

Development and Evaluation of a Math mCSCL Platform with Aggregated Content-Based Filtering and Guessing Detection

Rex P. Bringula,¹ Francis Arlando L. Atienza,² and John Paul P. Miranda³

¹College of Computer Studies and Systems, University of the East, Manila, Philippines

²School of Engineering, Computing and Architecture – IT Department, National University-Philippines, Manila, Philippines

³College of Computing Studies, Pampanga State University, Mexico, Philippines

(Received 27 April 2025; Revised 13 September 2025; Accepted 04 December 2025; Published online 04 January 2026)

Abstract: This study reported the development and evaluation of an intelligent math mobile computer-supported collaborative learning (mCSCL) system for solving fractions. The software included a recommender system and a guessing detector. The aggregated model of a content-based filtering (CBF) algorithm for a group was used to build the recommendation function of the software. The dice coefficient was utilized to determine the dissimilarity of the selected game settings vis-à-vis the problem space PPP. Meanwhile, the guessing detector comprised the Rasch model and computational fluency (CF). CF, in turn, had two components: response time and accuracy. All three indicators had to be flagged as true to classify a student as guessing. Fifty-five Grade 5 students from the elementary departments of four universities in Manila participated in the study. Results showed that the software was able to detect the guessing behavior of each student and provided individualized feedback to those who exhibited guessing. The software also generated recommendations (i.e., game settings), confirming the effectiveness of the CBF and dice coefficient. Content analysis revealed that the software received favorable remarks from the students, implying its relevance to their mathematics learning. However, despite these capabilities, students still exhibited under- and over-practice of certain skills. Limitations of the software were also discussed.

Keywords: adaptive; arithmetic; artificial intelligence; educational technology

1. INTRODUCTION

Computer-supported collaborative learning (CSCL) is an approach where computers and other digital tools support and enhance group learning activities [1]. With the proliferation of digital devices like cell phones and tablets, mobile CSCL (mCSCL) has emerged [2]. Mathematics education is one of the fields that leverage the functionalities of mCSCL. It has long been established that it has a positive impact on both mathematics academic performance and social aspects (e.g., group cohesion, group interactions, confidence, motivation, interest, satisfaction, and others) [3].

Although math mCSCL has existed for some time, it is noticeable that the focus on developing math mCSCL intelligence has been overlooked. In the study conducted by Bringula and Atienza [3], they found that the math mCSCL developed from 2007 to 2021 did not incorporate artificial intelligence (AI). Due to the current state of existing math mCSCL applications that adopt a one-size-fits-all approach, various research gaps have emerged. The current math mCSCL treats all learners the same, without being adaptive to individual and group proficiency levels or pace. Without AI functionality in math mCSCL, it is difficult to monitor the group and provide effective support to facilitate group collaboration (e.g., balancing participation, detecting guessing, and detecting over-/under-practice). Students are also not provided with timely prompts and feedback suited to their individual and group performance. Finally, educators may have limited insights into what is happening

in the math mCSCL environment, which consequently prevents them from providing appropriate manual interventions.

This study was designed to address the issues identified in the earlier discussions. Its objectives were to manage under- and over-practice, detect instances of guessing during the game, and recommend problems based on the performance of the group. Various algorithms were implemented in the development of the math mCSCL, specifically for solving fractions. The developed software was called *Ibigkas! Math*, which is subsequently referred to interchangeably as “the software” or “the game.” This study reported the design, development, and evaluation results of the software.

The succeeding sections are organized as follows. The Related Work section reviews earlier studies. The Design and Development of the Software section explains the development process. The Testing, Software Usage, and Evaluation section describes the testing procedures, user experiences, and evaluation metrics. The Results of the Content Analysis are then presented, followed by the Discussion section. The paper ends with the Conclusion, Recommendations, and Future Works, which summarize the results, suggest practical actions, and indicate directions for future research.

2. RELATED WORK

A. MOBILE CSCL

Several studies have examined the scope of research on mCSCL. Amara *et al.* [4] reviewed 12 studies published between 2005 and

Corresponding author: Rex P. Bringula (e-mail: rex.bringula@ue.edu.ph).

2013 that focused on group formation strategies. Learners were most often grouped by personal characteristics such as age, gender, interests, preferences, and prior learning experiences, including performance scores. In addition, more than 20 types of learning behaviors were used to form groups, and context information was also a common basis for grouping students.

Sung *et al.* [5] conducted a meta-analysis of 48 peer-reviewed journal articles and doctoral dissertations published between 2000 and 2015. Their results showed that mCSCL produced an above-average impact on academic performance (i.e., an overall mean effect size = 0.516). Learning achievement was the most frequently assessed outcome (66%), followed by learning attitudes (22%) and interaction (12%). Participants were primarily college students (35%) and elementary students (33%). Only five studies focused on mathematics. Group sizes most often included mixed groups (29%) and triads (21%). Interventions usually lasted one to four weeks (33%), most often in classroom settings (73%). A notable limitation was that more than half of the studies (56%) did not report group composition. Although students worked collaboratively, rewards were given individually in most cases (81%).

More recently, Peramunugamage *et al.* [6] reviewed 48 studies on mobile collaborative learning (CL) in engineering education. The most frequently addressed topics were mobile application development and agent-based systems. Common research designs included mixed methods, case studies, and experimental approaches. Reported sample sizes varied widely, ranging from 26 to 1,121 participants.

B. MATHEMATICS mCSCL

CL enhances students' understanding of mathematical concepts by encouraging interaction and knowledge exchange [7]. The use of mobile devices extends these benefits by creating interactive environments that facilitate collaboration. In mCSCL, students engage with mathematical ideas in real-world contexts through situational learning, which makes abstract concepts more accessible [8].

Mobile tools also promote negotiation and strengthen social interaction skills, as students discuss and solve mathematical problems together [9]. Unlike desktop computers, handheld devices support face-to-face communication, enabling learners to coordinate more effectively and to understand their peers' perspectives [10]. These advantages highlight the potential of mathematics mCSCL. However, research in this field remains scarce, and existing studies provide only partial insights without conclusive evidence.

C. DESIGN IMPROVEMENTS FOR mCSCL

The design of CSCL and mCSCL systems presents several challenges that limit their benefits. One issue concerns the balance between over-practice and under-practice of skills [11]. Groups often select tasks that they already find manageable, which narrows opportunities for growth and reinforces existing strengths rather than addressing weaknesses [11,12]. This behavior is partly shaped by group conformity, where members align with collective choices even when those choices are suboptimal for learning [13]. Another challenge lies in the limited ability of current systems to detect reduced participation, which can distort team outcomes [14]. Members may also overestimate their individual contributions, which creates a misleading sense of

productivity that does not match actual group performance [15]. Furthermore, these systems often lack the capacity to adapt to the evolving needs of groups or to assess real-time team underperformance [16].

