Towards Game-Based Formative Assessment

Towards Game-Based Formative Assessment

Kristian Kiili & Antero Lindstedt
Tampere University of Technology
Finland
kristian.kiili@tut.fi
antero.lindstedt@tut.fi

Abstract: This study sought to improve students' conceptual rational number knowledge by letting them play a math game with a tablet device. Additionally, the study evaluated the usefulness of in-game metrics and game features that support formative assessment. In the 3-hour intervention study, 52 Finnish fourth and sixth grade students studied conceptual rational number knowledge with a number line based game that included several scaffolding features founded on their playing behaviour. Students benefited significantly from the intervention (d = .64). This finding is consistent with recent numerical cognition research that has suggested that rational numbers should be trained with number lines. This research also provides evidence about the usefulness and fairness of game-based assessment. In general, the results indicated that game-based formative assessment can be a fair assessment approach and most of the students seems to be willing to adopt game-based learning solutions in schools.

Introduction

The use of game-based tasks has proven to be an effective approach for learning in the domain of mathematics education, in particular, to engage and motivate users (e.g. Moeller et al., 2015; Kiili et al, 2015; Kadosh et al., 2013; Devlin, 2011). The intrinsic appeal of games can be explained by their ability to satisfy basic psychological needs for competence, autonomy, and relatedness - which when experienced increase students' motivation and engagement (Przybylski, Rigby, & Ryan, 2010). Recent studies indicate that game-based cognitive tasks can increase also cognitive performance by increasing users' ability and willingness to train at their optimal performance level (e.g. Lumsden et al., 2016; Ninaus et al., 2015). Moreover, previous research has shown that the use of games in mathematics education can be beneficial also for affective outcomes, for example, reducing math anxiety and increasing enjoyment (Kiili & Ketamo, 2017; Castellar at al., 2014; Ke, 2008).

Games provide also an opportunity to combine learning and assessment processes in an engaging and effective way. According to Kim and Shute (2015) game-based assessment is fairly new in education and scant evidence exists regarding how to maximize the effectiveness and validity of game-based assessment without losing fun and engagement. Assessment that aims to improve learning is called formative assessment and assessment that is used to assess learning outcomes for purposes such as grading, promotion, and placement is called summative assessment. Games provide several possibilities especially for formative assessment. With respect to formative assessment, well designed games can be effective while they provide immediate feedback, adapt to players' competences, provide guidance, and can identify misconceptions and weaknesses in real-time. In general, game-based assessment can increase the validity of the assessment results by decreasing test anxiety.

Despite of positive promises that has been associated with game-based assessment, we should be aware that students are not always familiar with different kinds of digital devices, game mechanics, and user interface solutions that may also cause bias in assessment. In fact, Kim and Shute (2015) has emphasized that fairness should be addressed with a particular care in the game-based assessment because games may function differently across subgroups (e.g., male vs. female, gamers vs. non-gamers).

This paper reports the results of a study that explored the usefulness of game-based math intervention and game-based formative assessment. The main objective of the study was to explore the effectiveness of a short game-based rational number (fractions and decimals) intervention. We utilized our rational number game research engine to prepare the game that was used in the intervention. The Semideus School game is based on recent findings on numerical development and fraction processing (e.g. McMullen et al., 2014; Siegler, Thompson, & Schneider, 2011; Vamvakoussi, 2015) as well as theories that provide an account for the use of manipulatives in digital learning materials (Pouw, van Gog, & Paas, 2014). The assessment of players' conceptual rational number knowledge was embedded directly into the gameplay and several features that support formative assessment were included. We were interested in how much players use the personalized hints that the game provides and to what extend the overall game performance could be used in assessment. We explored also students' experiences about game-based math assessment in light of their general math motivation. Moreover, one aim of the study was to evaluate the fairness of the game-based assessment in terms of gender differences and students' engagement in playing entertainment games.

Method

Participants

Twenty fourth graders and 32 sixth graders participated in the study (N =52). Participants were from two Finnish public primary schools. Of the participants 28 were females, 24 males. Participants were 9-13 years old. Mean age was 11.11 years (SD = 1.03 years).

