A fun-accuracy trade-off in game-based learning

Simon Greipl[1], Manuel Ninaus[1,4][0000-0002-4664-8430], Darlene Bauer[2], Kristian Kiili[3], Korbinian Moeller[1,2,4]

[1] Leibniz-Institut für Wissensmedien, Tübingen, Germany
{s.greipl, k.moeller, m.ninaus}@iwm-tuebingen.de
[2] Department of Psychology, Eberhard-Karls University, Tuebingen, Germany
[3] TUT Game Lab, Tampere University of Technology, Pori, Finland
kristian.kiili@tut.fi
[4] LEAD Graduate School & Research Network, University of Tuebingen, Germany

Abstract. The present paper illustrates that the game-based implementation of a learning task - here to train basic math skills - entails benefits with strings attached. We developed a game for learning math with its core element based on the number line estimation task. In this task, participants have to indicate the position of a target number on a number-line, which is thought to train basic numerical skills. Participants completed both the game on a mobile device and a conventional paper-pencil version of the task. They indicated to have significantly more fun using the game-based environment. However, they also made considerably higher estimation errors in the game compared to the paper-pencil version. In this case, more fun in a math-learning task was ultimately bought at the expense of lower reliability, namely lowered accuracy of estimations in the learning game. This fun-accuracy trade-off between adding elements for enjoyment and clarity of content is discussed together with the consequences for game-design.

Keywords: game-based learning, reliability, enjoyment, number-line estimation, user-experience, mathematics

1 Introduction

Game-based learning is thought to create, amongst other positive effects, increased motivation and special situational interest (e.g., for an overview see [1]) in the game and therefore in the topic to be learned. There is ample empirical evidence indicating that (math) games for learning had positive educational effects in terms of better learning outcomes or higher academic achievement [2–10]. While motivation and interest are part of the foundations of game-based learning [11] and contribute, amongst others, to the beneficial effect of the use of games for learning, we may sometimes turn a blind eye to the fact that these benefits may not come without any strings attached.

Trade-offs in game-based learning? The current scenario does not refer to pragmatic setbacks like the high costs of developing and implementing a game in educational or similar non-profit settings. We rather want to delineate that the gamification or game-based setup of a task or educational content can be detrimental in various aspects regarding the very subject matter. For instance, the implementation of a storyboard to, for instance, create interest, identification or emotional value may be deployed at the expense of (learning-)time and energy as well as potentially distracting the player from the core topic.

An illustrative example showed that the implementation of games in education seemed to have elicited enjoyment during class as well as for the topic itself, but has not necessarily entailed a measurable cognitive advancement regarding the content to be conveyed [12]. Pittman used the commercial game *Portal 2*, which is basically a sequence of physics puzzles allowing to create individual experiments to teach physics in 11^{th} grade high school. The author repeatedly proclaimed the bilateral enjoyment of the lessons, the engagement of students, and the opportunity of such methods to create "memorable, teachable experiences"[12]. Nevertheless, inexperience with gameplay, mainly regarding the required handling of the first-person navigation, were the biggest setback and major obstacle. This in turn led to "unimaginative experiments" and problems even in simple tasks. Finally, exam results at the end of the year did not reveal additional general improvement in learning outcomes traceable to the use of the game.

In another study, researchers tried to implement the content to be learned in a game by tightly coupling a math task with the core game mechanic, namely in-game combat. What is called *intrinsic integration* aims to minimize extrinsically engaging elements and rather making them an intrinsically motivating part of the game [13]. However, the authors also observed a decline in accuracy due to the implementation/operationalization of their task. Nevertheless, compared to an extrinsic counterpart of the game (non-mathematical combat and math-quiz between levels) and a control condition, students who played the intrinsic version achieved the best learning results.

It is imaginable that the use of a complex, enriched (digital) learning environment can also state a negative influence on the content to be learned. Other than through handling and interface related obstacles, this seems also feasible through the accumulation of game-elements unrelated to the content and thereby blurring the subject matter or diverting player's attention. In other words, the relationship between (game elements used for) motivation/engagement and reliability seems reciprocal and may even be negatively correlated. In this vein, we would have to look for a *sweet spot* between engagement and reliability.

