workflow

Reflective abstraction was operationalized into four forms associated with separate Moodle questionnaires (Table 2). The common Tower of Hanoi context reduced variation in domain knowledge, while the item sets required different cognitive operations. Maintaining separate scores was important because improvement in rule recognition does not necessarily imply improvement in transfer.

The workflow comprised six stages: student characterization; translation of profiles into design

decisions; introduction of rules with Genially; interaction and feedback through GamifyPlay; collective responses through Mentimeter; and administration and export of questionnaire data through Moodle. The interaction layer (Genially and GamifyPlay), participation layer (Mentimeter), and assessment layer (Moodle) were therefore connected to the preceding profiling stage.

Each questionnaire was normalized independently from 0 to 100. Mentimeter did not replace the assessment instruments; it provided contextual information about perceived ease, difficulty, and interaction problems. This distinction allowed quantitative and exploratory qualitative evidence to complement each other without being merged into a single effectiveness score.

4. Evaluation Method

4.1. Context, Participants, and Procedure

The system was evaluated with sixth-grade students at a public educational institution in Valle del Cauca, Colombia, using a one-group pretest–posttest design. The analytical dataset contained 30 students with complete records at both measurement moments, and the student was the unit of analysis. Restricting the analysis to paired observations ensured that each posttest score was compared with the corresponding student's pretest score.

The evaluation was conducted in three sessions. Interiorization and Coordination were administered in the first pretest session, and Encapsulation and Generalization in the second. Students then completed the visual, interactive, and feedback activities. In the third session, the four posttest questionnaires were administered in two blocks. Moodle stored item responses and normalized scores.

Mentimeter prompts gathered perceptions and difficulties as exploratory evidence. Their purpose was to identify activities that students considered easy or difficult and to detect possible conceptual or interface-related problems. Because no formal qualitative coding protocol was defined, these comments were summarized descriptively and were not converted into an effectiveness score.

4.2. Data Analysis

Mean scores and absolute changes summarized performance across moments. The Wilcoxon signed-rank test was applied separately to each form because observations were paired and parametric distributional assumptions were not adopted. Exact two-sided p-values and counts of positive and negative changes were retained. Statistical significance was evaluated at $\alpha = .05$.

The analysis examines whether the system registered differentiated behavior across the four theoretically defined forms. It does not claim that the intervention alone caused the changes because there was no control group and no comparison between personalized and non-personalized versions.

5. Results

5.1. Student Perceptions

Figure 2 shows 16 responses to a prompt about the easiest and most difficult aspects of the activities. Students frequently described choosing answers, ordering elements, or completing the first questionnaire as relatively easy. Difficulties concerned true/false judgments, deciding whether movements were valid, manipulating pieces, and responding to particular later questions. One response also described completing a difficult questionnaire with a classmate.

The comments identify useful redesign points. Choice-based visual interaction appeared accessible, whereas validation and manipulation were not uniformly easy. These tasks require coordination of constraints and intermediate states, but difficulty may also arise from interface behavior. Future evaluations should therefore distinguish conceptual errors from usability problems. The peer-support response is consistent with the social interaction included for the predominant Socializer profile, although an isolated response cannot establish a systematic effect.

