Examining Competitive, Collaborative, and Adaptive Gamification in Young Learners’ Math Learning
Abstract—This paper presents the results of an empirical study conducted on three different types of gamified learning activities—namely competitive, collaborative, and adaptive—in lower primary mathematics classes. The participants were students from two second-grade and one third-grade classes who used tablet computers and digital learning lessons for learning mathematics. The study included a non-gamified and competitive, adaptive, and collaborative gamified conditions, which were integrated into lesson plans. The collected log data were used to calculate the changes in performance levels through the dimensions of task completion and time under each condition, and the data were further analyzed and compared across conditions. The quantitative analysis results were triangulated with interview data from the students. Overall, the results show that gamified activities contributed to increased student performance levels in math learning. Significantly higher performance levels appeared in a gamified condition combining competition, a narrative, and adaptivity with individual performance game elements. Although the highest performance levels appeared in conjunction with the most incorrect attempts by the students, the total number of correct attempts was unaffected. Our findings suggest that whether gamification works or not is not the result of individual game elements but rather the consequence of their balanced combination.

Keywords—Gamification; learning technologies; mobile learning; student performance; primary school

1 INTRODUCTION

Recently, gamification, which refers to the use of game mechanics and game elements in a non-game context, has attracted a significant amount of attention and has been applied in a wide range of fields in order to motivate to and engage people in performing certain activities and solving different problems (Kapp, 2012). Applications of gamification can be found in diverse situations and contexts, such as shopping, marketing, social networking, leisure, fitness, recycling, and learning (Tome Klock et al., 2015; Werbach & Hunter, 2012).

As an educational tool, gamification is used to facilitate learning, to encourage motivation and engagement, to improve learner participation and lesson interactivity, and to stimulate learners to expand their knowledge (Kapp, Blair, & Mesch, 2014). When implemented properly, gamification can increase intrinsic motivation and engagement (Villagrana, Fonseca, Redondo, & Duran, 2014) and represents a powerful tool for teachers at all levels in the educational system (Buckley & Doyle, 2014).

Despite much interest and potential, however, the efficacy of gamification in education remains insufficiently explored (Hanus & Fox, 2015). Among the existing research studies, some have explored the use of gamification in context of the classic "pen & paper" learning process (Kuo & Chuang, 2016), while surveys involving digital gamification are often conducted involving university-level participants (Buckley & Doyle, 2014; de-Marcos, Garcia-Lopez, & Garcia-Cabot, 2016; Garcia, Copiaco, Nufable, Amoranto, & Azcarraga, 2015; Villagrana et al., 2014) or as an extension of e-learning platforms (Domínguez et al., 2013; Muntean, 2011).

Some researchers argue that gamification alone is not sufficient to bring about the desired learning outcomes and that the efficacy of various game elements and their implementation need further exploration and empirical evidence (Hanus & Fox, 2015; Sailer, Hense, Mayr, & Mandl, 2017). In other words, they believe that additional empirical work should be conducted to identify efficient ways to implement different game elements that can increase learners’ engagement, motivation, and
performance. Furthermore, in gamification, it is agreed that one size does not fit all. While one game element (e.g., leaderboard) may work for some students, it could result in the opposite effect for other groups of students, even in the same class (Werbach & Hunter, 2012).

Recognizing the aforementioned challenges for incorporating gamification within educational contexts, this study explores the use of different types of gamification for young learners regarding their math learning processes and performance. In particular, we aim to contribute to the existing body of research concerning the impact of different gamification elements on learning. We concur with Sailer et al. (2017) that the notion of gamification is not a generic construct and that different design elements and game mechanisms can result in varying effects. To unpack the efficacy of different gamification types in learning processes and performance, we conducted a quasi-experimental study with three different configurations of gamification, namely competitive, collaborative, and adaptive gamification.

Using a custom mobile learning platform and digital lesson authoring tools (Authors, 2015), a series of experiments involving second- and third-grade primary school classes were conducted to compare the efficacy of gamification techniques on learning processes and outcomes in mathematics. In total, four sessions (referred to in the text as four experimental conditions) were conducted, each lasting for 15 minutes. Each session used a different gamification approach (i.e., competitive, adaptive, and collaborative) and implemented its own set of gamification elements.

The main research questions explored in this study were: (a) how do different types of gamification affect primary school students’ performance levels and (b) how do game elements used in different conditions contribute to student performance in gamified activities? This study will also discuss the implications of using different gamification types as well as their drawbacks.

2 THEORETICAL BACKGROUND

Games are meant to be fun and entertaining and create desirable experiences even in less-interesting activities, thus increasing intrinsic motivation and making activities more enjoyable and engaging (Deterding, Khaled, Nacke, & Dixon, 2011; Flattla, Gutwin, Nacke, Bateman, & Mandryk, 2011; Kuo & Chuang, 2016; Monerrat, Élise Lavoué, & Sébastien George, 2014). Well-designed games can be used as learning tools that support deep and meaningful learning (Shute & Ke, 2012). Today, there is a prevailing consensus in the literature that so-called serious games (games whose primary objective is not mere entertainment) have a positive effect on learning process and outcomes by increasing engagement, interest, and immersion (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Hamari et al., 2016; Kapp, 2012; Kuo & Chuang, 2016; Poondej & Lerdpornkulrat, 2016).

