

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Computers &amp; Education

journal homepage: [www.elsevier.com/locate/compedu](http://www.elsevier.com/locate/compedu)

## Evaluating the effectiveness of a game-based rational number training - In-game metrics as learning indicators

Kristian Kiili<sup>a,\*</sup>, Korbinian Moeller<sup>b,c,d</sup>, Manuel Ninaus<sup>b,c</sup><sup>a</sup> *TUT Game Lab, Tampere University of Technology, Pori, Finland*<sup>b</sup> *Leibniz-Institut für Wissensmedien, Tuebingen, Germany*<sup>c</sup> *LEAD Graduate School, Eberhard-Karls University Tuebingen, Germany*<sup>d</sup> *Department of Psychology, Eberhard Karls University, Tuebingen, Germany*

## ARTICLE INFO

## Keywords:

Interactive learning environments  
Elementary education  
Game-based learning  
Mathematics  
Rational numbers

## ABSTRACT

It was argued recently that number line based training supports the development of conceptual rational number knowledge. To test this hypothesis, we evaluated training effects of a digital game based on the measurement interpretation of rational numbers. Ninety-five fourth graders were assigned to either a game-based training group ( $n = 54$ ) who played a digital rational number game for five 30-min sessions or a control group ( $n = 41$ ) who attended regular math curriculum. Conceptual rational number knowledge was assessed in a pre- and posttest session. Additionally, the game groups' playing behavior was evaluated. Results indicated that the game-based training group improved their conceptual rational number knowledge significantly more strongly than the control group. In particular, improvement of the game-based training group was driven by significant performance increases in number magnitude estimation and ordering tasks. Moreover, results revealed that in-game metrics, such as overall game performance and maximum level achieved provided valid information about students' conceptual rational number knowledge at posttest. Therefore, results of the current study not only suggest that aspects of conceptual rational number knowledge can be improved by a game-based training but also that in-game metrics provide crucial indicators for learning.

## 1. Introduction

Mathematics proficiency is crucial for educational, vocational, and personal life prospects in today's Western knowledge societies. Importantly, [Parsons and Bynner \(2006\)](#) stated that on an individual level, insufficient mathematical competencies may be even more detrimental to career prospects than spelling or reading deficiencies. Moreover, on a societal level, mathematical deficiencies can lead to immense costs ([Gross, Hudson, & Price, 2009](#)). Therefore, effective, innovative and engaging ways to teach basic numerical skills but also more complex mathematical capabilities, such as deep understanding of rational numbers, are needed to foster mathematical achievement.

Recent studies indicated that understanding the meaning of rational numbers and acquiring knowledge about rational numbers (i.e., fractions, decimals, and percentages) is crucial in working life and societal practices ([ACME, 2011](#)). In particular, research showed that proficiency with fractions is associated with students' success in algebra, which has been argued a gateway to STEM professions (e.g., [Hansen, Jordan, & Rodrigues, 2017](#); [Siegler, Duncan, Davis-Kean, Duckworth, Claessens, Engel, et al., 2012](#)). Moreover, in everyday life, numerous instances, in cooking, interpreting shopping deals, and calculating loan rates etc., require

\* Corresponding author.

E-mail address: [kristian.kiili@tut.fi](mailto:kristian.kiili@tut.fi) (K. Kiili).

appropriate mastering of rational numbers. Importantly, however, rational numbers are one of the most difficult concepts to learn in primary school and even adults frequently fail to process them correctly (Gigerenzer, 2002; Siegler, Fazio, Bailey, & Zhou, 2013 for a review).

Given the difficulties that many children and even adults face with reasoning about rational numbers, traditional instructional methods should be reconsidered and complemented by new tools for enhancing rational number knowledge. In fact, previous research has indicated that digital learning games can support mathematics learning (e.g., ter Brugte et al., 2017; Bakker, van den Heuvel-Panhuizen & Robitzsch, 2015; Riconscente, 2013; Li & Ma, 2010 for a review) as well as facilitate engagement in mathematics (Castellar, Van Looy, Szmalec, & De Marez, 2014; Ke, 2008; Kiili & Ketamo, 2017) – provided they are properly designed.

For instance, Devlin (2011) argued that well designed digital games can support numerical development and mathematical proficiency. However, many published mathematics games primarily address math facts and procedural knowledge, a focus that tends to be found in classroom practices as well. Procedural knowledge refers to sequences of actions that can be carried out to solve specific numerical/mathematical problems (Rittle-Johnson & Alibali, 1999). In the domain of rational numbers, knowing how to add fractions with different denominators is one example of such procedural knowledge. In contrast, the current study addressed conceptual knowledge of rational numbers. According to Schneider and Stern (2010) conceptual knowledge can be defined as knowledge of central concepts and principles and their interrelations in a particular domain of knowledge. In the domain of rational numbers (i.e., fractions and decimals in this study), conceptual knowledge refers to a combination of the general properties of rational numbers, such as understanding (i) fraction and decimal number notation, (ii) rational numbers as a unified system of numbers that can be placed on a number line according to their magnitudes, (iii) that rational number magnitudes can be represented in an infinite number of ways (equivalence), (iv) that there is an infinite number of rational numbers between any two rational numbers (density), and (v) the differences between whole number and rational number properties (e.g., Gabriel et al., 2012; McMullen, Laakkonen, Hannula-Sormunen, & Lehtinen, 2015; Siegler et al., 2013).

The number line estimation, magnitude comparison, and magnitude ordering tasks are common ways to study the development of conceptual rational number knowledge (e.g., Alibali & Sidney, 2015; McMullen et al., 2015; Siegler & Braithwaite, 2017). In the number line estimation task participants have to indicate the spatial position of a target fraction on a number line with only its endpoints specified (e.g., where goes  $1/4$  on a number line ranging from 0 to 1; e.g. Fazio, Kennedy, & Siegler, 2016; Link, Moeller, Huber, Fischer, & Nuerk, 2013; Siegler & Opfer, 2003). In fraction comparison tasks, participants are asked to either choose the larger or the smaller one of two fractions, to judge whether a statement about relative fraction magnitudes is true or false (e.g.,  $3/5 > 2/3$ ), or to compare the magnitude of a given fraction to a “standard” value (e.g., is  $3/7$  smaller or larger than  $3/5$ ; e.g., Alibali & Sidney, 2015). Magnitude ordering tasks usually include three numbers. For example, in fraction ordering tasks, participants have to put the numbers in order from smallest to largest:  $6/8$ ;  $2/2$ ;  $1/3$  (e.g., McMullen et al., 2015) or are asked in which one of given sets the three fractions are arranged from smallest to largest (e.g., Hansen et al., 2017).

The study reported in this paper investigated the effectiveness of above described number line estimation, magnitude comparison, and magnitude ordering tasks in a game-based training context. To be more precise, we used our rational number game engine, *Semideus*, to develop a digital game for the training intervention (cf. Ninaus, Kiili, McMullen, & Moeller, 2016, 2017). This beta version of the *Semideus* School game focused primarily on conceptual rational number knowledge. Before we outline the more detailed aims and hypotheses of the current study we will first give a brief overview of the difficulties that students tend to face when learning rational numbers. The aim of the developed *Semideus* School game is to help students to overcome these difficulties. A detailed description of the game is provided in section 2.

### 1.1. Difficulties in learning rational numbers

There is accumulating evidence that even after considerable mathematics instruction many children fail to perform adequately even in simple rational number tasks (e.g., Siegler, Thompson, & Schneider, 2011; Siegler et al., 2013; Stafylidou & Vosniadou, 2004). As regards the origins of these difficulties, research on mathematics education suggested that most of students’ difficulties with rational numbers can be attributed to inadequate instruction (Vamvakoussi & Vosniadou, 2010) that do not adequately take the recent developments in numerical cognition into account.

According to conceptual change theories, children form an initial conception of natural numbers as counting units before they encounter fractions and decimals. As a result, later on they draw heavily on this initial understanding of number magnitude to make sense of rational numbers (e.g., DeWolf & Vosniadou, 2015; Stafylidou & Vosniadou, 2004). The associated phenomenon called whole number bias originates from people’s false belief that all properties of natural numbers can also be applied to rational numbers (e.g.,  $1/3 + 2/5$  may be solved incorrectly by summing numerators and denominators, i.e.,  $1 + 2/3 + 5 = 3/8$ , cf. Ni & Zhou, 2005; Alibali & Sidney, 2015). Although an established understanding of natural numbers is crucial for the development of mathematical capabilities, the very same understanding of natural numbers may also interfere in mathematical reasoning when rational numbers are involved (Van Dooren, Lehtinen, & Verschaffel, 2015). The transition from a cardinal number system to a system that relies on relational properties is challenging and requires a considerable conceptual leap, in particular so, when new information to be learnt seems incompatible with existing conceptions (e.g., Siegler et al., 2011). Interestingly, the phenomenon of the whole number bias is not only observed in elementary and high school students, but also in adults and even in expert mathematicians (Alibali & Sidney, 2015).

The whole number bias has been found to cause difficulties in reasoning about the size of rational numbers (Van Hoof, Lijnen, Verschaffel, & Van Dooren, 2013), because children tend to treat rational numbers in terms of their whole number components. For instance, when comparing fraction magnitudes, children sometimes reason that the fraction that is constituted from the larger whole

numbers as denominator and/or numerator is the larger one (e.g.  $1/5$  is larger than  $1/3$ , because 5 is larger than 3, e.g., Alibali & Sidney, 2015), instead of comparing the relative magnitudes of the fractions. In the domain of decimal numbers, children were found to misinterpret the number of digits to the right of the decimal point to be indicative of the size of the number (e.g. 0.125 is larger than 0.7, because 0.125 contains 3 digits and 0.7 contains only one digit) as it would be for comparing natural numbers (e.g., 125 to 7, cf. Durkin & Rittle-Johnson, 2015). In fact, comparison tasks are a common way of detecting whole number bias. However, in order to be successful in estimating the magnitude of a rational number or comparing the magnitudes of different rational numbers, children need to understand that not all whole number properties can be applied to rational numbers.

