Number Navigation Game (NNG): Experience and Motivational Effects

Gabriela Rodríguez-Aflecht, Boglárka Brezovszky, Nonmanut Pongsakdi, Tomi Jaakkola, Minna M. Hannula-Sormunen, Jake McMullen, and Erno Lehtinen

Abstract Number Navigation Game-based learning environment (NNG) is a mathematical game-based learning environment designed to enhance students' adaptivity with arithmetic problem solving and to increase their motivation towards math. Fourth through sixth grade classrooms were randomly assigned into either an experimental group (students $n = 642$) which played NNG during a 10-week period or into a control group (students $n = 526$) which continued with a traditional textbook-based mathematics curriculum. The aims of the present study were to investigate the effects of the intervention on students' motivation and to explore how students' differing game experiences were related to changes in their motivation and, as an indicator of cognitive outcomes, their arithmetic fluency. Results indicate the intervention resulted in small decreases in the math motivation expectancy-values of interest, utility, and attainment value. Students had mixed game experiences which varied by gender and grade level. When looking at the role of these game experiences on post-test motivation and arithmetic fluency, corresponding pre-test values were the strongest predictive variables. Out of game experiences, only competence was a significant predictor of post-test motivational scores; however, no game experience variable was a predictor of post-test arithmetic fluency.

Keywords Game-based learning • GEQ • Motivation • Expectancy-value model • Arithmetic fluency

G. Rodríguez-Aflecht (✉) • B. Brezovszky • N. Pongsakdi • T. Jaakkola
M.M. Hannula-Sormunen • J. McMullen • E. Lehtinen
Department of Teacher Education, Centre for Learning Research, University of Turku,
Assistentinkatu 7, Turku 20014, Finland
e-mail: gabriela.rodriguez@utu.fi; bogbre@utu.fi; nopong@utu.fi; tomi.jaakkola@utu.fi;
mimarha@utu.fi; jake.mcmullen@utu.fi; erno.lehtinen@utu.fi

There has been an increasing interest in applying digital games in teaching mathematics. In many countries, teachers and educational authorities are concerned about a decreasing motivation to study mathematics, and educational games are considered to be a solution which would make mathematics education more fun and motivating. For example, a survey carried out amongst Finnish school teachers by Klemmetti and colleagues (2009) revealed that 99 % of respondents believed game-based learning environments would motivate students' learning. Despite this assumption, there is a lack of empirical studies providing evidence that game-based learning environments are able to significantly increase motivation towards learning (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012), at least not when compared to conventional instruction methods (Wouters, van Nimwegen, van Oostendorp, & van der Spek, 2013).

A prototype game-based learning environment, Number Navigation Game (NNG), was developed to enhance students' adaptivity with arithmetic (see NNG description by Lehtinen et al., in this volume). NNG is based on an integrated game design (Habgood & Ainsworth, 2011) in which the game mechanics offer extensive and situated practice through which flexible and adaptive arithmetic problem-solving skills can be strengthened. The tasks players have to solve during gameplay are increasingly demanding and substantially differ from the tasks used in regular mathematics education. In the prototype game-based learning environment used for this study there are still few external motivating elements so students' engagement derived from the gaming mechanics itself. The aim of the present study was to find out the effects of playing NNG on students' motivation towards mathematics as well as on their arithmetic fluency, and whether these effects are related to students' game experiences.

Expectancy-Value Model as a Comprehensive Measure for Studying Motivation

Motivation is a broad concept for which numerous and sometimes overlapping theories exist (De Brabander & Martens, 2014). Motivation is often described as a set of cognitive motives which, together with emotions (which are alternately considered as a subset of motivation or as something separate from it) influence behavior; these motives can include beliefs, values, expectancies, intentions, or goals (Wegge, 2001). In this study, motivation was looked at from the theoretical perspective of the expectancy-values model, more specifically, Eccles and colleagues' (2002) expectancy-values model. The expectancy-values model is particularly suitable for this study because its usefulness in predicting students' future performance, persistence, and task choice has been demonstrated in educational studies (Berger & Karabenick, 2011; Eccles & Wigfield, 2002; Wigfield & Cambria, 2010). In this model, the expectation to succeed and the value given to succeeding will determine a person's motivation to perform tasks (Wigfield, 1994). The development of expectancy-values is influenced by psychological, sociocultural, and contextual

factors, such as the feedback a child receives from parents, schools, and peers (Wigfield & Cambria, 2010). Expectancy-values are already distinct in young children (Wigfield & Eccles, 2000).

Expectancy refers to how well a person believes they will perform a task whereas values refer to a person's reasons for engaging in a task (Wigfield & Cambria, 2010). Expectancy is understood as a personal expectation or belief in one's own ability to succeed at a task. Value is task-subjective and composed of four aspects: intrinsic value, attainment value, utility value, and cost. Intrinsic value refers to interest, or how enjoyable a person finds the task, and is linked to Ryan and Deci's (2000) concept of intrinsic motivation (Wigfield & Eccles, 2000). Attainment value is determined by how important it is for a person's identity to perform well at the task. Utility value is defined as how useful the task is for a person's life. Cost, the least studied of these values, focuses on the perceived price a person feels they must pay in order to perform well on a task, both in terms of effort and time. Throughout the present study, motivation towards math is studied through the expectancy-value model, focusing on the variables of self-efficacy and interest, utility, attainment value, and cost.