The importance of games and play in human learning and personal cognitive development is rooted in Piaget's and Vygotsky's theories and has remained a subject of research (Nicolopoulos, 1993), resulting in the affirmation of the interdisciplinary field of game-based learning (GBL). GBL is specialized in exploring the connection and interaction between game and play and is described as a
type of game with defined learning outcomes (Shaffer, Squire, Halverson, & Gee, 2005).

The term gamification was first coined in 2003 by Nick Pelling, a British game developer (Werbach & Hunter, 2012), and subsequently has gained broader recognition since 2010 (Deterding et al., 2011). The most commonly used definition states that gamification is “the use of game design elements in non-game contexts” (Deterding et al., 2011, p. 2). A more detailed and outcome-related definition states that “gamification is using game-based mechanics, aesthetics, and game-thinking to engage people, motivate action, promote learning, and solve problems” (Kapp, 2012, p. 83). Conceptually, gamification significantly differs from serious games and game-based learning in that, while serious games immerse learners into gameplay and strive to hide real educational objectives, lessons that are gamified have one or more game elements added (e.g., points, leaderboard, or badges), with the educational objectives clearly visible to users (Naik & Kamat, 2016; Plass, Homer, & Kinzer, 2015).

In recent years, a number of noteworthy gamification implementations and experiments were conducted, mostly targeting online and e-learning platforms and/or higher education courses. For example, Garcia et al. (2015) investigated the efficacy of a gamified educational environment for a programming course. They included typical gamification elements like points, leaderboards, and badges (the PLB triad) and found that students preferred using the gamified system to its non-gamified counterpart while simultaneously improving their performance in the programming tests. In another study, de-Marcos et al. (2016) compared educational games and social networking approaches to gamification and social gamification. Interestingly, their results stressed social gamification as the approach that yielded the strongest impact on students’ learning performances. Similarly, an experimental design study that included 97 participants by Yildirim (2017) reported a positive impact of gamification on students’ achievements and attitudes toward learning. Huang and Hew (2015) also concluded that gamification can be an effective tool for improving student participation, as well as encouraging extracurricular learning.

On the other hand, some researchers have reported mixed or negative results concerning the efficacy of gamification. For example, Frost et al. (2015) implemented a PLB triad with a storyline and the concept of lives into their learning management system, later measuring the levels of the desired outcomes. They concluded that, although students appreciated some gamification aspects, gamification alone did not have significant effects on the students’ interest, motivation, or satisfaction levels. They interpreted this finding with the consideration of the nature of voluntary gameplay, in that gameplay is a voluntary activity, while the course in which they implemented gamification elements was obligatory; thus, by definition, there was no voluntary agreement in solving the given tasks. Similarly, Hanus and Fox (2015) reported decreased motivation and satisfaction over time in students taking a gamified university course when compared with non-gamified courses. They interpreted this as resulting from a lack of personalization, which reduced the intrinsic motivation in some of the already motivated students. They further concluded that future research should investigate and compare different
gamification elements and approaches, as opposed to simply measuring the overall effectiveness of gamification.

Flow is a pertinent concept used to explain engaging and immersive experiences in games and gamification. Csikszentmihalyi (2003) defines flow as “a state of mind characterized by focused concentration and elevated enjoyment during intrinsically interesting activities.” In principle, flow theory is grounded in the challenge–skill dynamic, in which flow experiences are enhanced when an individual uses their skills to meet a challenging task/goal. The level of challenge is key to increasing or decreasing one’s motivation and engagement, that is, when the level of challenge is too far below one’s skill, this situation results in a state of apathy and disengagement. On the contrary, when the level of challenge is too high, this results in a state of anxiety and stress. Finding an optimal level or path in challenge–skill dynamics, therefore, has been suggested as an important research topic in the field of gamification (Hamari et al., 2016).

Despite the increasing popularity of gamification in educational contexts, there is a dearth of empirical evidence investigating under what conditions and how gamification works or fails. A review of the existing research shows that the impact of gamification is often evaluated through self-reported measures concerning an individual’s perceptions and attitudes through survey instruments. Therefore, more research that empirically investigates learners’ behavioral data generated during gameplay or activities and which uses such user-generated data to make valid inferences should be conducted.

Another gap in the existing literature lies in treating gamification as a generic construct with overly behavioristic mechanisms such as points and badges as a reward for motivation. Nicholson (2015) cautions the danger inherent in such an excessive use of reward-based gamification elements, namely badges, levels, leaderboards, achievements, and points (BLAP). Work has been done to measure the different aspects of student performance in class via techniques such as interviews, focus groups, surveys/questionnaires, or pre- and posttest quasi-experimental approaches. Recently a number of studies took an approach of analyzing automatically collected log data that emerged during educational systems usage by a variety of stakeholders, with students being the most prominent example (Bhaskar & Govindarajulu, 2008; Sheard, Ceddia, Hurst, & Tuovinen, 2003; Stockwell, 2008). Data logs (e.g., web server access logs or activity logs) have a number of advantages in comparison with other data collection methods due to their less intrusive execution procedure (Federico, 1999) reliability and more objective results as compared with self-reporting instruments (Ingram, 2000; Shute & Ke, 2012). Nevertheless, to get a clearer overall picture on learners’ behavior, Ingram (2000) and Shute and Ke (2012) noted that one should not rely only on one metric and instead advocated supplementing log data with information from additional sources (e.g., individual or focus group interviews).

