




The added value of game elements: better training performance but comparable learning gains

Manuel Ninaus^{1,2,3}  · Rodolpho Cortez^{1,4} · Izabel Hazin⁴ · Kristian Kiili⁵ · Silke M. Wortha⁶ · Elise Klein^{7,8} · Elisabeth M. Weiss² · Korbinian Moeller^{3,6,9}

Accepted: 2 June 2023
© The Author(s) 2023

Abstract

Even though game elements can increase motivation and engagement, they also might distract learners and thereby decrease performance and learning outcomes. In the current study, we investigated the effects of intrinsically integrated game elements on performance and learning outcomes. In a pre-post training study, 85 adult participants were randomly assigned either to the game-based or non-game-based training condition. Participants trained their fraction magnitude understanding with digital number line estimation tasks on five consecutive days (à 15–20 min). The learning outcomes were evaluated using a paper-based number line estimation task. While participants in both the game-based and non-game-based condition improved their fraction magnitude understanding from pretest to posttest, their improvement did not differ significantly. However, during the training, participants in the game-based condition responded more accurately but were slower than those in the non-game-based condition. The current results suggest that game elements might increase participants' cognitive engagement and change their priorities or strategies (i.e., accuracy over speed) during learning. Nevertheless, better training performance did not lead to superior learning outcomes.

Keywords Game-based learning · Game elements · Learning · Math · Intrinsic integration · Engagement

Introduction

In the last ten years, the use of game-based learning has increased significantly (for a review, see Hwang & Chen, 2022). One line of research indicates that game-based learning (i.e., using game elements in learning materials) seems to encourage learners to invest more effort in an instructional task (e.g., Plass et al., 2015; for meta-analyses see Clark et al., 2016; Sailer & Homner, 2020), often referred to as engagement (Fredricks et al., 2004; Schwartz & Plass, 2020). This increased engagement might also improve learning outcomes, as demonstrated in meta-analyses (Sailer & Homner, 2020; Wouters et al., 2013). The use of game elements, however, might also be considered to be in contrast with

M. Ninaus and R. Cortez contributed equally to this work.

Extended author information available on the last page of the article

common multimedia learning principles, which typically suggest keeping learning material as minimalistic as possible to reduce the cognitive load and working memory demands (e.g., Mayer & Fiorella, 2014). That is, additional (game) elements embedded in the learning material might distract learners and/or divert attention away from the essential content of the learning material and thus impair learning and/or performance—such distractors are often referred to as seductive details (e.g., Rey, 2012). Although recent studies have shown important advances in this research field, more details about the use of game elements in learning material are still necessary (e.g., Zainuddin et al., 2020).

As such, the current study set out to evaluate whether the integration of game elements to an already established math learning task (i.e., number line estimation) moderates learning and performance. Accordingly, in the following, we will first provide a brief overview of different theoretical rationales arguing for and against the inclusion of game elements in learning tasks. Subsequently, we summarize and discuss different research design approaches in the field of game-based learning and briefly describe the theoretical basis of using the number line estimation task as a training tool within the context of basic research on numerical cognition.

Engagement, cognitive load, and emotional design in game-based learning

The use of game elements or game-based learning is often justified by increasing the engagement of learners (e.g., Garris et al., 2002; Greipl et al., 2021; Mekler et al., 2017; Ninaus et al., 2015; Plass et al., 2015). Typically, engagement refers to actively participating in a task or learning activity, as opposed to being disinterested, apathetic, or only superficially involved (Newmann, 1992). In other words, engagement is considered to be the active and focused investment of effort in a task or learning environment (Schwartz & Plass, 2020). While there may be some variations in how engagement is defined, the most common view is that it involves three dimensions: behavioral, emotional, and cognitive (Fredricks et al., 2004). Although these dimensions are separate, they are also interconnected. Here, we will focus on cognitive engagement, which is often referred to as the degree of mental investment that learners put into their learning activities, which also includes thoughtful and strategic thinking to overcome problems, master difficult skills, or comprehend complex ideas (Fredricks et al., 2004; Meece et al., 1998). Accordingly, cognitive engagement can be assessed by, for instance, the number of solutions generated or the amount of time spent on a task (e.g., Schwartz & Plass, 2020). In the context of game-based learning, game elements might moderate users' willingness to invest mental effort during task execution which leads to generative processing, i.e., cognitive processing aimed at making sense of the presented (learning) material (Mayer, 2019), and better performance. In game-based learning, different game elements or, following Bedwell's taxonomy (e.g., Bedwell et al., 2012), game attributes such as immersion, game fiction, and challenge can be used. When a learning task with game elements becomes a learning game or remains "only" a gamified learning task is still a matter of debate (Plass et al., 2020; Tokac et al., 2019) and is beyond the scope of the current article.