A further concern is the phenomenon of gaming the system (GTS), which remains underexplored in CSCL and mCSCL contexts. GTS involves students exploiting regularities in the software to succeed without engaging meaningfully with the material [17]. Guessing represents one common form of this behavior [18]. Studies in intelligent tutoring systems have examined GTS extensively, applying models such as decision trees, Bayesian methods, neural networks, and latent response models to detect it [18].

Researchers have also investigated design strategies to reduce GTS. Effective approaches include integrating pedagogical agents to discourage gaming [19], providing textual feedback [20], informing students that their actions are being monitored [21], and making students aware of peers' gaming behaviors [22]. Delaying access to hints has shown less promise [23], while withholding points from students who engage in off-task behaviors has also been proposed as a corrective measure [24].

Addressing these limitations may require system designs that are more adaptive and responsive to group dynamics. One promising direction is the integration of group collaborative filtering within recommender systems. This approach enables systems to suggest tasks aligned with desired learning states and group progress, encouraging balanced participation and improved engagement [25]. Another strategy involves applying Item Response Theory (IRT) to detect behaviors such as guessing. IRT can identify cases where students provide correct answers despite lacking the ability level needed for the task, indicating potential guessing [26,27]. The predictive accuracy of IRT may be enhanced by incorporating measures of computational fluency (CF), assessed through accuracy and response speed [28,29]. Combining these methods could strengthen system feedback and provide more reliable indicators of student progress, which can support adaptations that better align with both individual and group needs.

3. DESIGN AND DEVELOPMENT OF THE SOFTWARE

A. SOFTWARE DESCRIPTION

The software is a web-based, game-based learning application for Grade 5 students. It is a collaborative game that covers addition, subtraction, multiplication, and division of fractions. The application generates arithmetic problems, and it displays on one of the team members' mobile devices (Fig. 1). The player must read aloud the arithmetic problems. The answers are presented in multiple-choice format. The correct answer only appears on one of the team members' devices. The game could be played by 2 to 4 members [30].

The game settings included the types of arithmetic problems (addition, subtraction, multiplication, and division), difficulty levels (very easy, easy, medium, hard, and very hard), and speed (very slow, slow, medium, fast, and very fast). The game scores were based on the speed setting. The untimed speed had no equivalent points. The points for the other settings were as follows: very slow corresponds to 2 points, slow for 5 points, medium for 10 points, fast for 15 points, and very fast for 20 points.

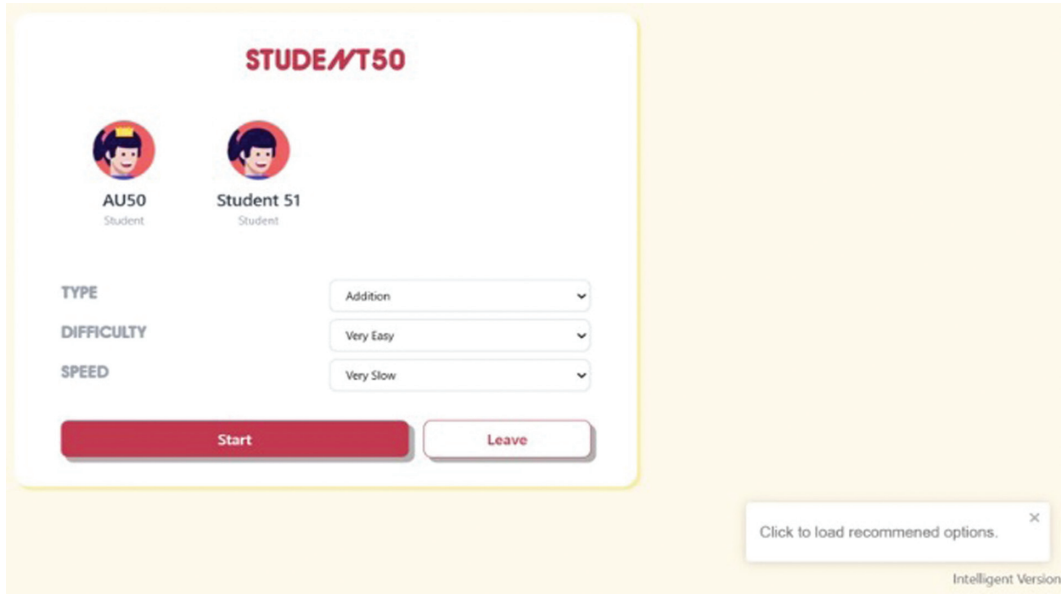


Fig. 1. Selected game settings of a student.

B. GUESSING DETECTION USING THE RASCH MODEL AND COMPUTATIONAL FLUENCY

“GTS” refers to a student’s intentional use of the properties of the software to complete a task, rather than engaging in deep thinking or understanding of the material [31]. Guessing is a form of GTS [18,32]. Prior works extensively investigated guessing behavior in an ITS and used different models (e.g., decision tree, Bayesian method, latent response model, etc.) to detect this behavior [18,33].

Guessing has been observed both at the individual and group levels. In an individual setting, guessing is related to the learner’s competency and intention [34]. At the group level, a member makes guesses because they want the team to succeed (i.e., to win a prize in a competition) [12]. Hence, the guessing detection of each member should be implemented in the software. The Rasch model (RM) of IRT is an appropriate model for detecting individual students at the group level.

The RM is a statistical approach that provides a probabilistic model that attempts to explain the response of a person to an item [35,36]. It has been applied to evaluate and improve the validity, accuracy, and reliability of items in a test [37]. It is based on the assumption that the performance of a person on a test item can be predicted (or explained) by latent variables like abilities [38]. In other words, a person may find the items on a certain instrument easy or difficult depending on their abilities relative to all those taking the same test. The relationship between the abilities of the person and the probability that the item can be answered correctly makes this model appropriate for detecting guessing. RM is mathematically defined below (Equation 1). Equation 1 is a one-parameter logistic model:

$$P_i(\theta) = \frac{e^{\theta - b_i}}{1 + e^{\theta - b_i}} \quad i = 1, 2, 3, \dots, n \quad (1)$$

where

- $P_i(\theta)$ is the probability that a randomly chosen student with ability θ answers item i correctly;
- θ is the ability of the student (measured in logits);

- b_i is the item’s difficulty parameter (measured in logits);
- n is the number of items in the test;
- e is Euler’s number with a value of 2.718 (rounded off to three decimal digits).

This model can detect misfits. Misfit occurs when the ability of the students does not match the difficulty of the item. If the students’ ability is less than the difficulty of the item, and yet that student correctly answered an item, this is considered a misfit [27]. This test behavior is construed as guessing [39].