3 METHODOLOGY

Based on the state-of-the-art in gamification, a quasi-experimental study that investigated the efficacy
of different gamification conditions on students’ learning processes and performance was conducted. In particular, an effort was made to move beyond the simple comparison between gamified and non-gamified conditions, with the belief that each gamification element has its own affordances. Hence, this study explores different combinations of various game elements such as narratives, adaptive algorithms, and simple collaboration, which are less explored areas thus far, by combining an in-app usage log and student opinions collected through focus group interviews as data sources. Performance is modeled on a task completion-time scale to examine changes in students’ actions during gamified educational activities.

3.1 Research Context and Participants

Participants (N = 54) in this study were second- and third-grade students (seven to eight years of age) in a Croatian primary school. There were 27 female and 27 male participants, with varying academic abilities. The students were from two second-grade classes (2A and 2B) and one third-grade class (3R). Each class had a dedicated teacher who actively helped in the process of conducting the experimental study. At the time of the experiment, the students had already had some prior exposure (i.e., two years) to using tablet computers and various digital learning lessons in the classroom.

3.2 Experimental Conditions

To address the identified gaps in the existing literature, we designed a quasi-experiment with four different conditions of gamification labeled as: non-gamified, competitive, adaptive, and collaborative. Table 1 gives an overview of each condition’s game elements, where the columns present all of the experimental conditions in the study and the rows present game elements used across the conditions. Each condition includes a set of well-known game elements and gamification strategies aligned with the curriculum and lessons conducted at the time of the experiment. The detailed set-up of each gamification condition, together with the included game elements, is described in the following section of this paper.

<table>
<thead>
<tr>
<th></th>
<th>Non-gamified</th>
<th>Competitive</th>
<th>Adaptive</th>
<th>Collaborative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-time feedback</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Reporting of individual score</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Reporting of others’ scores</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Reporting of group scores</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Adaptive to individual performance</td>
<td>-</td>
<td>-</td>
<td>+ (time)</td>
<td>-</td>
</tr>
<tr>
<td>Adaptive to group performance</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+ (virus score)</td>
</tr>
<tr>
<td>Competition</td>
<td>-</td>
<td>+ (in class)</td>
<td>+ (personal)</td>
<td>+ (group)</td>
</tr>
<tr>
<td>Collaboration</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Narratives</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Leaderboard</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+ (group)</td>
</tr>
<tr>
<td>Time limit awareness</td>
<td>-</td>
<td>+</td>
<td>+ (per condition &amp; per problem)</td>
<td>-</td>
</tr>
</tbody>
</table>
For this study, the students included participated in four different experimental conditions, which were implemented as separate sessions for each class. All of the conditions were implemented with the same group of students with comparable learning topics and lesson plans, allowing for comparisons and the identification of the effects of various game elements on student performance. Each condition lasted for 15 minutes, and the topics covered in them included mathematics lessons for both second and third grade individuals in alignment with their classes’ respective curriculum and lesson plans, such as addition and subtraction of the numbers 0 to 100; multiplication and division by the numbers 2, 3, 4, and 5; and multiplication by 0. The conditions were chronologically ordered as depicted in Figure 1. The experiment started with the non-gamified condition on the November 17, 2016 and ended with the collaborative condition on the April 12, 2017 (note that one class, 2A, had their first two experimental conditions reversed in order to perform the carryover effect analysis).

3.3 **Digital Learning Platform**

The experiments were conducted using a custom mobile learning platform that was designed to support the creation of simple yet highly dynamic media-rich and interactive digital lessons. Further enhancements of the lesson interactivity and flexibility were accomplished through custom modules (so-called widgets) that are easily customized embedded into any lesson. As shown in Figure 2, one of the technical setups in the experiments used a simple mathematics widget called “the Math Widget” for solving four main arithmetic operations.
Figure 2. The Math Widget.

The flow is as follows: the widget randomly picks a problem from the predefined problem set, the student attempts a solution, and the widget checks it. In the case of an incorrect solution, the student is prompted to solve the same problem again. If the solution is correct, the widget selects a new random problem to be solved.

For the purpose of the experiment, a time limit of 15 minutes was used for all three classes. The problem set was aligned with the math curricula goals, the grade and the lesson plan of the class in which the experiment was conducted.

3.3.1 NON-GAMIFIED DIGITAL LESSONS

The first experiment condition included no gamification. Students were presented with the digital lesson via the Math Widget and were asked to solve arithmetic tasks for 15 minutes. The time limit was not visible on the screen, and students were unaware of this limit (Figure 3).
3.3.2 Competitive Condition

In the second experiment condition, which was conducted two weeks after the first one, a similar exercise (the existing problem set was changed only slightly to stay in line with the lesson plan) was presented as a small in-class mathematics competition. For that purpose, a live leaderboard application was developed and shown on a class projector. For each correctly solved problem, students received 3 points, while, for each wrong solution, 1 point was deducted from the total score. To avoid the embarrassment of underachievers, the leaderboard showed only the six top-ranked students at any given time. The experiment lasted for 15 minutes and, this time, students were aware of the time countdown, which was shown in the upper right corner of the widget (Figure 4).

