Rodriguez-Aflecht, G., Jaakkola, T., McMullen, J., Hannula-Sormunen, M. & Lehtinen, E. (2017). Voluntary vs compulsory playing contexts: motivational, cognitive, and game experience effects. Simulation & Gaming. 48(1), 36–55. doi:10.1177/1046878116673679

Voluntary vs compulsory playing contexts: motivational, cognitive, and game experience effects

Abstract

Background. Serious games are often used in formal school contexts, in which students' lack of control over the playing situation may have repercussions on any motivational gains.

Aims and Method. The first aim was to investigate to what extent $n = 579$ fifth grade students in Mexico who received a mathematics serious game played it voluntarily. Then, we explored how students who played voluntarily ($n = 337$) differed from those who did not by either gender or pre-test mathematical skills or motivation. The second aim was to find out whether two play contexts, the group of voluntary players and a second group consisting of students playing at school as a compulsory part of their regular mathematics lessons ($n = 482$), differed in game experience, game performance, and cognitive and motivational outcomes.

Results. Students from the volunteer group who played had higher pre-test mathematical skills and math interest than those who did not play. Students in this group did not otherwise differ. Compared to students from the volunteer group who played, students in the school group played for longer, completed more tasks, and enjoyed playing the game more. However, their advanced mathematical skills did not improve as much.

Conclusion. Motivation did not improve regardless of play context, suggesting serious games should be implemented for their learning content rather than because they are assumed to be motivating.

Keywords

Mathematical skills, primary school students, serious games, voluntary play, interest, self-efficacy, game experience, game performance, adaptive number knowledge, arithmetic fluency

Serious Games are digital games designed to produce specific outcomes, such as learning or motivational gains. Regarding the latter, it has been argued that serious games are oxymoronic by nature (Abt, 1987). Jenkins (2011, p. xxv) describes some critics' concerns that games should not be "assigned responsibilities, but freely chosen irresponsibilities." Nevertheless the belief that serious games will increase student motivation towards learning persists (Klemmetti, Taimisto, & Karppinen, 2009) and is widely highlighted in educational policy documents (e.g. Claro, 2007). At the same time there is growing evidence that students' acceptance of digital games in the classroom cannot be taken for granted and that so far serious games have not been significantly more motivating compared to learning methods used in control groups (Bourgonjon, Valcke, Soetaert, & Schellens, 2010; Whitton, 2010; Wouters, van Nimwegen, van Oostendrop, & van der Spek, 2013).

Islas Sedano and colleagues (2013) have argued that different motivational results may be achieved with serious games if play were to be free and voluntary rather than a formal and prescribed school activity. Along with Wouters et al. (2013), they point out that when games are played in school it is usually teachers who decide what games will be played, for how long, and under which circumstances, and that this lack of autonomy (Ryan & Deci, 1985, 2000) may have repercussions on students' motivation. Autonomy is afforded to some extent within games through game design and in-game player choices (Ryan, Rigby, & Przybylski, 2006). Nevertheless, it is relevant to explore whether player control over when and for how long serious games are played could have an effect on the way games are played and experienced. This in turn could also have an effect on players' resulting outcomes, particularly motivational ones.

Research exploring the aforementioned issue is limited, perhaps because of the challenges of systematically collecting data on the type of play which could be called autonomous. Serious games are usually implemented in formal contexts such as schools or laboratories where logistic constraints might make truly free and voluntary play impractical. However, some important work has been carried out which studied the effects of serious games played outside of a formal context. For instance, Takatalo and colleagues (2010) investigated whether playing at either home or laboratory contexts had an effect on higher education students' game experiences with commercial games for entertainment. De Grove, Van Looy, Neys, & Jansz (2011) explored whether playing a serious game at home rather than at school would have different effects on adolescents' perceived learning and game experiences.

Even so, to our knowledge no studies have focused on the role of an informal and voluntary context on a serious game's effects on primary school children's game performance and game experience as well as their learning and motivational outcomes. In order to do this, the present study uses a mathematics serious game called the Number Navigation Game (NNG). Research shows NNG is successful at improving Finnish primary students' mathematical skills when compared to a control group (Brezovszky, 2016; Brezovszky et al., 2015), yet their motivation slightly decreases (Rodríguez-Aflecht et al., 2015). One of our aims is to find out whether students who receive a copy of NNG to play voluntarily will choose to do so. By stressing the voluntary aspect of play it is likely that not all students will, so additionally we are interested on how those who play and those who do not differ in pre-test motivation or mathematical skills. As gender differences have been reported not only in mathematics (Wigfield & Cambria, 2010; Wigfield & Eccles, 2002) but also in gaming (Bonanno & Kommers, 2008), it is also explored whether gender plays a role on students' decision to play or not to play. Our second aim is to compare the effects of NNG on mathematical skills and motivation when played in two different contexts: school, with play being a compulsory component of regular mathematics lessons, and the volunteer context, with play being free and taking place outside school, such as at home.

Background

Motivation

There are different and sometimes overlapping theories of motivation (de Brabander & Martens, 2013). Deci and Ryan's (2000) self-determination theory explicitly deals with autonomy and has also been applied to the study of digital games, such as to commercial games for entertainment (Przybylski, Rigby, Ryan, 2010; Ryan, Rigby, & Przybylski, 2006). According to their theory, activities foster intrinsic motivation by satisfying human needs, one of which is autonomy. Events or conditions that diminish a sense of choice, control, or freedom interfere with perceived autonomy and can undermine intrinsic motivation (Deci, Koestner, & Ryan, 1999). In this study, motivation is looked at from the theoretical perspective of Eccles and colleagues' expectancy-values model, as this model of achievement motivation has been proven useful in predicting a person's future performance, persistence, and task choice (Eccles & Wigfield, 2002; Eccles, Wigfield, & Schiefele, 1998; Wigfield, 1994; Wigfield & Cambria, 2010; Wigfield & Eccles, 2000). The motivation to perform tasks is determined both by a person's expectation to succeed and the subjective value they give to succeeding (Wigfield, 1994). Expectancy-values decrease across the school years, due to increased competition in schools and to children's improved ability to interpret feedback and self-assess (Wigfield & Eccles, 2002).

