Article

Videogames and open feedback systems to enhance probabilistic reasoning and engagement

Andrés Felipe Velasco-Hernández[1], David Sebastián Baldeón-Padilla[2] and Andrés Adolfo Navarro-Newball[3]

[1]Maths department, Colombo Británico School, Cali, Colombia; [2]Games and Interactive Experiences Center, Pontificia Universidad Javeriana, Cali, Colombia; [2]Electronic and computer science department, Pontificia Universidad Javeriana, Cali, Colombia
{ andres.velasco } @ ccbcali.edu.co; {david.baldeon, anavarro} @javerianacali.edu.co

Abstract

This paper aims to evaluate an innovative pedagogical strategy for teaching probabilistic reasoning skills and overcoming the widespread misconception that random events are solely based on "luck". It investigates the relationship between probabilistic reasoning, feedback types, and engagement levels across four sessions of a DGBL educational practice employing open-ended feedback systems. The study tested the hypothesis that involvement in DGBL sessions, accompanied by open feedback, enhances engagement and the development of probabilistic thinking. Open feedback and the DGBL environment were treated as independent variables, with academic engagement and probabilistic reasoning as dependent variables. The research featured the integration of the game Dicey Dungeons as an instructional tool and adopted a quantitative and correlational approach within a quasi-experimental design; it did not include a control group nor randomization in the group selection. Results indicate a dynamic nature of engagement and probabilistic reasoning, with significant correlations observed between open-ended feedback and probabilistic reasoning. While a correlation between engagement and probabilistic reasoning was observed in a single session, no consistent correlation was found for the entire study. These findings offer implications for teaching strategies and the development of DGBL sequences in secondary education, underscoring the significance of teacher-student interactions and open feedback systems.

1. Introduction

Mathematics is an area that has historically been regarded as fundamental knowledge, yet it has faced challenges in teaching owing to its perceived complexity. A notable mathematical skill is the ability to foresee and estimate the likelihood of events, enabling sound judgments

and decisions in uncertain situations, which is known as probabilistic reasoning [1]. Teaching this skill is hindered by the deep misconception that events are solely dictated by "luck", which can lead to misunderstandings, flawed generalizations [2] and difficulties in applying appropriate models to solve problems involving uncertainty [3].

Engagement holds particular interest, as its multifaceted dimension, with significant relevance in the classroom [4]. It depends on various factors including the instructional approach employed by teachers and pedagogical support, among others [5].

These challenges prompt us to explore opportunities for solutions that foster the development of pedagogical strategies rooted in didactics, teacher support, and students' learning experiences; thus equipping them with mathematical skills such as probabilistic reasoning. In this context, we must seek diverse alternatives within the classroom that actively engage students, requiring the use of specific skills and nurturing learning experiences that encourage motivation and commitment.

This study envisions Digital Game-Based Learning (DGBL) [6]–[8] as a didactic approach with the potential to enhance teaching and learning processes by significantly promoting engagement and skill development, compared to traditional pedagogical strategies [9]–[13]. It also aims to implement open feedback systems as a form of scaffolding mechanism to help students progress from their prior knowledge and toward the specific educational objectives of the didactic sequence [14]–[17].

Next, we will delve into probabilistic reasoning as a skill with the potential to enhance decision-making processes in people's lives, and engagement as a motivational factor conducive to meaningful learning experiences.

1.1   Probabilistic reasoning

One of the most highly valued mathematical skills is probabilistic reasoning, a crucial aptitude for addressing real-world issues and making informed decisions [18]–[20]. Probabilistic reasoning can be defined as the ability that allows one to make decisions based on given information, identify patterns, and utilize statistics to describe, model and interpret phenomena, ultimately enabling inferences [21].

As outlined by Batanero et al. [1], probabilistic reasoning is comprised of several components, including the ability to recognize random events, analyze the underlying conditions of these events, formulate models to represent these scenarios, develop mathematical models, and apply the appropriate methods and procedures to calculate probabilities and statistics.

Notably, Ricart and Estrada [20] emphasize the instructional aspect in nurturing this skill, as students who haven't received prior guidance on probability concepts tend to lack strategies for decision-making rooted in proportionality reasoning. In their study frequently cited in the field, Fischbein et al. [18] identify the difficulty of distancing oneself from the conventional notion of "luck" as a factor affecting probabilistic reasoning.

Having elucidated the concept of probabilistic reasoning and considering the various factors influencing learning, the subsequent section will delineate how engagement is conceived in this study.

1.2   Engagement

Engagement can be defined as a student's disposition toward a class or learning activity, recognized as a pivotal element in the learning process [4], [8], [13], [22]–[26]. It embodies the intricate interplay of a student's emotional, behavioral, and cognitive state [4], [27]–[30]. Fredricks et al. [4] point out the importance of adopting a multidimensional perspective when studying engagement, facilitating the exploration of patterns between behaviors, emotional state, and cognitive processes.

In the context of mathematics, Putwain and Wood [25] determined that to achieve better learning, some strategies that can be used are: improving fundamental mathematics skills, improving the quality of teacher feedback, and the use of digital tools. Interestingly, these findings and intervention proposals correlate with the dynamics of DGBL addressed in this study, reinforcing the pivotal role of the teacher feedback quality.

Several studies demonstrate that game-based pedagogies, can significantly boost student motivation and engagement in comparison to traditional classroom settings [31]. In examining the correlation between engagement and probabilistic reasoning, Furlan et al. [32] concluded that students who are inclined toward reasoning not rooted in notions of luck tend to be more engaged in tasks involving probabilistic reasoning. Additionally, Grotzer et al. [19] observed heightened engagement when students begin to find patterns, results and evidence in deterministic scenarios, motivating them to delve deeper into the subject. Furthermore, engagement flourishes when students are exposed to mechanisms illustrating that uncertainty and probability play a fundamental role in these situations

With the concepts of probabilistic reasoning and engagement clarified, and their research implications discussed, the subsequent section will define and explore the context for skill development and engagement, delving into game-based learning and open-ended feedback will subsequently be defined.

1.3   Digital game-based learning

Certainly, one emerging pedagogical strategy in recent years, has been the integration of games into educational settings [33]. When employed within the classroom, such methodologies have consistently demonstrated improved learning outcomes and academic performance compared to traditional activities that lack these interactive elements [9]–[13], [34].

This study employs a pedagogical approach that incorporates digital games into educational practices, specifically DGBL. Farah et al.[10], Huizenga et al.[16] and Moon and Ke [13] emphasize the effectiveness of the videogame format in enhancing engagement. Moreover, it is also well-documented that DBGL is an effective method for teaching probabilistic thinking [35]–[37]. However, to create DGBL environments that truly foster engagement, it is imperative to have feedback mechanisms in place during the learning activities [38]–[40].

Having established DGBL as the chosen pedagogical strategy for this study, the next consideration is the specific types of feedback that will be utilized to characterize the teacher's interventions.

1.4   Type of feedback

Hattie and Timperley [15] define feedback as the information provided by an agent to close the gap between what the learner understands and the learning objective. In this sense, feedback takes place in the interaction between the learner and the teacher, thus it is in that interaction where the type of feedback will be evidenced. For this, van Vondel et al.[41] propose five categories or levels: instruction, providing information, closed or comprehension questions, encouraging the continuation of verbalizations, and open-ended questions that encourage deep thoughts or high order skills. This scale is referred to as "openness scale" [42], [43] , valuing the degrees of freedom the adult allows the learner in their interaction. Encouraging verbalizations and open-ended questions allow students to develop critical thinking, as they promote analysis and evaluation of their reasoning, showing higher order thinking skills [41], [42].

For the purposes of this paper, open feedback will be understood as that involving the top two levels of the scale proposed by van Vondel et al. [41] : encouragement and stimulating follow-up. This interaction is beneficial in mathematics, given that in this area analytical and

procedural processes are generated, rather than declarative knowledge [14]; and generates better performance in students when faced with similar questions or situations [44], [45]. In addition, when games are used, the support provided to the student during the game influences the perception of their degree of involvement in the learning process [13].

While the impact of feedback within DGBL has been studied previously, the impact of involving open-ended questions within a game has not, at least in relation to academic outcome [46], [47]. Attali [14] concludes in his study, which is not developed in DGBL, that the benefits of this type of feedback are significant and therefore it is beneficial to involve these questions within educational environments.

Thus, the development of this study is justified on the basis of its contribution to the integration of methodologies that allow a better learning of mathematics, in particular, the development of probabilistic reasoning through better teacher interactions and involving innovative strategies that have proven to be effective for teaching. Likewise, it is intended to study strategies that allow maintaining a high level of engagement. At an academic level, there exists a gap in the literature studying the learning process in mathematics multidimensionally; this study contributes to filling said gap while providing information to improve mathematics teaching practices, as up to the time of the writing of this paper, no articles were found that related all the variables presented in this study.