![Figure 4. A student solving a problem under the competitive condition, which included a leaderboard with the top six students](image)

3.3.3 Adaptive Condition (Competing against personal limits in a narrative-based scenario)

This condition was different from the other conditions in that it incorporated two game elements: narratives and a personalized adaptive algorithm. First, the condition introduced the following story about a computer virus to better contextualize the gamified activity:

“A big bad computer virus attacked our servers and spread small viruses to all tablet computers in the classroom. The only way to defeat them is if every student competes against the virus on his/her own tablet and individually wins. In order to receive the full 3 points, you must solve mathematics problems faster than the virus; otherwise, the virus receives 3 points. An incorrect answer is −1 point for the player.”

The second game element was a personalized adaptive algorithm, which adaptively calculated the amount of time given to each student for every problem to be solved. If the student was outperforming the virus, the algorithm would reduce the available time by one second. Conversely, if the virus was catching up or even winning, more time was given to the student. In other words, the algorithm was designed to keep the students at the edge of their limits. This use of this element was consistent with the growth principle in flow theory, which states that, in order to reach a flow, an
individual should increasingly make progress with higher-level challenges and push their skills to the limit in pursuit of a challenging goal (Hamari et al., 2016). Towards the end of the game, during the last three minutes, the algorithm gave more time to all of the students and allowed them to beat the virus by the end of the game (Figure 5).

![Figure 5. Left: students during the adaptive condition. The progress bar displays the remaining time for each problem, the number of points won by the student, and the virtual opponent, as depicted. Right: a point in the activity where the participating students started showing increased levels of stress.](image)

### 3.3.4 Collaborative condition

Similar to the adaptive condition, the last experiment condition continued with the narrative; however, this time, collaboration instead of individual competition was encouraged among the students:

> “In the previous lesson, we fought and won a battle against small viruses on our tablets, but there is still the big mother-virus in our computer servers. Since it is too strong for each of us to beat individually, the whole class must work together to beat it. There is no time limit for each separate problem to be solved, but we still as a class receive +3 points for a correct answer or −1 point for a wrong solution.”

A simplified leaderboard showing only the sum of the points earned by the entire class and the points (supposedly) earned by the virus was employed (Figure 6). In addition to the narrative and the leaderboard, this condition included collaborative and adaptive components. The students still had to solve problems and collect points, but, this time, all of their individual results were summed up and presented as a score for the entire class. The objective was to collect more points than the virus, whose points were calculated as \(90 \pm \text{random (5)}\)% of the students’ points. That way, the virus was always a little behind the class, and the lesson outcome became somewhat uncertain until the very end.
3.4 Data Collection and Analysis

A verbose data log from each tablet computer was recorded in a database. This database stored information about each problem that was given to the student; the proposed solution; the start, end, and elapsed times; and, for the adaptive algorithm, the amount of time given to a correct or wrong answer, which was immediately sent through the network and stored in the event log on the remote database server. These data were then used to generate live leaderboard scores and were later used in the analysis of student performance and test results, as well as a means of feedback for the teachers on a particular problem (i.e., as a means to identify which tasks students had the most problems with). The log entries were generated after each student’s attempt to solve a problem.

The data collection in this research leveraged two main data sources: (a) learner-generated usage logs and (b) focus group interviews with the students. First, the collected log data were organized according to students, classes, and the four gamification conditions to facilitate further analysis. The analysis consisted of devising a model for quantitatively describing the rate of performance changes and subsequently applying analysis of variance (ANOVA) methods when comparing the performance rate changes amongst the four conditions and the task completion error rate, respectively. Second, focus group interviews including 15 students (10 girls and five boys) selected by their teachers from all three classes were conducted. There were three focus groups, with a typical focus group consisting of five students with varied prior performance, engagement, and academic levels as suggested by the teachers. The interviews were conducted in the school’s library by two project researchers and lasted for approximately 30 minutes. The students were asked to answer a variety of semi-structured questions, including their experience with using a tablet computer (e.g., do you like using a tablet computer?); learning approaches [e.g., if you got stuck during the learning activities with tablet computers, did you ask a student next to you (or another fellow student) to help you out?]; and engagement (e.g., during the session with tablet computers, did you lose interest as time went by or was the entire lesson equally interesting all the time?). Focus group participants were required to provide their individual answers in
a round-robin fashion and later, if applicable, to agree on a joint answer or reach a consensus.

4 RESULTS

4.1 MODELING PERFORMANCE CHANGE

The main data elements sourced from the learner logs were information on the exact time the students took to complete their tasks and the numbers of correctly and incorrectly completed tasks. On average, students had attempted to solve 86 tasks [standard deviation (SD): 36] in the 15 minutes of duration time, out of which, on average, there were 13 (SD: 11) incorrect attempts. The adaptive condition came with the highest average number of attempts (mean: 96, SD: 40), but also with the highest number of incorrect answers (mean: 20, SD: 9), while the other three conditions had comparable levels of total and incorrect attempts (Table 2).

To model the performance exhibited by the students, the total time (15 minutes) under each condition was split into three-minute intervals. For each interval, we calculated (a) the number of total tasks solved, (b) the number of correctly answered tasks, and (c) the number of incorrectly answered tasks.