Game Experience

Students' experiences during gaming depend on game features and on students' individual interpretations of these features. These interpretations may be related to age and gender. Interindividual differences in experiencing the game and gaming situations can mediate motivational and cognitive effects of educational games (Järvelä et al., 2000; Lowyck, Lehtinen, & Elen, 2004). Nevertheless, a common understanding of game experience has yet to be found (IJsselsteijn, de Kort, Poels, Jurgelionis, & Bellotti, 2007; Kiili, Lainema, De Freitas, & Arnab, 2014; Nacke & Drachen, 2011), though different frameworks and models have been proposed (see review by Nacke & Drachen, 2011). IJsselsteijn and colleagues (2007) argued this is because the field is relatively young, and the great variety of games makes it difficult to find a "one-size-fits-all" method to study all experiences elicited by games.

For the purposes of this study, game experience was considered from the framework developed by Poels and colleagues (2007). Their framework is composed of seven dimensions which are measured post-play through the Game Experience Questionnaire (GEQ). Although originally developed to measure users' experiences with commercial games for entertainment, the GEQ has also been used for game-based learning environments (De Grove, Van Looy, & Courtois, 2010; Gajadhar, Nap, De Kort, & IJsselsteijn, 2008; IJsselsteijn et al., 2007; Nacke, Stellmach, & Lindley, 2011; Oksanen, 2013; Poels, IJsselsteijn, de Kort, & Van Iersel, 2010). The seven dimensions of this framework are competence, challenge, flow, (sensory and imaginative) immersion, negative affect, positive affect, and tension.

The first four of the dimensions are highly interconnected. Flow is a key aspect of most existing game experience frameworks, and it is deeply related to the dimensions of challenge and competence. In Csikszentmihalyi's flow theory (1991), one of the characteristics of flow is the balance between challenge and ability. Challenge

must be neither too low, which would result in boredom, nor too high, which would result in frustration. In the flow theory, the balance of "challenge-skill" is considered to be one out of several components of flow and is crucial in educational games (Kiili et al., 2014). IJsselsteijn and colleagues (2007) characterize flow as a form of immersion which results from a player feeling there is balance between how challenging the game is and how competent they are. Jennett and colleagues (2008) distinguish between immersion and flow in that the latter leads to optimal experiences whereas the former does not necessarily do so. However, they admit these concepts may overlap. According to Ermi and Mäyrä (2005), immersion is seen as a broad concept in which (a) flow, or challenge-immersion, is separate from (b) sensory immersion and (c) imaginative immersion. Sensory immersion refers to audiovisual characteristics of the game, such as graphics or sound, whereas imaginative immersion refers for instance to the game's narration or characters. Whereas Ermi and Mäyrä (2005) look at these three types of immersion independently from one another, in the framework of Poels and colleagues (2007) used for the present study both sensory and imaginative immersion together conform one sole dimension, which refers to the absorption a player might feel towards game features such as story, game world, graphics, or sound.

The other dimensions included in the GEQ—positive affect, negative affect, and tension—focus on post-play affective states which indicate how enjoyable the game experience was. Nacke and Lindley (2009) argue that affect is an essential part of game experience, as it influences the cognitive decisions players take while playing. Based also on physiological responses, they report a correlation between flow and positive affect. Complementing this framework of game experience, our study included an additional dimension of "positive value," which measures students' belief that the game is helpful to them. Whitton (2010b) has argued that in order to benefit from game-based learning, users must first believe in the positive value of these games. While she was speaking of adult learners, it is here considered equally relevant for children who are playing games for educational purposes in school.

It is necessary to understand the types of game experiences students have when playing NNG, as it is clear that positive game experiences foster engagement while negative ones hinder it. An unengaged student might stop playing or only continue reluctantly (Oksanen, 2013), which might have repercussions on the effectiveness of NNG in enhancing students' motivation expectancy-values and arithmetic skills. Thus, examining the effect of students' game experiences on NNG's motivation expectancy-values and arithmetic fluency was part of the study's main aim.

Arithmetic Skills

In testing motivational effects of game-based learning, it is also important to compare them with possible cognitive gains. The objective of NNG is to enhance students' adaptivity with arithmetic problem solving and adaptive number knowledge (see Brezovszky et al., in this volume; Lehtinen et al., in this volume). However, in

the present study about the motivational effects of the game, arithmetic fluency measures are used as an indicator of arithmetic skill development because the adaptive number knowledge results are not yet available. Arithmetic fluency refers to the quick and accurate retrieval of basic number facts and combinations and is a requisite for further conceptual and procedural development (Baroody, Bajwa, & Eiland, 2009; Canobi, 2009). While enhancing arithmetic fluency is not the main goal of NNG, given its relation to adaptive number knowledge (McMullen, Brezovszky, Rodríguez-Aflecht, Pongsakdi, & Lehtinen, 2015), it is used in the present study as a proximal indicator of the game's mathematical impact.