Based on what has been previously presented, the purpose of this study is to analyze the relationship between probabilistic reasoning, the type of feedback, and engagement in a DGBL educational practice in eighth grade students.

2. Methods and Material

Central to our investigation is the question: How do different types of teacher feedback in a DGBL environment impact the levels of academic engagement and probabilistic reasoning among eighth-grade students? In order to achieve the aim of this study, we identified four specific goals. Firstly, this study focused on (i) identifying the levels of academic engagement among the students. Secondly, we (ii) examined the levels of probabilistic reasoning exhibited by the participants. Thirdly, (iii) identifying the type of feedback provided by the teacher during the course of this study. Finally, we aimed to (iv) establish the type of correlation that existed between the levels of probabilistic reasoning, academic engagement, and the types of feedback across four class sessions. These sessions employed a DGBL methodology, utilizing the game Dicey Dungeons as an instructional tool.

This study adopted a quantitative and correlational approach within a quasi-experimental design. It did not include a control group and did not involve randomization in the selection of groups. In this study, we tested the hypothesis that involvement in DGBL sessions, accompanied by open feedback, enhances engagement and the development of probabilistic thinking. Open feedback and the DGBL environment were treated as independent variables with academic engagement and probabilistic reasoning as dependent variables.

The quasi-experimental design favors studying phenomena in classrooms by having previously established groups [48] and, using repeated-measures methodology throughout four classroom sessions, allows measuring the levels of probabilistic reasoning, academic engagement, and level of feedback dynamically, making it possible to study longitudinally the dynamic nature of the variables [49] . The sessions were recorded for later analysis.

2.1   Participants

A total of 80 8th-grade students, including 44 girls and 36 boys aged between 13 and 15 years, participated in the study. The mathematics teacher of the four participating groups facilitated

the sequence using the video game. The choice of four groups was influenced by the school's existing class structure. These were the groups readily available within the school's organizational framework, making them the most accessible and practical options for this research. While the groups were consistent with the school's class organization, it's important to note that these groups were heterogeneous in terms of student composition, which allowed to explore a diverse range of perspectives and experiences.

2.2   Materials

The Dicey Dungeons videogame [50] was used for the research work. In the game, players select a character and navigate through dungeon levels, facing various challenges and adversaries. What sets it apart is its reliance on dice as a resource for actions. Players roll dice to determine the actions they can take during battles, introducing an element of probability into their decision-making process. This design necessitates that players assess the likelihood of obtaining specific dice values and adapt their strategies accordingly, contributing to the development of probabilistic reasoning skills. The dynamic and strategic gameplay engages players, and feedback mechanisms within the game influence their in-game decisions and learning experiences.

In Dicey Dungeons, battles serve as the core interactive component of the game (Figure 1). During these encounters, players and their chosen character engage in turn-based combat against a variety of monsters and adversaries encountered in the dungeons. The key feature of these battles is the use of dice as a resource for actions. Each turn, players roll their dice to generate a set of values. These values then determine the available actions the player can take during that turn, such as attacking, using spells, or healing. The outcome of the dice roll directly influences the strategic choices that players can make. As players progress through the game, they encounter different opponents with unique abilities and characteristics, leading to a dynamic and ever-evolving tactical landscape. Players must carefully consider their dice rolls, the abilities of their character, and the strengths and weaknesses of the opponents to determine the most effective course of action. This strategic component introduces probabilistic reasoning, as players must assess the probability of achieving specific dice combinations and adjust their strategies accordingly. These battles, as a central element of the game, offer an engaging context for the study of how students develop probabilistic reasoning skills, make strategic decisions, and respond to feedback within a game-based learning environment.

During the game, players encounter a diverse cast of characters, each with unique abilities and strategies. The order in which these characters are introduced is a deliberate scaffolding of the game's difficulty and learning curve. Players begin their journey with the Warrior, a character known for simplicity, and gradually progress to more complex characters like the Thief, Robot, and Inventor. Each character's introduction builds upon the knowledge and skills acquired in the previous stage, allowing players to steadily adapt to new mechanics and challenges. This thoughtful progression not only eases players into the game's intricacies but also encourages a deeper understanding of the strategic possibilities offered by each character.
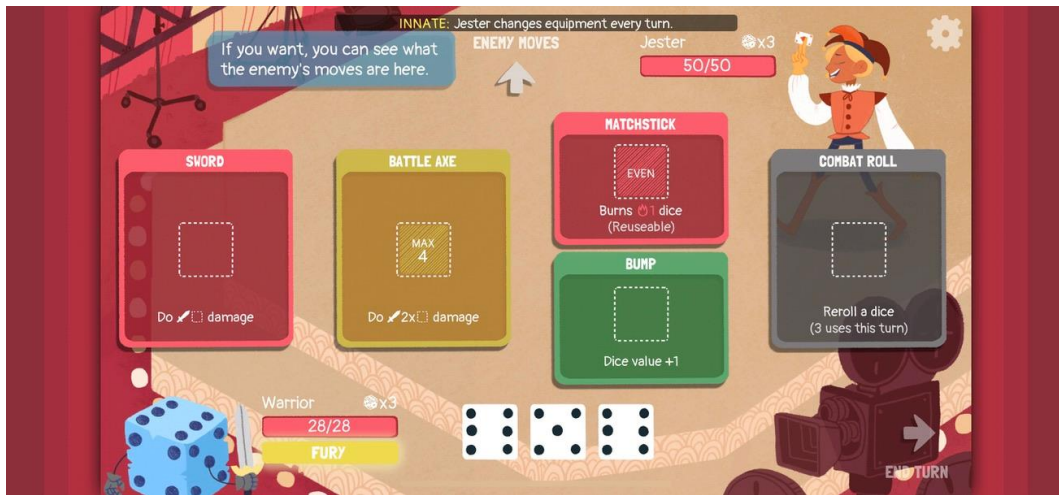
Figure 1. Battle screen.

The choice of this game was made after evaluating the cognitive demand that it proposes to the players. This was done based on the task analysis of the videogame as a problem-solving situation [51], [52], presented in Table 1. In this table it is also explained how players navigate probabilistic reasoning to use the dice during the game using figure 1 as an example.

Table 1. Task analysis for Dicey Dungeons.

Task analysis	Description
Main components of the task	The Dicey Dungeons videogame was used to enhance probabilistic reasoning by challenging players to make optimal decisions within the game's dynamic conditions. This required players to base their decisions on situations influenced by both chance and statistical probability. Players were asked to justify their choices, utilizing proportionality models for decision-making in simple events and mental tree models for compound events.
Relationship between task components	Taking Figure 1 as an example, a few aspects of dice usage in the game involve players applying probabilistic reasoning while considering the conditions of their abilities. The abilities vary in terms of the number of results in a die that can be employed, which requires careful decision-making. Some abilities, such as matchstick, restrict players to using only even dice; that is, three out of six dice results (2, 4 and 6) equating to a 50% of theoretical chance of success. Others, like battle axe, allow use of up to four dice results, resulting in a 66.7% of theoretical probability of success, while abilities, sword, bump and combat roll permit the use of any dice with a 100% chance of success. The players, in their turn, must take into account the inherent randomness of the abilities, such as combat roll. For instance, deciding whether to use the 5 or the 6 for the sword ability involves assessing the desired outcome, favoring the higher amount of damage dealt to the opponents with a 6. Another option is to decide whether to use the 6 in matchstick, bump or combat roll. Using the 6 in combat roll ensures an extra die, enabling the use of battle axe for double damage, but the probability of obtaining a usable die is 66.7%. Another strategy is to use the 5 in combat roll allowing the use of battle axe and matchstick with the pre-existing 6 if an odd die is acquired. The use of the dice, which have been randomly assigned, and the corresponding decision-making, considering various factors and uncertainties, exemplifies the application of probabilistic reasoning.
What are the task demands?	During different moments of the game, students should engage several cognitive functions to complete the tasks presented by the game, particularly by applying probabilistic reasoning. In battles, students must calculate the probability of successfully using each battle ability, infer potential scenarios when they cannot be certain about the dice they will receive in each turn, compare the probabilities associated with each battle ability, and asses the optimal dice allocation. The latter is an example of compound probability, where they not only need to understand the probability of using each battle ability but also estimate the potential outcomes of employing each die in a different

ability. Moreover, when selecting battle abilities, students must first conduct a comparative analysis to determine which abilities are likely to yield the best results in battles. Through the application probabilistic reasoning, they must decide which abilities offer the highest probability of success in upcoming battles.