Furthermore, children struggle to comprehend that multiple rational numbers actually refer to the same numerical magnitude or the very same position on a number line, respectively (Vamvakoussi & Vosniadou, 2010). In other words, rational number magnitudes can be represented in an infinite number of ways (e.g.,  $0.5 = 0.50 = 1/2 = 2/4 = 3/6 = \dots$ ). In this context, children need to understand that rational numbers do not have unique predecessors and antecessors just like natural numbers do. Thus, children need to develop an understanding of the density of rational numbers. In fact, children seem to have problems understanding that there is an infinite number of rational numbers between any two consecutive integers, such as zero and one (Hansen et al., 2017; McMullen et al., 2015).

Hecht and Vagi (2010) argued that in the early stages of developing fraction knowledge two forms of conceptual interpretation are most relevant. Learning fractions usually starts by interpreting part-whole relations, in which a fraction is understood as a part of an object or a subset of a group of objects. Part-whole understanding is typically taught with area models such as pie charts (e.g., Padberg, 2015). The second type of understanding is the measurement interpretation of fractions which reflects the cardinal size of the fractions, as for instance reflected by its position on a number line. This corresponds to the fact that fractions can be ordered according to their numerical magnitude (e.g.,  $1/4, 3/6, 7/8, 4/4$ ) – also requiring understanding of density properties of rational numbers. That is, rational numbers do not have a discrete sequence that can be reproduced. It has been hypothesized that the ability to choose or construct appropriate units is considered fundamental for the development of conceptual rational number knowledge, in particular, for the understanding of fractions as a measure (Vamvakoussi, 2015). In line with this, Fuchs et al. (2013) developed an intervention facilitating the understanding of fractions in which the measurement interpretation of fractions focused on representing, comparing, ordering, and placing fractions on a 0 to 1 number line. In general, the measurement interpretation is often indicated by using number lines. Importantly, the part-whole interpretation of fractions (e.g. pie chart) considers the concept of natural numbers (counting discrete pieces of a pie) and may thus compromise the understanding of fractions or rational numbers, respectively (for a review see Siegler & Braithwaite, 2017). Moreover, interventions on fraction understanding emphasizing measurement and number line interpretations of fractions are typically more effective than interventions emphasizing part-whole interpretations of fractions (e.g., Fuchs et al., 2013, 2014, 2016).

The number line estimation task has been widely used to assess as well as train individuals' representation of number magnitude (e.g. Fazio et al., 2016; Link et al., 2013; Ninaus, Kiili, McMullen, & Moeller, 2017; Siegler & Opfer, 2003). Performance in this task is not only associated with actual mathematical performance (e.g., Link, Nuerk, & Moeller, 2014) but can also predict future mathematical achievement (e.g., Booth & Siegler, 2006). Therefore, recent empirical studies emphasized that children's understanding of number magnitude can be enhanced by training to map numbers (including rational numbers) onto space as in the number line estimation task (e.g. Schneider & Stern, 2010; Siegler & Ramani, 2008). However, in number line estimation tasks children's ability to simultaneously consider the low (e.g. 0) and high ends (e.g. 1) of the scale (number line) and their relation to the number or fraction being estimated (e.g.  $7/9$ ), is probably also influenced by visual-spatial working memory as argued by Booth and Siegler (2006). Thus, not only conceptual knowledge and measurement knowledge in particular but also individual working memory capability may influence accuracy of number line estimates.

Importantly, for the case of fraction learning, Siegler et al. (2011) argued that number line estimation based training may be used to overcome whole number bias and thereby improve the conceptual measurement interpretation of fractions. However, the part-whole interpretation approach is usually emphasized in educational systems of many countries (Fuchs et al., 2013; Padberg, 2015). Yet, the disadvantage of focusing on part-whole interpretation is that it does not support the development of the density concept and children may also struggle to understand that rational numbers can be larger than one.

## 1.2. Aims and hypotheses

Despite increasing research interest in the domain of rational numbers, most existing studies are still focused on training students' basic numerical and arithmetical understanding of whole numbers. However, given the widespread difficulties children as well as adults face when dealing with rational numbers synced with their observed predictive value for later mathematics achievement (cf. Siegler et al., 2013 for a review), the current study aimed at evaluating the effectiveness of a game-based rational number training application called "Semideus School". A recent study demonstrated that performance when playing the Semideus game version that was designed for assessment purposes i) provided reliable and valid information about students' conceptual fraction knowledge in a summative assessment context (Ninaus et al., 2016; Ninaus et al., 2017) and ii) was comparable to paper-based assessments (Kiili & Ketamo, 2017). In the current article we report the results of a short game-based intervention ( $5 \times 30$ -min training) that was designed to train conceptual rational number knowledge at the beginning of rational number instruction in schools (i.e., Grade 4 in Finland). Several aspects of students' conceptual rational number knowledge were addressed in the game: magnitude understanding, equivalence, density, and different rational number representations.

Based on above literature review on rational number knowledge and the aforementioned aspects of our game-based training, we derived the following hypotheses: First, the game-based training should improve conceptual knowledge of rational numbers as

compared to performance changes of a control group (Hypothesis 1a). Moreover, we expected that the game-based training should enhance all four measurement aspects of conceptual rational number knowledge (i.e., estimation, comparing, ordering, and density understanding, Hypothesis 1b). Second, students who progressed further in the game should have benefited more from the game-based training, because they trained their conceptual rational number knowledge more intensively (Hypothesis 2a). Additionally, in-game metrics, such as overall game performance and achieved batches (i.e., stars and coins) should be correlated to and able to predict conceptual rational number knowledge after the training (Hypothesis 2b). Third, students' math achievement as indicated by their previous math grade should be positively correlated with their conceptual rational number knowledge in pre-as well as posttest scores (Hypothesis 3).

## 2. Mapping rational number research to game mechanics

The design of the game-based rational number training used in this study was based on seven key principles derived from recent rational number and serious games research. i) The core game mechanics were chosen to corroborate the measurement interpretation of rational numbers (Siegler et al., 2011). ii) Game mechanics also addressed magnitude comparison tasks that have been widely used to assess the whole number bias (Alibali & Sidney, 2015). iii) Fractions and decimals were introduced in parallel rather than sequentially as they have been introduced in most of the Finnish primary school math books. iv) The Game provided immediate feedback that should further support understanding of the learning content (Dunwell & de Freitas, 2011). v) Game mechanics were directly integrated into the learning objectives (Kiili, Devlin, & Multisilta, 2015). vi) Scaffolding features were provided for the players (e.g. Baalsrud Hauge, et al., 2015). vii) Success in the game reflected players' knowledge of conceptual rational numbers rather than chance (Shute & Kim, 2014).

In order to facilitate overcoming whole number bias, we based the core gameplay on tasks that require working with rational number magnitudes (Siegler et al., 2011), i.e. number line estimation, magnitude comparison and magnitude ordering. We implemented number lines as walkable platforms on a mountain. In the game, the player controls an avatar called Semideus, who tries to collect gold coins that a goblin stole from Zeus. Semideus has to discover the locations of the hidden coins, encrypted in mathematical symbols that are linked to number lines, order rational number stones according to their magnitudes, and must race the goblin to the mountaintop to retrieve all the coins.

Next we will describe the details of the different tasks types included in the game followed by the description of level progression and the main features of the Semideus school game.

### 2.1. Number line estimation tasks

The game included estimation tasks on number lines ranging from 0 to 1 and from 0 to 5. Fig. 1 (top left chart) illustrates an example of an estimation task in which the player has to locate the decimal number 0.90 on the number line to dig up hidden coins. The orange bar on the right side of the screen indicates virtual energy of the player. An estimation task may involve traps that need to be avoided by jumping over them (Fig. 1 top right chart; trap at position 0.33; coins at position 2/3). The location of the coins may be indicated as a symbolic fraction or decimal, a pie chart or a bar graph. Number line based bar graphs were implemented to bridge the understanding between part-whole interpretation and measurement interpretation of fractions (Fig. 1: bottom left chart). As can be seen, the magnitude represented by the pie chart (1/5) is also presented as a bar graph in a way that the number line is split into two horizontal lines depicting numerator (i.e. one green line section) and denominator (i.e. five white line sections) of a fraction separately. Moreover, estimation tasks may also include different kinds of graphical landmarks such as lianas, torches, rocks, or hatch marks that a player may use as help for estimation (lianas in Fig. 1: bottom left chart). However, numerical landmarks were shown only as feedback after the correct answer or last, i.e. third, attempt (Fig. 1: bottom right chart; visualization of fractions on the number line 1/7, 2/7, 3/7, etc.). because research has indicated that numerical landmarks can negatively influence the accuracy of estimations in certain situations (Siegler & Thompson, 2014).

For inaccurate estimates (i.e., estimates more than  $\pm 8\%$  away from the correct location) the avatar was struck by lightning and the player lost virtual energy. The player had three attempts on each task. When the location corresponding to the respective rational number was estimated accurately the player acquired 100 to 500 coins depending on the degree of correctness (i.e., over 98% = 500 coins; 95%–98% = 300 coins; 92%–94% = 100) and at the same time the respective accuracy percentage was shown. Moreover, after successful estimation or after the third inaccurate answering attempt, the correct location of the coin cache was shown by a green marker on the number line.

Because students' rational number estimation accuracy is usually lower on a 0 to 5 number line than on a 0 to 1 number line (e.g., Torbeyns, Schneider, Xin, & Siegler, 2015), a 2% larger estimation error was allowed on 0 to 5 estimation tasks than on the 0 to 1 estimation tasks ( $\pm 10\%$  estimation error on 0 to 5 tasks). When estimation accuracy was better than 98% the player got 500 coins, when it was 95%–98% the player got 300 coins, and when the accuracy was 90%–94% the player got 100 coins. Estimation tasks on 0 to 5 number lines may include graphical whole number landmarks (lianas, rocks, torches, or hatch marks located at points 1,2,3,4) dividing the number line into five equal parts. This whole number approach was used, because Siegler and Thompson (2014) indicated that whole number landmarks can increase accuracy of number line estimates for fractions larger than 1.