Game-based math-training. To shed some light on this relationship, the current study focused on the outcomes of a math game to objectively assess potential drawbacks of a game-based learning situation. Manipulating numbers in general is a necessity to deal with everyday life demands. For instance, the decision whether a purchase is still within budget or the percentage of savings of a discount requires addition/subtraction, multiplication and division and percentages as well as a general understanding of number magnitude. Difficulties in understanding and manipulating numbers have negative impacts on school career and can later lead to behavioural as well as societal problems like delinquency and in turn have "adverse consequences for cognitive function throughout life" [14]. In other words, numerical competences play a significant role in successful development.

For instance, Whyte and Bull [15] name nonverbal representation and manipulation of numbers as core competencies for developing adequate arithmetic abilities. Such basic numerical competencies in turn can be assessed by various tasks like enumeration, magnitude comparison, estimations, and the positioning of number magnitudes on a mental number-line [15]. The latter describes an often used metaphor to describe our mental representation of number magnitude according to which number magnitude is represented spatially along a number line in an analogue manner with magnitudes increasing from left to right (for an overview see [16]). It was observed repeatedly that arithmetic competencies can be predicted by more precise mental representations of number magnitude already in early childhood ages [17]. Therefore, fostering understanding of number magnitude is a crucial step in developing higher mathematical abilities.

In the current study, we examined the relationship between enjoyment and reliability of a well-known paper-pencil math task, the number-line estimation, and an in-house developed math game using the very same core task mechanic (see Fig. 1). In the following, we describe how the game was developed and which methods were employed to compare the two learning tasks to each other. Subsequently, the evaluation of subjective and objective measurements of task interaction is presented and lastly interpreted and discussed against the background of motivational, educational, and game design aspects.

2 Methods

We compared estimation accuracy as well as enjoyment between the game-based version and conventional number line estimation tasks on paper to investigate the relationship between enjoyment or fun, respectively, and answer accuracy.

2.1 Participants

18 adult student participants (10 females, mean age 22.72, *SD*=2.56, range 19 to 30) were randomly recruited at the library of the University of Tübingen.

2.2 Design

Paper-pencil and game version of the number line estimation task were randomly assigned to participants, so that 10 participants started with the paper-pencil version, while the other 8 participants started with the game *Shoot The Number Line*. The main dependent variable for comparison was estimation accuracy in terms of absolute estimation error.

2.3 Measurement

We created a game for the use on tablets called "Shoot The Number Line" to examine the relationship between enjoyment and reliability within game-based learning (see Fig. 1). As the name suggests, the game uses the so-called number-line estimation task as the core game mechanic. In this task, participants must indicate the spatial position of a target number by slinging a ball as accurately as possible to the correct position along a number-line with only its endpoints specified (e.g., where goes 28 on a number line ranging from 0 to 100).