However, RM only uses the final response data (right or wrong) to detect guessing. It does not account for the CF of the students. CF is an indicator of mathematics mastery. It is defined as the ability of the students to answer math problems quickly and accurately [40,41]. Based on this definition, CF was measured in terms of response time (RT) and accuracy (Acc). RT is the time spent by the students in answering an item test [42]. Rapid guessing is characterized by the RT to answer an item being very short relative to the amount of time required for the items [43]. However, RT alone does not indicate CF. The Acc of answers should also be included alongside RT. Acc is the ratio between correct attempts and the number of attempts [43].

Combining the results of the RM, RT, and Acc, the student’s guessing behavior is defined below (Definition 1). The guessing behavior of a student s (G_s) is a dichotomous classification based on the results of the RM, RT, and Acc. The authors determined in their separate study the threshold values for RT and Acc to detect if a student is guessing [30]. They found that if a student’s RT is less than 1.7 seconds and the average Acc is greater than 13%, then RM and Acc will be flagged as 1 (i.e., possibly guessing). All three indicators must be true for the software to determine that the student is guessing. Fig. 2 shows a screenshot of a student who was detected guessing.

Definition 1. The guessing behavior of a student s (G_s) is a dichotomous classification and is a function of guessing (RM), response time (RT), and accuracy (ACC) of computational fluency:

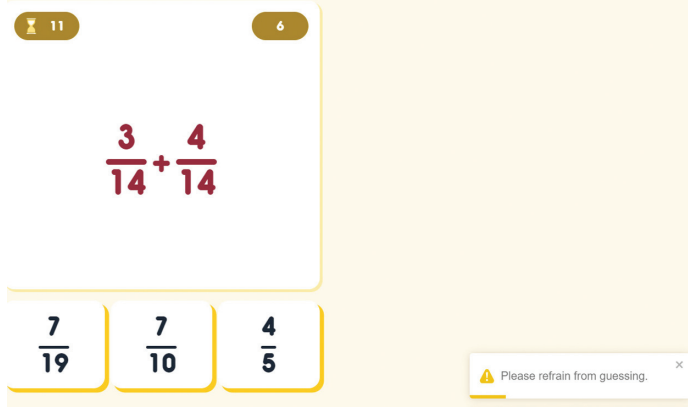


Fig. 2. A student was detected guessing and was reminded not to engage in a guessing behavior.

$$G_s = \begin{cases} \text{guessing,} & RM \wedge RT \wedge Acc = 1 \\ \text{not guessing,} & \text{otherwise} \end{cases}$$

C. AVOIDING UNDER-/OVER-PRACTICE OF SKILL THROUGH A RECOMMENDER SYSTEM

A recommender system was developed to avoid under- and over-practice of mathematics skills. The recommender system was developed based on the aggregated model of a content-based filtering (CBF) algorithm for a group (Fig. 3). As shown in Fig. 3, the selected problems of the individual user (u_i) comprised the user profile (up). A user profile was composed of problems solved (s_{ij}) by u_i . These user profiles were then combined using the aggregate model. The collected user profiles would form the aggregated profile (gp). The aggregated model of CBF for groups was then applied to the group profile to generate a recommendation (i.e., the problems to be solved by members of the group).

The aggregated CBF algorithm is a three-tuple $R = (U, S, P)$, where R is a recommendation task, $U = \{u_1, u_2, u_3, \dots, u_n\}$ is a set of users, $S_{ij} = \{s_{1j}, s_{2j}, s_{3j}, \dots, s_{ij}\}$ is a set of problems solved, and $P = \{p_1, p_2, p_3, \dots, p_n\}$ is a set of recommended problems to solve. The recommended problem P is a finite set of game settings. There are 100 possible game settings (4 arithmetic operations \times 5 difficulty levels \times 5 speed settings). To generate the recommended problem P game setting, it was based on the prior problems solved (s_{ij}) by the individual users (u_n). The recommended problem P was solved using the dice coefficient (Equation 2), where gp represents the group profile and p_n is the game setting. The dice coefficient can have a value from 0 to 1 (Equation 2; [25]). The lowest similarity value

between gp and p_n was recommended since this means that it was the least problem-solved among the user profiles up:

$$\text{similarity}(gp, p_n) = \frac{2 * |\text{categories}(gp) \cap \text{categories}(p_n)|}{|\text{categories}(gp)| + |\text{categories}(p_n)|} \quad (2)$$

Implementing the aggregate CBF, the software was able to provide a game recommendation setting (Fig. 1). This was generated based on the sample profiles of the users. To further clarify this concept, a sample computation is provided below (Table I, Table II, Fig. 4). For example, Student50 (u_1) and Student51 (u_2) used the game. Student50 hosted the game and initially selected the following game settings: Addition, Very Easy, and Very Slow (Fig. 1). These game settings are p_1 (Table I). Based on this selection, the group profile gp is generated (Table II). The aggregated CBF algorithm then used all possible game settings (Table II) and the group profile gp to compute the dice coefficient (Fig. 4). Based on the sample computation, the game settings Subtraction, Easy, and Slow will have the lowest (zero) similarity index. Thus, the software recommended these game settings (see Fig. 5).

It is worth noting that although the software provided recommendations, students still had the option to accept or reject them. This game setting was based on initial interviews with elementary math teachers.

Table I. Sample problem P (game settings)

Problem P	Game settings		
P ₁	Add	VE	VS
P ₂	Add	VE	S
P ₃	Add	VE	M
P ₄	Add	VE	F
P ₅	Sub	VE	VS
P ₆	Sub	VE	S
...
P ₃₂	Sub	E	S
...
P ₁₀₀	Div	VH	VF

Table II. Problems solved (s_{ij}) by student50 (u_1) generating the group profile gp

Domain	s_{ij}		gp
	u_1	u_2	
Add	x	-	x
Sub	-	-	-
Mul	-	-	-
Div	-	-	-
VE	x	-	x
E	-	-	-
M	-	-	-
D	-	-	-
VD	-	-	-
VS	x	-	x
F	-	-	-

Add – addition, Sub – subtraction, Mul – multiplication, Div – division, VE – very easy, E – easy, M – medium, D – difficult, VD – very difficult, VS – slow, S – slow, F – fast, VF – very fast.