2.3   Instruments

2.3.1   Probabilistic reasoning

To measure the levels of probabilistic reasoning shown by the students during the sessions, the Probabilistic Reasoning Rubric (PRR), presented in Table 2 was used. This rubric contains three components of probabilistic reasoning: identification of complementary events, compassion of probabilities of occurrence of events in situations of uncertainty, and decision making in situations of uncertainty, considering the possible outcomes of each decision [1]. Each criterion has four levels of performance.

Table 2. Probabilistic Reasoning Rubric. This rubric is constructed using as input the elements of probabilistic reasoning proposed by [1]

Probabilistic reasoning component	1	2	3	4
Classifies the abilities taking into account the ratio between dice that can be used to those that cannot.	The student considers that the abilities can be used freely.	The student identifies in the abilities the dice that can and cannot be used, attributing some wrong restrictions.	The student identifies in the abilities which dice can and cannot be used correctly.	The student classifies the abilities according to the universe of possibilities according to the dice that can and cannot be used.
Compares the capacity of use between the randomly assigned dice.	The student does not identify the possibilities of using the dice with the abilities.	The student partially identifies the possibilities of using the dice in some of the abilities.	The student fully identifies the possibilities of using the dice in the abilities.	The student compares the usability between the dice.
Argues and justifies decision making through comparisons.	The student randomly uses the dice.	The student argues and partially justifies the use of the dice, making some random decisions.	The student argues and justifies the use of dice in decision making.	The student argues and justifies the use of the dice, emphasizing the comparison of the consequences of one decision or another.

2.3.2   Open-ended feedback

To identify the type of feedback, the openness scale was used, referred by van Vondel et al.[41] and presented in Table 3.

Table 3. Level of feedback, using the openness scale [41].

Level	Description	Examples
Instruction	A task or a clarification of the task is given.	To make use of the dice in battle, you must drag them to the ability you want to use.
Providing information	Basic information is given to the student about some concept	You should take into account which ability is less likely to be usable.

Closed or knowledge-based questions	Questions are asked that have right or wrong answers, gives rise to little verbalization on the part of the student	Which dice can you use with this ability?
Encouragement	The student is asked to further explain an idea or to make their thinking visible by verbalizing their ideas.	Explain to me your thoughts on why it is better to use the dice in this way.
Stimulating follow-up	Questions are asked to encourage critical thinking, as well as higher order abilities: evaluating, analyzing, creating	How can you use your dice in such a way that you do more damage to the opponent this turn?

2.3.3 Engagement

To measure engagement, the Experience Sampling method (ESM) [53] was used during each session to preserve test-retest reliability. This method has been validated to measure engagement during classroom sessions [54], [55], in learning environments using mobile technology [56] and with adolescents [57]. In this study, a digital version of the ESM was used.

The ESM is a data collection technique that involves real-time prompts to capture participants' momentary experiences, thoughts, and behaviors during the gaming sessions. Participants received prompts at specific intervals, allowing them to provide self-reports on their engagement and related experiences as they occurred. The digital version of ESM utilized for this study was designed to be user-friendly and non-intrusive, ensuring minimal disruption to the gaming experience.

2.4 Procedures

2.4.1 Probability knowledge and videogame baseline phase

In order to help students have a similar baseline in terms of knowledge of probability concepts, two weeks of classes were held on this component, following the school curriculum. This made it possible to start from a similar base, even if it was heterogeneous. Standardized tests were not used, since what we wanted to encourage was an ability and not conceptual knowledge.

Similarly, to help in the learning curve of the video game, the students had a free play session, with the initial character from the tutorial, to understand the dynamics of the game and become better in the intervention sessions.

2.4.2 Application phase of DGBL sessions with Dicey Dungeons

During this phase, each of the four groups participating in the research engaged in four sessions of Dicey Dungeons game play at school. The sessions began with a brief explanation of the objective of the session, including guidance on which character to use. Each session lasted approximately 55 minutes. Students played individually on iPads and the teacher made interventions, providing individualized feedback to each student. The students were arranged in groups of four to facilitate a more effective learning curve of the game.

The intervention spanned around two weeks, following the school's schedule. During this period, all class time was dedicated to the intervention with the game, with no theoretical lessons in between.

The intervention protocol included a structured feedback framework for the teacher, addressing the different levels of the openness scale (table 3). While feedback ideally remains dynamic, the protocol was designed to guide teacher interventions, allowing them professional discretion to provide necessary support to the students.

Throughout the sessions, the teacher encouraged students to provide explanations through feedback, which could include both encouragement and follow-up questions. These

interactions introduced some complexity, but the teacher was encouraged to make the criteria for measuring probabilistic reasoning visible. Each student progressed at their own pace during the game, playing with the character chosen for each session.

At the end of each session, a plenary session took place. Photos of the game were shown, and questions related to the feedback system were asked. This provided all students with the opportunity to participate and answer questions, facilitating a comprehensive understanding of the feedback process.

2.4.3 Data collection phase

To measure engagement, the Emergent Motivation Questionnaire (EMQ) was used during the game sessions when at least 60% of the session had taken place. To facilitate this process and minimize interruptions due to the intrusive potential of the ESM [58], a QR code link was projected allowing students to easily and quickly complete a self-report via scanning. The ESM allowed captures of real-time experiences and responses of the participants by prompting them with specific questions or prompts related to their engagement during the gaming sessions.

The sessions were audio-visually recorded to classify feedback by type and students' verbalizations by their level of probabilistic reasoning. To conduct this assessment, the PRR and the openness scale were employed. The analysis of teacher-student interactions captured in the recordings centered on the evaluation of feedback types and the development of probabilistic reasoning, with particular attention to both the feedback type and the advancement of the probabilistic reasoning.

2.4.4 Data coding phase

To measure student performance during the sessions, student interactions were coded using the PRR and values were assessed on a scale of 1 to 4, where 1 represents the lowest level and 4 the highest level of probabilistic reasoning. A similar process applied to the openness scale and the teacher's feedback, with values assigned on a scale of 1 to 5, where 1 is the lowest level and 4 and 5 the highest levels, defined as open feedback. Additionally, the data obtained from the video recordings and the EMQ were processed and organized for subsequent analysis in relation to the engagement of the students during each session.

2.4.5 Data analysis phase

For this study, SPSS [59] was selected as the statistical analysis software due to its established reliability and widespread acceptance in academia [60], [61]. To perform descriptive analysis of the engagement, level of feedback, and probabilistic reasoning variables, the normality of the sample distributions was assessed using the Kolmogorov-Smirnov test. The results indicated non-normal distributions, prompting the use of the Kruskal-Wallis test for each variable. This analysis provided characteristic statistical data for the samples, including values of the H parameter, its significance and its average range.

To study correlations, Spearman's Rho tests were conducted in two methods, analyzing the correlations between engagement, feedback type and probabilistic reasoning. One method administered Spearman's Rho by session inclusive of all groups. The second analyzed each group individually, assessing the results of all four sessions collectively in a non-longitudinal manner.

3. Results

The results for each of the specific objectives addressed in this study are presented below, considering both data distribution and the statistical tests applied to achieve these objectives.

For the first specific objective, a descriptive analysis of the variable was conducted, segmenting by groups across the four DGBL sessions. Additionally, a combined group mean was calculated.

For the second and third specific objective, descriptive analyses of probabilistic reasoning and the level of feedback were performed across the four DGBL sessions, segmented by group. A mean value of the variables was also calculated longitudinally. Kruskal-Wallis tests were utilized to assess the variable's variation throughout the sessions.

For the fourth and last specific objective, Spearman's Rho tests were carried out to establish the correlations between engagement, probabilistic reasoning and the level of feedback in DGBL sessions. These tests were conducted longitudinally, discriminating by session, and groupwise, discriminating by group. The following sections present the results of these analyses.

