Geolocation and Gamification to Assess Mathematical Reasoning: Motivational and Innovative Strategy

Norka Bedregal-Alpaca*, Sidanelia Flores-Silva, Víctor Cornejo-Aparicio, Karim Guevara
Universidad Nacional de San Agustín de Arequipa, Arequipa, Perú

***Abstract*—The development of logical-mathematical reasoning skills is essential for fostering problem-solving and critical thinking. However, traditional teaching and assessment methods may limit the practical application of these skills. Exam-related tension often causes mental and physiological blocks that affect performance, making it necessary to design motivational assessment strategies that reduce anxiety and sustain engagement. This presents the use of geolocation and gamification as innovative, motivating strategies for implementing objective tests to evaluate mathematical reasoning. A correlational study was carried out with a quantitative approach and quasi-experimental design. Seventy students aged between 14 and 16 years (52.8% males and 47.2% females) were involved in the study. The validation of the objective test was carried out using the Expert Judgment strategy; the questionnaire administered had a Cronbach's α of 0.81. The reliability of the geolocated and gamified objective test was 0.92, and the difficulty and discrimination indices revealed results consistent with the hypothesis. Student perception was evaluated across intrinsic motivation, extrinsic motivation, and game characteristics. Results showed that 84% rated the experience as very good and 15% as good, with no students reporting fair or poor levels of motivation. Findings of this research indicate that the integration of gamification and geolocation can be effective in encouraging students' interest in the assessment of mathematical reasoning, and it is crucial to ensure that mathematical problems are challenging, but also accessible, and that the gaming experience is engaging and stimulating for students.**

***Keywords*—Evaluation; motivation for studies; practical examination; mathematic game; educational innovation**

I. Introduction

Developing logical-mathematical reasoning skills is crucial for developing other skills such as problem-solving, critical thinking, informed decision-making, abstract thinking, inductive and deductive reasoning. These skills have a positive impact on life both academically and professionally, as well as personally; they are valuable in an increasingly complex and information-based world.

Logical-mathematical reasoning involves a conscious way of thinking that allows solving problems and reaching appropriate conclusions. This encompasses skills like calculation, numerical thinking, understanding abstract concepts, and identifying relationships between objects, which are applicable beyond mathematics. These skills are developed throughout life and are supported by capacities like observation, imagination, intuition, and logical reasoning [1].

It is thus essential to review both teaching methods and approaches to assessing logical-mathematical reasoning. According to [2], the traditional approach often does not allow students to develop independent logical thinking nor apply their mathematical skills in real-world and temporal contexts. Teachers often face discrepancies between student performance in class and their results in assessments, a situation explained by comparing students' actions and thoughts with teachers' ideal expectations, which do not always consider factors like student tension and motivation. Reference [3] adds that "the way students are evaluated is undoubtedly one of the contextual factors influencing their motivation or demotivation towards learning."

In many cases, learning and evaluation problems are related to high levels of anxiety during exams [4], [5], [6]. Neuroscience supports the idea that excessive tension can produce mental to physiological blocks, making it difficult for the student to make conscious decisions [7]; in the context of logical-mathematical reasoning, this tension can increase even more. When a student is blocked, they risk failing in evaluation situations perceived as threatening. Exam tension is a serious problem that negatively affects performance and can be a significant factor in school failure, even when students do not have learning difficulties or capacity.

Considering this information, it is imperative to include motivational assessment dynamics that emphasize two types of blocks: mental and physiological and simultaneously contribute to maintaining student motivation during an assessment. Reference [8] has analyzed the use of geolocation in different educational situations. Technologies based on geolocation systems and augmented reality are useful didactic tools for addressing new inclusive, motivating, and intercultural learning models in schools [9].

On the other hand, the introduction of game design elements in non-playful contexts positively contributes to student motivation towards learning [10]. About a decade ago, there were over 10,000 mobile applications combining instructional design, educational content, and game elements for interaction [11]. In [12] it introduces the Implementation of Geo-referenced Pervasive Games where knowledge is applied in a real environment with the implementation of location, speed, and temperature sensors.

In this context, this article describes the experience of using geolocation and gamification as motivational strategies in developing an objective logical-mathematical reasoning test.

*Corresponding author.

The test is conducted in the context of a pirate game implemented for mobile devices. During the test, students move between different geolocated points; at each point, an objective test question is shown, and if they answer correctly, they receive rewards within the game's dynamics and are provided with a new geolocated point.

II. Related Works

There is limited direct research on the combined use of gamified and geolocated assessments, specifically for evaluating mathematical reasoning.

Studies such as [13, 14, 15, 16] have shown that gamification has a positive influence on students' mathematical competencies and assessment outcomes, making assessments more engaging and effective for the development of mathematical reasoning. Research by [17, 18] indicates that it can be used for both formative (continuous) and summative (final) assessment.

In an educational context, geolocation refers to the use of location-based technologies to trigger learning activities or assessments based on a student's physical location. While no concrete examples of geolocated assessments for mathematical reasoning have been identified, work by [19] highlights the motivational potential of combining gamification with geolocation, suggesting that this approach can further enhance student engagement and provide context for learning experiences.

III. Methodology

According to [20], the research begins with a deductive basis by establishing an initial hypothesis based on existing theory and then using the inductive method to explore how motivation and learning develop.

The research has a quantitative approach and a quasi-experimental design, as a control group is not available due to practical limitations. It is correlational as it analyzes students' motivational levels resulting from including geolocation and gamification elements in the application of the objective test to evaluate logical-mathematical reasoning.