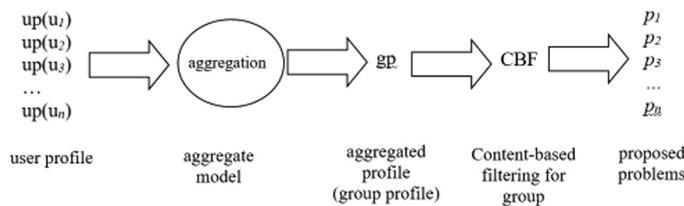


Fig. 3. Implementation of the aggregate model of content-based filtering algorithm for group.

$$\begin{aligned} \text{similarity}(gp, p_1) &= \frac{2 * |\text{categories}(gp) \cap \text{categories}(p_1)|}{|\text{categories}(gp)| + |\text{categories}(p_1)|} = \frac{2 * 2}{5 + 3} = \frac{4}{8} = 0.50 \\ \text{similarity}(gp, p_2) &= \frac{2 * |\text{categories}(gp) \cap \text{categories}(p_2)|}{|\text{categories}(gp)| + |\text{categories}(p_2)|} = \frac{2 * 3}{5 + 3} = \frac{6}{8} = 0.75 \\ \text{similarity}(gp, p_3) &= \frac{2 * 3}{5 + 3} = \frac{6}{8} = 0.75 \\ \text{similarity}(gp, p_4) &= \frac{2 * 2}{5 + 3} = \frac{4}{8} = 0.50 \\ \text{similarity}(gp, p_5) &= \frac{2 * 1}{5 + 3} = \frac{2}{8} = 0.25 \end{aligned}$$

Fig. 4. Sample computations using the dice coefficient.

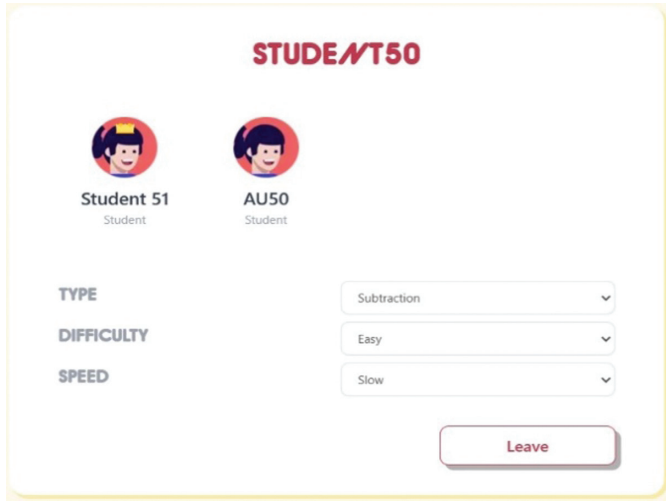


Fig. 5. The software provides game settings recommendations.

4. TESTING, SOFTWARE USAGE, AND EVALUATION

Two universities approved the ethics clearance for this study. After securing the ethics clearance, testing, software usage, and evaluation started. An alpha test was conducted [44]. The software was tested in two phases. The purpose of the first phase of testing is to determine if there are bugs in the software. The first part involved 80 software testers who underwent orientation. They utilized a testing protocol. The protocol involved testing which part of the module should be scrutinized. The testers need to report and describe the bug in the form, as well as their observations. Some of the observations reported included unsuccessful joining a team, not being able to see the group's names, and slow Internet connection. These observations were not necessarily bugs since they are Internet connectivity issues. Nonetheless, they are still worthy of note since they could affect the group's performance during the game.

After the first phase and all concerns were addressed, the second phase commenced. The second part of software testing involved checking whether the guessing and recommendation algorithms were performing correctly. The second part of software testing was also guided by a protocol (phase 2). Eight testers who

Table III. Open-ended questions (Q) asked of the participants after using the software

Q#	Question
Q1	What aspects of the game do you like?
Q2	What aspects of the game do you not like?
Q3	Any recommendations to improve the game?
Q4	Do you find the game useful in your studies?
Q5	Kindly explain your answer in the previous section.
Q6	What game settings do you usually choose?
Q7	Why do you choose those settings?
Q8	Are you comfortable playing with your teammates? Why or why not?

were not part of the 80 testers in the first testing evaluated the software. They utilized a testing form that was intended to catch possible bugs in the implementation of the algorithm. After this phase, it was shown that the guessing detection and the recommender system were working well. The forms for both phases can be downloaded here.

After the testing phases, actual participants utilized the software. Invitations were sent to the elementary education departments of five universities, and four agreed to participate. Fifty-five Grade 5 students from the elementary departments of four universities participated in a 5-day experiment. At the time the gaming sessions were conducted, the topic had only been introduced to them about a week earlier. They utilized the software for 15 minutes. The participants had an average age of 10 years ($SD = 0.47$) and were mostly male ($n = 31, 56\%$). A total of 19 teams consisting of 2 or 3 members participated in the study. Twenty trained facilitators assisted the students during the software utilization. The software was utilized during their math classes in their classrooms. All students utilized the software simultaneously. Five to six groups are simultaneously sharing an Internet connection. The interactions of the students (e.g., game setting selections, RT, selected responses, etc.) with the software were logged into the system.

After the experiment, students answered eight open-ended questions ([45]; Table III). The responses were tabulated in a spreadsheet. The responses were analyzed through content analysis [46]. In this method, the frequency of certain words across responses (e.g., how many students mentioned words directly answering the questions, such as fun, learn, easy, etc.) was counted [46]. The researcher encoded all unique terms in a word processor. Afterward, the same researcher interpreted the terms in light of the research questions and responses. Then, the second researcher read the interpretations independently. A meeting was then conducted to discuss the disagreement in the interpretations.

5. RESULTS OF THE CONTENT ANALYSIS

This study designed, developed, tested, and implemented a math mCSSL for fractions. Toward this goal, 55 students participated in a 5-day experiment. After the experiment, students were asked about their perceptions of the software using an open-ended questionnaire. Only 44 students responded to the questionnaire. Content analysis disclosed that there are 334 unique words identified through content analysis (Table IV). The five most frequent words per question and sample responses are also shown in Table IV.

Table IV. Top 5 frequently occurring words per question

Q#	Total unique words	Words (frequency, percentage)	Sample sentences
Q1	69	1. multiplayer (7, 10%) 2. game (6, 9%) 3. fun (5, 7%) 4. happy (4, 6%) 5. learn (3, 4%)	1. I like the multiplayer because I can play with my friends. 2. I like the questions and how the game is designed. 3. It is fun and educational. 4. I'm happy playing with my friends. 5. I learn math.
Q2	45	1. none (9, 20%) 2. timer (6, 13%) 3. fast (5, 11%) 4. questions (4, 9%) 5. hard (3, 7%)	1. None. 2. The timer is so fast. 3. The questions are hard.
Q3	43	1. none (7, 16%) 2. add (4, 9%) 3. fractions (4, 9%) 4. game (3, 7%) 5. play (3, 7%)	1. None. 2. Add a less difficult mode for other students. 3. Easier fraction problems. 4. More practice before the game. 5. Play more before the game.
Q5	71	1. help (6, 8%) 2. useful (5, 7%) 3. practice (4, 6%) 4. game (4, 6%) 5. improves (3, 4%)	1. It will help us study better in math. 2. I find it useful because it helped me. 3. It is a way for me to practice [my math]. 4. The game is useful because it helps me and my classmates with fractions in math. 5. It improves my self-confidence.
Q6	28	1. addition (14, 50%) 2. slow (9, 32%) 3. medium (9, 32%) 4. easy (7, 25%) 5. multiplication (5, 18%)	1. I usually choose addition and slow settings. 2. I usually choose addition and slow settings. 3. I selected a medium setting. 4. I solved easy multiplication problems. 5. I solved easy multiplication problems.
Q7	44	1. easy (12, 27%) 2. answer (5, 11%) 3. comfortable (4, 9%) 4. fun (3, 7%) 5. know (3, 7%)	1. I chose those problems because it is easy. 2. I know the answer. 3. I am comfortable with the problem. 4. It is fun. 5. I know the answer.
Q8	34	1. help (6, 18%) 2. comfortable (5, 15%) 3. fun (5, 15%) 4. enjoy (4, 12%) 5. together (4, 12%)	1. Yes, because we help each other. 2. Yes, because we are comfortable playing with them. 3. It is fun playing with them. 4. I enjoy playing with my classmates. 5. Solving problems together makes it easy.
Total	334		