3.1 Engagement

To measure engagement, the enjoyment, concentration and interest components of the QEM were used [62]. These variables are presented in Figure 2 for each group, and throughout the four sessions. Based on this data, several noteworthy observations have emerged. Firstly, a longitudinal examination of the Focus factor revealed an upward trend in Group 1, signifying a potential positive impact of the intervention on students' ability to concentrate during the gaming sessions. However, the analysis of Interest and Joy shows a certain level of variability across all groups and sessions, indicating that students' levels of interest and joy exhibited fluctuations throughout the intervention. This variability underscores the dynamic nature of students' emotional and motivational responses, warranting a deeper investigation into the factors contributing to these fluctuations. Lastly, the data reveals heterogeneity within each group, particularly in Interest and Joy scores, suggesting that individual students had diverse experiences and responses to the intervention.
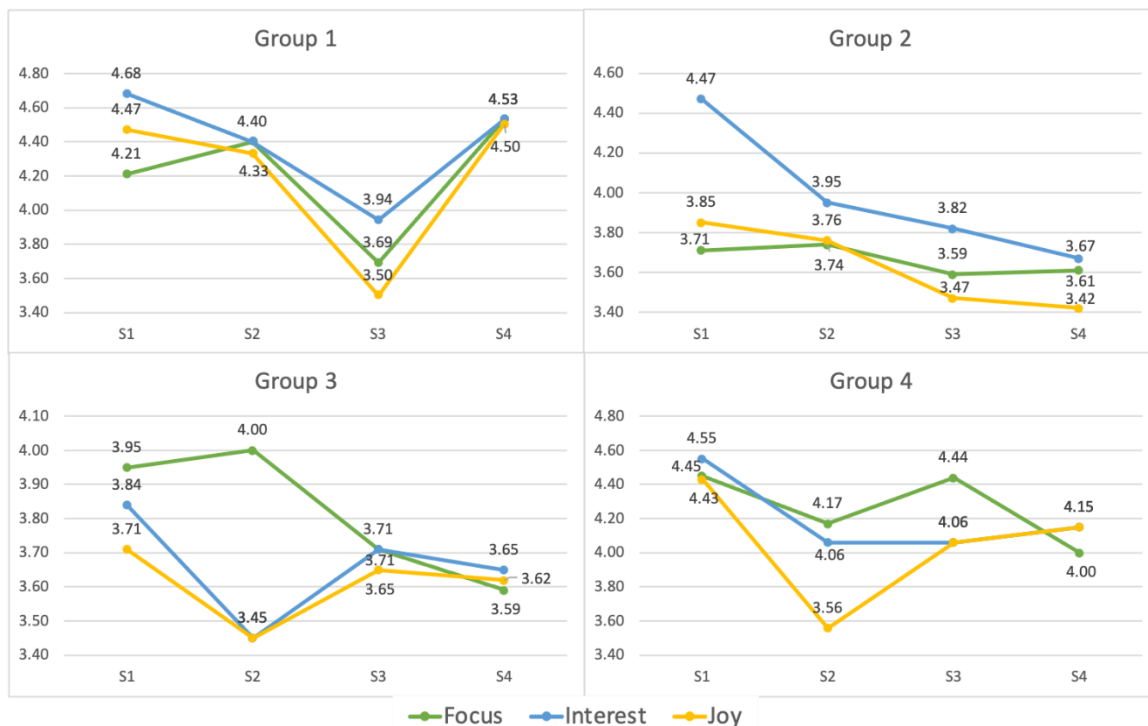

Figure 2. Components of Engagement for each group throughout the four intervention sessions.

Figure 3 shows the evolution of this variable for each of the groups and the mean of the groups, presenting the descriptive values found for the commitment of each group per session.
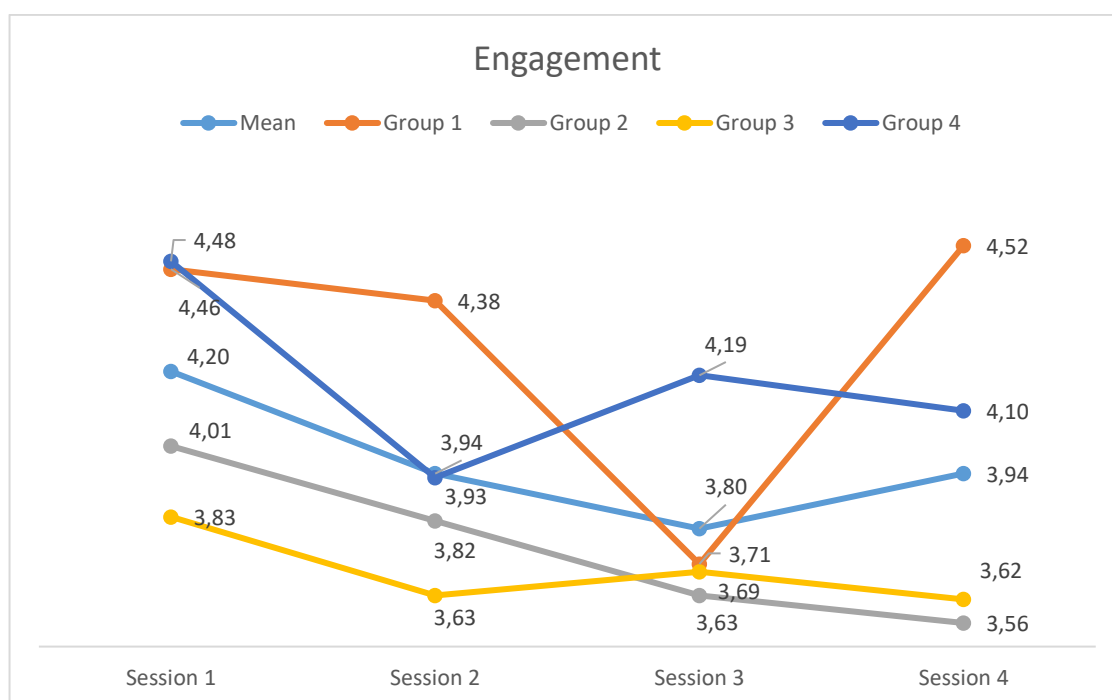

Figure 3. Engagement for each of the groups throughout the four DGBL sessions.

The analysis of engagement data reveals interesting trends and variations among the four groups across the four sessions. Group 1 consistently demonstrates the highest mean engagement levels throughout the study, suggesting a strong and sustained engagement with the learning activities. In contrast, Group 2 tends to exhibit lower mean engagement scores, indicating a potential need for interventions to enhance their engagement. Furthermore, the data illustrates that engagement is not static and varies over time. For instance, Group 4 shows an improvement in engagement from S1 to S4, indicating a positive impact of the learning experiences. These findings underscore the dynamic nature of student engagement and suggest that tailored strategies may be required to maintain and enhance engagement levels. Further statistical analyses are warranted to validate the observed differences and gain deeper insights into the factors influencing engagement in these learning contexts.

3.2   Probabilistic reasoning
The descriptive results of analyzing the probabilistic reasoning of the students are presented in Figure 4, which shows the values for each group longitudinally and also the mean of the groups. These values present variations throughout the four sessions and are not static. All groups, for session 3, present a decrease in the measurement of the variable.
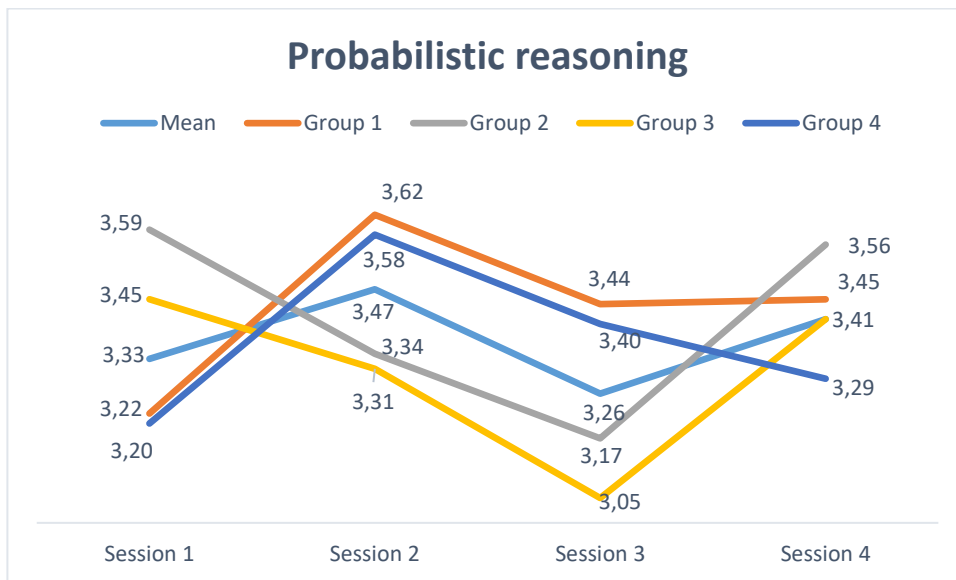
Figure 4. Probabilistic reasoning, scored for students during the intervention sessions differentiated by groups and on average.

To analyze the changes in probabilistic reasoning throughout the study, a Kruskal-Wallis test was conducted. The results of these tests are shown in Table 4. Here we provide the H value, along with its significance, degrees of freedom and the average ranges per session for each group. The results of this test reveal significant variations for each group longitudinally. Specifically, for groups one, two and three, the variation in probabilistic reasoning is highly significant ($p < 0.001$) while for group 4 it is significant ($p < 0.01$). Notably, the lowest range value for all groups, is observed in the third session, whereas the highest values vary among the groups: groups one and four exhibit the highest values in the second session, the second in the first session, and the third in the last session.

Table 4. Table of results of Kruskal-Wallis test applied to probabilistic reasoning for each of the groups.