Description of the used game

The Semideus School game was created for the intervention. The gameplay was founded on tasks that require working with number lines implemented as walkable platforms of a mountain. In the game the user controls a character, Semideus, who tries to collect gold coins that a goblin has stolen from Zeus and hidden on the trails of Mount Olympus. Semideus has discovered the locations of the hidden coins, encrypted in mathematical symbols, and must race the goblin to retrieve the coins from the trails of the mountain. Figure 1 shows examples of estimation and ordering tasks (for more details see YouTube video: https://youtu.be/dVatjMkk0_I). In the number line estimation tasks the user tries to locate a gold coin based on a given number. In ordering tasks, the user has to order the stones in ascending order with regard to the numerical magnitudes depicted on them.

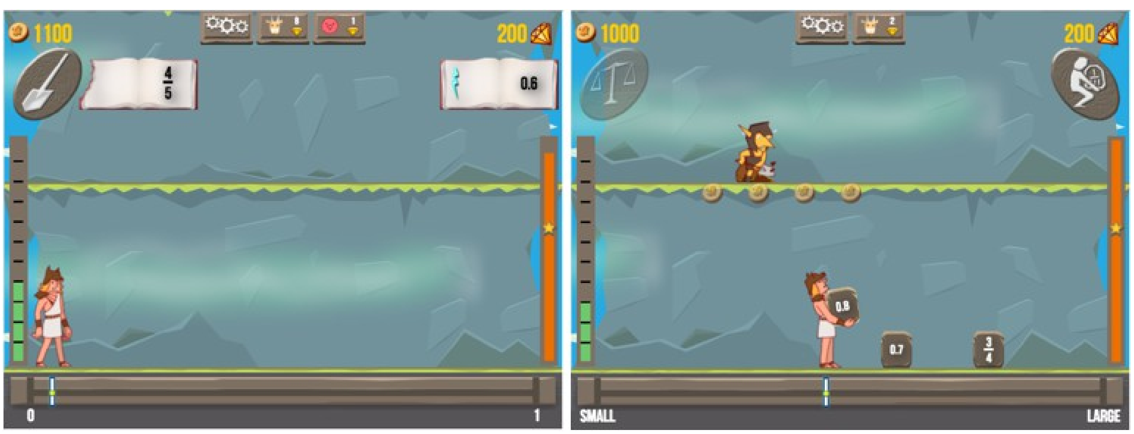

Figure 1. Examples of estimation (left) and ordering tasks (right)

The game consisted of 60 game levels focusing on conceptual rational number knowledge (number line estimation, magnitude comparison, and magnitude ordering tasks). The levels of the game were divided into eight game worlds: 1 Tutorial trails, 2 Show your skills I (pre-test in this study), 3 Lesson trails, 4 Beginner trails, 5 Show your skills II (post-test in this study), 6 Intermediate trails, 7 Show your skills III, 8 Expert trails. Players needed to perform at a certain level to open new levels and worlds. The pre-test and post-test levels could be played only once. In basic levels, a player could activate different kind of scaffolding features and aids by using earned diamonds.

Measures and procedure

The game continuously logged players' detailed playing behavior on a secured server according to a semantic model that is used to describe all the tasks of the game and related math competencies. The recorded playing behavior logs were used to create learning analytics and assessment reports. The player could follow his or her performance in real-time and the game provided also personalized hints to the player. The hints were based on identified misconceptions. In order to support conceptual change processes, links to game content related to misconceptions (lesson levels) were provided with the hints. Both in-game pre-test and in-game post-test consisted of 10 estimation tasks (number line 0-1) from which five were fraction number tasks and rest were decimal number tasks, 10 estimation tasks (number line 0-5) from which five were fraction number tasks and rest were decimal number tasks, 10 fraction and decimal comparing tasks, and 13 fraction and decimal number ordering tasks. Comparing and ordering task was designed to address the most common rational number misconceptions.