The analysis of Q1 reveals that the most common word is “multiplayer” (10%). Students were expressing their enjoyment of the multiplayer feature of the game. This is corroborated by the other four respondents, who said that they were happy playing with their classmates. The word “game” is the second most occurring word, reflecting their awareness of the nature of the software. Students feel “happy” while playing the game because they can interact with their friends. The word “learn” was mentioned three times, which suggests that the software was not only fun to use but also helped the students learn math.

When asked about what aspect they do not like about the game, the majority (20%) did not report anything. Issues relating to the timer (13%) and the game being fast (11%) were the second and third most reported issues, respectively. They also found that the questions (9%) are hard (7%). Other issues reported (not shown in the table) included that the game could not load properly due to a poor Internet connection. Regarding the second question, they were also asked what improvements could be made in the game. As expected, the majority reported nothing (16%) to be improved. The other recommendations align with the responses in Q2, such as adding easier modes (9%), providing simpler fraction problems

(9%), and conducting a practice game (7%) before the actual game session (7%).

In Q4, all students responded “yes” (100%) to the question of whether the software is useful for their studies. When asked to explain their response to Q4, the word “help” (8%) appeared the most. They find the game (5%) useful (7%) in their math studies since they could use it for practice (5%). It is worth noting that one respondent attributed the game to improving his/her self-confidence toward math.

Q5 validates the valuable support the application provides to students in studying math more effectively. The students provided a “help” (8%) response, that is, the application is helpful and “useful” (7%) for their learning math. Also, for them, it is a “practice” (6%) and a “game” (6%) at the same time. Lastly, it “improves” (4%) their self-confidence in dealing with Math fraction problems.

The responses in Q6 show that the most appealing words are “addition” (50%), “slow” (32%), “medium” (32%), “easy” (25%), and “multiplication” (18%). These words refer to the most selected game settings. These most selected game settings are easier to solve compared to subtraction and division of fractions. The responses to

Question 7 explained why students chose these game settings. The main reasons were that the problems were easy (27%), they already knew the answer (11%), or they felt comfortable answering (9%). Additionally, students selected easier problems because they found them fun to solve (7%).

In Q8, 43 out of 44 students responded positively to the question, that is, they find it comfortable working with their team members. The words “help” (18%), “comfortable” (15%), “fun” (15%), “enjoy” (12%), and “together” (12%) all indicate that they are at ease working with their classmates. These words are associated with their positive experience throughout the gaming sessions.

6. DISCUSSION

This study developed a math mCSCL. It incorporated a guessing detector and a recommender system. The guessing detector incorporated RM and CF, while the recommender system employed the aggregated CBF for the group. The developed software was then tested and evaluated by different sets of users. A series of tests disclosed that the software was successfully developed and could be deployed.

Actual users evaluated the software. It was shown that social interaction, enjoyment, and learning are the primary reasons why students were engaged in this platform. The most occurring terms in Q1 highlight the students’ engagement in the game. It is worth noting that the word “fun” is mentioned consistently throughout the three questions. The fun component of the game facilitated their positive learning experiences, implying that the software was effective for these participants. The frequent mention of this word and “multiplayer” reflects the core principles of game-based learning [47] and appreciates the collaborative aspect of the game. This suggests that students stay engaged because the game environment offers interactivity and enjoyment. Their responses are also to intrinsic motivation, as they choose to join the tasks for the satisfaction they gain from the software [48]. The word “multiplayer” also aligns with the sociocultural perspective, which sees learning as a result of social interaction [49,50]. This shows that the social features of the game help maintain interest and support meaningful learning through peer collaboration [51]–[54]. While the word “learn” appeared less frequently relative to the other words, its presence in the texts confirms that they recognize the educational value of the game. Therefore, the software effectively delivers both educational value and enjoyment.

The findings from Q2 show that some students are fully satisfied with the game. However, feedback from both Q2 and Q3 highlights areas for improvement—particularly in pacing, question difficulty, and overall game mechanics. Although students can choose the timer speed and difficulty level, many still find the game too fast and the “easy” mode still too challenging for their current mathematical competency. The timer feature seems to increase pressure, which may raise intrinsic cognitive load and contribute to anxiety [55]. This added stress may cause students to perceive the questions as more difficult than they are [55]–[57]. Although the content was validated by their teachers, the students still felt the content was difficult. This is understandable because the topic is relatively new to them. These reactions can also be interpreted as students seeking better management of cognitive load or striving for more experiences (i.e., more practice) that better support their need for competence. This suggests a need to adjust the time limits for each speed setting and the difficulty levels to better match students’ capabilities.

The students view the software as a study aid, as indicated in their responses in Q4 and Q5. Their responses to these questions are consistent with Q1. They perceived that the software reinforces math learning. One student said that the software helps both them and their classmates. This suggests that some students learn math better when they work with others [58,59]. In addition to the cognitive aspects of math learning, social aspects such as self-confidence building were also mentioned. The software is capable of boosting students’ confidence in solving math problems. This finding is consistent with a previous scoping review that math CSCL could improve social learning aspects [3].

The answers in Q6 and Q7 offer interesting insights about the students. They chose tasks that they felt confident and capable of answering. Learning was customized based on the group’s agreed consensus and based on their collective abilities. This confirms that peer interaction helps shape the learning environment. Students felt more comfortable and motivated when supported by teammates, which created a safe space for practicing skills [60]. Moreover, they prioritize success and comfort over challenge because they are still building foundation skills or confidence in math. They choose these game settings as a strategic way to help them gain fluency before moving on to more difficult problems. Students used the settings to regulate task difficulty and reduce pressure. This reflects the kind of self-regulated learning supported by adaptable scaffolding in mobile CSCL environments where learners benefit from tailoring tasks to match their confidence and competence [61]. For example, one student chose addition with a slow timer, which allowed more time to think and feel confident in solving problems. In other words, the high preference for easy-to-solve problem settings is an indication of self-aware learners.