Group	H	Sig.	df.	Session	Mean Rank
1	13.029	0.005***	3	1	104.72
				2	144.62
				3	119.84
				4	124.76
2	12.744	0.005***	3	1	104.72
				2	144.62
				3	119.84
				4	124.76
3	11.657	0.009***	3	1	104.72
				2	144.62
				3	119.84
				4	124.76

				1	104.72
				2	144.62
4	9.153	0.027**	3	3	119.84
				4	124.76

*p < 0,05; **p < 0,01; ***p < 0,001

3.3 Type of feedback

Given the importance of the level of feedback in this study, it was relevant to ensure that feedback defined as open, established as the two highest levels of the openness scale, did occur. The results of these coding are presented in Figure 5.
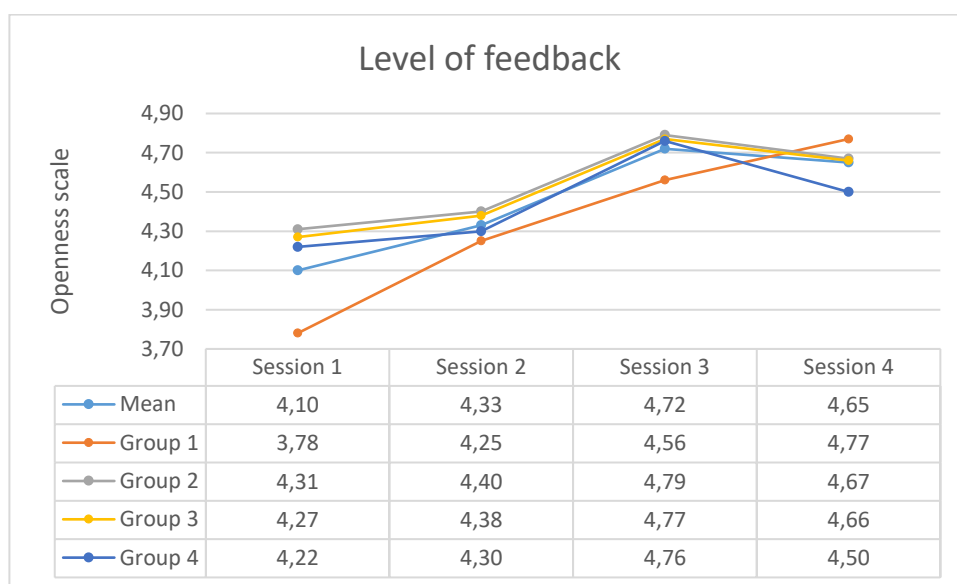

Level of feedback	Session 1	Session 2	Session 3	Session 4
Mean	4,10	4,33	4,72	4,65
Group 1	3,78	4,25	4,56	4,77
Group 2	4,31	4,40	4,79	4,67
Group 3	4,27	4,38	4,77	4,66
Group 4	4,22	4,30	4,76	4,50

Figure 5. Level of teacher's feedback during the intervention in the four groups.

To analyze changes in the type of feedback throughout the four sessions, a Kruskal-Wallis test was conducted, the results of which are shown in Table 5. The test results indicate high significance ($p < 0.01$) in the variation of feedback levels across all four sessions. This implies that the feedback provided in each session was not single-level but rather multitype and dynamic.

The results of the mean ranks reveal differences in feedback level values between groups for each session. Group one exhibited the highest levels of feedback in the last session and the lowest in the first session. In the second group the dataset yielded the lowest values compared to the other groups. Furthermore, this group had the lowest values of feedback in the first session and the highest in the third session. In the third group, the variations showed the highest level of feedback in the third session and the lowest in the first session. Lastly, in the fourth group analyzed, the session with the highest average range of feedback was the third session, while the session with the lowest level was the second session.

Table 5. Table of results of Kruskal-Wallis test applied to level of teacher's feedback.

Group	H	Sig.	df.	Session	Mean Rank
1	38.45	0.000***	3	1	91.73

				2	112.00
				3	144.97
				4	150.49
2	15.81	0.001***	3	1	69.78
				2	71.81
				3	96.88
				4	91.32
3	15.91	0.001***	3	1	95.88
				2	107.06
				3	137.26
				4	123.81
4	15.42	0.001***	3	1	103.45
				2	101.78
				3	139.86
				4	114.49

*p < 0,05; **p < 0,01; ***p < 0,001

3.4 Correlations between variables

To determine the correlation between variables, a Spearman's Rho test was performed. The results of the two segmentations discussed above are presented below.

3.4.1 Correlations segmented by sessions

Table 6 presents the results of the correlation between the variables engagement, probabilistic reasoning, and feedback during each of the sessions. The values of the correlation coefficient and their significance are reported.

Table 6. Results of Spearman correlation coefficients and bilateral significance during the intervention sessions.

		Engagement	Probabilistic reasoning	Feedback
Session 1	Engagement	1,00		
	Probabilistic reasoning	0,006 (0,961)	1,00	
	Feedback	0,013 (0,912)	0,524*** (0,000)	1,00
Session 2	Engagement	1,00		
	Probabilistic reasoning	-0,084 (0,466)	1,00	

		Engagement	Probabilistic reasoning	Feedback
	Feedback	0,021	0,382***	1,00
		(0,854)	(0,000)	
Session 3	Engagement	1,00		
	Probabilistic reasoning	0,313**	1,00	
		(0,010)		
	Feedback	0,011	0,013	1,00
		(0,929)	(0,857)	
Session 4	Engagement	1,00		
	Probabilistic reasoning	0,139	1,00	
		(0,270)		
	Feedback	0,211	0,159*	1,00
		(0,091)	(0,012)	

When segmenting the samples by sessions, it is found that for session one there is a moderate and statistically significant correlation, $r = 0.524$, $p < 0.01$, between the level of feedback and probabilistic reasoning. In session two there was also a moderate and statistically significant correlation between the same variables, $r = 0.382$, $p < 0.01$. These correlations differ in value, but both are moderate and positive, indicating that a shift toward a higher level of feedback will be accompanied by a higher level of probabilistic reasoning. For the third session, the moderate and significant correlation found was between probabilistic reasoning and engagement, $r = 0.313$, $p = 0.010$. Finally, for session four the correlation found is again between probabilistic reasoning and level of feedback. This is weak and marginally significant, $r = 0.159$, $p = 0.012$. The other correlations between variables not mentioned are not statistically significant.

3.5   Correlations segmented by groups

This section consolidates the results of the correlation between the variables throughout the intervention, discriminated by groups. Table 7 reports the values of the correlation coefficient and their significance.

In this segmentation, statistically significant correlations were found only between the level of feedback and probabilistic reasoning. No statistically significant correlations were found between the other variables. In sessions one and three, moderate and highly significant correlations were observed, $r = 0.60$, $p < 0.001$ and $r = 0.69$, $p < 0.001$ respectively. On the other hand, in sessions two and three, moderate and marginally significant correlations were identified, with correlation values $r = 0.39$, $p = 0.037$ and $r = 0.30$, $p = 0.036$, respectively.

Table 7. Results of Spearman's correlation analysis, reported with their significance, and discriminated by groups.

		Engagement	Probabilistic reasoning	Feedback
Group 1	Engagement	1,00		
	Probabilistic reasoning	0,26	1,00	
		(0,274)		

	Feedback	0,29	0,60***	1,00
		(0,237)	(0,000)	
Group 2	Engagement	1,00		
	Probabilistic reasoning	-0,18	1,00	
		(0,500)		
	Feedback	0,16	0,39*	1,00
		(0,553)	(0,037)	
Group 3	Engagement	1,00		
	Probabilistic reasoning	-0,38	1,00	
		(0,105)		
	Feedback	-0,16	0,69***	1,00
		(0,526)	(0,000)	
Group 4	Engagement	1,00		
	Probabilistic reasoning	0,35	1,00	
		(0,130)		
	Feedback	-0,06	0,30*	1,00
		(0,789)	(0,036)	

4. Discussion

The results of studying engagement longitudinally during the intervention show the changing and dynamic nature of this variable. This character can be observed both for the components of engagement (Figure 2) and for the descriptive values of engagement (Figure 3). This result agrees with previous findings, which suggest studying engagement as a multidimensional and changing aspect [4], [27]–[30]. This argument is reinforced since, on average across groups, the session with the highest engagement was the first (4.20) and the lowest the third (3.80), the latter extending for groups one (3.71) and two (3.63). Additionally, each group had its maximum engagement in different sessions. From the results it can be concluded that students had relatively high engagement, in the upper ranges of self-report, both for each group and on average. This is in agreement with the results of involving an DGBL in school environments [10], [13], [16], which predicted high student engagement.

Sessions that have high engagement may lead to students being more willing to teaching-learning experiences, which may be an option to solve the problems in teaching probabilistic reasoning. Now, in order to analyze which conditions can favor probabilistic reasoning from the results of this study, the next step will be to analyze this variable.