According to [21], the study population consists of students from a particular educational institution in Arequipa (Peru) offering initial, primary, and secondary education to 1,200 enrolled students.

The sample was obtained through intentional sampling and consisted of 70 students aged between 14 and 16 years (52.8% males and 47.2% females). The inclusion criteria were enrolment in the 2022 academic period in the third year of secondary school and regular attendance at class sessions.

To ensure the operability of the research process:

- An objective test was developed to assess the level of mathematical reasoning.
- Two surveys were designed and implemented, one to gather expert assessment of the objective test and the other to collect student perceptions regarding the implemented strategy.

IV. Educational Intervention

A. Development of the Objective Test

An assessment of "Algebraic Reasoning" was proposed, identifying specific competencies students should develop. Following the recommendations of [22] and [23], eight clear questions were formulated without grammatical discrepancies. The response alternatives were short, did not lead the student to guess, did not give clues to the correct answer, nor favored or harmed any type of student.

For each question, the statement, solution process, correct answer, response alternatives, skills developed with the solution, and competencies to which these skills contributed were listed. Subsequently, each question was reviewed by a group of experts, as for [24], it is necessary to review the content validity considering response processes (test administration) and internal structure (statistical and psychometric characteristics of the test).

B. Characteristics of the Geolocated Space

Reference [8] defines geolocation as the identification of an object at a specific point on the planet known as positioning, capable of returning location coordinates. In this case, an open space with an operational area of around 2400 m² with internet and GPS signal was used. Ten geolocated points were generated, distanced from each other so that there was no signal interference within 20 m around. Consequently, 360 different routes were generated for students to navigate without interrupting each other.

C. Game Description

The game involves the search and capture of educational problems forming a single-response test. Each problem is hidden in real-world locations within a specific programmed area. Players must physically move and actively participate in the game. Solving problems earns points and records time.

The game resembles a pirate treasure hunt where players follow a map route shown by a mobile application to find a hidden chest. When they reach the location, the chest opens and reveals the math challenge, which must be solved to proceed to the next secret location. Here, time, movement, and knowledge are key to success.

Reference [25] mentions that new technologies have contributed to the evolution of gamification, while [26] mentions gamification use: mechanics, elements, and techniques in the context of players, objects, goals, rules, conditions, and challenges, achieving a motivating digital experience. Following [27, 28], the dynamics, mechanics, and components of the game were defined.

The game rules require the participant to navigate the area following the map until reaching the location point that will activate the problem. It is crucial to solve the current problem before addressing the next, establishing a sequential progression. The student faces a limited number of opportunities previously set by the teacher who designed the test. Each opportunity represents a unique geographic node the student must address in order.

Starting from a specific point, the student advances through these nodes until reaching the last one, completing the predetermined number of exercises.

It's important to note that the students can restart the game, retain their current progress and allow them to continue from where they stopped, facilitating the successful completion of the game.

The reward is based on three key elements: the key count, the final score, and the resolution time. These factors contribute to the overall assessment of the participants' performance. The participant accumulates keys throughout the game; each associated with success in solving specific problems. The final score reflects the participant's overall efficiency and accuracy in solving all the problems. Additionally, the time taken to solve each challenge is considered, adding a temporal component that influences the final score and, consequently, the reward obtained at the end of the game.

The sequence to start the game begins with the installation of the APK package (Android Application Package), responsible for distributing and installing the bundled components that make up the game environment. Subsequently, the process continues intuitively by creating a username and password, without complicated parameters.

A menu is displayed: 1) Game, 2) Participants, 3) Questions, 4) Map Configuration, and 5) Monitoring. Additionally, there are help manuals available for download if necessary.

After registering the questions that will constitute the objective test, the base points for geolocation are configured. This selection can be created in two ways: completely randomly or by setting an initial point, always within a preselected set. Subsequently, the users are loaded.

When the game starts, an interface presents a story related to the title "Seeking Problems," which welcomes the student and invites them to enter their username and password. Once in the game, a character introduces a pirate story that immerses the user in the context, to involve them, assigning them a mission, and challenging their intellect to solve the game's challenges (Fig. 1).

Fig. 1. Game start: Story introduction and mission assignment.

Next, a map is shown, indicating the location of a treasure that moves as the player progresses in their search. When a question is activated, the statement is displayed, and the player is prompted to write their answer. If the answer is correct, the question will close, and the map will allow the activation of another point.

If the answer is incorrect, limited opportunities will be given according to the configuration set by the teacher. Once these opportunities are exhausted, the system will record the response and mark it as correct or incorrect (Fig. 2). Then, the game will proceed to the next point.

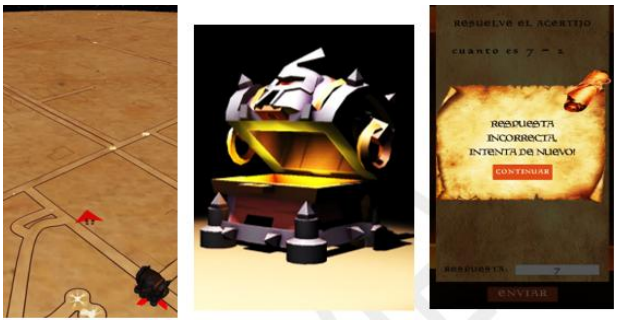

Fig. 2. Game start: Answer evaluation and attempt tracking.

At the end of the journey, the final result of the game is presented, which includes the player's score and a button to exit and end the game session. Subsequently, the game displays the monitoring section, where students are listed along with the questions answered correctly and incorrectly by each.