Previous research on emotional design demonstrated that adding game elements to learning and cognitive tasks can increase cognitive engagement, thereby modulating learning and performance (Bernecker & Ninaus, 2021s; Ninaus et al., 2015; Plass et al., 2015). However, using game elements may also have detrimental effects (for a review see Toda et al., 2018), such as lower scores in exams (de-Marcos et al., 2014), reduced motivation

(Hanus & Fox, 2015), or even distract learners from the actual learning task (Kocadere & Çağlar, 2015).

According to the seminal cognitive load theory (Sweller, 1988) and cognitive theory of multimedia learning (Mayer, 2005), any information entering our cognitive system is processed in working memory before it is transferred to long-term memory. As working memory resources are limited, learners' information processing capacity is limited and must be allocated among extraneous, essential, and generative processing (for a cognitive perspective on game-based learning, see Mayer, 2020). In this context, extraneous processing refers to cognitive processing that does not serve the instructional goal of the learning task or the game. Essential processing, on the other hand, is needed to hold and manipulate incoming information in working memory, while generative processing reflects making sense of this incoming information (learning material). Accordingly, instructional designers usually aim to optimize learning by decreasing extraneous processing, managing essential processing, and increasing generative processing. Importantly, it needs to be considered that this also means—at least according to the cognitive theory of multimedia learning (Mayer, 2005)—that although game elements increase extraneous processing, they may nevertheless foster generative processing by increasing engagement and the willingness to exert mental effort (Mayer, 2020). As such, a careful balance of game elements and instructional features is necessary not to hinder learning when using games. Intrinsic integration is a game design approach that is suggested to enhance learning (Habgood & Ainsworth, 2011; Kafai, 1996). It refers to the general idea that there must be an intrinsic association between the game's core mechanics and the learning content or learning mechanic (Habgood & Ainsworth, 2011). Further, according to the RETAIN model (Gunter et al., 2008) which was developed to design and evaluate educational games, the tight coupling of learning content and the game's fantasy and/or story (i.e., embedding content into the fantasy) is essential. In other words, learning effects might be enhanced if subject matter and learning content is reflected in the respective game actions in an engaging and pedagogically meaningful way. In line with this, Kiili et al. (2019b) have suggested that also learning domain-specific instructional knowledge is another relevant factor to be utilized in intrinsic integration. All in all, the aim is that intrinsic integration leads to an interaction that facilitates engagement and immersion in academic learning content (Gunter et al., 2008). However, there is emerging evidence that many learning games suffer from a lack of intrinsic integration between the learning content and the core game mechanics (Kiili et al., 2019b). It seems that either instructional or game design aspects dominate the design of learning games too strongly. However, the concept of intrinsic integration has not been studied extensively yet.

Nevertheless, Habgood and Ainsworth (2011) identified that an intrinsically integrated version of a math learning game improved learning significantly more than an extrinsically integrated version. The extrinsic game version was based on quiz mechanics, a convenient and thus popular way to integrate instructional material into a game. In this game version, participating children needed to answer a multiple-choice quiz solving divisions at the end of each game level in which zombie skeletons had to be defeated. In contrast, in the intrinsically integrated version of the game, math learning content was directly integrated into the game mechanics as players had to choose different types of attacks that represented a divisor (e.g., a single swipe of a sword to divide by 2, a shove with a shield to divide by 3) to defeat zombie skeletons with numbers printed on their chests by means of a division into equally sized portions. Interestingly, learning performance and long-term outcomes were better and children spent more time using the intrinsic version of the game.