When reviewing probabilistic reasoning, this varied during the study. On average, the session with the highest score was the second (3.47) and the third, the lowest (3.4), which was also obtained for each group individually. When examining the third session in the didactic sequence, the focus was on arguing the decisions and comparing them with each other, as proposed in the PRR. Given the low results of probabilistic reasoning in this session, it can be established that this component of probabilistic reasoning seems to be the most complex. Studies on metacognition show that making thinking visible is one of the most complex components of the type of studies that deal with verbalizations; this has been established in studies highlighting the difficulty in measuring metacognition at high reliability levels [63]–

[66]. Particularly, Shilo and Kramarski [67] state that, in mathematics, the success of making students reflect on their actions and thoughts in mathematics, to be able to go beyond procedural knowledge and likewise, depends on the development of strategies and protocols that allow evidencing and developing these abilities. This, within this study, implies the importance of strengthening the open feedback protocol to develop probabilistic reasoning.

When analyzing the results of the feedback level presented in Figure 5, it is observed that open feedback (levels 4 and 5) was maintained for each of the groups during all sessions, with the exception of session 1 for group 1. This variable also presents a variation, which is also supported and significant, according to the values presented in Table 5 for the Kruskal-Wallis test. Having this type of feedback for the sessions that were proposed would allow evidencing higher-order abilities in students [14], [41], [42] and would allow students to perform better in situations in which they face similar contexts [44], [45]. Another consequence of the results presented in Table 3, is that the feedback protocol turns out to be dynamic and not static at a single level, responding to the scaffolding proposals of current conceptions of education [14]–[17]. This model can be taken as a basis for the development of sequences intended to involve open-ended feedback. This result provides the opportunity to use this model and extend it to other educational practices in a broader way, again, because of its dynamic nature. Which will allow bringing students closer to the desired learning [15], [68]–[71].

With these results on the study variables: engagement, level of feedback and probabilistic reasoning individually, it is now of interest to analyze how they correlate with each other and what conclusions can be drawn from the data obtained. Thus, we will proceed to review the results presented in the study.

When examining the correlations between variables, segmented by sessions (see Table 6), the correlations between feedback and probabilistic reasoning stand out in three of the sessions. These correlations were positive in sessions 1, 2 and 4, with different levels of significance, which implies that a positive change in feedback was accompanied by a positive change in probabilistic reasoning; the difference in the degrees of significance implies that the effect of feedback was accompanied by a smaller change in probabilistic reasoning. These results demonstrate that in the DGBL sessions of this study higher levels of feedback were accompanied by changes in students' probabilistic reasoning. This suggests that the support provided to the student during the game sessions will also contribute to creating environments where students' probabilistic reasoning can develop. Fischer [72] discusses this aspect by mentioning how the environment in which expertise in abilities is developed is fundamental for successful development, which in the case of this study implies the importance in overt feedback.

With respect to the other variables, only a moderate correlation was found, with a coefficient of 0.313 (p = 0.010) for probabilistic reasoning and engagement in session 3, indicating that positive changes in probabilistic reasoning were accompanied by positive changes in engagement. This shows that with an appropriate level of mathematical ability, students will also have higher engagement [25] and that when students begin to use probabilistic reasoning, this being finding patterns, using models, and making decisions in situations of uncertainty based on probability concepts, students will be engaged in the activity [19], [32]. An academic implication of these results is that, in order to have high engagement in teaching-learning contexts in DGBL environments, promoting levels of probabilistic reasoning will be critical.

The results of segmenting the study by groups, presented in Table 5 for the correlations between variables, show correlations between the probabilistic reasoning variables and the level of feedback. These data show moderate and statistically significant positive correlations for groups 1 and 3, again highlighting the importance of good feedback during DGBL sessions to foster the development of probabilistic reasoning abilities. This suggests that useful and adequate feedbacks can lead to an improvement in the studied ability, which is consistent with the results of previous studies on the relationship between feedback and academic success [14],

[44], [45]. This result and analysis can also be extended for groups 2 and 4, even if their degree of significance is lower, implying a smaller change between variables.

What was found in this study allows affirming that the protocol was successful, to the extent that it allowed better scaffolding and thus better learning [68]–[71], [73]. Likewise, it can be established that it is a successful protocol for promoting probabilistic reasoning in DGBL environments [35]–[37].

The results obtained in this study allow us to establish that, in didactic sequences based on DGBL and involving open feedback systems, open feedback is positively correlated with levels of probabilistic reasoning, so that high levels of feedback will be accompanied by high levels of probabilistic reasoning, with a high significance in most cases. This result opens the door to open feedback systems and protocols that can generate new strategies for didactically approaching the development of abilities in the field of probability. Similarly, DGBL strategies were shown to maintain engagement in students above the lowest three levels of self-report during the intervention sessions, which is consistent with previous studies on how DGBL promotes high engagement [10], [13], [16], [31].

Between engagement and feedback, a positive correlation could only be evidenced during session 3, but in no other session nor segmenting by groups could this relationship be found; the above suggests the possibility that better feedback is accompanied by better engagement.

With respect to the correlation between level of feedback and engagement, none of the results support that this correlation was present throughout the study.

4.1   Limitations and future research

A proposal for future studies, which is worthwhile, is to have an intervention that addresses at least eight to ten sessions per group, with which more conclusive results could be obtained. However, this is difficult due to the academic and curricular conditions of secondary education institutions. Another possibility to improve the study is to use results obtained in studies on metacognition, to elaborate instruments and protocols that would allow the opportunity to obtain clearer and better-quality verbalizations from the students, in order to better and more comprehensively evaluate the probabilistic reasoning of the students.

5. Conclusions

The results obtained in this study have important implications in the field of DGBL and teaching, establishing correlations between engagement, probabilistic reasoning and open feedback systems. The study addresses the research question about the relationship between the variables by revealing positive correlations between open-ended feedback and probabilistic reasoning in DGBL environments, while showing a weak correlation between engagement and probabilistic reasoning in DGBL sessions.

Some practical applications of this study can be found in the feedback protocols developed for the research, which proved to be open-ended feedback associated with the enhancement of probabilistic reasoning. This opens the door to creating DGBL sessions with open feedback to promote learning and demonstrates how non-traditional strategies, such as DGBL and open feedback systems, can be effectively incorporated into formal subjects like mathematics, resulting in relatively high engagement and high levels of probabilistic reasoning.

In summary, this study offers valuable insights into the relationship between probabilistic reasoning and engagement, which when coupled with open-ended feedback, enhance students' probabilistic reasoning. These analyses also highlight the dynamic nature of the variables studied, emphasizing the positive correlation between feedback and probabilistic reasoning.

These findings underscore the importance of teacher-student interactions and provide significant implications for the development and planning of future didactic sequences.

Acknowledgments

References

[1] C. Batanero, E. J. Chernoff, J. Engel, H. S. Lee, and E. Sánchez, "Research on Teaching and Learning Probability," in *Research on Teaching and Learning Probability*, Cham: Springer International Publishing, 2016, pp. 1–33. doi: 10.1007/978-3-319-31625-3_1.

[2] E. J. Chernoff and G. L. Russell, "The Fallacy of Composition: Prospective Mathematics Teachers'Use of Logical Fallacies," *Canadian Journal of Science, Mathematics and Technology Education*, vol. 12, no. 3, pp. 259–271, Jul. 2012, doi: 10.1080/14926156.2012.704128.

[3] E. J. Chernoff and R. Zazkis, "From personal to conventional probabilities: from sample set to sample space," *Educational Studies in Mathematics*, vol. 77, no. 1, pp. 15–33, May 2011, doi: 10.1007/s10649-010-9288-8.

[4] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School Engagement: Potential of the Concept, State of the Evidence," *Review of Educational Research*, vol. 74, no. 1, pp. 59–109, 2004, doi: 10.3102/00346543074001059.

[5] Y. Zhang, S. Russell, and S. Kelly, "Engagement, achievement, and teacher classroom practices in mathematics: Insights from TIMSS 2011 and PISA 2012," *Studies in Educational Evaluation*, vol. 73, p. 101146, 2022, doi: https://doi.org/10.1016/j.stueduc.2022.101146.

[6] B. C. Douglas, S. V. Satyugjit, J. Barnes, and D. M. Adams, "Self-explanation and digital games: Adaptively increasing abstraction," *Computers & Education*, vol. 103, pp. 28–43, 2016, doi: 10.1016/j.compedu.2016.09.010.

[7] C. Haelermans and J. Ghysels, "The effect of individualized digital practice at home on math skills–Evidence from a two-stage experiment on whether and why it works," *Computers & Education*, vol. 113, pp. 119–134, 2017, doi: 10.1016/j.compedu.2017.05.010.