Items used to measure math beliefs and motivation were adopted from Berger and Karabenick (2011). Math interest was measured by three items, math self-efficacy was measured by three items, and cost of doing well in math was measured by two items. 5-point Likert-scale was used in math items. The quality of playing experience was evaluated in terms of flow experience (Csikszentmihalyi, 2002; Kiili et al., 2014). Flow experience was measured during the study with a modified version of Flow Short Scale (Engeser & Rheinberg, 2008). Flow Short Scale measures flow experience with ten items (7-point scale). We modified the statements of the scale from present tense to imperfect and added references to the game playing activity. Subjective learning effectiveness, learning method preference (game over paper-pencil tasks), playing motivation, and perceived playability was measured with 6-point Likert scale questions. Participants who played games almost daily and thought that they usually perform pretty well in digital games were classified as gamers and the rest as non-gamers.

Participants played the Semideus School game for approximately 2.5 hours (4-6 playing sessions). The participants were allowed to freely play the game during the sessions. At first, the participants completed the levels of the tutorial world. After the tutorial world the participants completed the Show your skills 1world that worked as a pre-test. There were 20 game levels between the pre-test world and the post-test world.

Results

Effectiveness of the intervention and learning analytics

The figure 2 shows means and standard errors of students' pre-test and post-test scores. A paired-samples t-test indicated that students' post-test scores were significantly higher (M = 68.52, SD = 19.06) than the pre-test scores (M = 56.24, SD = 19.6), t(51) = 8.45, p < .001, d = 0.64. The effect size was at medium level, which means that students clearly benefited from the short number line based game intervention. This finding is in line with the recent studies that have suggested that instructional interventions that aim to support development of conceptual rational number knowledge should target learners' interpretation of rational numbers as magnitudes by practicing them on number lines (e.g. Fuchs et al., 2013).

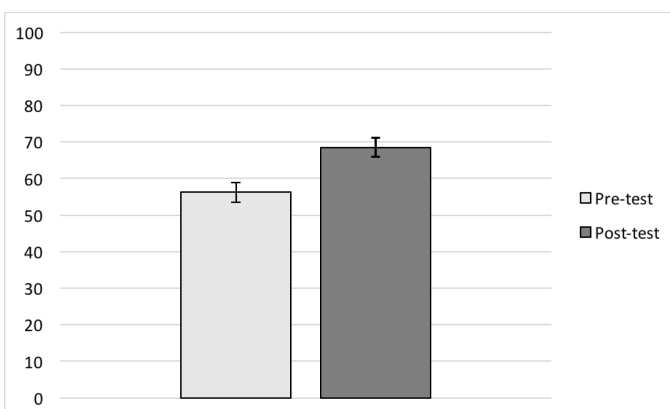

Figure 2. Means and standard errors of students' in-game pre-test and post-test scores.

Approximately 40 % of the students did not open at all game's note book page that shows personalized hints (based on the identified misconceptions) for the player. Most of these students were 6th graders. Students that used the book opened the page that provides hints 6 times on average during the intervention. 75% of these page views did not contain any hints and links to game levels, because students did not have any identified misconceptions when they opened the page. However, although hint/s were available, only 25% of the views ended up in a way that the student played the suggested game level addressing the identified misconception. In 23% of these occasions the student managed to understand the concept by playing the suggested game level and the misconception disappeared. These findings indicate that students can benefit from the hints and suggested game content, but the way in which they are provided to the player needs to be redesigned. In the next game version, the hints will be embedded directly to gameplay. For example, non-player game characters will be used to provide hints to the player in a relevant game context. Furthermore, personalized game levels will be provided through a new event system.

A correlation analysis was used to explore to what extend the overall game performance can be used in assessment. The correlation analysis indicated that there was a high positive correlation between overall game performance (M = 71.97%, SD = 11.58%) and post-test performance (M = 68.52%, SD = 19.06%), $r = 0.88$, $p < .001$. This suggests that teachers could assess students' competences based on overall game performance that makes the assessment process easier, because separated test levels are not needed. In fact, we have recently developed a learning analytics tool for teachers that makes possible to follow students' progress and development of competencies in real time. The usefulness of the tool will be studied in forthcoming studies.