Additionally, an individualized scrolling map is presented for each user, showing the route taken during the game. All this information is also available in an Excel document format (Fig. 3).

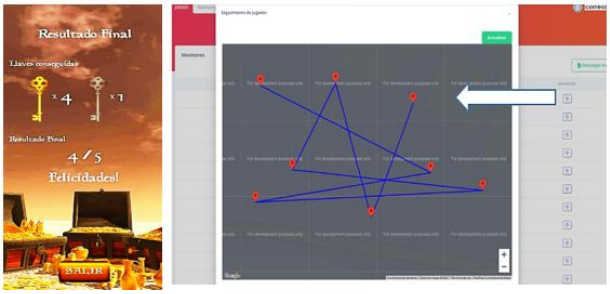

Fig. 3. Game start: Individual player route and progress.

V. RESULTS

A. Related to the Validation of the Objective Test by a Group of Experts

To validate the objective test, the 'Expert Judgment' strategy was followed, an approach used to evaluate content validity, focusing on estimating the level of agreement among experts.

Following [29, 30], eleven experts were identified, and their academic profiles were analysed. To evaluate the competence of the identified experts, the Expert Competence Coefficient (k) was calculated as the average of two quantitative indicators: knowledge coefficient (kc) and argumentation coefficient (ka). This coefficient is based on the self-assessment that experts make regarding their knowledge of

the research object, as well as the weighted assessment of specific sources that support and justify their rating. The proposal of [31] was used as a basis to identify and weigh these sources. The results of this expert rating process are shown in Table I.

TABLE I. LEVELS OF COMPETENCE OF THE SELECTED EXPERTS

Expert code	kc	ka	k	Level of competence
E-01	0.8	0.88	0.84	High
E-02	0.7	0.81	0.755	Medium
E-03	0.6	0.72	0.66	Low
E-04	0.9	0.93	0.915	High
E-05	0.8	0.87	0.835	High
E-06	0.8	0.83	0.815	High
E-07	0.9	0.91	0.905	High
E-08	0.7	0.88	0.79	Medium
E-09	0.9	0.96	0.93	High
E-10	1	0.92	0.96	High
E-11	0.9	0.91	0.905	High

Given the information derived from the coefficients of expert competence, the decision was made to exclude one of the experts. The elimination of this participant aimed to ensure the consistency and reliability of the evaluation process.

To establish and conceptualize the criteria used as evaluation parameters for the experts, the works of [32, 33] were used as references to evaluate the characteristics of the objective test. Likewise, the studies of [34, 35] were consulted to assess the mathematical content. As a result, a total of 15 criteria were defined and grouped into two dimensions, as presented in Table II.

To evaluate the reliability of the questionnaire and, with the purpose of giving meaning to the questions included in it, an analysis of internal consistency was carried out. The Cronbach's Alpha coefficient, developed by Cronbach in 1951, was calculated, which is based on the average correlation between the elements and assumes that the items, measured on a Likert scale, evaluate the same construct and are highly correlated [36]. An $\alpha = 0.81$ was obtained, which, following [37], leads to the affirmation that the consistency is good; therefore, the questionnaire is a good instrument for gathering the experts' opinions.

TABLE II. DIMENSIONS AND REAGENTS TO ASSESS THE OBJECTIVE TEST

Dimension 1: Characteristics of the objective test	
01	The subject matter can be assessed through multiple-choice questions.
02	The number of questions for each evaluated competency is proportional to the importance of the content to be assessed.
03	The questions are independent of each other and comprehensive in terms of the information necessary for their resolution.
04	Questions of varying degrees of difficulty are posed.
05	The questions are presented in straightforward and understandable language for the student.
Dimension 2: Relevance of the content for assessing algebraic reasoning	
06	The proposed exercises assume the mechanical application of algorithms.
07	The problems posed entail a challenge to delve deeper into the student's experience.
08	The exercises or problems posed allow for the appreciation of numerical relationships between the terms of an expression and between different expressions or equations.
09	The exercises or problems posed lead to the transformation of mathematical expressions.
10	The exercises or problems posed promote the development of knowledge about sets of mathematical objects, operations among them, properties of these objects and their operations, and properties of quantitative relationships.
11	The exercises or problems posed rely on the use of natural language and other non-analytical forms of expression.
12	The exercises or problems posed allow for the justification of procedures.
13	The exercises or problems posed require understanding concepts, structures, and principles that govern the procedures.
14	The exercises or problems posed involve processes of generalization.
15	The exercises or problems posed consider different levels of algebraization.

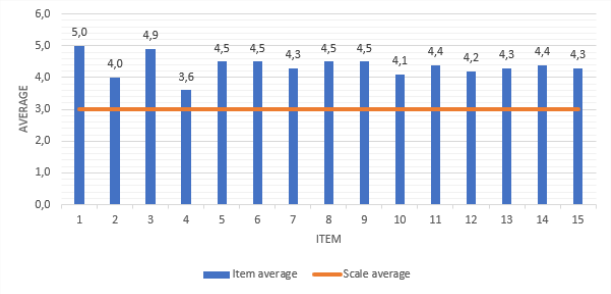

Fig. 4. Average Scores obtained in each item.

Fig. 4 illustrates how the average scores of each item behave compared to the overall scale average (represented by the dark horizontal line). It is observed that the average scores of each item are higher than the scale average. This reveals that items 1 and 3 were rated the highest by the experts. This suggests that the addressed topic can be effectively assessed through multiple-choice questions. Additionally, it is inferred that the questions posed in the assessment are independent of each other and comprehensively cover the information necessary for their resolution.