[8] J. Hutain and N. Michinov, "Improving student engagement during in-person classes by using functionalities of a digital learning environment," *Computers & Education*, vol. 183, p. 104496, 2022, doi: 10.1016/j.compedu.2022.104496.

[9] B. Brezovszky *et al.*, "Effects of a mathematics game-based learning environment on primary school students' adaptive number knowledge," *Computers & Education*, vol. 128, pp. 63–74, 2019, doi: 10.1016/j.compedu.2018.09.011.

[10] R. Farah *et al.*, "REVIEW OF INTERNATIONAL GEOGRAPHICAL EDUCATION Algebraic Lab: Pedagogical Tool to Teach and Learn Algebra through Game," *Review of International Geographical Education Online*, vol. 11, pp. 951–962, Sep. 2021, doi: 10.48047/rigeo.11.04.88.

[11] F. Fraga-Varela, E. Vila-Couñago, and A. Rodríguez-Groba, "Serious Games and Mathematical Fluency: A Study from the Gender Perspective in Primary Education," *Sustainability*, vol. 13, no. 12, 2021, doi: 10.3390/su13126586.

[12] A. Hershkovitz, M. Tabach, and A. Cohen, "Online Activity and Achievements in Elementary School Mathematics: A Large-Scale Exploration," *Journal of Educational Computing Research*, vol. 60, no. 1, pp. 258–278, 2022, doi: 10.1177/07356331211027822.

[13] J. Moon and F. Ke, "In-Game Actions to Promote Game-Based Math Learning Engagement," *Journal of Educational Computing Research*, vol. 58, no. 4, pp. 863–885, 2020, doi: 10.1177/0735633119878611.

[14] Y. Attali, "Effects of multiple-try feedback and question type during mathematics problem solving on performance in similar problems," *Computers & Education*, vol. 86, pp. 260–267, 2015, doi: /10.1016/j.compedu.2015.08.011.

[15] J. Hattie and H. Timperley, "The Power of Feedback," *Review of Educational Research*, vol. 77, no. 1, pp. 81–112, 2007, doi: 10.3102/003465430298487.

[16] J. Huizenga, G. T. M. Dam, J. Voogt, and W. Admiraal, "Teacher perceptions of the value of game-based learning in secondary education," *Computers & Education*, vol. 110, pp. 105–115, Mar. 2017, doi: 10.1016/j.compedu.2017.03.008.

[17] J. S. Tan and W. Chen, "Peer feedback to support collaborative knowledge improvement: What kind of feedback feed-forward?," *Computers & Education*, p. 104467, 2022, doi: 10.1016/j.compedu.2022.104467.

[18] E. Fischbein, M. S. Nello, and M. S. Marino, "Factors affecting probabilistic judgements in children and adolescents," *Educational Studies in Mathematics*, vol. 22, no. 6, pp. 523–549, Dec. 1991, doi: 10.1007/BF00312714.

[19] T. A. Grotzer, S. L. Solis, M. S. Tutwiler, and M. P. Cuzzolino, "A study of students' reasoning about probabilistic causality: Implications for understanding complex systems and for instructional design," *Instructional Science*, vol. 45, no. 1, pp. 25–52, Feb. 2017, doi: 10.1007/s11251-016-9389-6.

[20] M. Ricart and A. Estrada, "Combinatorial and Proportional Task: Looking for Intuitive Strategies in Primary Education," *Mathematics*, vol. 10, no. 8, 2022, doi: 10.3390/math10081340.

[21] O. G. Board, "PISA 2021 Mathematics Framework (first draft)." Stockholm, 2018.

[22] J. A. Fredricks *et al.*, "Using qualitative methods to develop a survey measure of math and science engagement," *Learning and Instruction*, vol. 43, pp. 5–15, 2016, doi: https://doi.org/10.1016/j.learninstruc.2016.01.009.

[23] T. González Ramírez and A. García-Hernández, "Estudio de los factores de estudiantes y aulas que intervienen en el engagement y rendimiento académico en Matemáticas Discretas," *Revista Complutense de Educación*, vol. 31, no. 2, pp. 195–206, 2020.

[24] E. A. Patall, K. A. Pituch, R. R. Steingut, A. C. Vasquez, N. Yates, and A. A. U. Kennedy, "Agency and high school science students' motivation, engagement, and classroom support experiences," *Journal of Applied Developmental Psychology*, vol. 62, pp. 77–92, 2019, doi: https://doi.org/10.1016/j.appdev.2019.01.004.

[25] D. W. Putwain and P. Wood, "Riding the bumps in mathematics learning: Relations between academic buoyancy, engagement, and achievement," *Learning and Instruction*, vol. 83, p. 101691, 2023, doi: https://doi.org/10.1016/j.learninstruc.2022.101691.

[26] Y. Zheng, S. Yu, B. Wang, and Y. Zhang, "Exploring student engagement with supervisor feedback on master's thesis: Insights from a case study," *Innovations in Education and Teaching International*, vol. 57, no. 2, pp. 186–197, 2020, doi: 10.1080/14703297.2019.1617181.

[27] J. A. Fredricks and W. McColskey, "The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments," in *Handbook of Research on Student Engagement*, S. L. Christenson, A. L. Reschly, and C. Wylie, Eds., Boston, MA: Springer US, 2012, pp. 763–782. doi: 10.1007/978-1-4614-2018-7_37.

[28] J. Glanville, "The Measurement of School Engagement Assessing Dimensionality and Measurement Invariance Across Race and Ethnicity," *Educational and Psychological Measurement - EDUC PSYCHOL MEAS*, vol. 67, pp. 1019–1041, Jun. 2007, doi: 10.1177/0013164406299126.

[29] F. H. Veiga, J. Reeve, K. Wentzel, and V. Robu, "Assessing students' engagement: A review of instruments with psychometric qualities," in *I Congresso Internacional Envolvimento dos Alunos na Escola: Perspetivas da Psicologia e Educação*, 2014, pp. 38–57.

[30] M.-T. Wang and J. Degol, "Staying Engaged: Knowledge and Research Needs in Student Engagement," *Child Development Perspectives*, vol. 8, no. 3, pp. 137–143, 2014, doi: https://doi.org/10.1111/cdep.12073.

[31] Y.-L. Lu and C.-J. Lien, "Are They Learning or Playing? Students' Perception Traits and Their Learning Self-Efficacy in a Game-Based Learning Environment," *Journal of Educational Computing Research*, vol. 57, no. 8, pp. 1879–1909, 2020, doi: 10.1177/0735633118820684.

[32] S. Furlan, F. Agnoli, and V. F. Reyna, "Intuition and analytic processes in probabilistic reasoning: The role of time pressure," *Learning and Individual Differences*, vol. 45, pp. 1–10, 2016, doi: https://doi.org/10.1016/j.lindif.2015.11.006.

[33] J. Byun and E. Joung, "Digital game-based learning for K–12 mathematics education: A meta-analysis," *School Science and Mathematics*, vol. 118, no. 3–4, pp. 113–126, 2018, doi: https://doi.org/10.1111/ssm.12271.

[34] F. Fraga, E. Vila Couñago, and E. Martínez Piñeiro, "Impacto de los juegos serios en la fluidez matemática: Un estudio en Educación Primaria," *Comunicar: revista científica iberoamericana de comunicación y educación*, vol. 69, pp. 125–135, 2021, doi: 10.3916/C69-2021-10.

[35] R. Capone, A. Pisa, M. Trerotola, and M. Del Sorbo, "CHALLENGE-BASED LEARNING AND GAME-BASED LEARNING TO IMPROVE MATHEMATICAL COMPETENCIES: AN ITALIAN CASE STUDY IN SECONDARY SCHOOL," Jul. 2019. doi: 10.21125/edulearn.2019.0206.

[36] E. Paparistodemou, M. Meletiou-Mavrotheris, and C. Vasou, "Designing and Playing Games in Scratch: Smart Pedagogy of a Game-Based Challenge for Probabilistic Reasoning," in *Smart Pedagogy of Game-based Learning*, L. Daniela, Ed., Cham: Springer International Publishing, 2021, pp. 57–70. doi: 10.1007/978-3-030-76986-4_4.

[37] P. Vankúš and M. Čujdíková, "OPPORTUNITIES FOR DEVELOPMENT OF MATHEMATICAL THINKING THROUGH PLAYING VIDEO GAME MACHINARIUM," Mar. 2022, pp. 352–361. doi: 10.21125/inted.2022.0164.

[38] B. D. Coller and M. J. Scott, "Effectiveness of using a video game to teach a course in mechanical engineering," *Computers & Education*, vol. 53, no. 3, pp. 900–912, 2009, doi: https://doi.org/10.1016/j.compedu.2009.05.012.