<table>
<thead>
<tr>
<th></th>
<th>Non-gamified</th>
<th>Competitive</th>
<th>Adaptive</th>
<th>Collaborative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to complete tasks</td>
<td>13.00</td>
<td>12.98</td>
<td>10.14</td>
<td>11.54</td>
</tr>
<tr>
<td></td>
<td>SD 8.26</td>
<td>5.90</td>
<td>3.90</td>
<td>3.87</td>
</tr>
<tr>
<td>No. of correctly attempted tasks</td>
<td>74.51</td>
<td>73.52</td>
<td>79.47</td>
<td>76.56</td>
</tr>
<tr>
<td></td>
<td>SD 34.59</td>
<td>38.15</td>
<td>30.59</td>
<td>32.23</td>
</tr>
<tr>
<td>No. of incorrectly attempted tasks</td>
<td>9.90</td>
<td>9.87</td>
<td>20.00</td>
<td>11.31</td>
</tr>
<tr>
<td></td>
<td>SD 11.84</td>
<td>10.08</td>
<td>9.28</td>
<td>9.06</td>
</tr>
</tbody>
</table>

4.2 THE CARRYOVER EFFECTS

The study presented in this paper used repeated measures (also referred to as a within-subjects) research design (Davidian & Giltinan, 2003; Lawal, 2014). In the repeated measures design, the outcome variable is measured for all subjects across time. The main benefit of such an approach, as compared with a between-subjects design, is that it exposes all participants of the study to all research treatments (different conditions). Not having an explicit control group increases the overall number of exposed participants, which in turn improves the statistical power of a study.

On the other hand, the main disadvantage of the within-subjects design is the potential carryover (residual) effects (Lawal, 2014), which occur when the effects of one experimental condition persist long enough to affect the results of the next or all of the following experimental conditions (Greenwald, 1976). There are two main types of carryover effects, specifically (1) the practice effect, in which participants improve their results in later conditions because they were trained by the previous
conditions, and (2) the fatigue effect, in which participants get bored or tired in later conditions.

There are several techniques available to minimize or counterbalance the carryover effect, from increasing the time period between the condition measurements to using the alternating treatments design (Barlow & Hayes, 1979). In this study, both of the aforementioned techniques were implemented. Prior to the intervention, the teachers indicated that setting the time between conditions at a minimum of three weeks should be enough to minimize the carryover effects. When compared to the relatively short time of each intervention (i.e., 15 minutes) and the difference in the contents of the tasks used in each of the conditions (due to the alignment of the conditions with the curriculum), the carryover effects were seen as negligible.

To ascertain the claim that the carryover effects were negligible, the first two conditions were designed as a crossover study, with two classes first receiving the base treatment (non-gamified condition) followed by the competitive gamified condition, while the third class received the conditions in the opposite order (i.e., first the competitive gamified condition and, after two weeks, the non-gamified condition). The data collected from these two conditions were used to check for the existence of carryover effects. A t-test between the performance change in the competitive gamified condition in the first order of conditions (non-gamified → competitive) (mean: 0.13, SD: 0.41) and the performance change in the competitive condition in the second order of conditions (competitive → non-gamified) (mean: −0.06, SD: 0.75) shows there is no significant difference in the performance changes exhibited by the two groups of students [t(48) = 1.138, p = 0.261]. Additionally, a t-test between the performance change in the non-gamified condition in the first order of conditions (non-gamified → competitive) (mean: −0.19, SD: 0.61) and the performance change in the non-gamified condition in the second order of conditions (competitive → non-gamified) (mean: −0.16, SD: 0.40) shows there is no significant difference in the performance changes exhibited by the two groups of students [t(48) = −1.171, p = 0.865]. Such results indicate that there are no observed carryover effects between the first two conditions of the study and support the assumption that the possible carryover effects were neutralized by the study design.

4.3 Comparing Completed and Correct Tasks Across the Conditions

By employing the presented model, the trend of performance change was calculated for each student in order to compare the general changes in performance across the gamified conditions. Such a comparison allows for the identification of the best performed condition during the math learning activity as well as the identification of underlying game elements supporting students’ persistent performance.

The four conditions were examined by correlating the number of total tasks solved and the number of correctly answered tasks for all 54 students. Table 3 presents the correlations among four different conditions in all three classes. It should be noted, that for all four conditions, the correlations between the number of total tasks solved and the number of correctly answered tasks were significant. This indicates that there is a trend that students who do well in one condition also do equally well in
other conditions (medium to high-power correlation coefficient).

### TABLE 3
**CORRELATIONS AMONG THE FOUR DIFFERENT CONDITIONS ACROSS ALL STUDENTS IN ALL THREE CLASSES**

<table>
<thead>
<tr>
<th></th>
<th>Non-gamified Correct</th>
<th>Competitive Correct</th>
<th>Competitive Sum</th>
<th>Adaptive Correct</th>
<th>Adaptive Sum</th>
<th>Collaborative Correct</th>
<th>Collaborative Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-gamified</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>0.940**</td>
<td>0.703**</td>
<td>0.675**</td>
<td>0.661**</td>
<td>0.639**</td>
<td>0.719**</td>
<td>0.693**</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Competitive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td></td>
<td>0.570**</td>
<td>0.588**</td>
<td>0.504**</td>
<td>0.504**</td>
<td>0.623**</td>
<td>0.625**</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adaptive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td></td>
<td>0.966**</td>
<td>0.732**</td>
<td>0.715**</td>
<td>0.694**</td>
<td>0.660**</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Collaborative</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td></td>
<td>0.685**</td>
<td>0.691**</td>
<td>0.631**</td>
<td>0.636**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Correlation coefficient, **p < 0.001*