Based on the characteristics of the objective test, it can be inferred that 10% of the experts consider that the number of questions for each evaluated competency is proportional to the importance of the contents to be assessed (item 2); therefore, it would be prudent to increase the number of questions. By doing so, item 4 would be addressed, and questions with different levels of difficulty could be formulated. On the other hand, attention needs to be paid to items 9, 13, and 15, as 10% of the experts consider that the evaluated aspects are not evident.

B. Related to the Characteristics of the Objective Test

Using the responses of the 70 students to the questionnaire, Cronbach's Alpha coefficient was calculated, obtaining a value of 0.92 (excellent reliability). Through statistical analysis, some psychometric characteristics of the objective test could be identified, which facilitated the evaluation of the adequacy of the items and the test as a whole. For each question of the objective test, the difficulty index and the discrimination index were calculated.

TABLE III. DIFFICULTY INDEX OF THE QUESTIONS IN THE OBJECTIVE TEST

Question	Nº evaluated (N)	Nº correct Answers (C)	Difficulty index.
1	70	68	0.029
2	70	69	0.014
3	70	63	0.100
4	70	70	0.000
5	70	67	0.043
6	70	64	0.086
7	70	66	0.057
8	70	61	0.129

Table III presents the difficulty indices for each question. It is observed that all students answered question 4 correctly, therefore, its difficulty index is zero. The question that was most difficult for the students was question 8, followed by question 3.

The convenience of calculating discrimination indices was evaluated, which expresses the extent to which a question discriminates, that is, contributes to placing a student in the superior or inferior group. The proposal of [38] was also applied, which defines the discrimination index as the proportion of correct answers in the upper group with respect to the total number of correct answers.

Table IV shows the discrimination indices 1 and 2 for each of the questions that made up the objective test.

From the data presented in Table IV, relevant information can be extracted. Question 4 can be considered easy (difficulty index = 0) since all students answered it correctly, as confirmed by its discrimination indices (Discrimination index 1 = 0 and Discrimination index 2 = 0.5).

Regarding question 2, its discrimination index Discrimination index 1 = 0.11 indicates that it barely discriminates, meaning it is very easy. However, the index Discrimination index 2 = 0.53 suggests that, although it is a very easy question, 53% of those who answered correctly are in the upper group. This implies that if a student fails to answer this question, they are likely to belong to the group with lower knowledge level.

C. Related to Student Motivation

A questionnaire was administered to gather student perceptions, with questions aimed at assessing three dimensions: intrinsic motivation, extrinsic motivation, and game characteristics (Table V).

The questions were to be answered using a Likert-type scaling with four categorical values (1 = never, 2 = rarely, 3 = often, 4 = always). On average, students took 10 minutes to complete it. Participation was mandatory, considering that high school students were being surveyed, they were asked to respond sincerely.

All questions, except question 11, received a positive evaluation, as the average score exceeded 3 points. Question 11, which addresses whether participants felt pressured knowing it was a mathematical reasoning test, received a score below the scale average, which was expected based on the results of the other questions.

TABLE IV. DISCRIMINATION INDEX 1 AND 2 OF THE OBJECTIVE TEST QUESTIONS

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8
Upper group hits (AS)	18	17	18	18	18	18	18	18
Lower group hits (AI)	17	15	9	18	9	13	13	10
Number of subjects in each group	18	18	18	18	18	18	18	18
Discrimination index 1	0.06	0.11	0.50	0.00	0.50	0.28	0.28	0.44
Discrimination index 2	0.51	0.53	0.67	0.50	0.67	0.58	0.58	0.64

TABLE V. RESULTS OF THE STUDENT PERCEPTION SURVEY

		Question	Never	Rarely	Often	Always	Mean	Deviation
Intrinsic motivation	1	Do you believe what you learned will always be useful?	0	0	34	36	3.51	0.50
	2	Did you feel satisfied with the type of assessment?	0	4	29	37	3.47	0.60
	3	Did you feel satisfied when solving the test questions?	0	6	29	35	3.41	0.64
	4	Did you feel satisfied when solving the test questions?	0	4	29	37	3.47	0.60
	5	Did the type of test make you want to learn more?	0	13	27	30	3.24	0.75
	6	Would you prefer to solve more mathematical reasoning problems with this type of test?	0	6	36	28	3.31	0.62
Extrinsic motivation	7	Did you try to solve the questions faster to complete the game?	0	0	42	28	3.40	0.49
	8	Did you aim to answer the questions correctly to earn rewards?	0	9	29	32	3.33	0.69
	9	Do you need encouragement to do mathematical reasoning exercises?	0	5	31	34	3.41	0.62
	10	Do you like to be praised when you do well on a test?	0	6	35	29	3.33	0.63
	11	Did you feel pressured knowing it was a mathematical reasoning test?	6	38	25	1	2.30	0.64
	12	Did you need help understanding the test questions?	0	6	38	26	3.29	0.61
Game features	13	Did you enjoy and have fun with the game?	0	9	27	34	3.36	0.70
	14	Was it easy to play the game?	0	6	30	34	3.40	0.64
	15	Did you like the game interface?	0	7	33	30	3.33	0.65
	16	Did you like the included elements?	0	6	34	30	3.34	0.63
	17	Did you find the game entertaining?	0	4	38	28	3.34	0.58
	18	Do you find the game interesting as a form of assessment?	0	14	28	28	3.20	0.75
	19	Did the game keep you focused on completing the test?	0	8	24	38	3.43	0.69
	20	Did you like having to move around different places during the game?	0	6	26	38	3.46	0.65

Levels and ranges were established to assess questions related to student motivation, and the results are presented in Table VI.