Research Questions

This study focuses on the following research questions:

1. What is the effect of the playing NNG on students' motivation towards mathematics, as framed by the expectancy-value model?
   Prior evidence on the motivational effects of game-based learning environments is mixed (Connolly et al., 2012). Thus, we assume that there is no strong overall development in math motivation, particularly when taking into account that the game-based learning environment used is a prototype that includes few externally motivating elements typical for commercial games. However, as gaming itself is already different from regular mathematics education, we assume that this, together with the nonstandard tasks to be solved within NNG, will produce a novelty effect and result in a slight increase in interest in mathematics.
2. What are students' experiences with the game and how do these experiences differ by gender and grade level?
   Considering that NNG is a prototype still lacking many externally motivating features common in the commercial games children are accustomed to (sound, advanced graphics, etc.), we expect students to rate their experiences close to the scales' midpoints. Gender differences have been much examined, often focusing on frequency of play, types of games preferred, and self-efficacy beliefs (Bourgonjon, Valcke, Soetaert, & Schellens, 2010; Carr, 2005; Jenson & de Castell, 2010). While NNG is meant to be gender-neutral, based on findings of earlier studies (e.g., Lucas & Sherry, 2004) we assume boys might report more positive game experiences. As the same version with the same difficulty level of NNG was used for children in different grade levels and different stages of their arithmetical development, we expect differences by grade level in game experiences, particularly those of challenge and competence.
3. How are students' game experiences with NNG related to changes in (a) motivation expectancy-values and (b) arithmetic fluency?
   Empirical studies directly analyzing the relationship between game experience and motivational effects are still rare. Based on earlier studies (e.g., the seminal work of Lepper & Malone, 1987) Paras and Bizzocchi (2005) concluded that if

games foster play and challenge, which produces a state of flow, then gameplay can result in increases in motivation, which supports the learning process. Students who have more positive experiences with the game are probably more engaged, and this might result in positive motivational and cognitive consequences. There was no systematic scaffolding or teacher support in this experiment and the "energy maps" (see NNG description by Lehtinen and colleagues, in this volume) are demanding, which can have diverse effects on students' mathematics self-efficacy beliefs. Students who did not experience competence during gameplay might report lower mathematics self-efficacy at post-test, whereas those who felt competent during gameplay might report stronger mathematics self-efficacy.

Method

Participants

In this study, 1168 students from 61 fourth through sixth grade classrooms spread across four cities in Finland participated. Participation was voluntary both for teachers and students, and informed consent was acquired in writing from the parents of all participants. Ethical guidelines of Turku University were followed. From the total, 546 participants were female, 620 were male, and there was missing data on the gender of two participants. As for grade level, 135 participants were fourth graders, 606 were fifth graders, and 427 were sixth graders. The mean ages for the fourth, fifth, and sixth grade participants were 10 years and 2 months, 11 years and 2 months, and 12 years and 3 months, respectively. Classes were randomly assigned into control and experimental groups, with 642 participants belonging to the experimental group and 526 to the control group.

Procedure

During the spring term 2014, the experimental group played NNG for a 10-week period as part of their regular math classes and curriculum, while the control group continued only with their regular textbook-based mathematics curriculum. Afterwards, conditions were reversed. While the present study only encompasses this first phase of the experiment, it is relevant to mention the reversal of conditions not only because it would have been ethically questionable to deny participants in the control group the chance to play, but also because the control group's knowledge about the upcoming play sessions could have an impact on some post-test measures. It was asked that students play for at least 10 h. Teachers were invited for a training session in which they were informed about NNG's learning aims and play mechanics. As part of their training, teachers were told sessions needed to last at

least 30 min in order to give their students enough time to make significant progress in the game. Nevertheless, teachers were free to decide how long play sessions would extend, how to space these sessions throughout the intervention, what kind of support they would provide their students, and whether students would play individually or in pairs. There were no instructions for teacher support during the gaming processes. In case students played in pairs, teachers chose the criteria under which pairs would be formed.

Measures

Data used for this study was collected by questionnaires and math tests completed by students before the 10-week intervention and immediately after the 10 weeks. Students in the experimental group received a copy of the game immediately upon completing the pre-test, and class teachers were free to schedule game sessions as they saw fit. Game log data was collected upon completing the post-test, but it is not analyzed in the current study. Both pre- and post-tests were rigorously timed and structured, and were imparted by trained testers following standardized procedures. Both pre- and post-questionnaires were filled out by students during regular class time under the guidance of their teachers. The pre-questionnaire was identical for all participants, containing demographic items and items measuring their math motivation expectancy-values, while the experimental group's post-questionnaire included additional items concerning their game experiences playing NNG.

Math Motivation Expectancy-Values: Fourteen items measuring math expectancy-values were completed before and after the intervention by all participants. The test was modified on the basis of the motivation scale used by Berger and Karabenick (2011), with items being translated into Finnish and adapted to the ages of respondents. Three items were used to measure interest (for example, "I like math"). Three items measured utility ("Math is useful for me in everyday life"). Three items measured attainment value ("It is important to me to be a student who is good at math"). Two items were used to measure cost ("I believe that success in math requires that I give up other activities that I enjoy"). Three items were used to measure self-efficacy, ("I am certain I can do difficult math tasks"). Participants responded to each item using a 5-point Likert scale ranging from 1 (*completely disagree*), 2 (*disagree*), 3 (*neutral*), 4 (*agree*), to 5 (*completely agree*). These items were studied through principal component analysis with varimax rotation. Five separate factors (interest, utility, attainment value, self-efficacy, and cost) were found, upholding the 5-factor model developed by Eccles and Wigfield (2002). The explained variance of the model was 75.90 %. Data was adequate for factor analysis with a 0.90 Kaiser–Meyer–Olkin Measure, and Barlett's test of sphericity showed a significance of $p < 0.001$. All but one factor were shown to have good internal consistency and to be reliable across the two tests. At pre-test: interest Cronbach's $\alpha = .91$, utility: $\alpha = .80$, attainment value: $\alpha = .82$, cost: $\alpha = .50$, and self-efficacy $\alpha = .81$. At post-test: interest Cronbach's $\alpha = .91$, utility: $\alpha = .79$, attainment value: $\alpha = .83$, cost: $\alpha = .58$, and