Our specific focus is on the motivational aspects of self-efficacy and interest. The expectancy or self-efficacy part of the model is similar to Bandura's (1997) efficacy expectation construct, in that it refers to a person's expectations for their future success in a specific task. It is also related yet distinct from Deci and Ryan's concept of competence (1985), or a person's beliefs in their own ability (Wigfield & Eccles, 2000, 2002). Interest is one of the *value* parts of the model. Interest or intrinsic value indicates how enjoyable a person finds a task. It is comparable to Hidi and Renninger's (2006) notion of individual interest and related to constructs of other theories of motivation, such as Deci and Ryan's (1985) intrinsic motivation (Wigfield & Eccles, 2002). In short, out of the many different perspectives on motivation, this study focuses on the expectancy-values model's concepts of self-efficacy and interest.

Mathematical skills

NNG was designed to improve students' general adaptivity with arithmetic (Lehtinen et al., 2015), which involves being able to select and use the most appropriate arithmetic problem solving strategies for a particular person, situation, and problem (Verschaffel, Luwel, Torbeyns, & Van Dooren, 2009). McMullen and colleagues (2016) have argued that the two foundations of adaptivity with arithmetic are Adaptive Number Knowledge (ANK) and arithmetic flexibility. NNG is particularly targeted at enhancing students' ANK, which refers to students' ability to recognize and utilize the knowledge of numerical characteristics and relations within an arithmetic problem and requires having a well-connected network of numerical relations (McMullen et al., 2016). Arithmetic fluency is necessary for further conceptual and procedural development, and refers to how quickly and accurately one is able to retrieve basic number facts and combinations (Baroody, Bajwa, & Eiland, 2009; Canobi, 2009).

Game performance and experience

Game performance, i.e. how long and how successfully a game is played, is tightly connected to the way a person experiences a game. For instance, if a person plays for such a short time that they do not make it past the tutorial stage of a game, they might feel more confused or frustrated compared to a person who already understands the basic mechanics of gameplay. Game performance has also been shown to mediate learning outcomes (Rodríguez-Aflecht et al., 2016). Game experiences influence players' cognitive decisions as they play (Nacke & Lindley, 2009). Many different models of game experiences have been proposed (see Nacke & Drachen, 2011). In the present study, we use the framework developed by Poels and colleagues (2008). However, we focus only on three particular game experiences: positive affect, competence, and positive belief. Positive affect refers to a post-play affective state which indicates how pleasant the game experience was to players. Competence refers to a person's beliefs in their own ability, in this case, as a player. Previous research suggests that out of all the game experiences in the framework of

Poels and colleagues (2008), competence seems to be the only significant predictor of post-test motivation (Rodríguez-Aflecht et al., 2015). Positive belief refers to a student’s belief that a game’s learning content is useful to them.

Research Questions

1To what extent do primary school students voluntarily play a mathematics serious game? Do students who play voluntarily differ from those who do not in mathematical skills, motivation, and gender?

Nielson (2011) investigated autonomous e-learning and reported that it results in non-participation when it lacks human support, while Fryer, Bovee, and Nakao (2014) found that higher education students were consistently amotivated to engage with the e-learning component of a language course. However, this study focused on a different type of technology, learning content, and age group.

Lai, Wang, and Lei (2012) found that gaming self-efficacy did not predict technology adoption. Therefore, we anticipate that gaming self-efficacy will not differ amongst students from the volunteer group who play and those who do not. However, we hypothesize that students who play will show higher pre-test mathematical skills, math interest, math self-efficacy, and gaming interest than those who do not play. Gauthier, Corrin, and Jenkinson (2015) reported that gender had no effect on voluntary use of an online study aid tool amongst medical students, so we do not expect to see gender differences.

2) Does play context (voluntary play or compulsory play at school) have an effect on students' game performance and experiences?

It has been reported that participants playing both serious games and commercial games for entertainment at home play significantly longer than those playing in a formal context such as a school or a laboratory (De Grove, Van Looy, Neys, & Jansz, 2011; Takatalo, Häkkinen, Kaistinen, & Nyman, 2010). Based on these results we believe that students from the volunteer group will play longer than students playing at school and thus complete more tasks. We do not expect the in-game success of these two groups to differ. As for game experiences, prior research shows that students playing at home report higher scores of enjoyment (De Grove et al., 2011) and report higher competence than those who played in a lab (Takatalo et al., 2010). Thus we believe that students playing voluntarily will report higher positive affect and competence than students playing at school. We expect that since teachers are making time in their regular mathematics lessons for students to play at school, this enforcement might influence students into thinking the game is useful and will help them improve at mathematics, and so these students will have higher scores on the belief that the game will have a positive value in their lives.

3) Does the play context have an effect on students' mathematical skills and/or on their motivation?

De Grove and colleagues (2011) found that students playing at home felt they learned more compared to students playing at school. However, this was a small effect moderated by the fact that students playing at home played for significantly longer. Also, their study focused on perceived rather than actual learning. Previous studies on the effects of NNG on mathematical skills and motivation show that the game has a positive effect on the former but not on the latter (Brezovszky, 2016; Brezovszky et al., 2015; Rodríguez et al, 2015). Based on these findings, we hypothesize that NNG will have a positive effect on the mathematical skills of all students, regardless of condition. We assume that participants' lack of control over the gaming condition/context in those prior studies could have moderated the motivational benefits of the game, and expect the motivation of students playing voluntarily will improve compared to those playing at school.