TABLE VI. FREQUENCY DISTRIBUTION OF THE MOTIVATION VARIABLE

Motivation levels	Range	Absolute freq.	Relative Freq. (%)
Poor	0-12	0	0
Fair	13-24	0	0
Good	25-36	11	15.71
Very Good	37-48	59	84.29
Total		70	100

It is observed that approximately 84% of students rate the motivation generated by the geolocated and gamified test as very good, while just over 15% consider it good. The categories of fair and poor were not perceived by the students.

A similar analysis was conducted regarding the characteristics of the game; the results are detailed in Table VII.

TABLE VII. FREQUENCY DISTRIBUTION OF THE GAMIFICATION VARIABLE (GAME CHARACTERISTICS)

Level of game acceptance	Range	Absolute Freq.	Relative Freq. (%)
Poor	0--7	0	0
Fair	8--15	0	0
Good	16--23	7	10
Very Good	24--32	63	90
Total		70	100

It is observed that 10% of students perceive the gamified activity as good, while the remaining 90% consider it very good.

To determine the relationship between students' perceptions and their performance on the algebraic reasoning test, Spearman's correlation coefficient was applied. The findings revealed a moderately high correlation between the two variables (r_s = 0.51). This indicates that as students demonstrate a more favorable perception of the assessment, they tend to achieve higher scores on the algebraic reasoning test.

VI. DISCUSSION

We agree with [39], who stated that "knowing how to teach is not about transmitting knowledge, but about creating possibilities for its own production or construction." Similarly, with [40], who argue that motivating students to study Mathematics through games and playful activities can be the starting point for the construction of meaningful knowledge. The results of this research show that games and playful activities allow us to take advantage of moments of pleasure and imagination to promote not only logical-mathematical reasoning skills but also affective and physical development.

The drafting of assessment instruments should be clear and precise, considering ethical and democratizing aspects [41, 42]. Hence, it was essential to consider the language used in formulating the questions, preferably employing familiar words for students [43]. This entailed ensuring that the exercises or problems posed led to the transformation of mathematical expressions, required the understanding of concepts, structures, and principles guiding the procedures, and considered different levels of algebraization. These findings underscore the need to increase the number of questions.

Regarding the results on student perception, the absence of frequencies in the "poor" and "fair" categories is a methodological artifact stemming from the initial cut-off points and the failure to reverse-score the relevant items. In future studies, this instrument could be replicated by adjusting the scales to the actual achievable ranges (with minimums of 12 and 8, respectively) and applying the appropriate reverse-coding to ensure construct validity.

Validating the relevance of an objective test is necessary to ensure its reliability, accuracy, fairness, and utility in performance assessment and educational decision-making. We concur with [44, 45] that the "Expert Judgment" technique provides a solid and well-founded evaluation of the suitability and relevance of the questions it comprises, and the mathematical content involved. Supplementing the results of the Expert Judgment with an estimation of the expected correlation with a similar test (Cronbach's Alpha equal to 0.92) suggests that, in similar objective tests, students would have been classified similarly.

It is important to evaluate the psychometric characteristics of an objective test; adequate values increase the utility of the test by providing precise and relevant information for educational decision-making, such as identifying areas for improvement, curriculum adaptation, and student performance assessment. Therefore, for the calculation of the difficulty index, the guidelines of [46, 47, 48, 49] were considered. Consequently, the ratio of the number of students who answered correctly to the number of students who took the test was calculated. This value was compared with a difficulty index, which refers to the numerical expression of the degree to which a question is difficult to answer correctly for the group to which it is applied. For the calculation of the discrimination index, the position of [38] was taken into account, who states that "There may be questions that discriminate well but are difficult (and some from the upper group fail) or are easy (and some from the lower group get them right)" and worked with an index defined as the proportion of correct answers in the upper group relative to the total number of correct answers.

To optimize the design of the game, identify areas for improvement in teaching, and evaluate the effectiveness of the game as an educational tool, it is important to know the students' perception; results that can reveal aspects of the teaching-learning process that could be adjusted. According to [50], gamification motivates and establishes a connection between the student and the content, changing the student's perspective either for better knowledge acquisition or for the improvement of a skill. Consequently, knowing the students' perception can provide valuable information on whether the game is achieving its educational objectives, such as improving mathematical reasoning or fostering critical thinking. If students find the game useful for their learning, they are more likely to engage in the learning process and improve their skills.

Students often perceive a predominance of traditional assessment methods [51], while simultaneously noting a low quality in assessment practices [52]. These findings suggest that students perceive a gap between the educational offering and their expectations of quality. The results of this research demonstrate that the combination of gamification and geolocation can be effective in motivating students in the assessment of mathematical reasoning. The key lies in ensuring that mathematical problems are challenging yet accessible and that the gaming experience is enjoyable and motivating for students. In the described experience, a game has been created that unfolds within an immersive narrative incorporating mathematical problems as challenges that students must overcome to advance in the story, using geolocation to unlock new parts of the story when students solve problems correctly. The experience has been adequately evaluated, as the assessment of electronic educational resources should be conducted systematically and rigorously, supported using instruments that address both the objective and subjective aspects of their quality, integrating concepts of usability and learning [53].