In line with the cognitive theory of multimedia learning (Mayer, 2005), extrinsically integrated game elements might increase extraneous processing, as these game elements do not directly serve the instructional goal. In contrast, intrinsically integrated game elements or games might reduce extraneous processing and ensure that players process essential learning content, a prerequisite for learning. The desired motivational effects of intrinsically integrated game elements are assumed to facilitate generative processing. Further, one might speculate that intrinsic integration should also support managing essential processing if relevant instructional approaches are successfully utilized in the integration. However, research on this conceptual issue is sparse, particularly to what extent intrinsic integration interacts with essential processing goes beyond the current study and needs to be investigated with well-controlled experimental manipulations in other studies. For instance, studies might want to systematically manipulate the degree of intrinsic integration and assess the different types of cognitive processing.

Another issue of the game-based learning research field seems to be the lack of systematic investigations of the specific benefits of game elements (cf. Boyle et al., 2016). In particular, a meaningful comparison between game-based and “conventional” learning is not a trivial task to achieve. Often, a game-based learning approach is compared to regular classroom activities, such as listening to lectures or working with books and PowerPoint slides (for meta-analyses see Sailer & Homner, 2020; Wouters et al., 2013). Importantly, however, this ignores many factors such as the underlying instructional approach, feedback opportunities, interpersonal interaction, use of technology, or instructor bias (Chen et al., 2021; Liao et al., 2010; Mayer, 2014). Consequently, even when positive/negative effects are found for game-based learning, it is often not clear whether this originated from the use of a game or specific game elements or might have other reasons (e.g., the presence of real-time feedback in game-based learning vs. delayed or even no feedback in regular classroom teaching). Accordingly, a well-matched and thus well-comparable control condition needs to be implemented to investigate the effects of game-based learning more specifically.

Number line estimation as a basic learning mechanic

Fractions are considered a particularly challenging topic in mathematics instruction in schools (Benbow & Faulkner, 2008; Booth & Newton, 2012), but even adults frequently fail to process them correctly (for a review, see Siegler et al., 2013). One of the major difficulties when dealing with fractions is the actual understanding of fraction magnitude. Importantly, fraction understanding was observed to be predictive for later more complex mathematical skills such as algebra (e.g., Booth & Newton, 2012). A lack of proper fraction understanding often persists into adulthood (Stigler et al., 2010).

A (mental) number line is a frequently employed metaphor for the mental representation of number magnitude. Consequently, the so-called number line estimation task is an often-used approach to measure and train the understanding of (fraction) number magnitude in children (e.g., Fazio et al., 2016) as well as adults (e.g., Sidney et al., 2019a). While one might assume that the majority of research was done on children (cf. Schneider et al., 2018 for a review), age ranges vary considerably depending on the research question at hand, this means from children in kindergarten/preschool (e.g., Praet & Desoete, 2014) over primary (e.g., Link et al., 2013) and secondary school (Gross et al., 2018, see also Nuraydin et al., 2023 for evidence from large-scale data) to adults (e.g., Gallagher-Mitchell et al., 2018; Sullivan et al., 2011) and elderly (e.g., Greipl et al., 2020; Matthews, et al., 2022) or even brain-damaged participants (e.g., Mihulowicz et al., 2015).

In this well-evaluated and established task (Link et al., 2013; Siegler & Opfer, 2003), the spatial position of a target number (e.g., $3/4$) on a number line (e.g., ranging from 0 to 1) needs to be estimated (e.g., where goes $3/4$ on a number line from 0 to 1?). It was repeatedly found that learners become more accurate (e.g., Fazio et al., 2016) and/or faster in their responses (e.g., Obersteiner et al., 2013; Wilson et al., 2009) when estimating or comparing the magnitude of numbers with training (for a review see Moeller et al., 2015). Importantly, number line-based training has also been shown to be effective in improving (conceptual) magnitude understanding of fractions and rational numbers (Fazio et al., 2016; Gunderson et al., 2019; Opfer et al., 2016; Siegler et al., 2013; van't Noordende et al., 2016). Interventions using number line estimation as a basic learning mechanic have shown higher and better outcomes as compared to other instructional methods, such as area models (Hamdan & Gunderson, 2017). Moreover, compared to interventions working with part-whole and area models, only number line-based training approaches also led to transfer effects onto untrained tasks such as fraction magnitude comparisons and fraction arithmetic (Gersten et al., 2017; Hamdan & Gunderson, 2017; Sidney et al., 2019b). Consequently, number line estimation constitutes a well-established and validated learning task with simple learning mechanics.