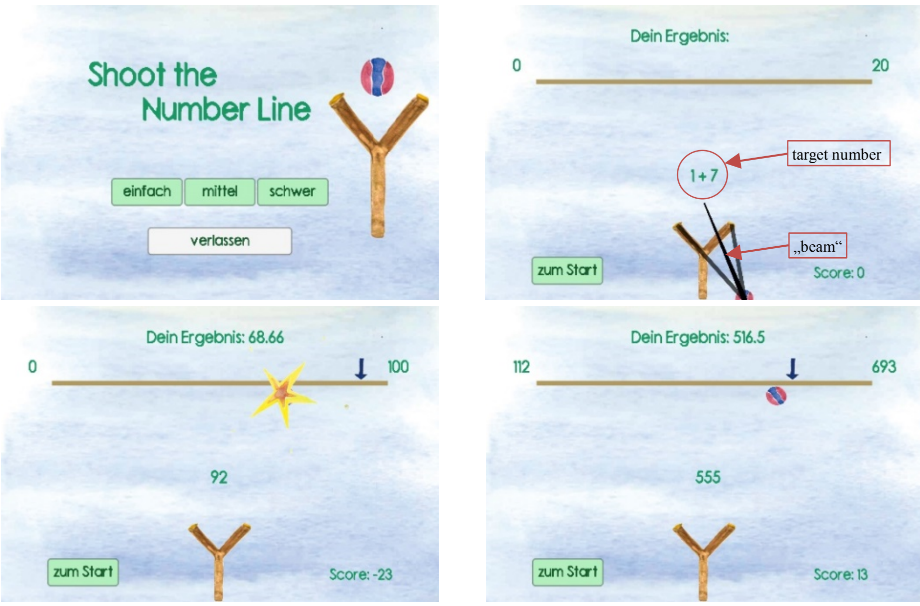

Fig. 1. Screenshots of "Shoot The Number Line" with target numbers displayed above the slingshot: *top-left:* startup screen with level choice; *top-right:* level "easy" with stretched slingshot, showing the "beam" as an indicator of shooting direction; *bottom-left:* level "medium" and negative feedback; *bottom-right:* level "hard" and successful estimation.

Accordingly, the current game focusses on the mental representation of magnitude on a number line. Usually, number line estimation tasks are implemented as paper-based tests. Accordingly, learners use a pen to mark the position of target numbers (e.g. with a cross or a simple stroke) on several number-lines printed on a multi-page document. Even though training number line estimation in such a way was found to be successful [18], the procedure itself seems rather tedious. Yet, the number-line estimation task is a suitable candidate for embedding into a game-based learning environment (see also e.g.[19, 20]) aiming at making the task more enjoyable and engaging. Likewise, through a comparison between in-game training and conventional paper-pencil methods, we can directly compare both methods in detail. *Shoot The Number Line* was developed in *Unity 3d*, which takes care of basic processes like model animation and physics interaction and is at the same time platform independent. The basic mechanics of the game comprise a slingshot with a ball that the player must shoot at the position of the respective target number – displayed right above the slingshot – on the number-line with highest possible accuracy. The moment the player touches the ball and stretches the slingshot, a beam is displayed to indicate the position where the ball will land on the number line. For a trial to be successful players need to hit the position of target numbers accurately. The maximum deviation from the target position/number allowed is held constant at 7%. Target numbers can vary from a simple number to a calculation problem. After a start-up menu, the player can choose between three difficulty levels *easy, medium,* and *hard*. On the easy level, the target number is a simple calculation within the 0 to 20 number range. On the medium level, participants had to estimate the position of a target number in the 0 to 100 range. Finally, the hard level employed the number range from 0 to 1000 and starting and endpoint of the range varied from trial to trial. In the current study, only the medium level was used. When the target was hit with sufficient accuracy, the player scored 20 points minus the ADT (absolute deviation from target). In case of a miss, the player's score is reduced by the ADT.

The paper-pencil version comprised a multiple-page document with number lines and target numbers between *0* and *100*. The game ran on a 10" Medion Touch-Netbook operated on Windows 8. Additionally, a questionnaire assessing user experience of the game was developed and had to be answered by participants (Example question: *How much fun did you have playing the game? – [1: not at all -> 5: very much];* see Appendix A).

2.4 Procedure

The experiment took about 10 minutes. After receiving oral and written instructions, participants were introduced with an exercise to either the digital game-based or paper-pencil version of the number line estimation task to get used to task requirements (onboarding phase). The game was played on medium difficulty level only with target numbers covering the whole range of the number line (0 to 100). Correspondingly, the same procedure was used in the paper-pencil version of the task. Each participant then had to indicate the position of 23 different target numbers in each version (in the order of presentation, paper-pencil targets: 56,16, 49, 3, 23, 95, 45, 31, 14, 73, 54, 91, 82, 76, 37, 51, 2, 98, 62, 69, 87, 28,8; game targets: 48, 39, 65, 81, 67, 52, 4, 12, 97, 1, 21, 43, 93, 26, 58,19, 89, 96, 35, 53, 74, 6, 78). After completion of both tests, the questionnaire had to be filled out. There was no time limit during the whole procedure. For the paper-pencil version, estimates on the number line were measured with a ruler and converted to the corresponding relative number to evaluate the estimation error. Afterwards, the single-factor-design allowed for *t*-tests to analyse all comparisons of interest.