This manuscript focuses on evaluating and presenting the educational contribution, pedagogical impact, and functional characteristics of the application in the context of learning algebraic reasoning, rather than its software architecture. Comprehensive details regarding the system architecture, data persistence, the distribution of responsibilities (mobile/server), and technical parameters related to geolocation—such as precision, activation radius, GPS sampling frequency, and tolerance for connectivity failures—have already been documented, evaluated, and published separately in a previous paper that focused strictly on the technical efficiency of the development.

VII. CONCLUSIONS

This study showed that the use of geolocation and gamification, through a pirate treasure-hunt activity, can be an interesting alternative for assessing algebraic reasoning among secondary school students. The experience was carried out with 70 students between 14 and 16 years of age in Arequipa, Peru, who responded positively to the proposal. In particular, 84.29% of the students reported feeling motivated during the activity,

while 90% expressed a favourable opinion of the game's characteristics and implementation.

The instruments used in the study also demonstrated a high level of reliability. The algebraic reasoning test obtained a Cronbach's alpha of 0.92, the expert evaluation instrument achieved 0.81, and the student perception questionnaire reached 0.95. These results indicate that the instruments were consistent and appropriate for collecting the study data.

Beyond the statistical results, the experience allowed students to approach mathematical assessment in a more dynamic, participatory, and engaging way. The combination of challenges, movement, collaboration, and game-based elements helped reduce the tension commonly associated with mathematical assessments and encouraged students to demonstrate greater willingness to solve algebraic problems.

In this regard, the proposed methodology represents a complementary strategy that can enrich mathematics teaching and assessment. However, the results should be interpreted considering certain limitations, including the sample size, the participation of students from a single educational context, and the absence of a control group. Therefore, future studies should involve larger and more diverse samples, compare this methodology with traditional assessment methods, and examine its long-term effects on academic achievement, motivation, and mathematics anxiety.

ACKNOWLEDGMENT

The authors express their gratitude to Universidad Nacional de San Agustín de Arequipa for its support of educational research and hope that the findings of this study will contribute to improving the university's educational quality.

VIII. DECLARATIONS

The authors declare that they have no conflicts of interest.

The study was conducted in accordance with the ethical principles applicable to educational research. The participants' identities were kept confidential, and the information collected was used exclusively for academic purposes.

The intervention was authorized by the participating school and approved by its administrative authorities. Informed consent was obtained from the participants' parents or legal guardians, and the students participated voluntarily. The confidentiality and anonymity of all participants were guaranteed.

Declaration on Generative AI and AI-Assisted Technologies: The authors used an artificial intelligence-assisted tool to identify and correct possible grammatical, stylistic, and translation errors in the English version of the manuscript. The tool was not used to generate research data, analyses, results, or scientific conclusions. The authors reviewed and approved all modifications and assume full responsibility for the final content.

REFERENCES

[1] L. Gómez and M. Villegas, "Laboratorio de matemática recreativa para el desarrollo del pensamiento lógico-matemático," 2007. [En línea]. Disponible en: http://biblioteca.ucp.edu.co/OJS/index.php/entrecei/article/viewFile/1643/1552

[2] O. Chapman, "Supporting the development of mathematical thinking," in *Proc. 35th Int. Conf. Psychology of Mathematics Education*, vol. 1, B. Ubuz, Ed., Ankara, Türkiye, 2011, pp. 69–75.

[3] J. Alonso, *Motivar para el aprendizaje: Teoría y estrategia*. Barcelona, España: Edebé, 1997.

[4] K. Reusser, "From text to situation to equation: Cognitive simulation of understanding and solving mathematical word problems," in *Learning and Instruction*, vol. 2, H. Mandl, E. De Corte, N. Bennett, and H. F. Friedrich, Eds. Oxford, U.K.: Pergamon, 1990.

[5] S. Arcas-Guijarro and A. Cano-Vindel, "Procesos cognitivos en el trastorno de ansiedad generalizada, según el paradigma del procesamiento de la información," *Ansiedad y Estrés*, vol. 5, no. 2-3, pp. 119–137, 1999.

[6] E. Bausela, "Ansiedad ante los exámenes: Evaluación e intervención psicopedagógica," *Educere*, vol. 9, no. 31, pp. 553–557, 2005.

[7] M. A. Calatayud Salom, "Hacia una cultura neurodidáctica de la evaluación: La percepción del alumnado universitario," *Revista Iberoamericana de Educación*, vol. 78, no. 1, pp. 67–85, 2018, doi: 10.35362/rie7813212.

[8] C. Bejarano-Parra and P. Vacca-Castro, *Revisión sistemática para integración de datos en geolocalización*. Bogotá, Colombia: Universidad Católica de Colombia, 2017.

[9] J. Leiva and N. Moreno, "Tecnologías de geolocalización y realidad aumentada en contextos educativos: Experiencias y herramientas didácticas," *DIM: Didáctica, Innovación y Multimedia*, no. 31, pp. 1–18, 2015. [En línea]. Disponible en: https://raco.cat/index.php/DIM/article/view/291534

[10] J. Palazón-Herrera, "Motivación del alumnado de educación secundaria a través del uso de insignias digitales," *Opción*, vol. 31, no. 1, pp. 1059–1079, 2015.

[11] A. Torres-Toukoumidis, M. Ramírez-Montoya, and L. Romero-Rodríguez, "Valoración y evaluación de los aprendizajes basados en juegos (GBL) en contextos e-learning," *Education in the Knowledge Society*, vol. 19, no. 4, pp. 109–128, 2019, doi: 10.14201/eks2018194109128.