### 2.2. Comparison and ordering tasks

In magnitude comparison and ordering tasks, players had to arrange stones in ascending order according with the numerical

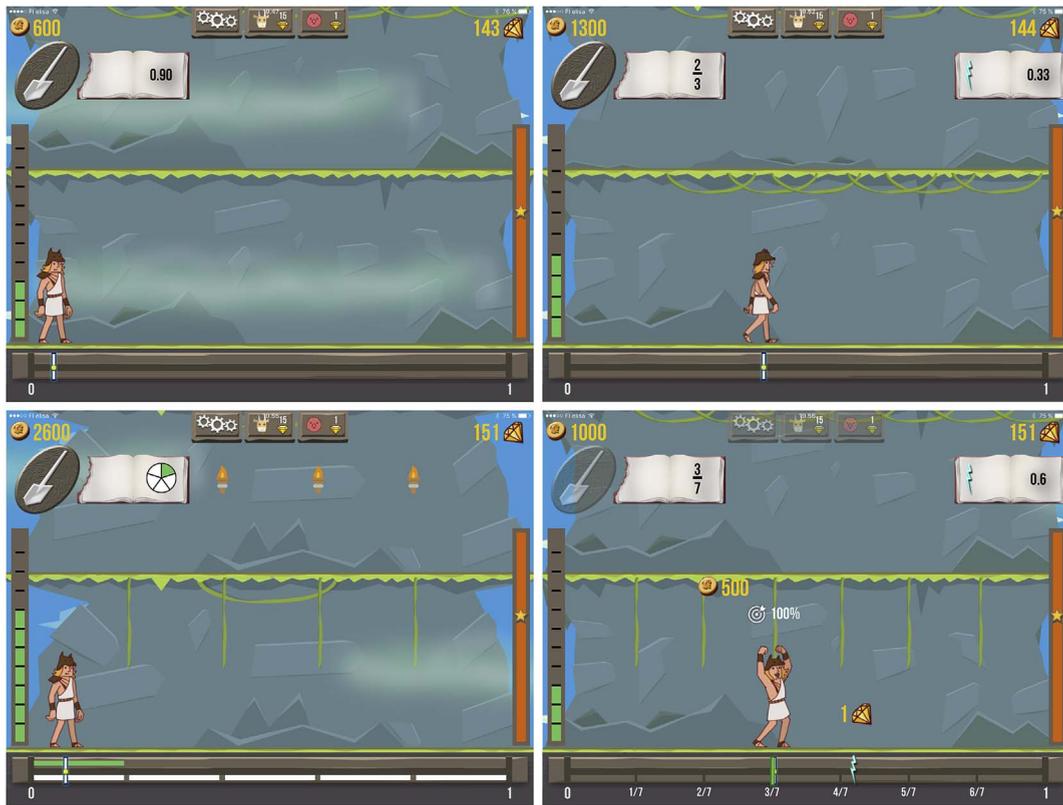


Fig. 1. Top left chart: An example of a number line estimation task in which the player should dig up a coin cache at the location reflecting 0.90. Top right chart: Example of a task in which the location of the coin cache is specified by a fraction and the location of a trap as a decimal number. Bottom left chart: Example of a number line estimation task that bridges the understanding between part-whole interpretation and measurement interpretation of fractions. Bottom right chart: Example of numerical landmarks as a feedback channel. Companion buttons (goat and bird) are located on top center of the screen and could be tapped in certain levels to receive help.

magnitudes depicted on them (Fig. 2: left chart). The exact location on the number line (ranging from smaller numbers on the left to larger numbers on the right) did not matter as long as the order of the stones was correct. Players were also able to pile up stones in case he or she thinks that the magnitudes were equivalent. Comparison tasks included two stones whereas ordering tasks included 3 to 4 stones. Some of the ordering tasks addressed density aspects of conceptual rational number knowledge. Basically, the density task is similar to the ordering task, but fractions or decimals involved in the task are selected specifically to induce cognitive conflicts between whole number and rational number properties (Fig. 2: right chart). For instance, when asked to order  $1/3$ ,  $2/3$ ,  $0.5$  and  $0.4$ , players using componential comparison strategies might conclude that there cannot be additional numbers between  $1/3$  and  $2/3$ , because 1 (nominator of  $1/3$ ) is followed by 2 (nominator of  $2/3$ ) in the context of whole number ordering. Therefore, such a strategy may result in a wrong answer to the task.

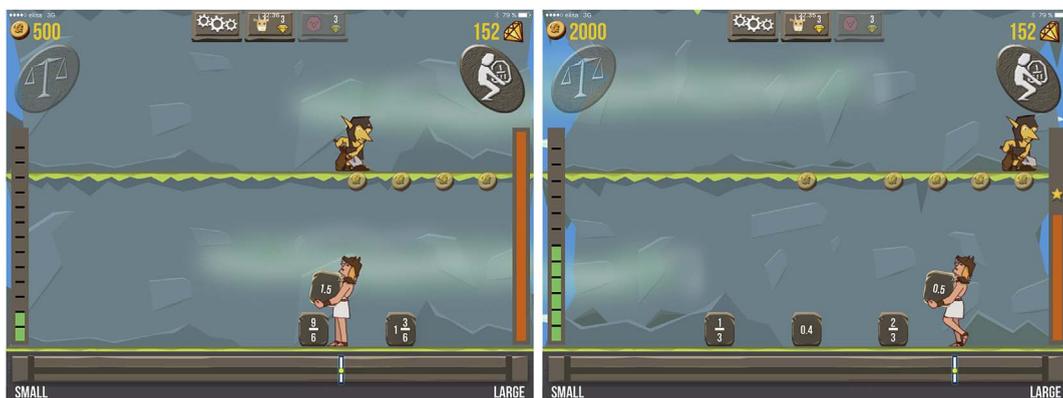


Fig. 2. Left chart: An example of an ordering task addressing equivalence. Right chart: An example of an ordering task addressing density.

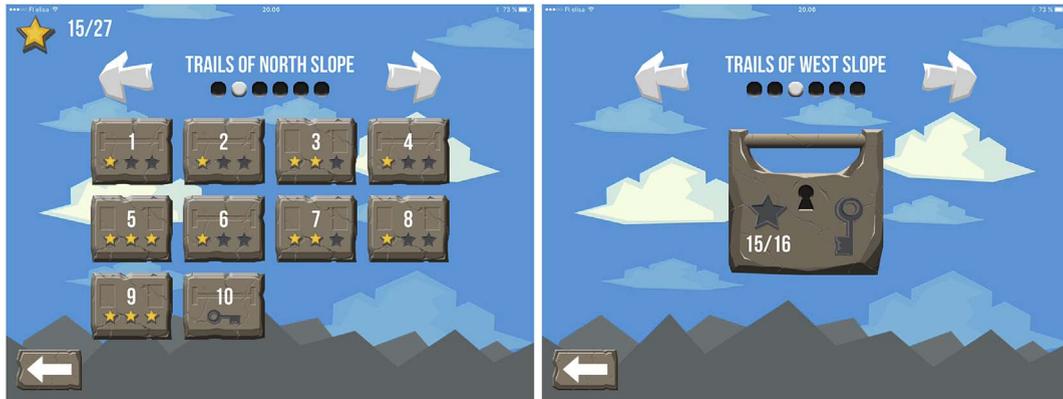


Fig. 3. Level menu of the game. Left chart: The player completed levels 1 to 9 of world 2 (Trails of north slope). Right chart: One more star and the key (acquired in key level 10) needs to be earned by the player to open world 3 (Trails of west slope).

In comparison and ordering tasks, the number of coins awarded to the player (100–500 coins) depended on the time taken to answer. The goblin was used to visualize the running of the time to the players. At the beginning of a task there were five coins on the platform (referring to a reward of 500 coins) and the goblin is digging up coins at a certain time interval. The aim of the time-based rewarding system was to create some tension and in this way maximize players' attention on game activities. However, the rewarding system was configured in a way that it provided players plenty of time to arrange the stones with the maximum number of coins (9500 ms). In previous studies the average comparison time was 6140 ms (Ninaus et al., 2017) and 6800 ms (Kiili & Ketamo, 2017). After the first coin the goblin keeps up digging coins in 4500 ms intervals. The goblin does not dig up the last coin and thus players always earned at least 100 coins in case they managed to arrange the stones in the correct order.

### 2.3. Level progression and learning analytics

The Semideus School beta version used in the current study featured six game worlds including 62 levels. Two of the worlds consisted of 11 levels and the rest consisted of ten levels. In the beginning, only the first level of the first world was available or open. Each level represents a trail to the mountaintop consisting of either 10 or 12 platforms (i.e. tasks). After completing a task, the player moves up to the next platform towards the mountaintop. In each level, the player had to perform well enough to reach the mountaintop and open the next level. The game was configured in a way that players were allowed to do (without dying) 5 mistakes in one level (i.e., 20 units virtual energy loss per mistake; 100 virtual energy units in the beginning). After completing a level (reaching the mountaintop) the player received additional feedback: 1 to 3 stars and earned coins were shown (i.e., one star for completing the level, one star for collecting enough coins, and one star for completing the level within the energy loss limit). Additionally, a bonus was given based on remaining energy (max bonus was 500). In order to open the next game world players had to pass each level of the current world and achieve 60% of the stars of the world. For instance, Fig. 3 shows a level menu in a situation in which the player has played through world 1 and has completed levels 1 to 9 of world 2 (Trails of north slope). As can be seen the player earned 15 out of 27 possible stars from the levels of world 2. However, the player still needs to earn one more star from world 2 and complete level 10 (key level) in order to open world 3 (Trails of west slope). The last level of each world is a key level that focuses on one of the main rational number competences trained in the other levels of the world. For example, the main competence of world 2 is fraction magnitude estimation on number lines ranging from 0 to 5.