Methods

Participants

The participants were 1061 fifth grade students (*n* = 508 girls; *M* =10.13 years, *SD* = 0.41) from the state of Nuevo León in northeastern Mexico. Students came from 53 classes and 39 different schools. Participation was voluntary and all participants had written consent from their parents. Ethical guidelines of the University of Turku were followed.

Procedure

Students were randomly assigned by class into either the volunteer group (*n* = 579, *n* = 278 girls, *n* = 32 classes) or the school group (*n* = 482, *n* = 230 girls, *n* = 21 classes). Students' mathematical skills and motivation was measured before and after an intervention period. These pre- and post-tests were administered by trained testers in the presence of the class teacher during regular 45-minute class periods. For the intervention period, all students received memory sticks containing NNG. The intervention took place in spring 2015 and lasted a little over two months (*M* = 61.38 days, *SD* = 5.90 days). Parents, teachers, and students involved in the volunteer group were instructed that play was voluntary. The school group played NNG as part of their regular curriculum-based mathematics lessons, and teachers were free to decide the number and length of play sessions, though it was recommended the play sessions last at least half an hour. At the time of the post-test, game log data from the memory sticks was retrieved.

All participating teachers received a brief guide for the intervention. The guide included NNG’s learning aims, game mechanics, and some pedagogical tips, such as how to introduce the game, provide support, and debrief students. Additionally, teachers received both a link to an online tutorial video and a document they could distribute amongst their students regarding the game. This document for students provided information on the research project, described the game’s objectives and mechanics, had a Frequently Asked Questions section, and gave information on who to contact in case of questions or issues.

Instrument

NNG. NNG is a PC strategy game in which a player controls a ship and navigates across maps, which represent archipelagos superimposed on a grid of numbers corresponding to a hundred square. Inputting location-dependent arithmetic operations through a number pad and a command box will take the ship from one number in the hundred square to another. The objective of the game is to retrieve items with which to build villages and towns. There are altogether 64 maps divided into four levels of increasing difficulty. Based on the quality of gameplay students earn bronze, silver, or gold coins upon completing a map. For example, a player completing a map with a top score would earn a gold coin. For a full description see Lehtinen and colleagues (2015).

Measures

Interest and self-efficacy. Math interest (for example "I like math") and self-efficacy ("I am certain I can do difficult math tasks") were measured at pre- and post-test through an adapted version of the mathematics expectancy-values scale used by Berger and Karabenick (2011). The items were translated into Spanish and adapted to the ages of our respondents, and were adapted to also measure gaming interest ("I like to play videogames") and gaming self-efficacy items ("I am confident in my ability to play videogames"). All measures included three items, to which participants responded to using a five point Likert-scale ranging from 1 (*completely disagree*) to 5 (*completely agree*), with the mid-scale 3 being neutral. The reliabilities of factors is as follows: at pre-test: math interest Cronbach's α= .78 and composite reliability (CR)= .86, math self-efficacy α= .60/CR= .78, gaming interest α= .87/CR= .91, and gaming self-efficacy α= .61/CR=.78. At post-test: math interest Cronbach's α= .80/CR= .89, math self-efficacy α= .64/CR= .80, gaming interest α= .88/CR= .92, and gaming self-efficacy α= .69/82. The reason both Cronbach's alphas and CR are used to examine the internal consistency and reliabilities of these and subsequent measures is that the former has been found to somewhat underestimate true reliability, and it has been suggested that both may be used interchangeably (Peterson & Kim, 2013).

Arithmetic fluency. A slightly adapted version of the Mathematical Fluency test of the Woodcock-Johnson Tests of Achievement (Woodcock, McGrew, & Mather, 2001) was used to measure arithmetic fluency. In this timed paper-and-pencil test, students have three minutes to answer as many simple arithmetic operations out of 160 as they can. The variable used is a total score of correct answers.

ANK. The Adaptive Number Knowledge (ANK) Task, a timed paper-and-pencil test, has been developed specifically to measure ANK and has been used previously (see McMullen et al., 2016). Students are given four to five numbers they need to combine using the four basic arithmetic operations to reach a given target number; they have 1.5 minutes to come up with as many solutions as they can. An example and two tasks were given at each measurement point. Although the numbers used in the items were changed from pre- to post-test, effort was put into guaranteeing that similar types of number-operation combinations would be required for solving the problem correctly. The ANK Task was scored looking at two criteria: total of correct solutions (number of solutions that were mathematically correct and followed the instructions), and complex solutions, i.e., solutions in which students used two or more different arithmetic operations to reach the target number; a sum score is here used as the ANK variable. At the pre-test ANK test's Cronbach's α= .67/CR= .82 and in post-test α= .71/CR= .87.

Game performance. Variables related to game performance were recorded on game log data files and were analyzed post-play. These include time played and the number of complete and incomplete tasks. Complete tasks refer to each map in NNG that a student played through, while incomplete tasks refer to maps which were abandoned or restarted. By dividing the total amount of gold coins earned by the total amount of maps or tasks completed, we created a variable called the proportion of tasks completed with the top score.

Game experience. Game experience was measured post-play using Oksanen's (2013) adapted version the Game Experience Questionnaire developed by Poels and colleagues (2008), which was translated into Spanish. Items consisted of a statement and a 1-5 scale to indicate level of agreement, with answers ranging from 1 (*not at all*) to 5 (*extremely*), with the mid-scale 3 being neutral. Measures had good reliabilities: competence ("I was good at playing") α = .76/CR= .82, positive affect ("I thought playing was fun") α =.87/CR= .88. The additional variable belief in the positive value of the game was measured with three self-developed items ("This game helped me learn math"), which also had good reliability, $\alpha = .77$/CR= .84.