[12] J. Arango López, GeoPGD: Metodología para la implementación de juegos pervasivos georreferenciados apoyados en Linked Open Data. Granada, España: Universidad de Granada, 2019.

[13] N. Dayo, M. M. Asad, F. H. Mangi, N. A. Bhellar, and U. Alvi, "Enhancement of Polya's problem-solving skills by gamifying mathematics lessons: An experimental study," in Innovative Pedagogical Practices for Higher Education 4.0: Solutions and Demands of the Modern Classroom, London, UK: Taylor & Francis, 2024, ch. 4, pp. 52-68.

[14] H. P. Arias-Flores, M. Solis, and M. Zapata-Rodríguez, "Playful Thinking as a Strategy to Assess Mathematical Skills in Primary School," in Smart Trends in Computing and Communications (Lecture Notes in Networks and Systems, vol. 645), Singapore: Springer, 2023, pp. 251-258.

[15] S. Bulubaan, D. Juandi, and N. Putri, "Gamification for the Development of Higher-Order Thinking Skills in Mathematics Learning: A Systematic Literature Review," in AIP Conference Proceedings, vol. 3432, no. 1, 2026, Art. no. 070023, doi: 10.1063/5.0327966.

[16] S. A. A. Freitas, E. D. Canedo, C. L. Frinhani, M. V. Fernandes, and M. C. Silva, "Evaluation of an Automatic Essay Correction System Used as an Assessment Tool," in Universal Access in Human–Computer Interaction. Human and Technological Environments (Lecture Notes in Computer Science, vol. 10278), Cham, Switzerland: Springer, 2017, pp. 210–222.

[17] K. Khawar, S. Munawar, and N. Naveed, "Bridging Gamification and Assessment: Proposed AI-Powered Model for Summative Evaluation," in AI-Powered Educational Games and Simulations, S. Munawar and N. Naveed, Eds., Hershey, PA, USA: IGI Global, 2026, ch. 3, pp. 69–116, doi: 10.4018/979-8-3373-0035-1.ch003.

[18] M. Larrosa, L. Wives, and V. Rodés, "Gamification Strategies as Formative Assessment Methods: A Systematic Review," in Proceedings of the 18th Latin American Conference on Learning Technologies

(LACLO 2023) (Lecture Notes in Educational Technology), S. Berrezueta, Ed., Singapore: Springer, 2023, pp. 145–156, doi: 10.1007/978-981-99-7353-8_12.

[19] S. Flores-Silva and V. Cornejo-Aparicio, "Gamification and Geolocation as Elements That Promote Motivation for the Use of Educational Software," RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, no. E50, pp. 211–224, 2022.

[20] R. Hernández Sampieri, C. Fernández Collado, and M. P. Baptista Lucio, *Metodología de la investigación*, 6.ª ed. Ciudad de México, México: McGraw-Hill Interamericana Editores, 2014.

[21] J. Valenzuela and M. Flores, *Fundamentos de investigación educativa*, vol. 2. Monterrey, México: Editorial Digital Tecnológico de Monterrey, 2012.

[22] T. M. Haladyna and S. M. Downing, "Functional distractors: Implications for test-item and test design," presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, USA, 1988.

[23] D. Morse, "The relative difficulty of selected test-wiseness skills among college students," *Educational and Psychological Measurement*, vol. 58, no. 3, pp. 399–408, 1998.

[24] M. Sánchez and A. Martínez, *Evaluación del y para el aprendizaje: Instrumentos y estrategias*. Ciudad de México, México: Universidad Nacional Autónoma de México, 2020.

[25] N. Bedregal and D. Tupacyupanqui, "Integration of active methodologies and virtual classroom in the teaching-learning processes of discrete mathematics," in *Proceedings of the LACCEI International Multi-Conference for Engineering, Education and Technology*, vol. 2018-July, 2018, Art. no. 81, doi: 10.18687/LACCEI2018.1.1.81.

[26] J. Roa González, A. Sánchez Sánchez, N. Sánchez Sánchez, and V. Basilotta Gómez-Pablos, "Uso de la gamificación como metodología activa en la educación secundaria española durante la pandemia por COVID-19," *REIDOCREA*, vol. 11, no. 10, pp. 105–119, 2022.

[27] K. Werbach and D. Hunter, *For the Win: How Game Thinking Can Revolutionize Your Business*. Philadelphia, PA, USA: Wharton Digital Press, 2012.

[28] A. Vaibhav and P. Gupta, "Gamification of MOOCs for increasing user engagement," in *Proc. 2014 IEEE Int. Conf. on MOOCs, Innovation and Technology in Education (MITE)*, 2014, pp. 290–295.

[29] J. Escobar-Pérez and A. Cuervo-Martínez, "Validez de contenido y juicio de expertos: Una aproximación a su utilización," *Avances en Medición*, no. 6, pp. 27–36, 2008.

[30] J. Cabero and M. C. Llorente, "La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información y la comunicación (TIC)," *Eduweb. Revista de Tecnología de Información y Comunicación en Educación*, vol. 7, no. 2, pp. 11–22, 2013.

[31] D. Tupacyupanqui-Jaén, *Diseño e implementación de un aula virtual para aplicar el modelo mixto en la asignatura Matemática Discreta*. Tesis doctoral, Unidad de Posgrado, Facultad de Ciencias de la Educación, Universidad Nacional de San Agustín de Arequipa, Arequipa, Perú, 2021.