Students' math beliefs and experiences about game-based math intervention

Table 1 shows descriptive statistics of students' math beliefs, flow experience, and opinions about the game. As we can see, students' were quite interested in math and their math self-efficacy was high. Both 4th graders and 6th graders were motivated to perform well in the game (4th: M = 5.45, SD = 0.76; 6th: M = 5.07, SD = 0.98) and they experienced reasonable high flow experience while playing the game (4th: M = 5.11, SD = 0.94; 6th: M = 4.73, SD = 1.13). Moreover, an independent-samples t-test indicated that 4th graders thought that they have to sacrifice more time to perform well in math (M = 3.13, SD = 0.18) than 6th graders (M = 2.14, SD = 0.17), $t(51) = 3.853$, $p < .001$.

Table 1. Descriptive statistics of students' math beliefs, flow experience, and opinions about the game

(Range of the scale)	4th graders	6th graders
Math interest (1-5)	M = 3.90, SD = 0.85	M = 3.51, SD = 1.06
Math self-efficacy (1-5)	M = 4.13, SD = 0.60	M = 4.18, SD = 0.69
Cost of performing well in math (1-5)	M = 3.13, SD = 0.18	M = 2.14, SD = 0.17
Flow experience (1-7)	M = 5.11, SD = 0.94	M = 4.73, SD = 1.13
Perceived effectiveness of the game (1-6)	M = 5.35, SD = 1.04	M = 4.36, SD = 1.7
Prefer game over paper-based learning (1-6)	M = 5.10, SD = 1.21	M = 3.36, SD = 1.99
Playability (1-6)	M = 5.60, SD = 0.82	M = 3.93, SD = 1.84
Playing motivation (1-6)	M = 5.45, SD = 0.76	M = 5.07, SD = 0.98

An independent-samples t-test revealed that 4th graders believed significantly more on learning effectiveness of the game (M = 5.35, SD = 1.04) than 6th graders (M = 4.36, SD = 1.7), $t(51) = 2.31$, $p = .025$. In line with this, 4th graders are more willing to study rational numbers with a game (M = 5.10, SD = 1.21) than 6th graders (M = 3.36, SD = 1.99), $t(51) = 3.49$, $p < .001$. Furthermore, 4th graders also appreciated the playability of the game more (M = 5.60, SD = 0.82) than 6th graders (M = 3.93, SD = 1.84), $t(51) = 3.785$, $p < .001$. The observations and discussions with students revealed that 6th graders demanded more features and interaction possibilities than 4th graders that were engaged with the core game mechanics.

Table 2 presents the correlations between measured variables. Correlation between math interest and math self-efficacy was high and statistically significant, $r = .70$, $p < .001$. In general, students who believed themselves to be more capable in mathematics and who were more interested in mathematics tended to experience higher flow in the game and perceived that the game was an effective tool for learning rational numbers. The correlations showed an association between students who preferred to learn rational numbers with the game and interest in math ($r = .42$), cost of performing well in math ($r = .47$), level of experienced flow ($r = .45$), perceived playability ($r = .48$), and perceived effectiveness of the game ($r = .68$). Moreover, there was a statistically significant correlation between perceived effectiveness and flow experience (r = .66). In general, these findings show that there is a positive relation between general attitude against math and engagement in game-based math activities.

Table 2. Correlations between math beliefs, flow, playing motivation, and opinions about the game

	Interest	Self-efficacy	Cost	Flow	Playing motivation	Playability	Perceived effectivenes	Prefer game
Interest	1							
Self-efficacy	.70***	1						
Cost	.23	-.05	1					
Flow	.57***	.51***	.28	1				
Playing mot.	.44**	.20	.34*	.32*	1			
Playability	.11	-.08	.52***	.31*	.15	1		
Perceived effect.	.48***	.29*	.44**	.66***	.24	.36**	1	
Prefer game	.42**	.11	.47**	.45**	.16	.48**	.68***	1

Correlation is significant at the 0.05*, 0.01**, 0.001*** level (2-tailed).