self-efficacy $\alpha = .81$. The poor reliability of cost compared to the other measures can be due to this dimension only having two items—consequently, the expectancy-value of cost was not used for any further analyses. Correlations of the variables at pre-test can be found in the Appendix.

Game Experience: Only participants in the experimental group were asked to fill the Game Experience Questionnaire (GEQ) after the intervention. The Finnish translations used by Oksanen (2013) were used, although the questionnaire was further modified by removing 15 of the 42 items and changing some of the phrasings to better suit our game and the age of our participants. Each item 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. The factor structure of the 31 items of GEQ was studied through principal component analysis with varimax rotation. Data was adequate for factor analysis with a 0.95 Kaiser–Meyer–Olkin Measure and Barlett's test of sphericity showed a significance of $p < .001$. Seven separate factors were found and used as basis for the subscales. The explained variance of the model was 69.60 %. The reliability of subscales was as follows: Challenge, (e.g., "I thought playing this game was hard"), $\alpha = .66$. Competence, ("I was good at playing"), $\alpha = .81$. Flow, ("I forgot everything around me when I played"), $\alpha = .79$. Immersion, ("I felt imaginative when I played"), $\alpha = .77$. Negative affect, ("I thought playing was boring"), $\alpha = .78$. Positive Affect, ("I thought playing was fun"), $\alpha = .92$. Positive value, ("This game helped me learn math"), $\alpha = .82$. Tension, ("I felt irritable when I played"), $\alpha = .77$. The reliability for challenge is low, which could be due to the removal of three items, although Oksanen (2013) also reported similar results. However, according to Clark and Watson (1995), this reliability is within the limits which can be used. As suggested by the theory, there are high correlations between flow, challenge, and competence (Appendix).

Arithmetic Fluency: Students' fluency in solving basic arithmetic tasks is used in this study as a basic measure of cognitive outcomes. The test was adapted from the Mathematical Fluency test of the Woodcock-Johnson Tests of Achievement (Woodcock, McGrew, & Mather, 2001), in which students have three minutes to answer as many simple arithmetic problems as they can. The minimum score was 0 and the maximum score 160. Changes made to the original test include replacing the multiplication symbol with its equivalent in Finland, that is, · was used instead of x.

Results

Results are organized into three subsections. The first subsection presents the effects of the intervention on math motivation expectancy-values. The second subsection describes the experimental group's game experiences with NNG as well as gender and grade level differences in these experiences. Finally, the third subsection explores how the experimental group's game experiences with NNG related to changes in their (a) math motivation expectancy-values and (b) arithmetic fluency.

The Effects of Intervention on Math Motivation and Arithmetic Skills

Descriptive statistics on pre- and post-test math expectancy-values subscales are presented in Table 1. A repeated measures ANOVA analyzing the effects of time (pre- and post-test) and condition (experimental or control) on math motivation expectancy-values measures was conducted. Overall, there was no main effect of time on interest or utility but there was a slight decrease in attainment value, $F(1, 1166) = 8.80$, $p = .003$, $\eta p^2 = 0.01$ and in self-efficacy, $F(1, 1166) = 5.14$, $p = .02$, $\eta p^2 = 0.004$. There was a small interaction effect of time and condition on interest, $F(1, 1166) = 13.21$, $p = .000$, $\eta p^2 = 0.011$, on utility, $F(1, 1166) = 7.15$, $p = .008$, $\eta p^2 = 0.01$, and on attainment value, $F(1, 1166) = 7.51$, $p = .006$, $\eta p^2 = 0.01$, showing a small decrease in these motivational aspects amongst the experimental group when compared with the control group.

There was no difference in the arithmetic fluency scores in the pre-test (experimental group $M = 70.56$; control group $M = 70.23$). Arithmetic fluency scores increased significantly in the post-test (experimental group $M = 80.15$; control group $M = 77.87$), $F(1,1166) = 589.52$, $p = .000$. There was a significant interaction effect of time and condition, $F(1,1166) = 5.99$, $p = .015$, $\eta p^2 = 0.01$, showing a small positive intervention effect on arithmetic fluency.

Game Experiences with NNG

Mean scores (Table 2) for the different dimensions of the GEQ averaged between 2 and 3. An independent samples *t*-test was run to determine whether there were differences in game experiences between girls and boys. Results showed significant differences between girls' and boys' experiences of challenge and competence while playing NNG; however, effect sizes were small (Table 2). Girls rated the game as more challenging than boys did, which means they were more likely to find the game difficult and that they had to make an effort when playing. Boys had higher competence scores than girls did, which means they were more likely to report feeling good, successful, and skillful while playing.