[32] P. Rodríguez de Miñón, M. I. Barbero, M. J. Navas, J. C. Suárez, F. P. Holgado, A. Villarino, and P. Recio, *Recomendaciones para la elaboración de pruebas objetivas de evaluación y la interpretación de sus puntuaciones*. Madrid, España: Facultad de Psicología, Universidad Nacional de Educación a Distancia, 2003.

[33] G. Gutiérrez Berumen, "Diseño, desarrollo y evaluación de una prueba objetiva fundamentada en la taxonomía de Marzano y Kendall para medir el aprendizaje del inglés en preescolares," *Revista Didáctica, Innovación y Multimedia*, no. 30, 2014.

[34] T. P. Carpenter, L. Levi, and M. L. Franke, "Algebra in elementary school: Developing relational thinking," *Zentralblatt für Didaktik der Mathematik*, vol. 37, pp. 53–59, 2005, doi: 10.1007/BF02655897.

[35] J. D. Godino, L. P. Aké, M. Gonzato, and M. R. Wilhelmi, "Niveles de algebrización de la actividad matemática escolar: Implicaciones para la formación de maestros," *Enseñanza de las Ciencias*, vol. 32, no. 1, pp. 199–219, 2014.

[36] S. Welch and J. Comer, Quantitative Methods for Public Administration: Techniques and Applications. Chicago, IL, USA: Dorsey Press, 1988.

[37] D. George and P. Mallery, *SPSS for Windows Step by Step: A Simple Guide and Reference*, 4th ed. Boston, MA, USA: Allyn & Bacon, 2003.

[38] P. Morales, *Análisis de ítems en las pruebas objetivas*. Madrid, España: Facultad de Ciencias Humanas y Sociales, Universidad Pontificia Comillas, 2009.

[39] P. Freire, *Pedagogía de la autonomía*. Ciudad de México, México: Siglo XXI, 1997. Trabajo original publicado en 1996.

[40] A. Souza, N. Costa, and E. Brito, "Psychometric properties in instruments evaluation of reliability and validity," *Epidemiologia e Serviços de Saúde*, vol. 26, no. 3, pp. 649–659, 2017, doi: 10.5123/S1679-49742017000300022.

[41] V. Navarro-Adelantado, M. L. Santos Pastor, F. Buscá-Donet, L. Martínez Mínguez, and L. F. Martínez Muñoz, "La experiencia de la red universitaria española de evaluación formativa y compartida: Proceso y abordaje," *Revista Iberoamericana de Educación*, vol. 52, no. 7, pp. 1–13, 2010, doi: 10.35362/rie5271766.

[42] M. González Guitián and M. Molina Piñeiro, "La evaluación de la ciencia y la tecnología: Revisión de sus indicadores," *ACIMED*, vol. 18, no. 6, 2008.

[43] N. Bedregal-Alpaca, "Tutoría virtual y blended learning en el posgrado: Orientaciones y resultados de una experiencia," in *Proc. 17th LACCEI Int. Multi-Conf. for Engineering, Education, and Technology*, 2019, Art. no. 220, doi: 10.18687/LACCEI2019.1.1.220.

[44] C. E. Bruna, V. A. Villarroel, D. V. Bruna, and J. A. Martínez, "Experiencia de diseño y uso de una rúbrica para evaluar informes de laboratorio en formato publicación científica," *Formación Universitaria*, vol. 12, no. 2, pp. 17–28, 2019, doi: 10.4067/S0718-50062019000200017.

[45] A. Souza, C. Bezerra, and C. Silva, "Jogos didáticos como auxílio para o ensino-aprendizagem da matemática," *Revista Científica Multidisciplinar Núcleo do Conhecimento*, año 5, ed. 3, vol. 7, pp. 57–68, 2020, doi: 10.32749/nucleodoconhecimento.com.br/educacao/jogos-didaticos.

[46] S. S. Carey, *Measuring and Evaluating School Learning*, 2nd ed. Boston, MA, USA: Allyn and Bacon, 2001.

[47] F. H. Canales, E. L. de Alvarado, and E. B. Pineda, *Metodología de la investigación: Manual para el desarrollo de personal de salud*, 2nd ed. Washington, D.C., USA: Pan American Health Organization, 2005.

[48] E. García-Cueto, *Introducción a la psicometría*, 2.ª ed. Madrid, España: Pirámide, 2005.

[49] J. Hurtado de Barrera, *Metodología de la investigación: Guía para la comprensión holística de la ciencia*, 4.ª ed. Bogotá, Colombia: Quirón Ediciones, 2018.

[50] M. Hanus and J. Fox, "Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance," *Computers & Education*, vol. 80, pp. 152–161, 2015.

[51] F. Martínez, "Investigación empírica sobre el impacto de la evaluación formativa: Revisión de literatura," *Revista Electrónica de Investigación Educativa*, vol. 14, no. 1, pp. 1–15, 2012.

[52] F. Trillo and M. Porto, "La percepción de los estudiantes sobre su evaluación en la universidad: Un estudio en la Facultad de Ciencias de la Educación," *Innovación Educativa*, no. 9, pp. 55–75, 1999.

[53] M. Pinto, C. Gómez-Camarero, A. Fernández-Ramos, and A. Vinciane-Doucet, "Evaluareed: Desarrollo de una herramienta para la evaluación de la calidad de los recursos educativos electrónicos," *Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información*, vol. 31, no. 72, p. 227, 2017, doi: 10.22201/iibi.0187358xp.2017.72.57831.