3 Results

Estimation accuracy differed significantly between the game-based and the paper-pencil version of the number-line estimation tasks [$t(17)=11.41$, $p<.001$]. The results showed, that participants estimated the target numbers more accurate and with less dispersion on the number line when they were performing the paper-pencil version [paper-pencil: $M= 2.16$, $SD=0.05$; game: $M=5.28$, $SD= 1.88$; see Fig. 2]. Conforming to this objective difference, participants also judged their own performance between the two trainings to differ significantly [$t(17) = 1.82$, $p = .043$]. In particular, they thought that their outcome was better in the paper-pencil version [$M= 3.67$, $SD=0.97$] than in the game [$M= 3.11$, $SD=1.08$] which was in consonance with the accuracy data. According to the questionnaire, participants seemed to have significantly more fun completing the game-based version [$M= 4.06$, $SD=0.83$] than the paper-pencil version [$M= 2.28$, $SD=0.83$; $t(17)=8.59$, $p<.001$]. Moreover, design [$M= 3.83$, $SD=0.62$] as well as clarity of the game [$M= 4.56$, $SD=0.62$] were positively evaluated by participants. Overall, participants indicated a few problems operating the slingshot [$M= 3.12$, $SD=1.04$]. This was substantiated by qualitative feedback. Some participants reported that they encountered problems handling shots close to the ends of the number line (see also Fig.2). Feedback that the game provided was primarily perceived as helpful rather than disturbing [$M= 3.89$, $SD=1.02$, see Fig. 3].

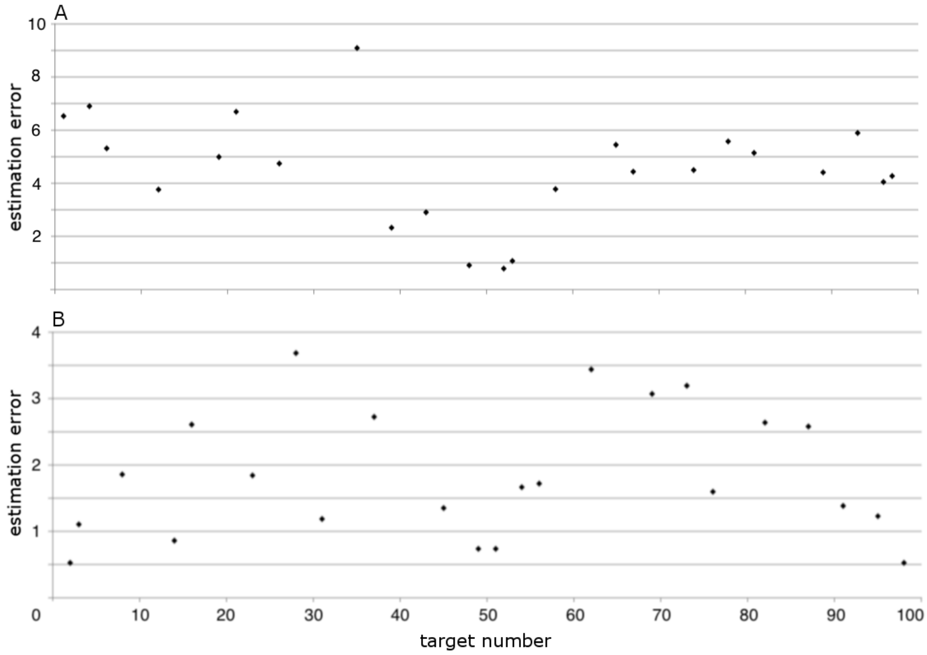

Fig. 2. Estimation errors in the game (top/A) and the paper-pencil test (bottom/B). These errors illustrate the average accuracy per target number across subjects. Higher values mean greater deviation from the target position on the number-line.