Present study

The use of game elements or game-based learning in education needs to be investigated carefully (Plass et al., 2020). On the one hand, some theoretical approaches argue that the use of game elements might increase extraneous processing (e.g., Mayer, 2005; Sweller, 1988). On the other hand, approaches motivated by the emotional design perspective (e.g., Plass et al., 2015; for a meta-analysis see Wong & Adesope, 2021), argue for the use of game elements to increase different aspects of engagement. Furthermore, the way game elements are integrated or how learning content is reflected by the game's core mechanic (i.e., intrinsic integration) seems to be another important aspect to be considered in game-based learning research (e.g., Habgood & Ainsworth, 2011; Gunter et al., 2008; for a systematic review see Kiili et al., 2019b). Finally, a lack of appropriate control conditions in game-based learning research further impeded clear recommendations on the use of game-based learning so far (for meta-analyses see Sailer & Homner, 2020; Wouters et al., 2013).

Therefore, the present study employed a simple yet effective math learning task (i.e., number line estimation) as the basic mechanic when designing a learning game to allow for intrinsic integration. That is, we utilized number line estimation as the game's core mechanic and compared it to a conventional number line estimation task without game elements included. In particular, we compared differences in learning effects from a pretest to a posttest and performance during the training between a game-based version of the number line estimation task and its non-game-based equivalent.

Overall, we expected that both the game-based as well as the non-game-based version of the number line estimation training should yield significant learning effects. That is, participants in both training conditions should increase their fraction estimation accuracy from a pretest to a posttest using a paper-pencil version of the number line estimation task (Hypothesis 1) (Fazio et al., 2016).

On the other hand, as suggested from an emotional design and intrinsic integration perspective, we expected that learning and thus accuracy gains from pretest to posttest should be more pronounced for the game-based as compared to the non-game-based intervention (Hypothesis 2) (Habgood & Ainsworth, 2011; Wong & Adesope, 2021).

Finally, in line with the previous hypothesis and the assumption that the use of game elements leads to increased (cognitive) engagement (e.g., Bernecker & Ninaus, 2021; Plass et al., 2015), participants might invest more mental effort in solving the tasks in the game-based training. Accordingly, we expected that participants will perform better in the game-based version of the number line estimation during the actual training as compared to its non-game-based version (Hypotheses 3), as reflected by higher accuracy and faster response duration.

Methods

Participants

For this study, 90 students from a German university were recruited and randomly assigned to the game-based ($n=41$) or the non-game-based training ($n=49$). We excluded 5 participants from further analyses. Two participants (non-game-based training) were excluded because of extremely poor performance (more than 4 SD below the sample mean). Three participants (all in the non-game-based condition) were excluded because of problems with data acquisition (i.e., pretest items not recorded for one participant; training sessions missing for two participants).

Consequently, 85 adult students were considered in the analyses (game: $n=41$; non-game: $n=44$). All of these participants were German native speakers with normal or corrected to normal vision and reported no history of psychiatric or neurological disorders or drug abuse. Due to technical problems, records on the age of 37 participants were lost. The remaining participants had a mean age of 23.60 years ($SD=3.56$, range 18 to 33 years). As all participants were recruited from a student population, it may well be assumed that age should be distributed comparably across those with missing data. The local Ethics Committee approved the study. All participants provided written informed consent before the study and received monetary compensation.

Study design

The experiment was conducted as a pre-post training study with five consecutive days of training between the pretest and posttest. For each training session, which lasted for approximately 15–20 min, depending on the participants' performance, participants came to the lab at the university. The participants were randomly assigned to either the game-based ($n=41$) or the non-game-based ($n=44$) training condition. A paper-pencil-based number line estimation task was used as a pretest and posttest (see more below).