A one-way ANOVA was carried out to look at grade level differences in game experiences (Table 3). There were significant differences between the class levels in Flow, Immersion, Negative Affect, Positive Value, and Tension. Post hoc comparisons using Bonferroni correction indicated that fifth graders' experiences of immersion was significantly higher than sixth graders' (Mean difference = 0.18, $p = .033$). This means that fifth graders reported higher feelings of being imaginative, liking the story, and being able to explore.

The scores of students' belief in the positive value of the game also significantly differed between fourth graders and sixth graders, with fourth graders seeing more benefit in playing NNG (Mean difference = 0.30, $p = .049$) and feeling the game

Table 1 Descriptive statistics on motivation expectancy-values

Variable	All participants (N = 1168)					Experimental (n = 642)					Control (n = 526)				
	M	SD	Range	Skew.	Kurt.	M	SD	Range	Skew.	Kurt.	M	SD	Range	Skew.	Kurt.
Pre-test interest	3.13	0.96	1–5	−0.74	−0.15	3.14	0.94	1–5	−0.03	−0.07	3.11	0.98	1–5	−0.12	−0.25
Post-test interest	3.11	0.93	1–5	−0.09	−0.03	3.05	0.96	1–5	−0.04	−0.22	3.18	0.88	1–5	−0.13	0.28
Pre-test utility	4.19	0.73	1–5	−1.31	2.51	4.18	0.70	1–5	−1.13	2.05	4.21	0.78	1–5	−1.47	2.84
Post-test utility	4.20	0.69	1–5	−1.07	1.49	4.14	0.70	1–5	−0.96	1.20	4.28	0.67	1–5	−1.25	2.099
Pre-test attainment value	3.50	0.83	1–5	−0.29	−0.00	3.53	0.80	1–5	−0.29	0.12	3.46	0.87	1–5	−0.27	−0.15
Post-test attainment value	3.42	0.83	1–5	−0.23	−0.33	3.40	0.84	1–5	−0.24	0.37	3.45	0.82	1–5	−0.22	0.29
Pre-test self-efficacy	3.70	0.80	1–5	−0.58	0.51	3.68	0.78	1–5	−0.53	0.52	3.72	0.82	1–5	−0.65	0.53
Post-test self-efficacy	3.65	0.79	1–5	−0.53	0.65	3.63	0.82	1–5	−0.49	0.49	3.67	0.76	1–5	−0.57	0.88

Table 2 Game experiences of the experimental group, by gender, and gender differences

Variable	All experimental students (n=642)					Girls (n=299)		Boys (n=343)		T-test of gender effects		
	M	SD	Range	Skew.	Kurt.	M	SD	M	SD	t(640)	p	Cohen's d
Challenge	2.34	0.82	1–5	0.37	0.05	2.43	0.79	2.25	0.85	2.75	.006	0.22
Competence	3.06	0.90	1–5	−0.07	−0.14	2.88	0.85	3.22	0.92	−4.78	.000	−0.38
Flow	2.07	0.79	1–5	0.71	0.38	2.08	0.74	2.06	0.83	0.42	.68	0.03
Immersion	2.04	0.87	1–5	0.75	0.19	2.03	0.84	2.05	0.91	−0.32	.75	−0.03
Negative affect	3.03	0.96	1–5	0.25	−0.53	3.00	0.94	3.05	0.99	−0.55	.58	−0.04
Positive affect	2.31	1.00	1–5	0.56	−0.27	2.27	0.93	2.35	1.06	−0.91	.36	−0.07
Positive value	2.31	0.91	1–5	0.50	−0.02	2.29	0.84	2.33	0.96	−0.63	.53	−0.05
Tension	2.52	1.20	1–5	0.59	−0.55	2.50	1.16	2.54	1.23	−0.45	.65	−0.04

Table 3 Game experiences of experimental group by grade level

| | Groups | | | | | | | |
| | Fourth grade (n=63) | | Fifth grade (n=309) | | Sixth grade (n=270) | | | |
Variable	M	SD	M	SD	M	SD	F(2, 639)	p
Challenge	2.37	0.87	2.34	0.81	2.33	0.84	0.05	.949
Competence	3.05	0.94	3.15	0.90	2.97	0.90	2.87	.057
Flow	2.22	0.90	2.11	0.79	1.98	0.75	3.22	.041
Immersion	2.12	1.03	2.12	0.88	1.94	0.81	3.45	.031
Negative affect	2.67	0.94	3.03	0.92	3.11	1.00	5.25	.005
Positive affect	2.48	1.12	2.36	1.00	2.21	0.97	2.60	.075
Positive value	2.50	1.03	2.37	0.91	2.20	0.86	4.33	.014
Tension	2.16	1.15	2.62	1.19	2.48	1.21	4.25	.015

helped them learn math. Scores for negative affect significantly differed both between fourth graders, with fifth graders having a higher score in negative affect (Mean difference=0.36, p=.021) and between fourth and sixth graders, with sixth graders having a higher score in negative affect (Mean difference=0.44, p=.004). That is, the youngest students were less likely to report feelings of boredom while playing. However, fourth graders were more likely to feel annoyed or irritable while playing than fifth graders, with the scores for tension significantly differing between fifth and fourth graders (Mean difference=0.46, p=.014). There were no significant differences between grade levels in the experiences of challenge, competence, or positive affect.