Analysis

To determine whether multilevel analyses were necessary given the nested structure of our data, we ran variance component analyses on SPSS. The Intraclass Correlation Coefficients (ICC) of all variables at pre-test were lower than .04, which indicated multilevel analyses were unnecessary (Kline, 2011). To answer the first research question and sub-question, we used descriptive statistics, a chi-square test for association between gender and playing voluntarily at home, and a series of ANCOVAs to compare pre-test mathematical skills and motivation of students who played and those who did not, whilst also controlling for gender. As we were interested in looking at pre-test differences which could explain why some students felt interested enough to even give the game a chance, here any player who played for at least one minute is considered as having played the game.

From this point onward, we selected participants from the school group and the volunteer group who played on the basis that they completed at least five maps. We use this threshold value as we consider that in order to study their game performance, game experience, and above all, the effects of the game, it is necessary to ensure that students made significant progress to make these comparisons possible. The second research question was answered through ANCOVAs comparing the game performance and game experiences of students in the volunteer group who played and those in the school group while controlling for gender effects. Finally, to answer the third research question, ANCOVAs were conducted to determine statistically significant differences on post-test mathematical skills and motivation between students from the volunteer group who played, students from the volunteer group who did not play, and students from the school group. Gender was controlled for and the corresponding values at pre-test were used as co-variates. When necessary, post-hoc comparisons were carried out.

Results

Voluntary play

Out of the students in the volunteer group (n = 579), log data indicates that n = 337 played for longer than a minute. On the other hand, there was no viable data for n = 242 participants. Students with no viable data either failed to return the memory stick containing the game (n = 135) or else they did turn in the memory stick but it was damaged, malfunctioned, carried a virus, or contained no data (n = 107). Memory sticks containing no data could mean either that students did not play or that the log files were wiped. Upon being asked at post-test whether they had played game, n = 166 out of those n = 242 participants without viable data either admitted they had not played and/or selected reasons why they had not played. These students (n = 166) are therefore the students that can be confidently described as not having played. When asked to select the reason they had not played, all except three selected at least one option. The largest group of students (n = 39) said they did not have a computer with Windows at home. Others (n = 24) said their memory stick did not work, broke down, or got lost. Some students (n = 26) said they did not have the time, while others (n = 7) admitted they did not like the game. Though some students (n = 10) thought the game was too difficult, others (n = 21) thought it was too easy. Another group of students (n = 28) selected "other" as the reason for not playing. A few students (n = 9) gave more than one reason, out of which seven listed not having a computer with Windows at home as one of them. Perhaps other obstacles may or may not have overcome had there been easy technological access at home.

There was no association between gender and playing ($X^2(1) = 2.82, p = 0.11$). Amongst students who did play, n = 173 were girls and n = 164 were boys, while amongst students who did not play, n = 72 were girls and n = 94 were boys. Table 1 shows descriptive statistics on the pre-test mathematical skills and motivation of students in the volunteer group who played and those who did not play. Both pre-test ANK and math interest was higher in students who played than those who did not, $F(1,499)$=10.99, $p < .01$,

$np2 = .02$ for ANK and $F(1,434)= 5.37$, $p < .05$, $np2 = .01$ for math interest. In the case of ANK, there was also a gender effect showing boys outperformed girls in pre-test ANK, $F(1,499) = 11.95$, $p < .01$, $np2 =.02$. While condition had no effect on gaming interest or on gaming self-efficacy, there was a gender effect. Boys had significantly higher interest towards gaming than girls did at pre-test, $F(1,433) = 23.14$, $p < .001$, $np2 =.05$, and they also had higher pre-test gaming self-efficacy, $F(1,434) =9.46$, $p < .01$, $np2 =.02$. There were no significant differences in arithmetic fluency or math self-efficacy, nor were there interaction effects between condition and gender

[Table 1]

Game performance and experiences

Table 2 shows descriptive statistics on game performance and game experiences of students from the school group ($n = 290$) and those from the volunteer group who completed at least five tasks ($n = 197$). Students in the school group played for significantly longer, $F(1,483) = 24.85$, $p < .001$, $np2 = .05$), and completed more tasks, $F(1,483) = 21.12$, $p < .001$, $np2 = .04$, while also leaving a lot more tasks incomplete, $F(1,483)=69.33$, $p < .001$, $np2 = .13$. The proportion of tasks completed with the top score did not vary. As for game experiences, while competence did not vary by play context, students in the school group reported higher positive value, $F(1,457)=4.55$, $p < .05$, $np2 = .01$, and positive affect, $F(1,457)=11.04$, $p < .01$, *np2* = .02, than students in the volunteer group. Furthermore, there was a gender effect on tasks completed, $F(1,483) = 5.76$, $p < .05$, *np2* =.01, on the proportion of tasks completed with a top score, $F(1,483)=23.08$, p < .001, *np2* = .05, and on feelings of competence, $F(1,456)=8.04$, $p < .01$, *np2* =.02; in all three, boys reported higher scores than girls. There were no interaction effects of gender and condition.