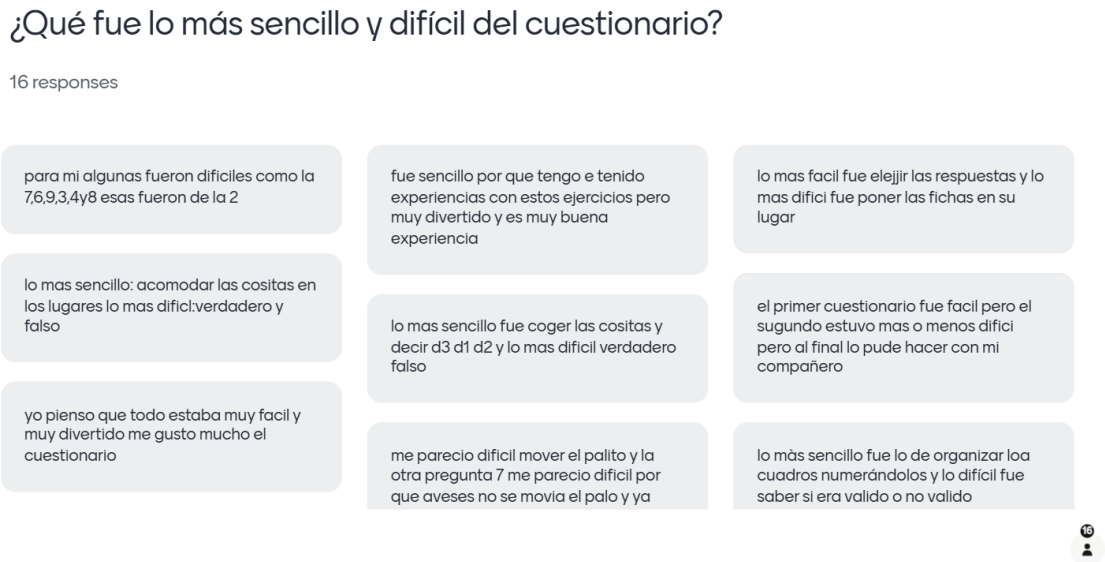

Figure 2: Mentimeter feedback on easy and difficult aspects of the activities (responses in Spanish).

Table 3
Pretest–posttest results by form of reflective abstraction

Form	Pre	Post	Δ	C^+/C^-	W	Exact p
Interiorization	42.8	57.5	+14.8	23/7	72.5	0.000555
Coordination	49.1	61.2	+12.1	20/10	107.0	0.008705
Encapsulation	36.6	47.0	+10.5	22/8	77.5	0.001431
Generalization	16.1	21.8	+5.8	16/12[a]	146.5	0.197307

[a]Two tied cases. C^+ and C^- denote positive and negative changes; W is the reported Wilcoxon statistic.

This evidence remains exploratory. Its principal value is formative: it indicates where students experienced friction and where additional explanation, feedback, or interface support may be needed. It should not be interpreted as causal evidence of increased motivation or learning.

5.2. Reflective Abstraction Outcomes

Table 3 integrates descriptive and inferential results. Mean scores increased in all four forms. Positive changes outnumbered negative changes in Interiorization, Coordination, and Encapsulation, and the corresponding Wilcoxon tests were statistically significant. Generalization increased from 16.1 to 21.8, but the difference did not reach statistical significance.

Interiorization showed the largest mean increase, followed by Coordination and Encapsulation. This ordering is consistent with the intervention because students repeatedly encountered the rules, visual relations among disks, legal movement constraints, and intermediate states. The resources provided several opportunities to recognize and coordinate these relationships in different representations.

Generalization followed a different pattern. The system offered fewer explicit opportunities to transfer the solution structure beyond the practiced instance. Generalization requires learners not only to complete a solution, but also to identify an invariant pattern and relate the number of disks to the minimum number of moves. The result therefore identifies a specific limitation in the current activity sequence.

6. Discussion

GamifyPlay integrates resources that are often used as separate classroom tools. Profiling informs interaction design; Genially and GamifyPlay expose visual and manipulable problem states; Mentimeter

surfaces perceptions and difficulties; and Moodle produces structured assessment records. The system contribution lies in coordinating these functions around a shared competence model rather than in presenting any individual platform as novel.

The significant changes in Interiorization, Coordination, and Encapsulation suggest alignment with rule identification, sequence construction, move validation, and interpretation of states. Students encountered these relationships through visual explanations, interactive practice, feedback, discussion, and assessment. The differentiated results also support maintaining four separate scores: a single global abstraction score would have concealed the weaker outcome in Generalization.

Generalization requires transfer rather than the successful resolution of one familiar case. Future versions should sequence two-, three-, and four-disk problems, request predictions before interaction, compare alternative strategies, and prompt students to formulate the $2^n - 1$ relationship. Tasks could also ask learners to explain why the recursive pattern is preserved or estimate the number of moves for an unseen configuration.

The qualitative responses highlight an additional methodological issue. Difficulties with validating movements or manipulating pieces may reflect conceptual reasoning, interface usability, or both. Future logging should capture attempts, illegal moves, response time, hint use, and strategy changes. Relating these indicators to questionnaire scores would help distinguish conceptual obstacles from interaction problems and improve the interpretation of the assessment.

Personalization made the design rationale explicit, but the evaluation did not compare personalized and non-personalized conditions. The findings therefore support the feasibility of the integrated ecosystem rather than the isolated effectiveness of personalization. A subsequent study could compare alternative interaction designs and examine whether profile-informed resources produce different engagement or performance patterns.