In general, the difficulty level of the tasks gradually increased according to players' progress in the game. Appendix A provides a table that describes some design details and examples of tasks of the first three worlds (31 levels). In the first two levels, fractions were introduced with additional visualizations that bridge the understanding between part-whole interpretation and measurement interpretation with number line based bar charts (Fig. 1: bottom left chart). Later on in the game such area models are used only as hints presented for example after the second incorrect answer on a task. Moreover, some fraction estimation tasks of the first world were randomly represented as pie chart instead of symbols. Graphical landmarks such as torches and lianas were available during the so-called onboarding phase (first five estimation levels). After the fifth level landmarks were shown only after the first incorrect attempt.

Players could track their performance on a personalized statistics/analytics page. Accuracy on different kinds of number line estimation tasks is shown in bar graphs and correctness in different kinds of comparison and ordering tasks in pie charts. Additionally, the game generated personal hints for the player with respect to identified misconceptions (Fig. 4). In the used game version simple accuracy based rules were used to identify misconceptions. For example, a hint addressing the meaning of a zero right after the decimal point was shown when players' comparison accuracy in such comparison tasks (e.g. 0.08 vs 0.8) was lower than 50%.

### 2.4. Playing companions

Players also had two playing companions in the game, a goat and a bird. Players could use diamonds (in-game currency) that they

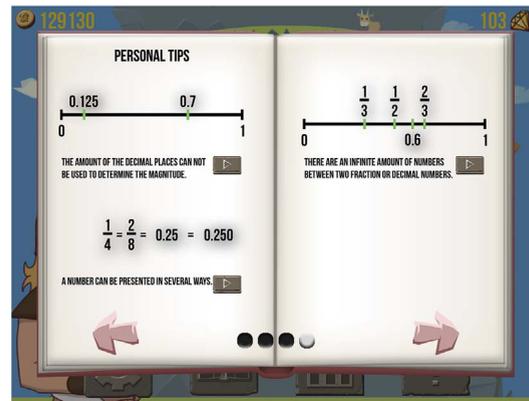


Fig. 4. Personalized statistics/analytics page and examples of personal hints generated according to identified misconceptions.

earned previously to demand help from these companions. Diamonds could be earned by avoiding traps. In estimation tasks, the goat would show the right location of the coin cache to the player. The price of this feature is set high (i.e. 15 diamonds) in order to avoid that players can use it frequently. In comparison and ordering tasks, the goat drives away the goblin so that players can use as much time as they need to arrange the stones (cost: 3 diamonds). In estimation tasks, players can use birds as markers in the same way as one might draw hatch marks to the number line when one is solving paper-based estimation tasks (cost: 1 diamond). Additionally, in some comparison tasks the player had the possibility to ask the bird to expand or reduce the fractions to common denominators (cost: 3 diamonds). Yet, the game also included several levels in which companions were not available and players had to cope on their own. When the playing companions were available the companion buttons were visible and the player could activate them by tapping either the goat or the bird button located on top center of the screen (see Fig. 1).

### 3. Method

A quasi-experimental between subject group design involving a game-based training group and a control group was used to evaluate the effectiveness of the game-based training of conceptual rational number knowledge. Participants of both groups were recruited from the same school in Finland. The beta version of the Semideus School game was used as a treatment and a paper-based rational number test as pre- and posttest (see below for a more detailed description). The Semideus training was part of students' regular math classes. The control group did not receive any training of rational numbers but had the opportunity to complete the Semideus training after the study. The study was approved by local school authorities. Parents were informed about the study and they approved all used data gathering methods including the logging of game behavior in aggregated form by providing written informed consent.

#### 3.1. Participants

Five Finnish fourth grade classes participated in the study. Three of the classes formed the training group that played the Semideus rational number game. Originally 68 students were recruited for the training group from which 54 (mean age = 10.24 years;  $SD = 0.43$ ; 25 males) followed the requested protocol (i.e. played the game and participated in both the pre- and posttests). 45 students were recruited for the control group, from which 41 (mean age = 10.02 years;  $SD = 0.27$ ; 25 males) participated in both the pre- and posttest. Math achievement was measured by participants' previous math grade (following the Finnish classification scheme 10 reflects the best and 4 the lowest grade). Math grades of the training ( $M = 8.20$ ;  $SD = 0.88$ ) and control group ( $M = 8.07$ ;  $SD = 0.96$ ) did not differ significantly [ $t(93) = 0.690$ ,  $p > .05$ ]. Before the study both groups only had some introductory lessons about rational numbers focusing mainly on part-whole interpretation of fractions and symbol notation of fractions and decimals.

#### 3.2. Measures

##### 3.2.1. Pre- and posttest

For pre- and posttest assessment of conceptual rational number knowledge we used the same test. This rational number test consisted of 28 items. The test consisted of four types of problems: Estimation of rational number magnitudes, comparison of rational numbers, ordering of rational numbers, and density of rational numbers. The maximum score students were able to achieve was 30.

The test included 6 estimation tasks from which four were on 0 to 1 number line (with target numbers:  $3/5$ ,  $2/3$ , 0.25, 0.75) and two were on 0 to 5 number line (target numbers:  $7/5$ , 3.75). Each item was scored as correct or incorrect (10% estimation error as a limit) with a maximum score of six for the estimation part of the test.

All 14 comparison items were multiple-choice questions including i) five items comparing two fractions (e.g. "Indicate the larger fraction. When the numbers are equal indicate both.", e.g.,  $2/9$  vs.  $1/3$ ), ii) four items comparing two decimals (e.g. "Indicate the larger decimal. When the numbers are equal indicate both.", e.g., 1.66 vs. 1.125), and iii) five items comparing fractions and decimals

(e.g. “Indicate the larger number. When the numbers are equal indicate both.”, e.g.,  $7/10$  vs.  $0.7$ ). Items were designed to address both, whole number ordering consistent (e.g.,  $3/4$  vs.  $1/4$  and  $2.65$  vs.;  $2.79$ ) and inconsistent comparisons (e.g.,  $2/9$  vs.  $1/3$  and  $0.4$  vs.  $0.40$ ) as well as different representations (e.g.,  $0.25$  vs  $3/4$ ) and equivalence (e.g.,  $3/6$  vs  $0.5$ ). Each item was scored as correct or incorrect with a maximum score of 14 for the comparison part of the test.

Ordering items required short answer responses including i) four items ordering fractions (e.g. “Put the numbers in order from smallest to largest”:  $3/3$ ;  $1/4$ ;  $6/8$ ) and ii) two items ordering decimals (e.g. “Put the numbers in order from smallest to largest”:  $4.782$ ;  $4.3$ ;  $4.94$ ). Each item was scored as correct or incorrect with a maximum score of six for the ordering part of the test.

Finally, the test also included two density items in the short answer format (e.g. “How many numbers are there between  $0.7$  and  $0.8$ ? Select the most suitable answer.” a) None, b) Numbers:  $0.71$ ,  $0.72$ ,  $0.73$ ,  $0.74$ ,  $0.75$ ,  $0.76$ ,  $0.77$ ,  $0.78$ ,  $0.79$ , c) Infinite number of decimal numbers., d) Infinite number of numbers including decimal and fraction numbers.). Each density item was scored as incorrect (answer a: 0 points), partially correct (answer b: 1 point), or correct (answers c and d: 2 points). Correct responses were those that displayed the infinite nature of rational numbers. Partially correct responses displayed some understanding that there are numbers between the two rational numbers. Incorrect responses displayed no understanding of the density of rational numbers, stating that there are no numbers in between the asked two numbers. The maximum score for the density part of the test was four.

### 3.2.2. Corsi Block-Tapping task

To assess whether children's visuo-spatial working memory may influence their accuracy in number line based tasks as suggested by Booth and Siegler (2006) and thereby confounding training results, students in the game-based training condition also completed a custom-made digital version of the Corsi Block-Tapping Task. The Corsi Block-Tapping test (Schellig, 1993) assessed the “immediate block span”, which is associated with visual short-term memory capacity. The task was performed on an iPad and consisted of nine blue rounded squares with white outline ( $105 \times 105$  px) on a black background ( $892 \times 717$  px). Participants performed the test individually. The test started from level two (sequence of two squares). A trial was presented as a sequence of yellow-flashing squares (1000 ms interval). After the sequence ended, participants had to reproduce the sequence by tapping the squares in the correct order. When a square was tapped, it lit up to confirm that the device detected the tap. Before each trial the number of blocks and the trial number was shown as text, for example in level three, “3 blocks will appear. First round starts.”. An onscreen text (“Repeat”) signaled the end of the sequence and instructed the participant to repeat the sequence in the same serial order. There were three trials per sequence of the same length. In case, at least two of these were repeated correctly, participants levelled up and the next three trials of a sequence of an increased length were generated, otherwise the test was terminated. When participants had repeated the sequence the program provided textual and spoken feedback (“Right”/“Wrong”). Sequences for level two were adopted from the study of Capitani, Laiacina, and Ciceri (1991) and sequences for levels 3 to 9 were the same as reported in Schellig (1993). The Total Corsi Score that equaled the number of successfully repeated sequences was used in the analyses.