[Table 2]

Learning and motivational outcomes

Descriptive statistics on post-test mathematical skills and motivation after controlling for their corresponding pre-test scores are presented in Table 3. Regarding mathematical skills, there was a significant effect of condition on post-test arithmetic fluency, $F(1,646)= 11.46$, $p < .001$, *np2* = .03 and on post-test ANK, $F(1,646) = 6.32$, $p < .01$, *np2* = .02 after controlling for their equivalent scores at pre-test. Post-hoc comparisons using Bonferroni correction were carried out. These show that students in the volunteer group who did not play had lower post-test arithmetic fluency scores than students in the volunteer group who played (Mean difference= -5.26, $p < .001$). They also had lower post-test arithmetic fluency scores than students playing at school (Mean difference= -5.63, $p < .001$). As for ANK, students in the volunteer group who did not play had lower post-test

ANK scores than those in the volunteer group who played (Mean difference= -.99, $p < .01$). No gender effects were found.

Condition did not have an effect on post-test math interest, $F(1,551) = 2.90$, $p = .06$, *np2* = .01 after controlling for pre-test math interest. However, condition did have an effect on post-test gaming interest, $F(1,549) = 8.42$, $p < .001$, *np2* = .03, post-test math self-efficacy, $F(1,551) = 3.35$, $p < .05$, *np2* = .01, and post-test gaming self-efficacy, $F(1,550) = 5.46$, $p < .01$, *np2* = .02 after controlling for the corresponding variables at pre-test. In the case of gaming interest, students in the volunteer group who did not play scored significantly lower than both volunteers who did (Mean difference = -0.30, $p < .01$) and students playing at school (Mean difference= -0.34, $p < .001$). There was also a gender effect, $F(1, 549) = 20.49$, $p < .001$, *np2* = .04, with girls having lower post-test gaming interest than boys in all conditions. As for math self-efficacy, students in the volunteer group who did not play had lower post-test scores than students who played either voluntarily (Mean difference = -0.20, $p < .05$). Finally, in the case of post-test gaming self-efficacy, students in the volunteer group who did not play had lower post-test gaming self-efficacy than students in school group only (Mean difference = -0.27, $p < .01$). There was also a gender effect, $F(1,550)=14.00$, $p < .001$, *np2* = .03, with boys reporting higher gaming self-efficacy than girls in all conditions. There were no interaction effects between gender and condition.

[Table 3]

Discussion

The two aims of this study were: 1) to investigate how primary school students who were given a mathematics serious game and played it voluntarily differed from those who decided not to play and 2) to compare volunteer contexts vs. organized school situations regarding the effects of the game on game experience, game performance, and cognitive and motivational outcomes. Our most important finding is that playing has an overall positive effect on mathematical skills. In addition, it seems that voluntary play has a positive influence on advanced mathematical skills (ANK). As for motivation, it remains stable, although students who did not play had lower post-test scores on gaming interest, gaming self-efficacy, and math self-efficacy.

Over half of participants in the volunteer group actually played and there was no prominent reason why the rest did not; the most frequently cited one was the lack of a computer that would run the game. Students in the volunteer group who played differed from those who did not in that their advanced arithmetic skills (ANK) were higher at pre-test. However, there were no differences in basic arithmetic skills (fluency). Students who played also had a higher pre-test math interest than those who did not. This is in line with previous research reporting that students look for opportunities to engage in tasks which support their interest (Hidi & Renninger, 2006). Notably, there were no differences in gaming interest between the groups. It seems that when it comes to serious games, interest towards the subject matter itself is more important than interests towards games. However, there can be no direct conclusions that these results are related to students' willingness to play because there could be many other explanations related for instance to socioeconomic factors. Socioeconomic conditions and access to technology at home could be an issue limiting the opportunities for many (Barron, Walter, Martin, & Schatz, 2010). Socioeconomic conditions too could possibly explain the lower advanced arithmetic skills amongst students in the volunteer group who did not play. In the future, when studying voluntary play outside of schools it is necessary to also collect data on participants' access to technology. It is also worth noting that independent from condition, gender had an effect on pre-test ANK, gaming interest, and gaming self-efficacy, with boys scoring higher in all three.

Unlike results reported by De Grove et al. (2011), students playing at school played for longer and completed more tasks than students playing at home. A possible explanation is that the skills NNG aims to develop are more evidently related to the school curriculum than those of the game used by De Grove and colleagues, so perhaps teachers felt they could justify spending more time on the game. Students in the school group also had a higher proportion of tasks completed with the top score, but they had a higher number of incomplete tasks. They also reported more positive game experiences than students who played voluntarily at home. There are two interpretations for the result. First, in the school group there were fixed playing times, so at the end of a session some students might have been forced to stop mid-task. This group's more positive game experiences could perhaps be due to students comparing the game to the activities in their regular mathematics lessons as well as students playing voluntarily at home having other engaging activities to spend their time on. A second explanation could be that at school there were more distractions, leading to less focused gameplay. This would differ from the conclusion reached by Takatalo et al. (2010), who found that participants in the home group concentrated less. However, their result might be related to their collection of physiological data from participants in the laboratory context. Compared to their laboratory context in which physiological measurement instruments were used, our school class context certainly seems less formal. In this study, players from the volunteer group lacked the distractions of a social school environment. This could support the idea that volunteer players were more concentrated in the content. Simultaneously, it could also be the reason why students playing voluntarily at home or elsewhere did not report as high positive affect or positive value as students who played at school.

Play context had an effect on post-test mathematical skills when controlling for their equivalents at pre-test. Students who played (either at school or voluntarily at home) had higher post-test arithmetic scores compared to students in the volunteer group who did not play, which is in line with previous studies done with NNG in Finland (Brezovszky et al., 2015; Rodríguez-Aflecht et al., 2015). In the case of ANK, students in the volunteer group who played had higher post-test scores than students who did not play. So although students in the school seemed to enjoy the game experience the most, students playing voluntarily at home showed the most improvement in their mathematical skills.