Training procedure

We used two different versions of a fraction number line estimation training (i.e., game-based vs. non-game-based training version). In both versions of the training, participants had to indicate the correct spatial position of a target fraction (e.g., $9/20$) on a number line ranging from 0 to 1. That is, through the training, participants should improve in

understanding the number magnitude of fractions. Both versions of the training employed the same task and numerical content. However, the game-based training utilized typical game elements (cf. Bedwell et al., 2012), such as visual aesthetics (i.e., game-like visual design), game fiction (i.e., game narrative), as well as a virtual incentive system (i.e., coins & life points). The non-game-based version of the training did not include any of these elements. The items in both versions were identical and used fractions with a numerator ranging from 1 to 25 and a denominator ranging from 2 to 30. Each training session employed the same 48 items which were completed twice (i.e., 96 trials) and in randomized order. In total, one training session consisted of 12 levels, each one involving eight items. Participants had to come to the lab for each training session and work on the items using a computer. Participants were instructed to answer as quickly and accurately as possible.

To evaluate potential performance improvement across all five training sessions, we calculated the mean Percentage Absolute Error (PAE; $[\text{estimate} - \text{actual location}] / \text{numerical range}] * 100$; cf. Booth & Siegler, 2006) and response duration for the first answer (time of item onset to response) on each trial for each session (see below). The following will give a more detailed description of the game-based and non-game-based training tasks.

Game-based number line estimation training

For the game-based training version, we used the *Semideus* rational number environment, which was already successfully utilized in previous studies for evaluating fraction knowledge in Finnish and German schoolchildren (Kiili et al., 2018a, 2018b; Ninaus et al., 2021).

In the game-based version, the number line from 0 to 1 was implemented and visualized as a walkable platform in the game (see Fig. 1). Regarding the game's fantasy context, participants controlled the character Semideus, who tries to recover gold coins that a goblin has stolen from Zeus and hidden along the trails of Mount Olympus. Semideus has discovered notes that mark (in the form of fractions) where the gold coins are hidden along the path. Regarding the core mechanic of the game, participants could move the avatar along the walkable platform—i.e., the number line—to the correct position with the help of the arrow keys of a computer keyboard. By walking on the path, participants could experience fraction magnitudes in a meaningful context (walking distance on the path). After reaching the estimated position, participants should press the space bar to lock that position and dig coins from the ground. Participants had 10 s to estimate the correct position. In case they failed to answer or answered incorrectly (i.e., 94% accuracy or less), the avatar was struck by lightning and lost virtual life points, which were represented by an orange bar on the right of the screen. For correct answers, participants found coins, i.e., earned points—the more precise their answer, the more points they earned (i.e., 100–99% accuracy would earn 500 coins, 98–97% accuracy would earn 300 coins, 96–95% accuracy would earn 100 coins). Moreover, as feedback, the correct position (100% accuracy) was shown with a vertical, green line (see also Fig. 1, Panel C). Participants had to estimate the location of each fraction/trial until it was estimated correctly before they could proceed to the next task. On the left of the screen, a green bar showed participants how far they had progressed in the level they were currently working on. At the end of each level, participants also received delayed feedback about their performance (3-star rating). For completing the level, they earned one star; for collecting at least 2000 points, they could earn a second star; and for maintaining at least 80% of the life points, they could earn a third star.