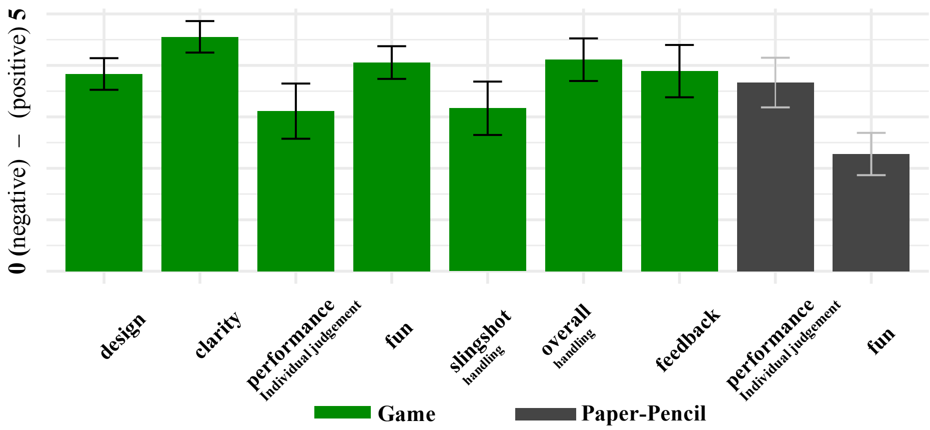

Fig. 3. Mean results of the questionnaire on various facets regarding user experience of the game-based as well as the paper-pencil learning task. Lower values (0) represent negative while higher values (5) represent positive experience. Error-bars at the top represent standard errors.

4 Discussion & Conclusion

The current study compared the performance of participants in two versions of the number line estimation task. Even though participants had more fun using the game-based version, accuracy was significantly higher in the paper-pencil based version of the task. In the following, we discuss these findings and the implications of enriching learning tasks with game elements.

Design of the game and its overall handling did receive very positive user-feedback. Yet, participants sometimes stated that there is still room left for improvement. For instance, they found targets on the very left or right side on the number-line explicitly hard to hit (see also Fig. 2 A). This seems to be reflected in the only low positive evaluation of the slingshot-handling.

Obviously, there are some pitfalls that come along with an enriched digital learning environment. While there are a few issues – we already mentioned for instance problems with the slingshot handling – we want to focus on those that became apparent in the comparison between the game and paper-pencil training. Most strikingly, there was a difference in estimation errors participants made during the task. They obviously had more problems accurately hitting the target in the game-based than in the paper-pencil version, which seemed particularly apparent for target numbers close to start and endpoint of the number line (see Fig. 2). Because of very accurate performance in the paper-pencil-training, we must assume that participants have decent representations of number magnitudes. Therefore, we may infer that the significantly worse performance in terms of estimation accuracy in the game-based version originated from a property of the game itself. The shooting mechanism, in particular the rather short visualization of the beam in the game, is a very likely reason causing these performance deviations. Accordingly, the visual distance to target positions in the middle of the number line is smaller as compared to positions towards the ends, requiring more visual extrapolation of the trajectory of the slingshot ball.

Importantly, while this mechanic increases some levels of uncertainty and consequently difficulty it also decreases the reliability of the game with respect to estimation competence. Uncertainty is an important factor in engaging and motivating (e.g. [21]) players but may also improve learning outcomes due to its related release of dopamine (e.g., [22, 23]). Some degree of reliability however is needed to determine and foster players understanding of number magnitude. Importantly though, participants rated the game to be considerably more fun than the paper-pencil-task. Consequently, while the rather short visualization of the beam might have been a major contributor to participants' enjoyment, it might have also led to the worse performance in the game-based task as compared to the paper-pencil version. Similarly, Kiili and Ketamo [24] found that although participants performed significantly better in a paper-pencil based math test than in a game-based math test, the game-based test was more engaging.

Thus, the current results indicate that we may swapped a part of task accuracy and reliability, respectively, in favour of task enjoyment. It is additionally conceivable that an unknown amount of attention during the game is unwantedly shifted away from the core task – and presumably towards the shooting mechanism. Future studies may consider more exercises before the actual training to minimise the risk of handling-biases originating from participants' inexperience with a new interface. In line with this, Kiili and Ketamo [24] have proposed that an appropriate onboarding phase should be included in learning games to decrease mistakes caused by unfamiliar user interfaces, which should lead to smaller assessment errors in turn. In our case, some inaccuracies may have been caused by the inexperience with the new interface (i.e., slingshot mechanic), in particular targeting numbers on both ends of the number line (see Fig. 2. A). We may at this point hint to a similar mobile math training game, *Semideus*, that needs an avatar to walk to the right place along the number line using tablet-based tilt-control [25]. The game therefore implemented the same core task – the number line estimation – but uses different control mechanics that enable high accuracy and thereby maintaining reliability of the task. Although the developers faced a similar effect as the tablet-based mechanics led to overall longer response times, hallmark effects of number magnitude processing were successfully replicated [25, 26]. Based on the study in which Semideus was used to train rational numbers, Lindstedt and Kiili [27] showed that user interface checkpoint tasks revealing participants' true controlling ability through trivial tasks can be used to increase the validity of math assessments by reducing effects of the user interface. The latter supports the validity of this instantiation of a game-based number line task as an assessment and learning tool [19].