Like in previous studies, intervention with NNG does not seem to result in motivational improvements, and in fact seems to be quite stable (Rodríguez-Aflecht et al., 2015, 2016). Contrary to expectations, allowing students to control where, when, and for how long they would play mostly did not have an effect on motivation. There was no effect of condition on math interest. However, students in the volunteer group who received the game but did not play had lower post-test scores on gaming interest, math self-efficacy, and gaming self-efficacy. Students in this group may have redefined or reevaluated their gaming interest and self-efficacy in light of their unwillingness or inability to play, though it seems that their math interest was left intact. Gender effects were found in gaming interest and self-efficacy, with boys reporting higher feelings at post-test than girls. However, gender was not related to the changes of motivation and mathematical skills in different conditions.

Limitations

One limitation of this study is that it is difficult to have control over the setting and know what exactly the students in the voluntary play group really did; it may have been possible for students to cheat, using a calculator or getting someone else to play for them. However, based on their improved mathematical skills, it is unlikely this was the case. It also is unclear why some students did not return their memory sticks, so in the end there was a relatively small amount of participants we could use. The role of both parents and teachers is another important limitation. It is to be expected that parents' and teachers' attitudes as well as the support or encouragement they offered to students greatly varied, and this is problematic especially in the case of the volunteer group. Furthermore, while teachers of the school group received a short guide with pedagogical suggestions on how to support their students before, during, and after play sessions, it is unknown to what extent teachers familiarized themselves with this material and used it. Some teachers might have passively watched their students play while others could have been enthusiastic and active in reflecting with their students on their learning.

Conclusion

Despite these limitations, we believe this study offers important insight into the opportunities of voluntary play. Playing voluntarily outside of formal environments results in higher mathematical skills, but many students are unwilling or unable to participate. Students who played voluntarily showed higher advanced mathematical skills and interest towards mathematics compared to those who did not. Further studies should focus in more detail on how different explicit and implicit motivational interpretations and background factors are related to students' decisions regarding whether they start playing or not in a volunteer play situation. Students playing at school report a more positive experience, and while their mathematical skills do not improve as much as students playing voluntarily despite longer exposure to the game, they still improve when compared to students who do not play at all. While NNG is successful at improving mathematical skills, motivation remains quite stable regardless of the conditions in which playing takes place. It remains to be seen to what extent can an improved game design bring different results. In the future, it is important to consider how serious games are introduced in formal educational contexts and find ways in which some features of voluntary play could be integrated. Some recommendations for future studies are to collect data also from parents and teachers in order to gain a comprehensive understanding of the role these other stakeholders play. Finally, it is important that serious games are implemented because of their valuable learning content and not simply because they are assumed to be motivating.

Acknowledgements

We warmly thank all our participants. We appreciate all the help received from all the staff at the Consejo Nacional de Ciencias y Tecnología for Children, and want to thank specially Dr. Adriana Elizondo and Mayra Rosales for their invaluable assistance in the implementation of this study.

References

Abt, C. C. (1987). *Serious games*. University Press of America.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York, NY: W.H. Freeman.

Baroody, A.J., Bajwa, N.P., & Eiland, M. (2009). Why can't Johnny remember the basic facts? *Developmental Disabilities*, *15,* 69-79. doi: 10.1002/ddrr.45

Barron, B., Walter, S.E., Martin, C.K., Schatz, C. (2010) Predictors of creative computing participation and profiles of experience in two Silicon Valley middle schools. *Computers & Education, 54*(1), 178-189. doi:10.1016/j.compedu.2009.07.017

Berger, J., & Karabenick, S. A. (2011). Motivation and Students' Use of Learning Strategies: Evidence of Unidirectional Effects in Mathematics Classrooms. *Learning and Instruction, 21*(3), 416–428. http://dx.doi.org/10.1016/j.learninstruc.2010.06.002

Bonanno, P., & Kommers, P. A. M. (2008). Exploring the influence of gender and gaming competence on attitudes towards using instructional games. *British Journal of Educational Technology*, *39*(1), 97-109.

Bourgonjon, J., Valcke, M., Soetaert, R., & Schellens, T. (2010). Students' perceptions about the use of video games in the classroom. *Computers & Education, 54*(4), 1145–1156. http://dx.doi.org/10.1016/j.compedu.2009.10.022

Brezovszky, B. (2016). *The Effects the Number Navigation Game-Based Training on the Development of Primary School Students' Adaptive Number Knowledge*. Manuscript in preparation.

Brezovszky, B., Rodríguez -Aflecht, G., McMullen, J., Veermans, K., Pongsakdi, N., Hannula-Sormunen, M.M., & Lehtinen, E. (2015). Developing adaptive number knowledge with the Number Navigation Game-based learning environment. In J. Torbeyns, E. Lehtinen, & J. Elen (Eds.), *Developing competencies in learners: From ascertaining to intervening* (pp. 45-61). New York, NY: Springer.

Canobi, K.H. (2009). Concept–procedure interactions in children's addition and subtraction. *Journal of Experimental Child Psychology*, *102,* 131–149. doi:10.1016/j.jecp.2008.07.008

Claro, M. (2007). *OECD Background paper for OECD-ENLACES expert meeting: Video games and education*. Retrieved from: http://www.oecd.org/edu/ceri/39548100.pdf

De Brabander, C., & Martens, R.L. (2014). Towards a unified theory of task-specific motivation. *Educational Research Review, 11*, 27-44. doi: 10.1016/j.edurev.2013.11.001

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin, 125*(6), 627-668.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self determination in human behavior*. New York, NY: Plenum Press.

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry, 11*(4), 227-268.