The main limitations are the one-group design, one school, one principal problem, and analysis restricted to complete paired records. The perception data were exploratory and were not analyzed through a formal qualitative protocol. These limitations prevent strong causal claims and restrict generalization to other contexts. Nevertheless, the evaluation demonstrates that the ecosystem can generate distinct evidence for four forms of reflective abstraction in an authentic school setting and can identify concrete priorities for redesign.

7. Conclusions

GamifyPlay combines learner profiling, visual scaffolding, interactive problem solving, collective feedback, and Moodle-based assessment within a traceable workflow. Its main contribution is the alignment between a gamified learning experience and an operational model of reflective abstraction. Rather than reporting only an overall achievement score, the system distinguishes Interiorization, Coordination, Encapsulation, and Generalization.

With 30 students, mean posttest scores increased in every form, with statistically significant differences in Interiorization, Coordination, and Encapsulation. Generalization improved descriptively but remained the most demanding form. This pattern provides an actionable basis for redesign through explicit transfer missions, multiple problem sizes, prediction tasks, usability measures, and richer interaction logs.

The study should be understood as an initial system evaluation rather than definitive causal evidence. Future work should validate the instruments more extensively, compare personalized and non-personalized conditions, incorporate process indicators, and replicate the evaluation with additional groups and problem contexts. Overall, GamifyPlay shows how engagement-oriented resources can be connected with differentiated and traceable evidence of computational thinking.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT to support language editing and manuscript restructuring. The authors reviewed and revised the resulting text and take full responsibility for the publication's content.

References

[1] A. Piatti, G. Adorni, L. El-Hamamsy, L. Negrini, D. Assaf, L. M. Gambardella, F. Mondada, The ct-cube: A framework for the design and the assessment of computational thinking activities, Computers in Human Behavior Reports 5 (2022) 100166. doi:10.1016/j.chbr.2021.100166.

[2] R. F. Zúñiga Muñoz, , J. A. H. Alegría, C. A. Collazos, H. Fardoun, Childprogramming evolution, a method to increase the computational thinking skills in school, in: Human-Computer Interaction: 4th Iberoamerican Workshop, HCI-Collab 2018, Revised Selected Papers, Springer, Cham, 2019, pp. 57–69. doi:10.1007/978-3-030-05270-6_5.

[3] U. Cakiroglu, I. Cevik, A framework for measuring abstraction as a sub-skill of computational thinking in block-based programming environments, Education and Information Technologies 27 (2022) 9455–9484. doi:10.1007/s10639-022-11019-2.

[4] E. Gun-Tosik, T. Guyer, The effect of computer science unplugged on abstraction as a sub-component of computational thinking, Thinking Skills and Creativity 53 (2024) 101552. doi:10.1016/j.tsc.2024.101552.

[5] S. Deterding, D. Dixon, R. Khaled, L. Nacke, From game design elements to gamefulness: Defining gamification, in: Proceedings of the 15th International Academic MindTrek Conference, 2011, pp. 9–15. doi:10.1145/2181037.2181040.

[6] K. M. Kapp, The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education, Pfeiffer, San Francisco, 2012.

[7] A. Mora, G. F. Tondello, L. E. Nacke, J. Arnedo-Moreno, Effect of personalized gameful design on student engagement, in: IEEE Global Engineering Education Conference (EDUCON), 2018, pp. 1925–1933. doi:10.1109/EDUCON.2018.8363431.

[8] L. A. Wijsman, N. Saab, J. Schuitema, J. H. van Driel, P. M. Westenberg, Promoting performance and motivation through a combination of intrinsic motivation stimulation and an extrinsic incentive, Learning Environments Research 22 (2019) 65–81. doi:10.1007/s10984-018-09273-8.

[9] S. Atmatzidou, S. Demetriadis, Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences, Robotics and Autonomous Systems 75 (2016) 661–670. doi:10.1016/j.robot.2015.10.008.

[10] M. Saritepeci, Developing computational thinking skills of high school students: Design-based learning activities and programming tasks, The Asia-Pacific Education Researcher 29 (2020) 35–54. doi:10.1007/s40299-019-00480-2.

[11] C. Angeli, N. Valanides, Developing young children's computational thinking with educational robotics: An interaction effect between gender and scaffolding strategy, Computers in Human Behavior 105 (2020) 105954. doi:10.1016/j.chb.2019.03.018.

[12] J. Piaget, Studies in Reflecting Abstraction, Psychology Press, Hove, 2001.

[13] A. Marczewski, Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design, CreateSpace Independent Publishing Platform, 2015.