In sum, these examples show that game-based task realisations can inherit a trade-off. In the present case this was reflected by heightened enjoyment at the expense of accuracy. Careful and sophisticated game-design can cushion or minimize such a fun-accuracy trade-off and simultaneously keep the clarity as regards content. Ultimately, the person in charge of training or assessment is obliged to evaluate in how far positive effects may outweigh negative ones, if existent, and which audience he/she intends to reach. Training and assessment comprise the main applications of game-based learning. and these two applications may come with different requirements. It would be rather inacceptable when reliability issues bias the outcome of a game-based (math) test, for instance, when introducing game elements for enjoyment may cause uncertainty with respect to performance, which may ultimately be reflected in an official examination (grade). In a training, however, such problems may be irrelevant when they evidently do not impede actual learning. For instance, even if such a fun-accuracy trade-off occurs in game-based learning, studies showed that the learning achievement was not necessarily lowered [12], or it was regardless the most successful learning strategy [13].

The latter even speaks in favour of the perspective that motivation – especially when intrinsically integrated – may potentially outweigh reliability issues. *Semideus* is again an example where the constraints caused by the game-based implementation of a task does not always lead to vital restrictions affecting the quality of the training nor the assessment power of the game [25]. Last, as mentioned before, for some audiences the motivational and engaging benefit originating from games designed for learning can provide the critical impulse that tips the scales in advance for a topic that is negatively connoted, tedious, boring, or otherwise difficult to approach (see e.g. [11]). To conclude, in some situations, depending on audience and/or game-design, trading reliability for fun can actually be a good deal.

The present work outlines that game-based learning environments can induce a setback that has to be evaluated carefully. In our case, however, there are some restrictions hindering a more detailed evaluation of this phenomenon. We already mentioned, for instance, handling issues, mainly regarding the slingshot and its beam. It is also likely that the onboarding phase was too short to make sure that participants were prepared enough to minimize interface induced biases. Because the paper-pencil method produced high accuracies despite a short introduction to the task, our findings suggest to employ an appropriate adjustment of the onboarding phase when interacting with a new interface. Other criticism can be pointed out about the rather small sample size and the age of participants. In particular, we did not assess participants previous experience with tablet/touch-devices or even digital games, which might have affected our results.

Future studies will have to clarify the fun accuracy trade-off present in this and in other games and its theoretical and practical relevance. This entails the direct comparison of games for learning to conventional learning methods individually as well as in long term and for different skill/interest levels in the respective learning domain. Different skill and interests may have differential implications with respect to a continuum, in which seemingly a part of precision/reliability is given up for motivation or engagement. For instance, an engaging but less accurate learning game may act as a door opener for a child who struggles in the respective domain, but a child with high intrinsic motivation may not benefit from – or even be demotivated by – such a learning environment.

References

1. Plass, J.L., Kaplan, U.: Emotional Design in Digital Media for Learning. In: Emotions, Technology, Design, and Learning. pp. 131–161. Elsevier (2016).
2. Boyle, E.A., Hainey, T., Connolly, T.M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., Pereira, J.: An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. Computers & Education. 94, 178–192 (2016).
3. Hainey, T., Connolly, T.M., Boyle, E.A., Wilson, A., Razak, A.: A systematic literature review of games-based learning empirical evidence in primary education. Computers & Education. 102, 202–223 (2016).
4. Gee, J.P.: What video games have to teach us about learning and literacy. Palgrave Macmillan, New York (2003).
5. Prensky, M.: Digital game-based learning. Paragon House, St. Paul, MN (2007).
6. Wouters, P., van Nimwegen, C., van Oostendorp, H., van der Spek, E.D.: A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology. 105, 249–265 (2013).
7. Li, M.-C., Tsai, C.-C.: Game-Based Learning in Science Education: A Review of Relevant Research. J Sci Educ Technol. 22, 877–898 (2013).
8. Wouters, P., van Oostendorp, H., ter Vrugte, J., Vandercruysse, S., de Jong, T., Elen, J.: The effect of surprising events in a serious game on learning mathematics. British journal of educational technology. 48, 860–877 (2017).
9. Vandercruysse, S., ter Vrugte, J., de Jong, T., Wouters, P., van Oostendorp, H., Verschaffel, L., Moeyaert, M., Elen, J.: The effectiveness of a math game: The impact

of integrating conceptual clarification as support. Computers in human behavior. 64, 21–33 (2016).