De Grove, F., Van Looy, J., Neys, J., & Jansz, J. (2011, October). Playing in school or at home? An exploration of the effects of context on educational game experience. In *Proceedings of the 7th European Conference on Management Leadership and Governance: ECGBL 2011* (p. 137). Academic Conferences Limited.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109–132.

Eccles, J.S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon (Series Ed.), N. Eisenberg (Vol. Ed.), *Handbook of child psychology* (5th ed., Vol. III, pp. 1017–1095). New York, NY: Wiley.

Fryer, L. K., Bovee, H. N., & Nakao, K. (2014). E-learning: Reasons students in language learning courses don't want to. *Computers & Education, 74*, 26-36. doi.org/10.1016/j.compedu.2014.01.008

Gauthier, A., Corrin, M., & Jenkinson, J. (2015). Exploring the influence of game design on learning and voluntary use in an online vascular anatomy study aid. *Computers & Education, 87*, 24-34. doi:10.1016/j.compedu.2015.03.017

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111-127. doi.org/10.1207/s15326985ep4102_4

Islas Sedano, C., Leendertz, V., Vinni, M., Sutinen, E., & Ellis, S. (2013). Hypercontextualized Learning Games: Fantasy, Motivation, and Engagement in Reality. *Simulation & Gaming*, *44*, 821-845. doi:10.1177/1046878113514807

Jenkins, H. (2011) Foreword. In deFreitas, S. & Maharg, P. (Eds.). (2011). *Digital Games and Learning.* (pp. xxiii-xxvi) New York, NY: Continuum

Klemetti, M., Taimisto, O., & Karppinen, P. (2009). The Attitudes of Finnish School Teachers Towards Commercial Educational Games. In Kankaanranta, M. & Neittaanmäki, P. (Eds.), *Design and Use of Serious Games* (pp. 97-106). Jyväskylä, Finland: Springer.

Kline, R. B. (2011). Convergence of structural equation modeling and multilevel modeling. In M. Williams & W. P. Vogt (Eds.), *Handbook of methodological innovation in social research methods* (pp. 562-589). London, England: Sage.

Lai, C., Wang, Q., & Lei, J. (2012). What factors predict undergraduate students' use of technology for learning? A case from Hong Kong. *Computers & Education, 59*(2), 569-579. doi:10.1016/j.compedu.2012.03.006

Lehtinen, E., Brezovszky, B., Rodríguez -Aflecht, G. Lehtinen, H. Hannula-Sormunen, M. M., McMullen, J. Pongsakdi, N., Veermans, K., & Jaakkola, T. (2015). Number Navigation Game (NNG): Design principles and game description. In J. Torbeyns, E. Lehtinen, & J. Elen (Eds.), *Developing competencies in learners: From ascertaining to intervening* (pp. 45-61). New York, NY: Springer.

McMullen, J., Brezovszky, B., Hannula-Sormunen, M. M., Rodríguez-Aflect, G., Pongsakdi, N., & Lehtinen, E. (2016). Adaptive number knowledge: Exploring the foundations of adaptivity with whole-number arithmetic. *Learning and Individual Differences*, *47*, 172-181. doi:10.1016/j.lindif.2016.02.007

Nacke, L. E., & Drachen, A. (2011). Towards a framework of player experience research. In *Proceedings of the Second International Workshop on Evaluating Player Experience in Games at FDG* (Vol. 11). Retrieved from: http://hciweb.usask.ca/uploads/230-NackeDrachenPXFramework

Nacke, L.E., & Lindley, C.A. (2009). Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience. *Loading*, *3*. Retrieved from: http://journals.sfu.ca/loading/index.php/loading/article/view/72/71

Nielson, K.B. (2011). Self-study with language-learning software in the workplace. *Language Learning & Technology*, *15*(3), 100-129. Retrieved from: http://llt.msu.edu/issues/october2011/nielson.pdf

Oksanen, K. (2013). Subjective experience and sociability in a collaborative serious game. *Simulation & Gaming*, *44*, 767-793. doi: 10.1177/1046878113513079

Peterson, R.A., & Kim, Yeolib. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology*, *98*(1), 194-198. doi: 10.1037/a0030767

Poels, K., De Kort, Y. A. W., & IJsselsteijn, W. A. (2008). *The game experience questionnaire: Development of a self-report measure to assess player experiences of digital game* (FUGA Deliverable D3.3. Technical Report). Eindhoven, The Netherlands: TU Eindhoven.

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of general psychology*, *14*(2), 154. doi:10.1037/a0019440

Rodríguez-Aflecht, G., Brezovszky, B., Pongsakdi, N., Jaakkola, T., Hannula-Sormunen, M.M., McMullen, J., & Lehtinen, E. (2015). Number Navigation Game (NNG): Experience and motivational effects. In J. Torbeyns, E. Lehtinen, & J. Elen (Eds.), *Describing and studying domain-specific serious games* (pp. 171-189). New York, NY: Springer.

Rodríguez-Aflecht, G., Jaakkola, T., McMullen, J., Pongsakdi, N., Hannula-Sormunen, M., Brezovszky, B., Lehtinen, E. (2016). *Predictors, outcomes, and trajectories of* *situational interest during a digital mathematics game*. Manuscript submitted for publication.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, *30,* 347–363. doi:10.1007/s11031-006-9051-8

Takatalo, J., Häkkinen, J., Kaistinen, J., & Nyman, G. (2010). User experience in digital games: Differences between laboratory and home. *Simulation & Gaming*, *42*(5), 656-673. doi: 10.1177/1046878110378353

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education*, *24*(3), 335–359. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract

Whitton, N. (2010). Game Engagement Theory and Adult Learning. *Simulation & Gaming*, *42*, 597-610. doi: 10.1177/1046878110378587

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, *6*(1), 49-78.