In the last question, the repeated mention of the words “help” and “together” shows that students were engaged in teamwork. It further indicates a strong level of group cohesion. The keywords “comfortable”, “fun”, and “enjoy” imply positive social dynamics throughout the game sessions. Therefore, positive peer collaboration is another social benefit the software could facilitate. This insight shows that peer interaction contributes not only to comfort but also to students’ willingness to engage in the task. When learners feel supported by their teammates, they are more likely to participate and persist [5]. This is again consistent with the findings of several studies on the social benefits of math mCSCL [3].

7. CONCLUSION, RECOMMENDATIONS, AND FUTURE WORKS

Based on the findings of the study, the software was able to detect whether the student was exhibiting a guessing behavior. It was also able to recommend game settings based on the group’s math performance within the software. Hence, the core algorithms of the software function effectively. The successful deployment of the software across users validates the functionalities of the software. Therefore, it can be concluded that the software performed its intended purpose.

Meanwhile, the software successfully combines both learning and enjoyment. As such, students perceived the game as a learning aid at both the individual and group levels. It fostered positive peer interaction, strong group cohesion, and effective teamwork. It also encouraged students to be self-aware of their current capabilities, striking a balance between enjoyment and the desire to improve their mathematics skills. In short, the software supports mathematics learning, builds confidence, and promotes collaboration. Therefore, it can be concluded that the software effectively fosters both academic and social aspects of learning.

As suggested by the students, the timer settings need to be adjusted, and easier problems should be included. In addition to improvements in the software, the game mechanics also need to be enhanced. Students recommend providing more practice time before starting the actual game sessions. Another major challenge is the Internet connection, so it is suggested to limit the number of groups (about 2 to 3) using the game at the same time.

Additionally, the incorporation of a hybrid filtering algorithm may be considered in the future. For instance, if Group A is performing well in addition to using the medium-speed setting, and Group B has a similar group profile, the system could recommend the same settings to Group B. Finally, the plans include a collection of more user feedback and integrate it into the system. This feedback will be analyzed using natural language processing (NLP) techniques to better understand the students' sentiments.

CONFLICT OF INTEREST STATEMENT

The author(s) declared no potential conflicts of interest in the publication of this article.

REFERENCES

- [1] G. Stahl and K. Hakkarainen, "Theories of CSCL BT," in *International Handbook of Computer-Supported Collaborative Learning*, U. Cress, C. Rosé, A. F. Wise, and J. Oshima, Eds. Cham: Springer International Publishing, 2021, pp. 23–43. DOI: [10.1007/978-3-030-65291-3_2](https://doi.org/10.1007/978-3-030-65291-3_2).
- [2] C.-K. Looi *et al.*, "Leveraging mobile technology for sustainable seamless learning: a research agenda," *Br. J. Educ. Technol.*, vol. 41, no. 2, pp. 154–169, Mar. 2010, DOI: <https://doi.org/10.1111/j.1467-8535.2008.00912.x>.
- [3] R. P. Bringula and F. A. L. Atienza, "Mobile computer-supported collaborative learning for mathematics: A scoping review," *Educ. Inf. Technol.*, vol. 28, pp. 4893–4918, 2022, DOI: <https://doi.org/10.1007/s10639-022-11395-9>.
- [4] S. Amara *et al.*, "Group formation in mobile computer supported collaborative learning contexts: A systematic literature review," *J. Educ. Technol. Soc.*, vol. 19, no. 2, pp. 258–273, 2016. [Online]. Available: https://www.researchgate.net/publication/284550462_Group_Formation_in_Mobile_Computer_Supported_Collaborative_Learning_Contexts_A_systematic_literature_review
- [5] Y.-T. Sung, J.-M. Yang, and H.-Y. Lee, "The effects of mobile-computer-supported collaborative learning: Meta-analysis and critical synthesis," *Rev. Educ. Res.*, vol. 87, no. 4, pp. 768–805, Apr. 2017, DOI: <https://doi.org/10.3102/0034654317704307>.
- [6] A. Peramunugamage, U. W. Ratnayake, and S. P. Karunanayaka, "Systematic review on mobile collaborative learning for engineering education," *J. Comput. Educ.*, vol. 10, no. 1, pp. 83–106, 2023, DOI: <https://doi.org/10.1007/s40692-022-00223-1>.
- [7] D. Mullins, N. Rummel, and H. Spada, "Are two heads always better than one? Differential effects of collaboration on students' computer-supported learning in mathematics," *Int. J. Comput. Support. Collab. Learn.*, vol. 6, no. 3, pp. 421–443, 2011, DOI: <https://doi.org/10.1007/s11412-011-9122-z>.
- [8] H. Sollervall and M. Milrad, "Theoretical and methodological considerations regarding the design of innovative mathematical learning activities with mobile technologies," *Int. J. Mob. Learn. Organ.*, vol. 6, no. 2, pp. 172–187, 2012, [Online]. Available: <https://dx.doi.org/10.1504/IJMLO.2012.047595>
- [9] I. Boticki, L.-H. Wong, and C.-K. Looi, "Experiences in implementing and using a technological framework for mobile collaborative learning of mathematics and Chinese," *Int. J. Mob. Learn. Organ.*, vol. 6, no. 1, pp. 79–93, May 2012, DOI: <https://doi.org/10.1504/IJMLO.2012.046884>.
- [10] H. Järvenoja, S. Järvelä, and J. Malmberg, "Supporting groups' emotion and motivation regulation during collaborative learning," *Learn. Instr.*, vol. 70, p. 101090, 2020, DOI: <https://doi.org/10.1016/j.learninstruc.2017.11.004>.
- [11] G. Zurita and M. Nussbaum, "Computer supported collaborative learning using wirelessly interconnected handheld computers," *Comput. Educ.*, vol. 42, no. 3, pp. 289–314, 2004, DOI: <https://doi.org/10.1016/j.compedu.2003.08.005>.
- [12] R. Bringula and M. M. Rodrigo, "Can 'stag-and-hare hunt' behavior be modeled using interaction data from a mobile-supported collaborative learning application?," in *International Conference on Computers in Education*, M. Chang *et al.*, Ed., Kenting, Taiwan: Asia-Pacific Society for Computers in Education (APSCE), pp. 770–773, Dec. 2019, DOI: <https://doi.org/10.58459/icce.2019.831>.
- [13] G. Osman *et al.*, "Learning through collaboration: student perspectives," *Asia Pac. Educ. Rev.*, vol. 12, no. 4, pp. 547–558, 2011, DOI: <https://doi.org/10.1007/s12564-011-9156-y>.
- [14] C. Phielix, F. J. Prins, and P. A. Kirschner, "Awareness of group performance in a CSCL-environment: Effects of peer feedback and reflection," *Comput. Hum. Behav.*, vol. 26, no. 2, pp. 151–161, 2010, DOI: <https://doi.org/10.1016/j.chb.2009.10.011>.
- [15] W. Stroebe, M. Diehl, and G. Abakoumkin, "The illusion of group effectivity," *Pers. Soc. Psychol. Bull.*, vol. 18, no. 5, pp. 643–650, 1992, DOI: <https://doi.org/10.1177/0146167292185015>.
- [16] X. Ying *et al.*, "Group laziness: The effect of social loafing on group performance," *Soc. Behav. Pers. An Int. J.*, vol. 42, no. 3, pp. 465–471, 2014, DOI: <https://doi.org/10.2224/sbp.2014.42.3.465>.
- [17] R. Baker, A. Mitrović, and M. Mathews, "Detecting gaming the system in constraint-based tutors," in *User Modeling, Adaptation, and Personalization*, P. De Bra, A. Kobsa, and D. Chin, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 267–278.
- [18] J. A. Walonoski and N. T. Heffernan, "Detection and analysis of off-task gaming behavior in intelligent tutoring systems," in *Intelligent Tutoring Systems*, M. Ikeda, K. D. Ashley, and T.-W. Chan, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 382–391.
- [19] R. S. J. D. Baker *et al.*, "Adapting to when students game an intelligent tutoring system," in *Intelligent Tutoring Systems*, M. Ikeda, K. D. Ashley, and T.-W. Chan, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 392–401.
- [20] I. Arroyo *et al.*, "Repairing disengagement with non-invasive interventions," in *Proceedings of the 2007 Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, R. Luckin, K. R. Koedinger, and J. Greer, Eds., NLD: IOS Press, 2007, pp. 195–202.
- [21] T. M. Nunes *et al.*, "Discouraging gaming the system through interventions of an animated pedagogical agent," in *Adaptive and Adaptable Learning*, K. Verbert, M. Sharples, and T. Klobučar, Eds., Cham: Springer International Publishing, 2016, pp. 139–151.
- [22] I. Verginis *et al.*, "Guiding learners into reengagement through the SCALE environment: An empirical study," *IEEE Trans. Learn. Technol.*, vol. 4, no. 3, pp. 275–290, 2011, DOI: <https://doi.org/10.1109/TLT.2011.20>.
- [23] J. E. Beck, "Engagement tracing: Using response times to model student disengagement," in *Proceedings of the 2005 Conference on Artificial Intelligence in Education: Supporting Learning through*