Effects of Game Experiences on Math Motivation Expectancy-Values and Arithmetic Fluency

In order to determine how the experimental group's game experiences related to their post-test scores for (a) math motivation expectancy-values and (b) arithmetic fluency, multiple linear regression analyses were conducted. As dependent variables, the post-test sum scores of the math motivation expectancy-values of interest, utility, attainment value, and self-efficacy were used, as well as arithmetic fluency post-test scores. For each dependent variable, its corresponding variable at pre-test as well as all game experience variables and arithmetic fluency were included as independent variables. Table 4 provides the results of these analyses.

Models were sufficient for explaining the post-test (Total R^2s > .33). For all the math motivation expectancy-values of interest, utility, attainment value, and self-efficacy, the corresponding pre-test variables proved to be the strongest predictive variables. In all these cases, the only other significant predictor of post-test scores was the game experience of competence (βs > .12). In the case of arithmetic fluency, $F(9, 997)=236.82$, $p<.001$, $R^2=.68$, only pre-test arithmetic fluency was a predictor ($\beta=.82$, $p<.001$), with game experiences not playing any role on post-test results.

Table 4 Regression analyses on post-test math motivation expectancy-values

	β			
	Interest	Utility	Attainment	Self-efficacy
Pre-test value	.58***	.48***	.48***	.47***
Challenge	−.01	.04	.04	−.03
Competence	.15***	.12**	.21***	.33***
Flow	.00	−.03	.06	−.02
Immersion	.01	.00	−.05	−.09
Negative affect	−.05	−.02	−.06	−.04
Positive affect	.04	.03	.01	−.01
Positive value	.05	.09	.06	.06
Tension	−.04	−.06	.00	−.00
R^2	.52	.33	.40	.45
F	113.78***	53.01***	72.04***	86.34***

Note: **$p<.01$, ***$p<.001$

Discussion

The aim of the present study was to investigate the effects of gameplay on students' math motivation as measured by using the expectancy-value model, and to explore how students' differing game experiences were related to the changes in their math motivation and arithmetic fluency.

There was a slight decrease in two of the motivation dimensions, attainment value and self-efficacy, for all participants in both experimental and control groups, which is in line with previous research reporting a general decrease of expectancy-values throughout the school term (Berger & Karabenick, 2011; Wigfield & Cambria, 2010). The control group's interest and utility slightly increased from pre-test to post-test. This could possibly be explained by the fact that participants in the control group anticipated the reversal of conditions and their interest and beliefs in the utility of math were sparked by the upcoming NNG intervention, although data is not sufficient to determine this. For the most part, playing NNG did not have a large impact on students' math motivation expectancy-values. Compared to the control group, the experimental group showed a slight decrease in three of the dimensions of the expectancy-value model (interest, utility, and attainment value), but all these effects were quite small. In spite of this, there was a slightly positive intervention effect on arithmetic fluency.

When focusing on the game experiences of the experimental group, results show that participants rated their game experiences in a predominantly negative or only slightly positive way. The game experience with the lowest mean score was immersion while the one with the highest mean score was competence. Participants reported higher feelings of competence than of challenge. Thus, the challenge-skill balance necessary to produce the game experience of flow, which leads to positive affect, was not reached. The high scores for competence suggest that the game might have been perceived as too simple, although it is unclear whether this was due to the game's form (gameplay) or content (arithmetic strategies needed) or both.

Results support Whitton's (2010a) claim that it cannot be assumed that a game dynamic will automatically make something interesting to learners who have no interest in the subject itself. More studies are needed to explore the impact of background factors, such as interest in math and in games, on game experiences. It is also important to further study to what extent improving the design of the game can foster positive game experiences.

There were some differences in students' game experiences depending on gender and grade level. Girls had higher mean scores for the dimension of challenge while boys had higher mean scores for the dimension of competence. However, it is not clear whether this is due to different perceptions of math skills or gaming skills between genders. Previous research suggests that boys have higher competence beliefs than girls for math even when controlling for skill level, although the gap in gender differences narrows with age (Wigfield & Eccles, 2002). Altogether, competence and challenge significantly differ by gender but not by grade level. Other dimensions such as immersion, positive value, negative affect, and tension differ by grade level, with the mean scores of fourth graders showing the most substantial differences when compared with students from other grade levels. It seems that younger students overall had more positive game experiences than older students. Here, it is important to note the much smaller number of participants in the fourth grade. Finally, the dimension of positive affect is the only dimension of game experience that shows no significant differences by gender or grade level.

Pre-test math motivation expectancy-values played a larger role in predicting post-test math motivation expectancy values than the different dimensions of game experiences, suggesting that there was little change in expectancy-values due to gameplay. Amongst game experiences, competence was the strongest predictor of post-test expectancy-values in all cases. However, this raises some questions about the validity of the dimension of competence as currently measured by the GEQ. It is not clear whether students interpreted the competence items of the GEQ as referring to their math competence or their gaming competence or both. If the former, then it seems that the items for competence and self-efficacy might have overlapped. However, as the correlation between the variables is only moderate ($r = .44$), this needs to be further explored. Similarly, game experiences did not predict post-test arithmetic fluency, suggesting that game experiences may not play a role in mathematical learning outcomes.