10. Seaborn, K., Fels, D.I.: Gamification in theory and action: A survey. International Journal of human-computer studies. 74, 14–31 (2015).

11. Plass, J.L., Homer, B.D., Kinzer, C.K.: Foundations of Game-Based Learning. Educational Psychologist. 50, 258–283 (2015).

12. Pittman, C.: Teaching with portals: The intersection of video games and physics education. LEARNing Landscapes. 6, 341–360 (2013).

13. Habgood, M.P.J., Ainsworth, S.E.: Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games. Journal of the Learning Sciences. 20, 169–206 (2011).

14. Beddington, J., Cooper, C.L.: The mental wealth of nations. 455, 3 (2008).

15. Whyte, J.C., Bull, R.: Number games, magnitude representation, and basic number skills in preschoolers. Developmental Psychology. 44, 588 (2008).

16. de Hevia, M.D., Girelli, L., Macchi Cassia, V.: Minds without language represent number through space: origins of the mental number line. Frontiers in Psychology. 3, (2012).

17. Opfer, J., Siegler, R.: Representational change and children's numerical estimation. Cognitive Psychology. 55, 169–195 (2007).

18. Geary, D.C., Hoard, M.K., Nugent, L., Byrd-Craven, J.: Development of Number Line Representations in Children With Mathematical Learning Disability. Developmental Neuropsychology. 33, 277–299 (2008).

19. Kiili, K., Moeller, K., Ninaus, M.: Evaluating the effectiveness of a game-based rational number training - In-game metrics as learning indicators. Computers & Education. 120, 13–28 (2018).

20. Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., Gälli, M., Martin, E., von Aster, M.: Mental number line training in children with developmental dyscalculia. Neuroimage. 57, 782–795 (2011).

21. Howard-Jones, P.A., Demetriou, S.: Uncertainty and engagement with learning games. Instructional Science. 37, 519–536 (2009).

22. Fiorillo, C.D.: Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons. Science. 299, 1898–1902 (2003).

23. Lisman, J.E., Grace, A.A.: The Hippocampal-VTA Loop: Controlling the Entry of Information into Long-Term Memory. Neuron. 46, 703–713 (2005).

24. Kiili, K., Ketamo, H.: Evaluating Cognitive and Affective Outcomes of a Digital Game-Based Math Test. IEEE Transactions on Learning Technologies. 1–1 (2017).

25. Ninaus, M., Kiili, K., McMullen, J., Moeller, K.: Assessing fraction knowledge by a digital game. Computers in Human Behavior. 70, 197–206 (2017).

26. Schneider, M., Siegler, R.S.: Representations of the magnitudes of fractions. Journal of Experimental Psychology: Human Perception and Performance. 36, 1227–1238 (2010).

27. Lindstedt, A., Kiili, K.: Evaluating playing experience and adoption of a math learning game. In: Proceedings of the 1st International GamiFIN Conference. pp. 39–46 (2017).

Appendix

Table A. Items to examine user experience

Item		1	2	3	4	5	
How appealing was the design of the app?	not at all	O	O	O	O	O	very much
Was the app clearly arranged?	not at all	O	O	O	O	O	very much
How good do you think your performance was in the game?	not at all	O	O	O	O	O	very much
How much fun did you have playing the game?	none	O	O	O	O	O	very much
How well did you manage to handle the slingshot?	not at all	O	O	O	O	O	very much
How well did you in general manage to handle the app?	not at all	O	O	O	O	O	very much
Was the feedback (explosions, arrows etc.) helpful?	not at all	O	O	O	O	O	very much
How good do you think was your performance in the paper-pencil test?	not at all	O	O	O	O	O	very much
How much fun did you have doing the paper-pencil test?	none	O	O	O	O	O	very much
Do you have other remarks?							