- Intelligent and Socially Informed Technology*, C.-K. Looi, G. McCalla, B. Bredeweg, and J. Breuker, Eds., NLD: IOS Press, 2005, pp. 88–95.
- [24] E. E. Kraemer *et al.*, “A comparison of the mystery motivator and the get ‘em on task interventions for off-task behaviors,” *Psychol. Sch.*, vol. 49, no. 2, pp. 163–175, Feb. 2012, DOI. <https://doi.org/10.1002/pits.20627>.
- [25] A. Felfernig *et al.*, *Group Recommender Systems: An Introduction*. Cham, Switzerland: Springer, 2018. [Online]. Available: <https://link.springer.com/book/10.1007/978-3-319-75067-5>
- [26] J. Johns and B. Woolf, “A dynamic mixture model to detect student motivation and proficiency,” in *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1*, in AAAI’06. AAAI Press, 2006, pp. 163–168.
- [27] A. Edwards and L. Alcock, “Using Rasch analysis to identify uncharacteristic responses to undergraduate assessments,” *Teach. Math. Its Appl. Int. J. IMA*, vol. 29, no. 4, pp. 165–175, 2010, DOI. <https://doi.org/10.1093/teamat/hrq008>.
- [28] A. D. Miller and W. L. Heward, “Do your students really know their math facts? Using daily time trials to build fluency,” *Interv. Sch. Clin.*, vol. 28, no. 2, pp. 98–104, Nov. 1992, DOI. <https://doi.org/10.1177/105345129202800205>.
- [29] National Council of Teachers of Mathematics, *Principles and standards for school mathematics*. 2000.
- [30] R. P. Bringula, “Ibigkas! math 2.0: An intelligent computer-supported collaborative learning application for grade 5 mathematics BT - adaptive instructional systems,” in *International Conference on Human-Computer Interaction*, R. A. Sottilare and J. Schwarz, Eds., Cham: Springer Nature Switzerland, 2023, pp. 287–297.
- [31] R. S. J. D. Baker *et al.*, “Modeling and studying gaming the system with educational data mining,” in *International Handbook of Metacognition and Learning Technologies*, R. Azevedo and V. Aleven, Eds., New York, NY: Springer New York, 2013, pp. 97–115, DOI. [10.1007/978-1-4419-5546-3_7](https://doi.org/10.1007/978-1-4419-5546-3_7).
- [32] C. Peters *et al.*, “Predictors and outcomes of gaming in an intelligent tutoring system,” in *Intelligent Tutoring Systems*, R. Nkambou, R. Azevedo, and J. Vassileva, Eds., Cham: Springer International Publishing, 2018, pp. 366–372.
- [33] R. S. Baker, A. T. Corbett, and K. R. Koedinger, “Detecting student misuse of intelligent tutoring systems,” in *Intelligent Tutoring Systems*, J. C. Lester, R. M. Vicari, and F. Paraguaçu, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 531–540.
- [34] A. Al Fraidan, “Interplay of language proficiency, gender, test anxiety, and cognitive strategies: The spectrum of guessing behaviors in multiple-choice assessments among Saudi EFL learners,” *Edelweiss Appl. Sci. Technol.*, vol. 8, no. 6, pp. 1148–1161, Oct. 2024, DOI. <https://doi.org/10.55214/25768484.v8i6.2217>.
- [35] B. Culligan, “Item Response Theory, Reliability and Standard Error,” 2011, [Online]. Available: https://www.wordengine.jp/research/pdf/IRT_reliability_and_standard_error.pdf
- [36] M. E. Umobong, “The one-parameter logistic model (1PLM) and its application in test development,” *Adv. Soc. Sci. Res. J.*, vol. 4, no. 24, pp. 126–137, 2017, DOI. <https://doi.org/10.14738/assrj.424.3938>.
- [37] J. Kean and J. Reilly, “Item response theory,” in *Handbook for Clinical Research: Design, Statistics, and Implementation*, F. M. Hammond, J. F. Malec, T. G. Nick, and R. M. Buschbacher, Eds., New York, USA: Gasch Printing, 2015, pp. 195–197.
- [38] R. K. Hambleton, H. Swaminathan, and H. J. Rogers, “Fundamentals of item response theory,” *Fundamentals of item response theory*. In Measurement methods for the social sciences series, Vol. 2. Sage Publications, Inc., Thousand Oaks, CA, US, pp. x, 174–x, 174, 1991.
- [39] B. Sumintono, “Rasch Model Measurements as Tools in Assessment for Learning,” in *Proceedings of the 1st International Conference on Education Innovation*, 2017, pp. 38–42. DOI. <https://doi.org/10.2991/icei-17.2018.11>.
- [40] M. Carr and N. Alexeev, “Fluency, accuracy, and gender predict developmental trajectories of arithmetic strategies,” *J. Educ. Psychol.*, vol. 103, no. 3, pp. 617–631, 2011. DOI. <https://doi.org/10.1037/a0023864>.
- [41] M. Vasilyeva, E. V. Laski, and C. Shen, “Computational fluency and strategy choice predict individual and cross-national differences in complex arithmetic,” *Dev. Psychol.*, vol. 51, no. 10, pp. 1489–1500, 2015, DOI. <https://doi.org/10.1037/dev0000045>.
- [42] S. Araneda *et al.*, “Exploring relationships among test takers’ behaviors and performance using response process data,” *Educ. Sci.*, vol. 12, no. 2, Article 104, 2022, DOI. <https://doi.org/10.3390/educsci12020104>.
- [43] R. P. Bringula *et al.*, “Modeling ‘stag and hare hunting’ behaviors using interaction data from an mCSSL application for grade 5 mathematics,” *Multimodal Technol. Interact.*, vol. 7, no. 4, 2023, DOI. <https://doi.org/10.3390/mti7040034>.
- [44] S. Hai-Jew, *Designing Instruction For Open Sharing*. Cham, Switzerland: Springer, 2019. [Online]. Available: <https://link.springer.com/book/10.1007/978-3-030-02713-1>
- [45] M. M. T. Rodrigo *et al.*, “The iterative development of a mobile collaborative game for building phonemic awareness and vocabulary,” *Comput. Learn. Context*, vol. 1, no. 1, pp. 28–42, 2019, [Online]. Available: <https://www.upenn.edu/learninganalytics/CBLC/issue-1-1/CBLC-2019-1-1-3.pdf>
- [46] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, California: SAGE Publications, Inc. 2019, DOI. <https://doi.org/10.4135/9781071878781>.
- [47] A. Knox, “Game-based learning design optimized for cognitive load,” in *Reimagining Education: Studies and Stories for Effective Learning in an Evolving Digital Environment*, D. Cockerham, R. Kaplan-Rakowski, W. Foshay, and M. J. Spector, Eds., Cham: Springer International Publishing, 2023, pp. 239–250, DOI. https://doi.org/10.1007/978-3-031-25102-3_20.
- [48] R. M. Ryan and E. L. Deci, “Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions,” *Contemp. Educ. Psychol.*, vol. 61, p. 101860, 2020, DOI. <https://doi.org/10.1016/j.cedpsych.2020.101860>.
- [49] D. L. Ott, “Social learning theory,” in *Elgar Encyclopedia of Cross-Cultural Management*, A. I. Mockaitis and C. L. Butler, Eds., Cheltenham, UK: Edward Elgar Publishing, 2024, pp. 133–134. DOI. <https://doi.org/10.4337/9781803928180.ch33>.
- [50] P. Shih, G. M. Velan, and B. Shulruf, “Shared values and socio-cultural norms: E-learning technologies from a social practice perspective,” *Issues Educ. Res.*, vol. 27, no. 3, pp. 550–566, 2017, [Online]. Available: <https://www.iier.org.au/iier27/shih.pdf>
- [51] M. Guinebert *et al.*, “Analysis of peer interactions features in Multi-Player Learning Games: semi-automatic approach and proof of concept,” in *Proceedings of the European Conference on Games Based Learning*, 2018, p. European Conference on Games Based Learning.
- [52] S. Burns *et al.*, “Improving young children’s peer collaboration in early educational settings: A systematic review,” *Rev. Educ.*, vol. 12, no. 2, p. e3484, Aug. 2024, DOI. <https://doi.org/10.1002/rev3.3484>.
- [53] M. A. Qureshi *et al.*, “Factors affecting students’ learning performance through collaborative learning and engagement,” *Interact. Learn. Environ.*, vol. 31, no. 4, pp. 2371–2391, May 2023, DOI. <https://doi.org/10.1080/10494820.2021.1884886>.