Implications

It seems that NNG's mechanics as such were not motivating for the majority of students, even though gameplay resulted in improvement in mathematical skills. Math motivation expectancy-values remained mostly stable, and although the experimental group showed a slight decrease in interest, utility, attainment value, and self-efficacy, these effect sizes were very small. The changes in expectancy-values of the experimental group could partly be predicted by their experience of competence during gameplay, which in turn differs by gender. As the basic mechanics of the game seems to work and resulted in improved mathematical skills,

a next step would be to analyze whether new features of later versions of the game will lead to meaningful improvements in gaming experiences. It is encouraging that regardless of the quality of gaming experience, NNG is still effective in improving arithmetic fluency.

Limitations and Future Directions

Conditions could somewhat vary between classrooms, and there is no detailed information what the role of the teacher was in, for example, debriefing, feedback, support activities, or reflection. However, giving teachers the freedom to use the game as they saw fit allowed for testing the effectiveness of the game in the most natural school settings possible. The detailed log data which will be analyzed in the future will give some information of these differences, but in future studies it is important to collect detailed data of teachers' roles during gameplay.

A major limitation of this study is its dependence on subjective and self-reported data. Informal feedback from teachers paints a different picture of students' experiences, as many teachers claimed their students were very engaged while playing and enjoyed the experience. This will be remedied in the future with the addition of a feedback feature within the game itself, which will make it possible for players to give feedback on their affective states upon completing a map, in a situated way that does not disrupt flow.

When looking at game-based learning environments, it's important to acknowledge their oxymoronic nature (Abt, 1987). Jenkins (2011) brings up the contradictory relationship of playing—a "freely chosen irresponsibility"—and learning—an "assigned responsibility." Along similar lines, it has been argued that having teachers decide what games will be played, for how long, and under which circumstances, will have repercussions on the level of control felt by students and consequentially on their motivation (Wouters et al., 2013). Different results may be achieved when play is free and voluntary, as opposed to a formal and prescribed school activity (Islas Sedano, Leendertz, Vinni, Sutinen, & Ellis, 2013), as it has been reported that playing in a different context, such as at home, increases players' enjoyment, identification, and learning experiences (De Grove, Van Looy, Neys, & Jansz, 2012). An important next step could be to study the effects on motivation and core gaming experiences of having students play voluntarily at their homes.

Appendix

Correlations of pre-test expectancy-values and the GEQ

	1	2	3	4	5	6	7	8	9	10	11	12
1. Pre-test interest	–											
2. Pre-test utility	.46**	–										
3. Pre-test attainment value	.60**	.57**	–									
4. Pre-test self-efficacy	.56**	.42**	.56**	–								
5. Challenge	-.04	-.01	-.11*	-.21**	–							
6. Competence	.29**	.10*	.32**	.41**	-.00	–						
7. Flow	.20**	.01	.18**	.07	.49**	.38**	–					
8. Immersion	.20**	.01	.17**	.12**	.45**	.47**	.73**	–				
9. Negative affect	-.22**	-.01	-.07	-.01	-.28**	-.25**	-.57**	-.55**	–			
10. Positive affect	.26**	.02	.18**	.14**	.44**	.53**	.77**	.79**	-.64**	–		
11. Positive value	.19**	.04	.08	.06	.54**	.44**	.66**	.72**	-.51**	.73**	–	
12. Tension	-.13**	-.07	-.07	.00	-.18**	-.21**	-.36**	-.39**	.69**	-.52**	-.39**	–

*$p<.05$, **$p<.01$

References

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

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.

Berger, J., & Karabenick, S. A. (2011). Motivation and students' use of learning strategies: Evidence of unidirectional effects in mathematics classrooms. *Learning and Instruction, 21*, 416–428. doi:10.1016/j.learninstruc.2010.06.002.

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

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. xx–xx). 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.

Carr, D. (2005). Contexts, gaming pleasures, and gendered preferences. *Simulation & Gaming, 36*, 464–482. doi:10.1177/1046878105282160.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of the empirical evidence on computer games and serious games. *Computers & Education, 59*, 661–686. doi:10.1016/j.compedu.2012.03.004.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience*. New York, NY: Harper Perennial.

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.

De Grove, F., Van Looy, J., Neys, J., & Jansz, J. (2012). Playing in school or at home? An exploration of the effects of context on educational game experience. *Electronic Journal of e-Learning, 10*, 83–208. Retrieved from http://files.eric.ed.gov/fulltext/EJ985422.pdf.

De Grove, F., Van Looy, J., & Courtois, C. (2010). Towards a serious game experience model: Validation, extension and adaptation of the GEQ for use in an educational context. In L. Calvi, K. Nuijten, & H. Bouwknegt (Eds.), *Playability and player experience* (Vol. 10, pp. 47–61). Breda, The Netherlands: Breda University of Applied Sciences.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132. Retrieved from http://www.rcgd.isr.umich.edu/garp/articles/eccles02c.

Ermi, L., & Mäyrä, F. (2005). Fundamental components of the gameplay experience: Analysing immersion. In S. de Castell & J. Jenson (Eds.), *Changing views: Worlds in play: Selected papers of the 2005 digital games research Association's second international conference* (pp. 15–27). Vancouver, British Columbia, Canada: Digital Interactive Games Research Association.

Gajadhar, B., Nap, H., De Kort, Y., & IJsselsteijn, W. (2008). Shared fun is doubled fun: Player enjoyment as a function of social setting. In P. Markopoulos, B. Ruyter, W. IJsselsteijn, & D. Rowland (Eds.), *Proceedings of the 2nd International Conference on Fun and Games* (pp. 106–117). Berlin, Germany: Springer.