[39] S. Perini, R. Luglietti, M. Margoudi, M. Oliveira, and M. Taisch, "Learning and motivational effects of digital game-based learning (DGBL) for manufacturing education –The Life Cycle Assessment (LCA) game," *Computers in Industry*, vol. 102, pp. 40–49, 2018, doi: https://doi.org/10.1016/j.compind.2018.08.005.

[40] J. Tay, Y. M. Goh, S. Safiena, and H. Bound, "Designing digital game-based learning for professional upskilling: A systematic literature review," *Computers and Education*, vol. 184. 2022. doi: 10.1016/j.compedu.2022.104518.

[41] S. van Vondel, H. Steenbeek, M. van Dijk, and P. van Geert, "Ask, don't tell; A complex dynamic systems approach to improving science education by focusing on the co-construction of scientific understanding," *Teaching and Teacher Education*, vol. 63, pp. 243–253, 2017, doi: https://doi.org/10.1016/j.tate.2016.12.012.

[42] H. B. Meindertsma, M. W. G. van Dijk, H. W. Steenbeek, and P. L. C. van Geert, "Assessment of Preschooler's Scientific Reasoning in Adult–Child Interactions: What Is the Optimal Context?," *Research in Science Education*, vol. 44, no. 2, pp. 215–237, Apr. 2014, doi: 10.1007/s11165-013-9380-z.

[43] A. W. Oliveira, "Improving teacher questioning in science inquiry discussions through professional development," *Journal of Research in Science Teaching*, vol. 47, no. 4, pp. 422–453, 2010, doi: https://doi.org/10.1002/tea.20345.

[44] A. C. Butler, N. Godbole, and E. J. Marsh, "Explanation feedback is better than correct answer feedback for promoting transfer of learning," *Journal of Educational Psychology*, vol. 105, no. 2, pp. 290–298, 2013, doi: 10.1037/a0031026.

[45] V. Law and C.-H. Chen, "Promoting science learning in game-based learning with question prompts and feedback," *Computers & Education*, vol. 103, pp. 134–143, 2016, doi: 10.1016/j.compedu.2016.10.005.

[46] L. Sun, H. Ruokamo, P. Siklander, B. Li, and K. Devlin, "Primary school students' perceptions of scaffolding in digital game-based learning in mathematics," *Learning, Culture and Social Interaction*, vol. 28, p. 100457, 2021, doi: 10.1016/j.lcsi.2020.100457.

[47] K.-H. Yang and B.-C. Lu, "Towards the successful game-based learning: Detection and feedback to misconceptions is the key," *Computers & Education*, vol. 160, p. 104033, 2021, doi: 10.1016/j.compedu.2020.104033.

[48] W. R. Shadish, T. D. Cook, and D. T. Campbell, *Experimental and quasi-experimental designs for generalized causal inference.* Houghton, Mifflin and Company, 2002.

[49] J. Frost, "Repeated Measures Designs: Benefits and an ANOVA Example," *Statistics By Jim*. Mar. 2017. [Online]. Available: https://statisticsbyjim.com/anova/repeated-measures-designs-benefits-anova-example/

[50] T. Cavanagh, "Dicey Dungeons." 2019.

[51] P. Blatchford, A. D. Pellegrini, and E. Baines, *The child at school: Interactions with peers and teachers*. Routledge, 2015.

[52] R. Puche Navarro and O. Ordoñez Morales, "Comprensión, resolución y formación de herramientas científicas en el niño," *R. Puche Navarro, Formación de herramientas científicas en el niño pequeño (2 ed., pág. 46). Cali, Colombia: Universidad del valle*, 2005.

[53] R. Larson and M. Csikszentmihalyi, "The Experience Sampling Method," in *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*, Dordrecht: Springer Netherlands, 2014, pp. 21–34. doi: 10.1007/978-94-017-9088-8_2.

[54] J. A. Schmidt, H. Z. Kackar-Cam, A. D. Strati, and L. Shumow, "The Role of Challenge in Students' Engagement and Competence in High School Science Classrooms: Hispanic and Non-Hispanic Whites Compared.," *NCSSS Journal*, vol. 20, no. 1, pp. 20–26, 2015.

[55] K. Xie, B. C. Heddy, and B. A. Greene, "Affordances of using mobile technology to support experience-sampling method in examining college students' engagement," *Computers & Education*, vol. 128, pp. 183–198, 2019, doi: https://doi.org/10.1016/j.compedu.2018.09.020.

[56] K. Xie, B. C. Heddy, and V. W. Vongkulluksn, "Examining engagement in context using experience-sampling method with mobile technology," *Contemporary Educational Psychology*, vol. 59, p. 101788, 2019, doi: https://doi.org/10.1016/j.cedpsych.2019.101788.

[57] S. Ochoa-Angrino, J. A. Montes-González, and T. Rojas-Ospina, "Percepción de habilidad, reto y relevancia como predictores de compromiso cognitivo y afectivo en estudiantes de secundaria," *Universitas Psychologica*, vol. 17, no. 5, pp. 1–18, 2018, doi: 10.11144/Javeriana.upsy17-5.phrr.

[58] S. Zirkel, J. A. Garcia, and M. C. Murphy, "Experience-Sampling Research Methods and Their Potential for Education Research," *Educational Researcher*, vol. 44, no. 1, pp. 7–16, 2015, doi: 10.3102/0013189X14566879.

[59] IBM Corp., "IBM SPSS Statistics for Windows." Armonk, NY: IBM Corp, 2021. [Online]. Available: https://hadoop.apache.org

[60] M. Nagaiah and K. Ayyanar, "Software for Data Analysis in SPSS On over view," Feb. 2016.

[61] A. Rahman and M. G. Muktadir, "SPSS: An Imperative Quantitative Data Analysis Tool for Social Science Research," vol. V, pp. 300–302, Oct. 2021, doi: 10.47772/IJRISS.2021.51012.

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

[63] A. Desoete and B. De Craene, "Metacognition and mathematics education: an overview," *ZDM*, vol. 51, no. 4, pp. 565–575, Aug. 2019, doi: 10.1007/s11858-019-01060-w.

[64] T. Duangnamol, T. Supnithi, G. Srijuntongsiri, and M. Ikeda, "Computer-Supported Meta-reflective Learning Model via mathematical word problem learning for training metacognition," *Research and Practice in Technology Enhanced Learning*, vol. 13, no. 1, p. 14, Sep. 2018, doi: 10.1186/s41039-018-0080-1.

[65] A. Kuzle, "Assessing metacognition of grade 2 and grade 4 students using an adaptation of multi-method interview approach during mathematics problem-solving," *Mathematics Education Research Journal*, vol. 30, no. 2, pp. 185–207, Jun. 2018, doi: 10.1007/s13394-017-0227-1.

[66] M. V. J. Veenman and D. van Cleef, "Measuring metacognitive skills for mathematics: students'self-reports versus on-line assessment methods," *ZDM*, vol. 51, no. 4, pp. 691–701, Aug. 2019, doi: 10.1007/s11858-018-1006-5.

[67] A. Shilo and B. Kramarski, "Mathematical-metacognitive discourse: how can it be developed among teachers and their students? Empirical evidence from a videotaped lesson and two case studies," *ZDM*, vol. 51, Dec. 2018, doi: 10.1007/s11858-018-01016-6.

[68] S. W. Chong, "Reconsidering student feedback literacy from an ecological perspective," *Assessment & Evaluation in Higher Education*, vol. 46, no. 1, pp. 92–104, 2021, doi: 10.1080/02602938.2020.1730765.

[69] C. Mercader, G. Ion, and A. Díaz-Vicario, "Factors influencing students' peer feedback uptake: instructional design matters," *Assessment & Evaluation in Higher Education*, vol. 45, no. 8, pp. 1169–1180, 2020, doi: 10.1080/02602938.2020.1726283.

[70] A. C. F. Tam, "Undergraduate students' perceptions of and responses to exemplar-based dialogic feedback," *Assessment & Evaluation in Higher Education*, vol. 46, no. 2, pp. 269–285, 2021, doi: 10.1080/02602938.2020.1772957.

[71] J. Wood, "A dialogic technology-mediated model of feedback uptake and literacy," *Assessment & Evaluation in Higher Education*, vol. 46, no. 8, pp. 1173–1190, 2021, doi: 10.1080/02602938.2020.1852174.

[72] K. W. Fischer, "Ciclos dinámicos de desarrollo cognitivo cerebral: cómo medir el crecimiento de la mente, el cerebro y la educación," in *El Cerebro educado: ensayos sobre la neuroeducación*, 1ª ed., A. M. Battro and P. Léna, Eds., Barcelona: Gedisa, 2016.

[73] G. Ion, A. S. Martí, and I. A. Morell, "Giving or receiving feedback: which is more beneficial to students' learning?," *Assessment & Evaluation in Higher Education*, vol. 44, no. 1, pp. 124–138, 2019, doi: 10.1080/02602938.2018.1484881.