### 4.4 PERFORMANCE TRENDS FOR FOUR EXPERIMENTAL CONDITIONS

The change in student performance was modeled using linear regression, with the steepness (slope) of the linear regression line representing the improvement of performance throughout the activity. For each condition, a linear regression line was used to illustrate the performance trend in the case of the total number of solved tasks, including both correctly and incorrectly answered tasks (Figure 7). Here, we implemented a usage trend as an indicator of students’ sustained performance in the process of solving gamified math activities. Figure 7 illustrates the negative trend in the non-gamified condition, meaning that students gradually solved fewer and fewer tasks as the lesson time progressed. Three other gamified conditions showed positive trends in terms of the number of solved tasks as the time progressed, with the adaptive condition being the most prominent, followed by the competitive and collaborative conditions. A similar analysis was performed at the class level. Figure 8 illustrates trends in the gamification conditions for all three classes as the time progressed. It is noted that, at the class level, the adaptive condition remained the most prominent in all three classes, while the collaborative and competitive conditions showed positive trends in two out of the three classes.
Figure 7. Calculated total trends summed across the three classes (2A, 2B, 3R) for the four different gamification conditions (X-axis: time [min], Y-axis: performance change [no. of task completion attempts]).

Figure 8. Calculated total trends for each class (2A, 2B, 3R) for the four different gamification conditions (X-axis: time [min], Y-axis: performance change [no. of task completion attempts]).

The non-gamified condition showed a borderline positive trend in 2B and negative trends in the other two classes, respectively. Collectively, the trend analysis results indicate that the adaptive gamification condition was the most effective for sustaining student performance for solving the gamified mathematics tasks.

4.5 COMPARING PERFORMANCE TRENDS IN THE FOUR CONDITIONS

Performance trends were calculated for all students with a normalized coefficient of their usage changes for each of the four conditions (slope). A performance trend represents a normalized estimate of students’ progress during an activity, and these coefficients were compared across the four conditions (Table 4). A two-way ANOVA was run on a sample of 54 participants to examine the effects of gamification and class on the students’ performance trends in the four gamification conditions. Notably, there was a
significant interaction between the effects of gamification and class on the performance trend \([F(6, 54) = 2.716, p = 0.0014]\). Simple main effects analysis showed that there was a significant difference in the performance trend exhibited by the students in both the non-gamified and gamified activities \((p = 0.002)\). However, there were no significant differences among the three classes in the performance trend exhibited by the students in the gamified activities \((p = 0.059)\). Although it was not statistically significant, there was absolute difference between the performance trends in the second and third grades, in that students in the third grade showed lower performance trends in the adaptive and collaborative conditions as compared with those in the second grade. Conversely, the performance trend of the third-grade students was higher in the competitive gamification condition (not significantly), indicating such an approach might be more appropriate for slightly older early primary school students.

**TABLE 4**

<table>
<thead>
<tr>
<th>(I) Gamification Type</th>
<th>(II) Gamification Type</th>
<th>Mean Difference (I–II)</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive</td>
<td>Collaborative</td>
<td>0.264</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>Competitive</td>
<td>0.207</td>
<td>0.301</td>
</tr>
<tr>
<td></td>
<td>Non-Gamified</td>
<td>0.463*</td>
<td>0.000</td>
</tr>
<tr>
<td>Collaborative</td>
<td>Adaptive</td>
<td>−0.264</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>Competitive</td>
<td>−0.057</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Non-Gamified</td>
<td>0.199</td>
<td>0.359</td>
</tr>
<tr>
<td>Competitive</td>
<td>Adaptive</td>
<td>−0.207</td>
<td>0.301</td>
</tr>
<tr>
<td></td>
<td>Collaborative</td>
<td>0.057</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Non-Gamified</td>
<td>0.256</td>
<td>0.095</td>
</tr>
<tr>
<td>Non-gamified</td>
<td>Adaptive</td>
<td>−0.463*</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Collaborative</td>
<td>−0.199</td>
<td>0.359</td>
</tr>
<tr>
<td></td>
<td>Competitive</td>
<td>−0.256</td>
<td>0.095</td>
</tr>
</tbody>
</table>

*p < 0.05

The two-way ANOVA procedure revealed that there was a significant difference between the non-gamified condition and the adaptive condition \((MD: 0.46, p < 0.05)\). However, the difference between the adaptive condition and the other gamified conditions (competitive and collaborative) was not significant (this excludes the non-gamified condition), showing that there is no significant difference in performance level increase among the gamified conditions.

In order to better understand the differences between students in terms of performance across different gamification conditions, an analysis of the impact of gender was performed. Two-way ANOVA was used to compare student performance across the conditions (adaptive, collaborative, competitive and non-gamified) and gender (male, female). The results indicated that neither gender \([F(1,209) = 0.920, p = 0.339]\) nor the combined effects of gender and gamification type \([F(3,209) = 3.085, p = 0.976]\) were significant predictors of performance in gamification conditions. Girls do generally perform better than boys across all conditions, with the greatest difference appearing in the competitive condition; however, the difference is not statistically significant. It should be noted that the sample size of both boys and girls per a single condition was small and that more research with a greater sample size is
warranted to more deeply investigate the role of gender in gamification.