Fairness of the game-based math intervention

We studied the fairness of the game-based assessment in relation to students' subjective playing skills and playing frequency (gamers vs. non-gamers) as well as to gender. Correlation analysis indicated that either students' general digital game playing frequency (r = .04, p = .778) or subjective digital game playing competence (r = -.11, p = .457) did not relate to overall game performance. For further analyses, based on these metrics we categorized the students as gamers and non-gamers. Independent samples t-test indicated that there were no statistically significant differences in experienced flow level, t(51) = 1.205, p = .234, preference of learning method, (51) = 1.065, p = .292, and motivation to play the intervention game, t(51) = -1.120, p = .269, between gamers and non-gamers. Moreover, independent samples t-test indicated that there were no statistically significant differences in experienced flow level, t(51) = -1.415, p = .165, preference of learning method, t(51) = -1.962, p = .057, motivation to play the intervention game, t(51) = 1.518, p = .137, and overall game performance, t(46) = -.723, p = .474, between girls and boys.

To summarize, these results indicate that game-based assessment can be a fair assessment approach and most of the students seems to be willing to adopt game-based learning solutions in schools. However, we have to remember that the results of this study can not be generalized to other games and more research with different kinds of games on the topic are needed.

Conclusion

In summary, the current study aimed at evaluating the usefulness of a digital math game with respect to learning gains and learning assessment. We conducted an intervention study, in which Finnish fourth and sixth grade students learned conceptual rational number knowledge with a number line based game that included several scaffolding features founded on students' playing behavior. The results indicated that students benefited significantly from the intervention and were engaged to play the game. The study showed that separated test are not necessary, but teachers could assess students' math competences based on overall game performance. The results also revealed that 4^{th} graders appreciated the game more than 6^{th} graders. In general, it seems that game-based formative assessment can be a fair assessment approach and most of the students seems to be willing to adopt game-based learning solutions in schools. To conclude, the use of games particularly in formative assessment seems very promising while the learning analytics of a game can be harnessed to support both teachers' and students' learning and teaching goals in non-intrusive manner. However, the utilization of game-based assessment is currently fairly new in education and more research is needed to developed more effective solutions and to support the diffusion of game-based assessment into daily practices.

References

Berger, J. L., & 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.

Castellar, E. N., Van Looy, J., Szmalec, A., & De Marez, L. (2014). Improving arithmetic skills through gameplay: assessment of the effectiveness of an educational game in terms of cognitive and affective learning outcomes. *Information sciences, 264*, 19-31.

Csikszentmihalyi, M. (2002). Flow: The psychology of optimal experience (2nd edition). Harper & Row, New York.

Devlin, K. J. (2011). *Mathematics Education for a New Era: Video Games as a Medium for Learning.* AK Peters Ltd.

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion. 32*(3), 158-172.

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., ... & Changas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology*, *105*(3), 683.

Kadosh, R. C., Dowker, A., Heine, A., Kaufmann, L., & Kucian, K. (2013). Interventions for improving numerical abilities: present and future. *Trends in neuroscience and education*, *2*(2), 85-93.

Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay? *Computers & Education, 51*(4), 1609-1620.

Kiili, K., & Ketamo, H. (2017). Evaluating Cognitive and Affective Outcomes of a Digital Game-Based Math Test, *IEEE Transactions on Learning Technologies*, http://doi.org/10.1109/TLT.2017.2687458

Kiili, K., Lainema, T., Freitas, S. D., & Arnab, S. (2014). Flow framework for analyzing the quality of educational games. *Entertainment Computing, 5(4), 367-377*.

Kiili, K., Devlin, K., Perttula, T., Tuomi, P., & Lindstedt, A. (2015). Using video games to combine learning and assessment in mathematics education. *International Journal of Serious Games*, *2*(4), 37-55.

Kim, Y. J. & Shute, V. J. (2015). The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. *Computers & Education*, *87*, 340-356.

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. *JMIR Serious Games*, *4*(2).

Moeller, K., Fischer, U., Nuerk, H. C., & Cress, U. (2015). Computers in mathematics education–Training the mental number line. *Computers in Human Behavior*, *48*, 597-607.

McMullen, J., Laakkonen, E., Hannula-Sormunen, M., & Lehtinen, E. (2015). Modeling the developmental trajectories of rational number concept (s). *Learning and Instruction*, *37*, 14-20.

Ninaus, M., Pereira, G., Stefitz, R., Prada, R., Paiva, A., Neuper, C., & Wood, G. (2015). Game elements improve performance in a working memory training task. *International Journal of Serious Games*, *2*(1), 3-16.

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