### 3.2.3. In-game metrics

The Semideus School game logged players' playing behavior as aggregated scores on a secured server according to a semantic model used to describe all the tasks of the game. Based on the semantics, each task was tagged with keywords that describe the task in terms of rational number competencies and properties. For example, the comparison task,  $2/9$  vs.  $1/3$ , was tagged with the following tags: comparison, fraction, whole number inconsistent, small distance, same side (both fractions are located on the same half of a 0 to 1 number line), and unit fraction included. The learning analytics engine of the game allows to fetch competence data based on the tags. The following in-game metrics were used in the analysis of this study:

- *Overall game performance* referred to the percentage of correctly solved number line estimation, comparison, and ordering tasks.
- *Estimation correctness*, *Comparison correctness*, and *Ordering correctness* indicated the percentage of correctly solved tasks. Estimation correctness of the tasks on 0 to 1 number line was based on  $\pm 8\%$  error limit and on 0 to 5 number line on  $\pm 10\%$  error limit.
- *Effective playing time* reflected the summed up time that a player took to complete all tasks. For each task the effective playing time was computed as time of players' answers minus time of showing the tasks. Thus, it excluded time associated with processing feedback, proceeding to the next platform/task, navigating in the game menus, and looking at playing statistics.
- *Maximum level achieved* referred to the highest level that a player has successfully completed. Maximum level could be 62 at most.
- *Number of played games* referred to the total number of games that the player played. It included games that the player won (reached the mountaintop) or lost (ran out of energy before reaching the mountaintop), but excluded unfinished games that players quitted manually before running out of energy or before reaching the mountaintop.
- *Star ratio* was computed as (stars a player earned/stars that a player could have earned) \* 100.
- *Collected coins* indicated the sum of all coins a player collected in the game over all the playing sessions.
- *Number line estimation accuracy* was computed as  $100 \cdot \text{abs}(\text{correct value} - \text{estimated value}) / \text{the numerical range of the number line}$ .

### 3.3. Procedure

The study was conducted during a four-week period. For the *game-based training group* the current study took seven math lessons ( $1 \times 20$  min pretest -  $5 \times 30$  min recommended playing sessions -  $1 \times 20$  min posttest). In the first lesson, students completed the pretest. Prior to the pretest, the experimenter explained the different task types of the test to students. Before the playing sessions,

students of the training group also performed the Corsi Block-Tapping task. In the second lesson, the experimenter introduced the training game to students. The experimenter explained the structure of the game to students and showed how to play the number line estimation, comparison, and ordering tasks (i.e., the game screen was projected on a wall). The experimenter also explained the details of the task types including the logic of right and wrong answers. After the introduction phase, students started to play the game individually in the classroom with their iPads (model: iPad Air). However, students were allowed to discuss with their classmates about their in-game progress during the playing sessions. The experimenter administered the first playing session and ensured that every student understood how the game works. The remaining four playing sessions were administered by the teacher. Although the game includes sound effects, it was played in mute state because students did not have headphones. Students did not get additional teaching on rational numbers at school, except the game, over the time course of the study. Students completed the posttest the day after the fifth playing session. The experimenter administered the posttest.

The *control group* completed only pretest and posttest administered by the experimenter. There were approximately four weeks between the two tests. The control group attended regular math lessons between tests which did not include any specific teaching on rational numbers.

## 4. Results

### 4.1. General playing behavior

This section provides an overview of students' playing behavior that helps to interpret the results presented in the next section focusing on the effectiveness of the game-based training. Only one student was able to finish all 62 levels. In fact, the aim was that most of the students would manage to play through the first four game worlds (41 levels) that address all the content of pre-and posttest. However, on average students only proceeded to level 21 (min 9; max 62). Despite this, game logs showed that all students tried to progress in the game indicating that the game may have got too difficult for some of the students as they had to earn 60% of the maximum number of stars of each world in order to open a new world. On average students achieved 73% ( $SD = 9\%$ ) of stars of completed worlds and collected 171941 ( $SD = 78392$ ) coins.

Quizzing players after the posttest further indicated that most of the students were motivated to achieve all stars in each level. Moreover, teachers reported that a great deal of students were discussing about their in-game progress (current level and achieved coins) after the playing sessions. Average effective playing time was 110 min ( $SD = 41$  min).

### 4.2. Effectiveness of the game-based training

To evaluate the differential effectiveness of the game-based training on students' conceptual rational number knowledge we conducted a multivariate analysis of variance (MANOVA) with the between-participant factor training condition (game-based training vs. control). Dependent variables were gain scores ([posttest score] - [pretest score]) of the rational number test subscores for i) estimation tasks, ii) comparison tasks, iii) ordering tasks, and iv) density tasks. The MANOVA on gain scores for conceptual knowledge of rational numbers revealed a significant main effect of the training condition [*Pillai-trace* = 0.225,  $F(4,90) = 6.53$ ,  $p < .001$ ,  $\eta_p^2 = 0.23$ ],<sup>1</sup> supporting our Hypothesis 1a. For subsequent univariate analyses, Bonferroni corrections for multiple comparisons were applied. Results suggested that this differential overall training effect seemed to stem from significant effects of the training on the estimation tasks [ $F(1,93) = 15.91$ ,  $p < .001$ ,  $\eta_p^2 = 0.15$ ] and the ordering task [ $F(1,93) = 8.73$ ,  $p < .01$ ,  $\eta_p^2 = 0.09$ ], supporting our Hypothesis 1b only partly. For these tasks large (estimation task) and medium (ordering task) sized effects were observed indicating that the game-based training group improved more strongly than the control group (see Fig. 5). Moreover, a marginally significant improvement in the density tasks was observed for the game-based as compared to the control group [ $F(1,93) = 3.53$ ,  $p = .06$ ,  $\eta_p^2 = 0.04$ ]. No significant difference was found for the comparison tasks. Note, however, students in the game-based training group showed significantly better performance than those in the control group already at pretest [ $t(93) = 6.46$ ;  $p < .001$ ].

Moreover, as a second step to follow up the MANOVA, we used a discriminant function analysis with the gain scores (estimation, comparison tasks, ordering, and density tasks) as dependent measures. The discriminant analysis revealed exactly one significant discriminant function [*Wilks-Lambda* = 0.775,  $\chi^2(4) = 23.194$ ,  $p < .001$ ; see also Fig. 6.]. The dependent measures allowed to classify participants as belonging to the game-based training or the control group with an classification accuracy of 68.4%. Standardized canonical discriminant function coefficients indicated that the discriminant function was mainly driven by the gain in the estimation task ( $b = 0.771$ ) and to a lesser degree by the ordering ( $b = 0.411$ ) and density ( $b = 0.419$ ) tasks. Smaller contributions resulted from the comparison task ( $b = 0.096$ ).

### 4.3. Detailed examination of the game group

#### 4.3.1. Exploring predictors of training success

We used correlation analyses to identify relations between in-game metrics and gain of conceptual rational number knowledge

<sup>1</sup> Please note that results did not change when also considering math grade as a covariate [group main effect *Pillai-trace* = 0.222,  $F(4,89) = 6.34$ ,  $p < .001$ ,  $\eta_p^2 = 0.22$ ; math grade effect *Pillai-trace* = 0.02,  $F(4,89) = 0.47$ , n.s.].

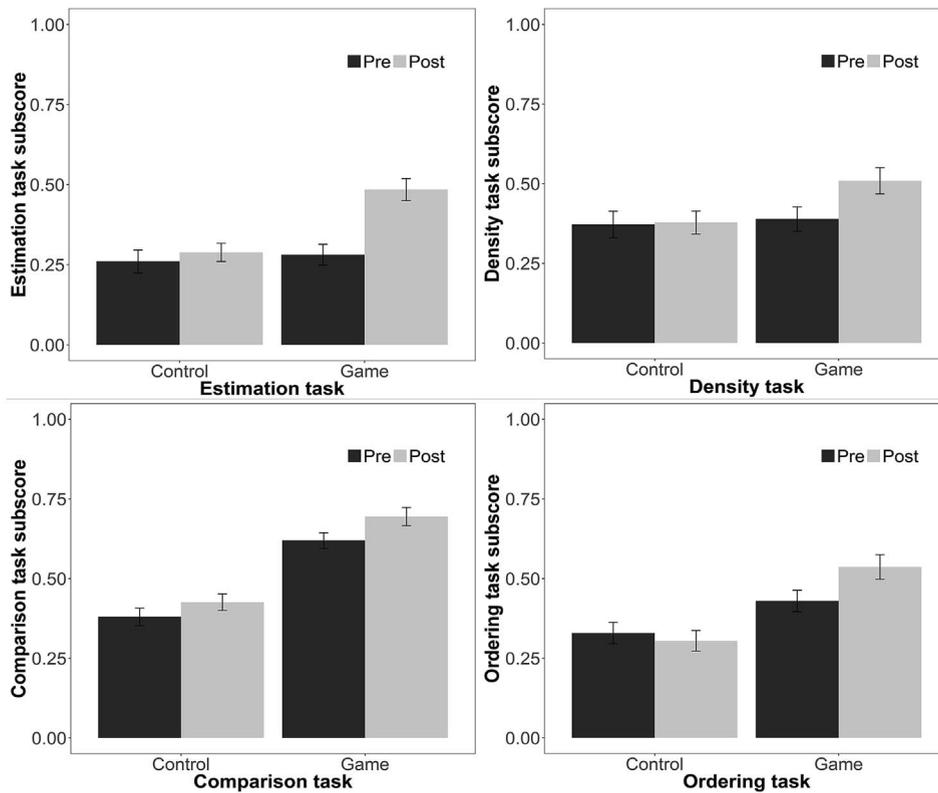


Fig. 5. Students' performance in different task types in the pre- and posttest for the game and the control group. On the y-axes mean correctness (0–1) on each of the task-types is represented. Error bars depict 1 standard error of the mean.

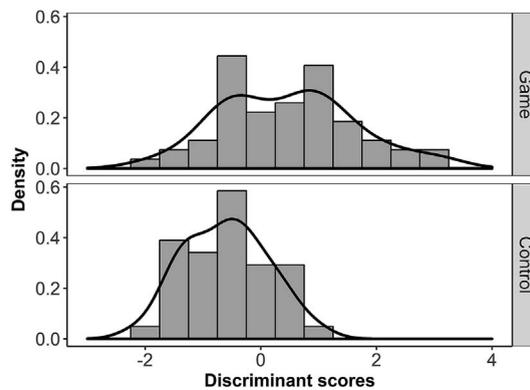


Fig. 6. Density histogram visualizing the separation achieved by the linear discriminant scores. On the x-axis discriminant scores of the only significant linear discriminant function are represented.