4.6 Incorrect Task Attempts in the Four Conditions

The incorrect attempt rate was calculated as the proportion of incorrect task completion attempts among the overall total number of attempts (Table 5). We assumed that an incorrect attempt rate could give more insight into the process of gamified problem-solving, possibly elaborating whether the increase in performance and the number of problems solved in time was related to the increased number of incorrect attempts. This information could be used to argue for the suitability of gamification as a digital lesson design element. Since the adaptive condition showed the highest absolute incorrect attempt ratio, a one-way ANOVA was conducted to compare the differences in the incorrect attempt proportion across the four conditions. There was a statistically significant difference across these conditions \[ F(3,206) = 3.085, p = 0.028 \]. A Tukey post-hoc test revealed that the proportion of incorrect task attempts was significantly higher in the adaptive gamified condition versus in the non-gamified condition (0.084 minutes ± 0.029 minutes, \( p = 0.008 \)). There were no statistically significant differences in the other conditions.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Incorrect attempt rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-gamified</td>
<td>13%</td>
</tr>
<tr>
<td>Collaborative</td>
<td>15%</td>
</tr>
<tr>
<td>Competitive</td>
<td>15%</td>
</tr>
<tr>
<td>Adaptive</td>
<td>22%</td>
</tr>
</tbody>
</table>

5 Discussion

This paper presents the results of an empirical study on different types of gamified learning activities, namely competitive, collaborative, and adaptive, conducted in lower primary mathematics classes. The two research questions that guided this study were: (a) how do different types of gamification affect primary school students’ performance levels and (b) how do game elements used in different conditions contribute to student performance in gamified activities? In this section, we will revisit each research question and discuss implications.

First, our findings indicate that using gamification could contribute to the sustainment and improvement of the performance levels of primary school students in digital mathematics lessons. All three gamification conditions in the experiment led to sustained or increased performance over time, with the adaptive gamified condition prompting the most significant improvement in performance. On the other hand, the activities in the non-gamified condition resulted in negative performance levels, with
students becoming bored and/or starting to lose focus as the noise levels and jitter in the classroom increased. These results are in line with the existing body of research that has reported a positive efficacy of gamification in comparison with the non-gamified conditions (Attali & Arieli-Attali, 2015; Iosup & Epema, 2014; Naik & Kamat, 2016; Sailer et al., 2017; Werbach & Hunter, 2012; Yildirim, 2017).

Second, we examined how the specific elements in the gamification conditions functioned. A comparison of the competitive and non-gamified conditions revealed nonsignificant differences in performance. Such findings, although not completely uniform amongst all classes in this study, may suggest that additional game mechanisms beyond the use of leaderboards and points are necessary to bring about positive outcomes. As such, the collaborative condition employed an enhanced version of leaderboard together with collaboration, narrative game element, and an adaptive mechanism to help group performance. Although adding these game elements was expected to advance the performance levels exhibited by the students, there was no significant difference observed in comparison with the non-gamified and competitive conditions. The results of this collaborative approach varied substantially across the classes, which warrants more experiments in order to fully understand the effects of adding different game mechanics into a competitive gamified activity.

The adaptive condition, similarly to the collaborative condition, utilized competition as a key gamification element, with the elements of narrative and adaptation to individual performance added. The main difference was that the condition used adaptive algorithms that pushed students to the limits of their abilities. The gamification in the adaptive condition led to the highest performance increase as compared with all other conditions. Furthermore, the difference between the adaptive and non-gamified conditions was statistically significant. However, it should be noted that the adaptive condition caused the greatest amount of stress amongst the students, as reported in the focus group interviews, and led to the greatest number of incorrect task completion attempts (significantly more than in other conditions). This observation is illustrated by the following quotes from the focus group interviews:

*When we don’t compete, we are not under stress. So, we don’t have so many wrong answers.* (Ivy, 3R)

*I didn’t like the part when we were fighting against the virus, because it was too hard, I needed more time to calculate the answer, and the virus was so fast. There was too much pressure.* (John, 2B)

Both the collaborative and the adaptive conditions shared narrative gamification elements. The focus group interviews with the students reveal that they were impressed with the story about the computer virus and seemed highly motivated and interested in competing against it, as seen in the quote below from a focus group interview. In particular, the narrative remained interesting even after its initial use and across lessons:

*A text message from one student’s mother to his teacher: Tom was ill and went to the doctor. The doctor found out Tom has a viral infection, to which he responded: “last week at school we were fighting against [a] virus on our tablets. I won, so maybe now the virus is taking revenge on me.”* (Tom, 2B)

Narratives and storytelling elements leveraging “[the] human brain and its natural affinity for narrative construction” (Kapp, 2012) showed successful impacts in this study, although it has to be
acknowledged that the two conditions (collaborative and adaptive) differed in terms of the performance levels exhibited by the students (the difference was not significant but was observable). It seems that the narrative element alone was not sufficient to improve performance, and that the real strength of gamification comes from the combination of carefully interconnecting different gamification elements.

During the lessons with the adaptive condition, researchers observed that students reacted in different ways to the adaptive gameplay. Since the adaptive algorithm modified the activity time limit so that students were challenged, extremely short time limits for solving a single problem (i.e., five seconds) occurred. Some of the students could not cope with the pressure and gave up on the whole activity altogether. In contrast, other students reacted more affirmatively. Once the algorithm allowed for more task completion time, they were motivated to start competing again. Such a setup significantly increased the incorrect attempts given by the students but did not reduce the absolute number of correct attempts in the adaptive condition.