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences, 20*, 169–206. doi:10.1080/10508406.2010.508029.

IJsselsteijn, W., de Kort, Y., Poels, K., Jurgelionis, A., & Bellotti, F. (2007). Characterising and measuring user experiences in digital games. *Proceedings from the 2007 International*

Conference on Advances in Computer Entertainment Technology, Workshop 'Methods for Evaluating Games—How to measure Usability and User Experience in Games'. ACM Digital Library. Retrieved from http://alexandria.tue.nl/openaccess/Metis215134

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.

Järvelä, S., Lehtinen, E., & Salonen, P. (2000). Socio-emotional orientation as a mediating variable in teaching learning interaction: Implications for instructional design. *Scandinavian Journal of Educational Research, 44*, 293–306. http://dx.doi.org/10.1080/713696677.

Jenkins, H. (2011). Foreword. In S. de Freitas & P. Maharg (Eds.), *Digital games and learning* (pp. xxiii–xxvi). New York, NY: Continuum.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, *66*, 641–661. doi:10.1016/j.ijhcs.2008.04.004.

Jenson, J., & de Castell, S. (2010). Gender, simulation, and gaming: Research review and redirections. *Simulation & Gaming, 41*(1), 51–71. doi:10.1177/1046878109353473.

Kiili, K., Lainema, T., De Freitas, S., & Arnab, S. (2014). Flow framework for analyzing the quality of educational games. *Entertainment Computing, 5*, 367–377. doi:10.1016/j.entcom.2014.08.002.

Klemetti, M., Taimisto, O., & Karppinen, P. (2009). The attitudes of Finnish school teachers towards commercial educational games. In M. Kankaanranta & P. Neittaanmäki (Eds.), *Design and use of serious games* (pp. 97–106). Jyväskylä, Finland: Springer.

Lehtinen, E., Brezovszky, B., Rodríguez-Aflecht, G., Lehtinen, H., Hannula-Sormunen, M., M., McMullen, J., … 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. xx–xx). New York, NY: Springer.

Lepper, M. R., & Malone, T. W. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction* (Cognitive and affective process analysis, Vol. 3, pp. 255–286). Hillsdale, NJ: Erlbaum.

Lowyck, J., Lehtinen, E., & Elen, J. (2004). Editorial: Students' perspectives on learning environments. *International Journal of Educational Research, 41*, 401–406. doi:10.1016/j.ijer.2005.08.008.

Lucas, K., & Sherry, J. L. (2004). Sex differences in video game play: A communication-based explanation. *Communication Research, 31*, 499–523. doi:10.1177/0093650204267930.

McMullen, J., Brezovszky, B., Rodríguez-Aflecht, G., Pongsakdi, N. & Lehtinen, E. (2015). *Adaptive Number Knowledge: Exploring the foundations of adaptivity with whole-number arithmetic.* Manuscript submitted for publication.

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-Nacke DrachenPXFramework

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.

Nacke, L. E., Stellmach, S., & Lindley, C. A. (2011). Electroencephalographic assessment of player experience: A pilot study in affective ludology. *Simulation & Gaming, 42*, 632–655. doi:10.1177/1046878110378140.

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

Paras, B., & Bizzocchi, J. (2005). Game, motivation, and effective learning: An integrated model for educational game design. *Proceedings of DiGRA 2005 Conference: Changing Views: Worlds in Play*. Retrieved from http://www.digra.org/wp-content/uploads/digital-library/06276.18065.pdf

Pavlas, D., Heyne, K., Bedwell, W., Lazzara, E. L., & Salas, E. (2010). Game-based learning: The impact of flow state and videogame self-efficacy. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54*, 2398–2402. doi:10.1177/154193121005402808.

Poels, K., de Kort, Y. A. W., & IJsselsteijn, W. A. (2007). It is always a lot of fun!—Exploring dimensions of digital game experience using focus group methodology. In B. Kapralos, M. Katchabaw, & J. Rajnovich (Eds.) *Proceedings of the 2007 Conference on Future Play* (pp. 83–89). ACM Press. Retrieved from http://alexandria.tue.nl/openaccess/Metis215146

Poels, K., IJsselsteijn, W., de Kort, Y., & Van Iersel, B. (2010). Digital games, the aftermath. Qualitative insights into post game experiences. In R. Bernhaupt (Ed.), *Evaluating user experiences in games*. Berlin, Germany: Springer.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54–67.

Wegge, J. (2001). Motivation, information processing, and performance: Effects of goal setting on basic cognitive processes. In A. Efklides, J. Kuhl, & R. M. Sorrentino (Eds.), *Trends and prospects in motivation research* (pp. 269–296). Dordrecht, The Netherlands: Kluwer. doi:10.1007/0-306-47676-2.

Whitton, N. (2010a). Game engagement theory and adult learning. *Simulation & Gaming, 42*, 597–610. doi:10.1177/1046878110378587.

Whitton, N. (2010b). *Learning with digital games: A practical guide to engaging students in higher education*. New York, NY: Routledge.

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.), *Advances in motivation and achievement. The next decade of research in 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). Academic. Retrieved from http://www.rcgd.isr.umich.edu/garp/articles/eccles02o

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

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*, 249–265. doi:10.1037/a0031311.