(see Table 1). Interestingly, effective playing time [ $r(52) = 0.00$ ; n.s.] and the mere number of played games [ $r(52) = 0.10$ ; n.s.] did not correlate significantly with the overall gain score. However, the maximum level achieved [ $r(52) = 0.36$ ;  $p < .01$ ] did correlate significantly with the gain score indicating that overall improvement in conceptual rational number knowledge was higher in students who proceeded further in the game supporting Hypothesis 2a. This seems reasonable, because students had to perform well to progress in the game. Moreover, those students who reached world four in the game got training on all competency areas that the pre-/posttest was measuring. This way the maximum level reached in the game reflects both the scope of the faced learning content and success in the rational number tasks.

Moreover, in line with our expectations math achievement (Hypothesis 3), as assessed by students' previous math grade, correlated positively and significantly with both the pretest [ $r(52) = 0.63$ ,  $p < .001$ ] and the posttest score [ $r(52) = 0.54$ ,  $p < .001$ ].

Importantly, we did not observe a significant correlation between the Corsi Block Score, assessing visuo-spatial working memory, and correctness of the number line estimation of the pretest [ $r(52) = 0.17$ , n.s.] or with in-game number line estimation accuracy [ $r$

**Table 1**

Correlations between control variables (CV), paper-based metrics (PBM), and in-game metrics (IGM) of the game-based training group.

		1	2	3	4	5	6	7	8	9	10	11	12	13
CV	1 Math grade	1												
	2 Corsi score	.38 <sup>a</sup>	1											
PBM	3 Pretest score	.63 <sup>a</sup>	.35 <sup>a</sup>	1										
	4 Posttest score	.54 <sup>a</sup>	0.21	.80 <sup>a</sup>	1									
	5 Est. correctness (pretest)	.43 <sup>a</sup>	.17	.76 <sup>a</sup>	.59 <sup>a</sup>	1								
	6 Gain score	.03	-.12	.07	.60 <sup>a</sup>	-.02	1							
IGM	7 Overall game performance	.58 <sup>a</sup>	.26	.79 <sup>a</sup>	.78 <sup>a</sup>	.54 <sup>a</sup>	.24	1						
	8 Star ratio	.19	.13	.36 <sup>a</sup>	.33 <sup>b</sup>	.22	.07	.58 <sup>a</sup>	1					
	9 Collected coins	.14	.24	.41 <sup>a</sup>	.43 <sup>a</sup>	.37 <sup>a</sup>	.17	.44 <sup>a</sup>	.50 <sup>a</sup>	1				
	10 Effective playing time	.09	.17	-.02	-.01	-.15	.00	.03	.20	.40 <sup>a</sup>	1			
	11 Est. accuracy (game)	.42 <sup>a</sup>	.20	.70 <sup>a</sup>	.73 <sup>a</sup>	.46 <sup>a</sup>	.27 <sup>b</sup>	.88 <sup>a</sup>	.47 <sup>a</sup>	.47 <sup>a</sup>	-.00	1		
	12 Max level achieved	.44 <sup>a</sup>	.40 <sup>a</sup>	.64 <sup>a</sup>	.72 <sup>a</sup>	.49 <sup>a</sup>	.36 <sup>a</sup>	.63 <sup>a</sup>	.34 <sup>b</sup>	.71 <sup>a</sup>	0.21	.54 <sup>a</sup>	1	
	13 Number of played games	-.02	.24	.05	.10	.07	.10	-.14	.05	.63 <sup>a</sup>	.47 <sup>a</sup>	-.14	.49 <sup>a</sup>	1

<sup>a</sup> Correlation is significant at the 0.01 level (2-tailed).<sup>b</sup> Correlation is significant at the 0.05 level (2-tailed).

(52) = 0.20, n.s.]. This is not in line with the assumed influence of visuo-spatial working memory on estimation accuracy claimed by Booth and Siegler (2006). However, it indicates that inter-individual differences in visuo-spatial working memory may not explain differential learning effects in our game-based training group. Further support for this interpretation comes from the non-significant correlation between Corsi Block Score and the overall gain score [ $r(52) = -0.12$ ; n.s.]. However, we observed positive correlations between Corsi Block Score and the maximum level [ $r(52) = 0.40$ ,  $p < .01$ ], pretest score [ $r(52) = 0.35$ ,  $p < .01$ ], and math grade [ $r(52) = 0.38$ ,  $p < .01$ ]. Analyses using Corsi Block Span instead of Corsi Block Score did yield identical results.

#### 4.3.2. Usefulness of basic in-game metrics in assessing conceptual rational number knowledge

To evaluate how well basic in-game metrics reflected students' conceptual rational number knowledge assessed with the paper-pencil based posttest correlation analyses were run as a first step (addressing Hypothesis 2b). These analyses are particularly relevant for investigating the applicability of our game-based environment as an assessment tool. As shown in Table 1, of the basic in-game metrics overall game performance, estimation accuracy, ratio of earned stars, collected coins, and maximum level reached correlated significantly with the posttest score, suggesting that these metrics provide important information about students' conceptual rational number knowledge supporting Hypothesis 2b. However, effective playing time and the number of played games did not correlate significantly with the posttest score.

Because these different in-game metrics reflect different aspects of the game such as correctness in the respective tasks, scope of the completed learning content and invested effort, a combination of in-game metrics might reflect students' conceptual rational number knowledge better than any single metric. Thus, in a second step, a forced-entry multiple regression analysis was run to predict students' conceptual rational number knowledge, assessed by the posttest score, based on overall game performance, maximum level achieved, star ratio, and number of coins earned. In-game estimation accuracy was excluded from this analysis, because it seemed to cover similar aspects as overall game performance (as indicated by their very high intercorrelation of  $r = 0.88$ ;  $p < .001$ ).<sup>2</sup> Results of the regression analysis indicated that overall game performance (standardized  $\beta = 0.62$ ,  $p < .001$ ) and maximum level (standardized  $\beta = 0.46$ ,  $p = .001$ ) explained 70% of the variance ( $F(4,49) = 31.71$ ;  $p < .001$ ; adjusted  $R^2 = 0.70$ ). Coins earned (standardized  $\beta = -0.10$ , n.s.) and star ratio (standardized  $\beta = -0.13$ , n.s.) did not account for a unique part of the variance of conceptual rational number knowledge. This indicated that better posttest conceptual rational number knowledge was predicted by better overall game performance as well as a higher maximum level reached.

## 5. Discussion and limitations

In this study we aimed at evaluating the effectiveness of a game-based training of conceptual rational number knowledge and investigating whether in-game metrics provided meaningful information on conceptual rational number knowledge and its acquisition. In the following, we will first discuss the results concerning the effectiveness of the game-based training before we will elaborate on in-game metrics and their association to training success and conceptual rational number knowledge.

<sup>2</sup> Including in-game estimation accuracy in the regression did not change the results of this analysis [ $F(5,48) = 25.62$ ,  $p < .001$ ; adjusted  $R^2 = 0.70$ ]. Importantly, in-game estimation accuracy did not account for a unique part of the variance of conceptual rational number knowledge beyond overall game performance (standardized  $\beta = 0.17$ , n.s.).

### 5.1. Game-based training partly improves conceptual rational number knowledge

A first objective of the current study was to evaluate the general improvement of students' conceptual rational number knowledge through game-based training. With respect to conceptual knowledge, we focused on the measurement interpretation of rational numbers because it was argued to be a crucial mechanism in explaining the development of rational number competences (e.g. Fuchs et al., 2013; Geary et al., 2008). Consistent with Hypothesis 1a and in line with previous studies (e.g. Fazio et al., 2016), fourth graders overall conceptual knowledge of rational numbers improved significantly with our game-based training as compared to performance differences of the control group. Interestingly, this differential overall training effect stemmed from specific beneficial effects of the training on rational number magnitude estimation and ordering tasks (Hypothesis 1b). In contrast, no specific training effect was observed for the comparison task while the improvement in the density tasks tended to be more pronounced for the game based training group as compared to the control group. Moreover, in line with this, gain in the estimation, ordering and density tasks contributed most to differentiating participants into the game-based training or the control group.

Students' performance in the comparison task did not improve differentially between groups. One might speculate that this may be due to the fact that students in the game-based training group showed considerably better performance than those in the control group already at pretest. This means that there was less possibility for improvement for the game-based training group - making the observation of a differential training effect less likely. It might also be possible that comparison performance of the game-based training group did not improve in comparison tasks because students did not proceed far enough in the game and thus they did not get enough training on all rational number aspects that the comparison items of the test were measuring. However, this seems unlikely, because even though students did not proceed as far as expected, estimation and comparison tasks were almost equally frequent within the average 21 levels played. To be sure, future studies should consider modifying the level design in a way that all players will get training on each aspect of the game irrespective of their proceeding speed. Moreover, because the game-based training extensively drew on number lines as basic learning/game mechanics training gains may be expected to be more pronounced for rather close evaluation tasks such as number line estimation. Additionally, one may also expect stronger training gains in those evaluation tasks, which benefit specifically from visualizations by means of a number line such as the ordering as well as the density task. In contrast, for simply comparing the magnitudes of two fractions, a number line may be less helpful, in particular because both comparison tasks in the game-based training as well as the tasks used in pre- and posttest did not require to specify the absolute magnitudes of the to-be-compared numbers but only their relative position to each other.

Contrary to recent claims (Booth & Siegler, 2006), we did not find significant correlations between visuo-spatial working memory and number line estimation performance. However, it seems that children with higher visuo-spatial working memory were able to reach higher levels in the game-based intervention, performed better in the pretest and had better math grades. As working memory is an important domain-general predictor for learning (e.g. Pickering, 2006), it is not implausible that higher Corsi Block Scores come with better performance in the game-based training. Moreover, working memory was also found to be a reliable predictor for mathematical skills (e.g. Gathercole, Alloway, Willis, & Adams, 2006) and may therefore be associated with the better math grades as well as higher performance in the pretest observed for students with higher visuo-spatial working memory capacity.