- [54] V. J. Kamalodeen *et al.*, “Designing gamification for geometry in elementary schools: Insights from the designers,” *Smart Learn. Environ.*, vol. 8, no. 1, p. 36, 2021, DOI. <https://doi.org/10.1186/s40561-021-00181-8>.
- [55] X. Shi *et al.*, “Cognitive processing features of elementary school children with mathematical anxiety: Attentional control theory-based explanation,” *J. Exp. Child Psychol.*, vol. 224, p. 105513, 2022, DOI. <https://doi.org/10.1016/j.jecp.2022.105513>.
- [56] S. M. M. de Mooij *et al.*, “Should online math learning environments be tailored to individuals’ cognitive profiles?,” *J. Exp. Child Psychol.*, vol. 191, p. 104730, 2020, DOI. <https://doi.org/10.1016/j.jecp.2019.104730>.
- [57] R. G. Pizzie, N. Raman, and D. J. M. Kraemer, “Math anxiety and executive function: Neural influences of task switching on arithmetic processing,” *Cogn. Affect. Behav. Neurosci.*, vol. 20, no. 2, pp. 309–325, 2020, DOI. <https://doi.org/10.3758/s13415-020-00770-z>.
- [58] F. S. Salim and N. A. Shukor, “The role of computer-supported collaborative learning (CSCL) in improving students’ learning performance in chemistry,” *Man India*, vol. 97, no. 12, pp. 203–211, 2017.
- [59] H.-S. Siller and S. Ahmad, “Analyzing the impact of collaborative learning approach on grade six students’ mathematics achievement and attitude towards mathematics,” *Eurasia J. Math. Sci. Technol. Educ.*, vol. 20, no. 2, p. em2395, 2024, DOI. <https://doi.org/10.29333/ejmste/14153>.
- [60] W. Niu *et al.*, “Impact of perceived supportive learning environment on mathematical achievement: The mediating roles of autonomous self-regulation and creative thinking,” *Front. Psychol.*, vol. 12, 2022, [Online]. DOI. <https://doi.org/10.3389/fpsyg.2021.781594>
- [61] F. Vogel *et al.*, “Adaptable scaffolding of mathematical argumentation skills: The role of self-regulation when scaffolded with CSCL scripts and heuristic worked examples,” *Int. J. Comput. Collab. Learn.*, vol. 17, no. 1, pp. 39–64, 2022, DOI. <https://doi.org/10.1007/s11412-022-09363-z>.