Wigfield, A., & Cambria, J. (2010). Expectancy-value theory: Retrospective and Prospective. In T. Urdan, & S.A. Karabenick (Eds.), *The decade ahead: Theoretical perspectives on motivation and achievement* (pp. 35-70). London, England: Emerald.

Wigfield, A., & Eccles, J.S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68–81. doi:10.1006/ceps.1999.1015

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In: A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 91–120). San Diego, CA: Academic Press.

Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson tests of achievement*. Itasca, IL: Riverside Publishing.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E.D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265. doi:10.1037/a0031311

Table 1

Pre-test means amongst the volunteer group

	Did not play						Played					
	Girls		Boys		Total		Girls		Boys		Total	
	n	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*
Arithmetic Fluency	72	51.40 (17.71)	94	51.44 (18.58)	166	51.42 (18.15)	173	51.58 (17.40)	164	54.79 (17.88)	337	53.14 (17.68)
ANK	72	0.35 (0.70)	94	0.75 (1.00)	166	0.58 (0.90)	173	0.73 (1.05)	164	1.13 (1.59)	337	0.93 (1.36)
Math interest	63	3.85 (1.13)	77	3.67 (0.95)	140	3.75 (1.04)	154	3.96 (0.86)	144	4.00 (0.85)	298	3.98 (0.85)
Gaming interest	62	3.76 (1.17)	77	4.24 (1.02)	139	4.02 (1.11)	154	3.89 (1.09)	144	4.41 (0.82)	298	4.14 (1.00)
Math self-efficacy	63	4.13 (0.75)	77	4.04 (0.80)	140	4.08 (0.77)	154	4.20 (0.69)	144	4.06 (0.71)	298	4.13 (0.70)
Gaming self-efficacy	63	3.68 (1.00)	77	3.95 (0.95)	140	3.83 (0.98)	154	3.79 (0.89)	144	4.07 (0.77)	298	3.93 (0.84)

Table 2

Game performance and game experience means by play context

	School						Volunteer (Played)					
	Girls		Boys		Total		Girls		Boys		Total	
	n	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*
Game Performance												
Total time played	145	341.46 (221.80)	145	335.20 (229.10)	290	338.33 (225.11)	96	220.34 (173.88)	101	254.05 (240.78)	197	237.63 (211.00)
Tasks completed	145	23.03 (24.34)	145	28.30 (25.99)	290	25.66 (25.27)	96	14.40 (12.76)	101	18.66 (16.03)	197	16.58 (14.65)
Tasks incomplete	145	112.67 (118.93)	145	145.92 (165.22)	290	129.29 (144.66)	96	40.14 (48.31)	101	40.38 (53.58)	197	40.26 (50.95)
Proportion of tasks completed with top score	145	0.13 (0.20)	145	0.22 (0.24)	290	0.17 (0.22)	96	0.11 (0.15)	101	0.20 (0.20)	197	0.15 (0.18)
Game Experiences												
Positive Value	134	3.99 (0.85)	138	3.97 (0.97)	272	3.98 (0.91)	92	3.81 (0.86)	97	3.79 (0.85)	189	3.80 (0.85)
Positive Affect	134	4.42 (0.74)	138	4.26 (0.88)	272	4.34 (0.82)	92	4.02 (0.98)	97	4.12 (0.81)	189	4.07 (0.89)
Competence	134	3.89 (0.76)	138	3.97 (0.85)	272	3.93 (0.80)	91	3.66 (0.78)	97	4.00 (0.74)	188	3.84 (0.78)

Table 3

Descriptive statistics by play context and gender when controlling for pre-test scores

	Girls		Boys		Total	
School	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*
Arithmetic Fluency	145	57.21 (17.65)	145	61.13 (19.17)	290	59.17 (18.50)
ANK	145	3.77 (2.73)	145	4.26 (3.04)	290	4.02 (2.90)
Math interest	124	4.03 (0.90)	128	3.92 (1.05)	252	3.97 (0.98)
Gaming interest	124	3.99 (1.06)	128	4.48 (0.82)	252	4.24 (0.98)
Math self-efficacy	124	4.16 (0.70)	128	4.10 (0.84)	252	4.13 (0.77)
Gaming self-efficacy	124	3.96 (0.73)	128	4.29 (0.78)	252	4.13 (0.77)
Volunteer (played)						
Arithmetic Fluency	96	63.34 (22.52)	101	64.76 (21.89)	197	64.07 (22.15)
ANK	96	4.78 (3.34)	101	5.30 (3.31)	197	4.05 (3.33)
Math interest	81	3.99 (0.85)	85	4.07 (0.82)	166	4.03 (0.83)
Gaming interest	81	3.90 (0.99)	84	4.54 (0.60)	166	4.23 (0.87)
Math self-efficacy	81	4.22 (0.65)	85	4.23 (0.54)	166	4.23 (0.60)
Gaming self-efficacy	81	3.81 (0.76)	85	4.22 (0.67)	166	4.02 (0.74)
Volunteer (did not play)						
Arithmetic Fluency	72	55.33 (18.67)	94	55.57 (21.29)	166	55.47 (20.13)
ANK	72	3.25 (2.80)	94	3.29 (2.87)	166	3.27 (2.83)
Math interest	63	3.66 (1.08)	77	3.72 (1.00)	140	3.69 (1.03)
Gaming interest	62	3.55 (1.25)	76	4.08 (1.05)	138	3.84 (1.17)
Math self-efficacy	63	4.08 (0.60)	77	3.91 (1.00)	140	3.99 (0.45)
Gaming self-efficacy	63	3.65 (1.01)	76	3.95 (0.94)	139	3.81 (0.98)

Note. Only students who completed at least five maps are included.