Taken together, the results of the current study indicated the effectiveness of a game-based training for aspects of fourth-graders' conceptual rational number knowledge on the measurement interpretation of rational numbers (Hecht & Vagi, 2010); thereby, extending the literature on rational number learning. This is particularly relevant as game-based interventions are used often in educational settings (e.g. Boyle, Hainey, Connolly, Gray, Earp, Ott et al., 2016). However, evaluations of game-based interventions are largely missing so far, which makes it difficult for teachers and educators to choose an appropriate intervention for their students. Furthermore, previous studies indicated that conceptual rational number knowledge is an important predictor of actual and future math achievement in general (e.g., Siegler et al., 2013 for a review). In line with this state of affairs, we not only observed a significant association between students' previous math grade and conceptual rational number knowledge as assessed by our pre- and posttest (Hypothesis 3) but also with in-game metrics such as overall game performance, estimation correctness, and the maximum level achieved. The former indicates that performance in pre- and posttest was higher for students with better math grades, which is in line with the identified link between actual mathematical performance and number line estimation (e.g., Link et al., 2014). The latter finding is important as it highlights that the observed relevance of conceptual rational number knowledge in mathematics education is also reflected by in-game metrics of our training. In the next section we will discuss influences of these in-game metrics in more detail.

### 5.2. In-game metrics as indicators for training success and conceptual rational number knowledge

A second aim of the current study was to better understand the improvement in conceptual rational number knowledge through our game-based training by evaluating students' playing behavior in more detail by considering specific in-game metrics. Contrary to our expectations, students only proceeded to level 21 of the game on average even though we tried to design the levels in a way that they should manage to play through the first 41 levels (4 game worlds) during the training in order to address all the content assessed by the evaluation tasks in pre- and posttest. However, only three students played through the first four game worlds. As a

consequence of our level design, most of the students solved more tasks including fractions than tasks including decimals. Moreover, most of the students did not get training on all rational number aspects that the game included. Therefore, future studies should better balance the occurrence of the different task types as well as the occurrence of different rational number aspects in the level design.

Digital learning environments should not only be beneficial to the learners, but also to teachers and educators. Accordingly, we were interested in to what extent in-game metrics may provide indications for training success as well as assessing and predicting students' conceptual rational number knowledge. In line with this idea, we observed that overall game performance and the maximum level achieved (Hypothesis 2a) were associated significantly with students' conceptual rational number knowledge. Additionally, these in-game metrics predicted substantial variance of fourth graders conceptual rational number knowledge (Hypothesis 2b). Interestingly, however, other in-game metrics such as star ratio and collected coins failed to explain unique additional variance. Importantly, we also replicated the positive correlation between in-game estimation accuracy and math grades identified in another recent study using the *Semideus* game (Ninaus et al., 2017). Therefore, in-game metrics seem to be indicators that teachers and educators can easily access and they might use them to evaluate students' performance during math classes or specific training. Moreover, using in-game metrics to assess conceptual rational number knowledge also seems to lower test anxiety compared to paper-based testing (Kiili & Ketamo, 2017). In summary, it seems that the level progression approach of our game-based training, in which levels have to be completed one by one based on playing performance not only worked well but also provided critical information about students' conceptual rational number knowledge as well as training gains. Therefore, in-game metrics should be considered informative and valid predictors of training success in future studies on game-based learning.

### 5.3. Limitations and perspectives

Although our results are well in line with theoretical assumptions on the training of conceptual rational number knowledge, our study also comes with some limitations that should be noted.

First, we used a control group who attended regular math classes (i.e., treatment as usual) while our training group completed the game-based training. Thereby, we aimed at controlling for performance differences due to (i) passage of time between the pretest and posttest, (ii) possible retest improvements by employing the conceptual rational number knowledge test twice, and (iii) improvements due to regular mathematic lessons not related to our training. Importantly, however, this approach does not allow for the conclusion that a game-based training of conceptual rational number knowledge is more effective than a non-game based training of the same content. Therefore, it is important that future studies involve a control group completing a non-game based training on rational numbers to substantiate our findings and evaluate the hypothesis whether game-based elements are particularly helpful for improving conceptual rational number knowledge beyond effects of a non-game-based training.

Second, related to this, we did not investigate possible long-term or far transfer effects of our game-based training. Follow-up measurements as well as curricular-based or standardized measures of math achievement (including fraction understanding) should be considered for future studies to allow for evaluating the stability of the observed positive effects of our game-based training on conceptual rational number knowledge in students, potential long-term effects on students' overall math achievement, as well as transfer effects to other (numerical) skills.

Third, we only got permission from participating schools to use aggregated data of students' learning behavior (e.g., mean estimation performance). Importantly, however, this prevented us from examining the learning process in greater detail based on item-level log data. The analysis of such detailed data reflecting the learning process might provide more detailed information about students' learning trajectories and faced difficulties. Thus, future studies should exploit log data that the game produces on the item level by means of learning analytics to model the development of conceptual rational number knowledge in more detail.

Finally, the current sample of students was rather high achieving as indicated by their math grades. Thus, future studies should investigate students with lower mathematical achievement in particular to investigate whether game-based approaches might be specifically beneficial for those students.

## 6. Conclusion

In summary, in the present study we evaluated the effectiveness of a game-based training of conceptual rational number knowledge. In particular, our research engine, *Semideus*, employed number line estimation, magnitude comparison, and magnitude ordering tasks on rational numbers. Results clearly indicated the usefulness and effectiveness of our game-based approach to train and improve aspects of students' conceptual rational number knowledge. We observed that the game-based training group improved their conceptual rational number knowledge more strongly over the training period as compared to a control group. Besides the effectiveness of the training, its game-based implementation also allowed to evaluate whether in-game metrics are associated with training success but also predict conceptual rational number knowledge at the posttest. Importantly, we found that in-game metrics reliably predicted posttest performance. Additionally, the maximum level students achieved while playing the game was associated positively with training gain in conceptual rational number knowledge. Taken together, these results indicated that a game-based training is effective to increase aspects of conceptual rational number knowledge in fourth graders. Moreover, they also suggest that in-game metrics may support learning assessment by providing important indications for training success and students' general conceptual rational number knowledge.

Appendix A. Level design details from the first three game worlds

World	Level	Type	NL/ Range	Number types	Landmarks on attempt	Hint after attempt	Task examples	Emphasis
1	1	Est	0-1	F, pie charts	0	1	3/4 as a pie chart & NL based bar chart	Area models
1	2	Est	0-1	F = D	0	1	2/4 = 0.5 & NL based bar chart	*Equivalence
1	3	Est	0-1	F	0	1	3/4	
1	4	Est	0-1	F	0	1	1/6	
1	5	Est	0-1	F   F	0	1	5/6	
1	6	Com	0-1	F   dots	-	-	1/2 vs 2/2	Same denominators
1	7	Com	0-1	F   dots	-	-	1/5 vs 3/5   5/8 vs 1/8	Same denominators
1	8	Com	0-1	F   dots	-	-	2/7 vs 2/3   3/4 vs 3/8	Same nominators
1	9	Com	0-1	F   dots	-	-	3/4 vs 3/7   4/5 vs 4/9	Same nominators
1	10	Est	0-1	F   D	1	2	4/5   0.33	
1	11	Com: key	0-1	F	-	-	1/8 vs 3/8   3/9 vs 3/4	Same nominators or denominators
2	12	Est	0-1	F   D	1	2	3/8   0.2	
2	13	Est	0-1	F   D	1	2	1/5   0.125	
2	14	Ord	0-1	F   FD	-	-	2/3 vs 1/3 vs 3/3   0.8 vs 0.7 vs 3/4	Density
2	15	Est	0-5	F	1	2	2 3/4	Mixed numbers
2	16	Com	0-7	F	-	-	2/3 vs 1 1/3   6 2/8 vs 6 2/4	Mixed numbers
2	17	Est	0-5	F	1	2	8/7   9/2	Improper fractions
2	18	Com	0-4	F	-	-	15/4 vs 15/3   8/2 vs 6/2	Improper fractions
2	19	Ord	1-8	F   FD	-	-	3/1 vs 3/3 vs 3.0   1 1/4 vs 4/1 vs 5/4	Improper fractions, mixed numbers
2	20	Com	0-4	F   D	-	-	3/4 vs 4/9   3.62 vs 3.99	NNB in F
2	21	Est: key	0-5	F	-	-	7/2   4/7   23/6	Improper fractions
3	22	Est	0-1	F   D	1	2	5/7   0.71	
3	23	Est	0-1	F   D	1	2	7/8   0.375	
3	24	Est	0-1	F   D	1	2	9/10   0.88	
3	25	Com	0-4	D	-	-	1.88 vs 1.125   0.06 vs 0.6   0.25 vs 0.250	NNB in D, equivalence
3	26	Com	0-3	D	-	-	0.90 vs 0.775   1.500 vs 1.5   0.5 vs 0.7	NNB in D, equivalence
3	27	Est	0-5	D	1	2	4.25   1.2   3.5	
3	28	Est	0-5	D	1	2	0.22   1.67   3.25	
3	29	Com	0-1	F	-	-	3/6 vs 1/6   4/7 vs 4/9	Same nominators or denominators
3	30	Com	0-7	F   FD	-	-	0.4 vs 4/10   0.66 vs 3/4   1/3 vs 2/6	Equivalence
3	31	Order: key	0-6	FD, D	-	-	1.6 vs 1.7 vs 1.655   3/2 vs 1.5 vs 1 2/4	Equivalence, density

Notes. Est = estimation, Com = comparison, Ord = ordering, NL = number line, F = fraction, D = decimal, FD refers to tasks that includes both fractions and decimals, NNB = Natural Number Bias. Landmarks on attempt 0 means that graphical landmarks are visible immediately and 1 means that they are shown after one wrong answer. Hints are shown after wrong answers based on attempt. \*Equivalence is shown in task, but does not necessarily require active processing of equivalence.