6 CONCLUSIONS

The aim of the present study was to unpack the specifics of gamification types and their impacts in the primary school context. Overall, the findings of this study are in line with flow theory, wherein the best engagement and performance appear when a game player is challenged at levels aligned with his/her skills. We found that early primary school students were engaged when they were challenged at a suitable difficulty level, but their performance dropped when the level of challenge became too high and they grew frustrated. We also found that not all students were motivated by the game elements to the same extent. Our findings are consistent with the prior research that showed that some gamification conditions yield a number of under- or non-achievers who appear to have little or no desire to complete a task or to compete with others (Hew, Huang, Chu, & Chiu, 2016).

The use of competition in early school educational activities is fairly complex, due to the large number of ways it can be implemented. One key finding of this study is that the efficacy of commonly-used game mechanisms (e.g., leaderboards, points) should not be assumed. We found that the use of leaderboards alone was not sufficient to sustain students’ performance levels. Instead, the level of student performance can be greatly enhanced via the integration and careful combination of different game elements, including not only leaderboards and badges but also elements such as narratives and adaptive mechanisms based on individual performance.

Further, our finding challenges the prevalent belief regarding the positive efficacy of collaborative learning methods. While collaborative learning is generally regarded as an effective way to promote interaction and engagement, our findings disagree with the existing research that reported the positive effects of social elements in the gamification context (de-Marcos et al., 2016), indicating the need for further empirical evidence. Collaboration is a mechanism that introduces social learning aspects into gameplay. In multiplayer video games, for instance, multiplayer interaction is a common
strategy wherein players collaborate to help each other reach a common goal (Domínguez et al., 2013). In the current study, however, achieving a group goal was not as effective as achieving an individual one. One potential speculation about these unexpected findings regarding the collaboration mechanism may be related to the system design, which did not provide recognition for individual contributions in collaborative efforts. It is possible that students would be more engaged and motivated if they can see their contributions to the group’s performance as well as if they were able to obtain recognition from other students.

Some limitations of this research should be noted. First, generalizing the findings from this study to other research contexts should be done with caution, due to the small sample size, the lack of randomization of experimental conditions, and the within-subjects design. Although significant, the results of the present study are executed using a mix of gamification elements that might not exist in other technology-based gamification learning environments, thereby potentially reducing the replicability of the study. Second, the measured number of total, correct, and incorrect task completion attempts per student comes with a large standard deviation. Third, this study used a rather short intervention time in a single session. It is necessary to conduct follow-up research that employs longer intervention times in a longitudinal manner. Lastly, this study focuses on the use of different game mechanics in digitalized math learning involving simple arithmetic calculations. Determining the efficacy of gamification in other subject areas (e.g., language learning) and in conjunction with higher-order complex skills requires additional empirical investigations.

Nonetheless, we believe that this study makes both theoretical and practical contributions to the field of gamification in education. First, the theoretical contribution is that this study provides empirical evidence of the impact of different gamification elements on learning performance. As mentioned earlier, the existing body of research tends to regard gamification as a generic construct and little is known about how and under what conditions the specifics of game elements and mechanics work or fail. This study provides important empirical evidence about the efficacy of adaptive elements in gamification and also challenges the prevailing positive assumption of collaborative elements. Next, the practical contribution is that the findings presented in this study can be utilized as a means of improving everyday class activities (e.g. mathematics practices). Drawing from this study’s results, simple game strategies and elements such as narratives, points, or badges can be introduced into both digitalized and non-digitalized lessons. Online tools such as OpenBadges1 and Classcraft2 can help teachers to easily and quickly create game-like environments for their students and to harness the benefits of gamification in the classroom. These services allow for the combination of online and offline activities as well as the definition of rules for virtual or real awards, without necessarily changing the structure of existing or planned lessons. Gamification can also be done completely offline by adding motivational narratives as a prequel to an activity or by awarding paper badges or medals for certain educational achievements.

---

1 https://openbadges.org/
2 https://www.classcraft.com/
The biggest challenge in implementing gamified educational activities in early primary school is sustaining gameplay engagement and student performance throughout lessons for a longer time period. Drawing from the main implications of this research, we suggest promising areas for future research. First, one potential for further research lies in designing suitable mechanisms for gameplay adaptation according to different student preferences. This will lead to the development of a more complete adaptive learning system with learner profiling that incorporates the dynamics of personalizable features. Another interesting area for future research would be the examination of the effects of differentiating challenges and skill levels based on flow theory (Csikszentmihalyi, 2014) to find an optimal experience or path wherein students reach a state of flow in gamified learning activities, without the triggering of too much anxiety or stress among young learners. Lastly, we suggest that future research should examine the hidden efficacy of incorrect attempts, wherein students possibly learn from their failed attempts. The notion of productive failure suggests that when students are engaged in problem-solving first and then are taught concepts and procerus (reversing the typical sequence of instruction), students may fail to generate correct solutions in the initial phase but that this failed experience can be productive to prepare them for future learning (Kapur, 2014). While we were not able to examine the efficacy of productive failure due to our study’s design (i.e., our students received the instruction first and then used the application), future research can reverse the order of direct instruction and problem-solving with the application to empirically examine whether students can learn from their incorrect attempts.

In conclusion, this research addresses the research gap and lack of experimental research design regarding the effects of different gamification elements. Future investigations should consider exploring the efficacy of different dimensions or combinations of game mechanics.
REFERENCES