## Acknowledgment

### Funding

The current research was supported by the Leibniz-WissenschaftsCampus "Cognitive Interfaces" and Leibniz-Competition Fund (SAW) supporting Manuel Ninaus (SAW-2016-IWM-3) and Academy of Finland (grant numbers 274240 and 289140) supporting Kristian Kiili.

## References

- ACME (2011). *Mathematical Needs: Mathematics in the workplace and in higher education*. The Royal Society.
- Alibali, M. W., & Sidney, P. G. (2015). Variability in the natural number bias: Who, when, how, and why. *Learning and Instruction, 37*, 56–61.
- Baalsrud Hauge, J., Stanescu, I., Arnab, S., Moreno, P., Lim, T., Serrano-laguna, A., et al. (2015). Learning analytics architecture to scaffold learning experience through technology-based methods. *International Journal of Serious Games, 2*(1)<http://doi.org/10.17083/ijsg.v2i1.38>.
- Bakker, van den Heuvel-Panhuizen, M., & Robitzsch, A. (2015). Effects of playing mathematics computer games on primary school students' multiplicative reasoning ability. *Contemporary Educational Psychology, 40*, 55–71.
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42*(1), 189–201. <http://doi.org/10.1037/0012-1649.41.6.189>.
- Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., et al. (2016). 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. <http://doi.org/10.1016/j.compedu.2015.11.003>.
- Capitani, E., Laiacona, M., & Ciceri, E. (1991). Sex differences in spatial memory: A reanalysis of block tapping long-term memory according to the short-term memory level. *The Italian Journal of Neurological Sciences, 12*(4), 461–466.
- 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.
- Devlin, K. J. (2011). *Mathematics education for a new Era: Video games as a medium for learning*. Florida: CRC Press.
- DeWolf, M., & Vosniadou, S. (2015). The representation of fraction magnitudes and the whole number bias reconsidered. *Learning and Instruction, 37*, 39–49.
- Dunwell, I., & de Freitas, S. (2011). *Four-dimensional consideration of feedback in serious games*. Continuum International Publishing Group.
- Durkin, K., & Rittle-Johnson, B. (2015). Diagnosing misconceptions: Revealing changing decimal fraction knowledge. *Learning and Instruction, 37*, 21–29.
- Fazio, L. K., Kennedy, C. A., & Siegler, R. S. (2016). Improving Children's knowledge of fraction magnitudes. *PLoS One, 11*(10), e0165243. <http://doi.org/10.1371/journal.pone.0165243>.
- Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., et al. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology, 105*, 683–700.
- Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Malone, A., Wang, A., et al. (2016). Effects of intervention to improve at-risk fourth graders' understanding, calculations, and word problems with fractions. *The Elementary School Journal, 116*(4), 625–651.
- Fuchs, L. S., Schumacher, R. F., Sterba, S. K., Long, J., Namkung, J., Malone, A., et al. (2014). Does working memory moderate the effects of fraction intervention? An aptitude-treatment interaction. *Journal of Educational Psychology, 106*(2), 499–514.
- Gabriel, F., Coché, F., Szucs, D., Carette, V., Rey, B., & Content, A. (2012). Developing children's understanding of fractions: An intervention study. *Mind, Brain, and Education, 6*(3), 137–146.
- Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A.-M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology, 93*(3), 265–281. <http://doi.org/10.1016/j.jecp.2005.08.003>.
- Geary, D. C., Boykin, A. W., Embretson, S., Reyna, V., Siegler, R., Berch, D. B., et al. (2008). In U.S. Department of Education (Ed.). *Report of the task group on learning processes* Washington, DC: U.S. Department of Education. The final report of the National Mathematics Advisory Panel <http://www.ed.gov/about/bdscomm/list/mathpanel/report/learning-processes.pdf>.
- Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon and Schuster.
- Gross, J., Hudson, C., & Price, D. (2009). *The long term costs of numeracy difficulties*. London: Every Child a Chance Trust and KPMG.
- Hansen, N., Jordan, N. C., & Rodrigues, J. (2017). Identifying learning difficulties with fractions: A longitudinal study of student growth from third through sixth grade. *Contemporary Educational Psychology, 50*, 45–59. <https://doi.org/10.1016/j.cedpsych.2015.11.002>.
- Hecht, S. A., & Vagi, K. J. (2010). Sources of group and individual differences in emerging fraction skills. *Journal of Educational Psychology, 102*(4), 843–859. <http://doi.org/10.1037/a0019824>.
- Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay? *Computers & Education, 51*(4), 1609–1620.
- Kiili, K., Devlin, K., & Multisilta, J. (2015). Editorial: Is game-based math learning finally coming of age. *International Journal of Serious Games, 2*(4)<http://doi.org/10.17083/ijsg.v2i4.109>.
- Kiili, K., & Ketamo, K. (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>.
- Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review, 22*(3), 215–243. <http://doi.org/10.1007/s10648-010-9125-8>.
- Link, T., Moeller, K., Huber, S., Fischer, U., & Nuerk, H. C. (2013). Walk the number line - an embodied training of numerical concepts. *Trends in Neuroscience and Education, 2*(2), 74–84. <http://doi.org/10.1016/j.tine.2013.06.005>.
- Link, T., Nuerk, H.-C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic competencies. *The Quarterly Journal of Experimental Psychology, 67*(8), 1597–1613. <http://doi.org/10.1080/17470218.2014.892517>.
- 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., Kiili, K., McMullen, J., & Moeller, K. (2016). A game-based approach to examining students' conceptual knowledge of fractions. *Proceedings of games and learning alliance conference (GALA 2016) - lecture notes in computer science, 10056, 37–49* Springer [http://doi.org/10.1007/978-3-319-50182-6\\_4](http://doi.org/10.1007/978-3-319-50182-6_4).
- Ninaus, M., Kiili, K., McMullen, J., & Moeller, K. (2017). Assessing fraction knowledge by a digital game. *Computers in Human Behavior, 70*, 197–206. <http://doi.org/10.1016/j.chb.2017.01.004>.
- Ni, Y., & Zhou, Y. D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist, 40*(1), 27–52.
- Padberg, F. (2015). *Didaktik der Bruchrechnung (Mathematik Primarstufe und Sekundarstufe I + II)* (4th ed.). Springer Spektrum.
- Parsons, S., & Bynner, J. (2006). *Does numeracy matter more?* London: National Research and Development Centre for Adult Literacy and Numeracy.
- Pickering, S. J. (2006). *Working memory and education*. Elsevier Academic Press.
- Riconcente, M. M. (2013). Results from a controlled study of the iPad fractions game Motion Math. *Games and Culture, 8*(4), 186–214.
- Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*, 175–189.

- Schellig, D. (1993). *Corsi block-tapping test. Wiener testsystem (Vienna test system)*. Mödling: Schuhfried.
- Schneider, M., & Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: A multimethod approach. *Developmental Psychology*, *46*, 178–192.
- Shute, V. J., & Kim, Y. J. (2014). Formative and stealth assessment. In M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.). *Handbook of research on educational communications and technology* (pp. 311–321). New York: Springer.
- Siegler, & Braithwaite (2017). Numerical development. *Annual Review of Psychology*, *68*, 187–213.
- Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., et al. (2012). Early predictors of high school mathematics achievement. *Psychological Science*, *23*(7), 691–697.
- Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2013). Fractions: The new frontier for theories of numerical development. *Trends in Cognitive Sciences*, *17*(1), 13–19. <https://doi.org/10.1016/j.tics.2012.11.004>.
- Siegler, R. S., & Opfer, J. E. (2003). The development of numerical Estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*(3), 237–243. <http://doi.org/10.1111/1467-9280.02438>.
- Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, *11*(5), 655–661. <http://doi.org/10.1111/j.1467-7687.2008.00714.x>.
- Siegler, R. S., & Thompson, C. (2014). Numerical landmarks are useful-except when they're not. *Journal of Experimental Child Psychology*, *120*(1), 39–58. <http://doi.org/10.1016/j.jecp.2013.11.014>.
- Siegler, R. S., Thompson, C., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology*, *62*(4), 273–296. <https://doi.org/10.1016/j.cogpsych.2011.03.001>.
- Stafylidou, S., & Vosniadou, S. (2004). The development of students' understanding of the numerical value of fractions. *Learning and Instruction*, *14*(5), 503–518. <https://doi.org/10.1016/j.learninstruc.2004.06.015>.
- Torbeyns, J., Schneider, M., Xin, Z., & Siegler, R. S. (2015). Bridging the gap: Fraction understanding is central to mathematics achievement in students from three different continents. *Learning and Instruction*, *37*, 5–13.
- Vamvakoussi, X. (2015). The development of rational number knowledge: Old topic, new insights. *Learning and Instruction*, *37*, 50–55.
- Vamvakoussi, X., & Vosniadou, S. (2010). How many decimals are there between two fractions? Aspects of secondary school students' understanding of rational numbers and their notation. *Cognition and Instruction*, *28*(2), 181–209.
- Van Dooren, W., Lehtinen, E., & Verschaffel, L. (2015). Unraveling the gap between natural and rational numbers. *Learning and Instruction*, *37*, 1–4.
- Van Hoof, J., Lijnen, T., Verschaffel, L., & Van Dooren, W. (2013). Are secondary school students still hampered by the natural number bias? A reaction time study on fraction comparison tasks. *Research in Mathematics Education*, *15*(2), 154–164.
- ter Vrugte, J., de Jong, T., Vandercruyse, S., Wouters, P., van Oostendorp, H., & Elen, J. (2017). Computer game-based mathematics education: Embedded faded worked examples facilitate knowledge acquisition. *Learning and Instruction*, *50*, 44–53 (Available online 23 November 2016).