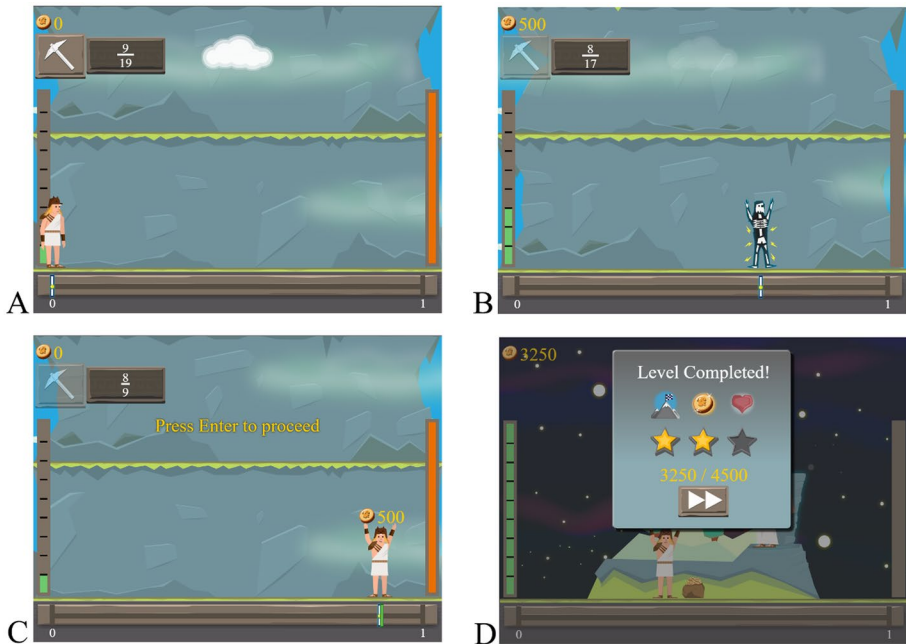


Fig. 1 Examples of the game-based version. The onset of the number line estimation task (A); negative feedback (B); positive feedback (C); overall feedback showed at the end of each level (D)

Non-game-based number line estimation training

For the non-game-based version, we again used the *Semideus* rational number environment but stripped from all game elements described above. That is, we created a conventional, digital version of the number line estimation task with minimalistic implementation. In particular, we used a simple grey background with a black number line from 0 to 1 (see Fig. 2). The non-game-based training did neither utilize the narrative of *Semideus* (fantasy context) nor any virtual incentives (i.e., coins, life points). Participants had to move a white slider along the number line to the position of the respective target fraction by using the arrow keys of the computer keyboard—identical to the game-based training. Again, participants had 10 s to estimate the correct position by pressing the space bar at the selected position. In case they failed to answer, or their estimation was not accurate enough (i.e., 94% accuracy or less), a red cross was shown, and—identical to the game-based training—they had to repeat the trial until it was solved correctly. For correct estimations (i.e., 100–95% accuracy), participants were rewarded by simply showing a green checkmark, and the correct position (100% accuracy) was shown with a vertical green line as feedback (see also Fig. 2).

Pre-post measures

Participants performed a paper–pencil version of a number line estimation task before (pre-test) and after (posttest) the training. The number line ranged from 0 to 1 with the start and endpoint specified and included 96 items/fractions. Participants had to indicate the correct

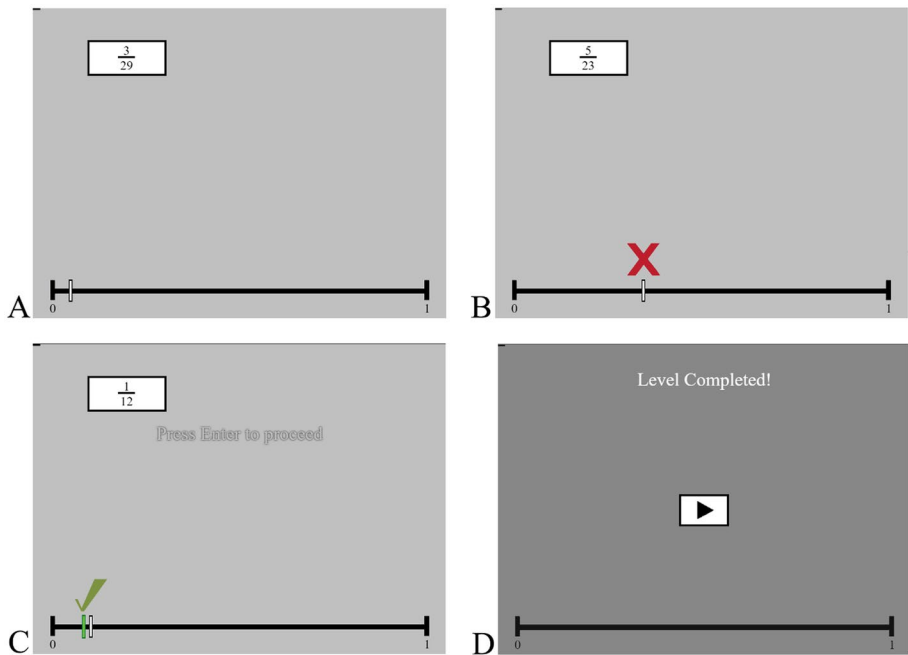


Fig. 2 Examples of the non-game-based version. The onset of the number line estimation task (A); negative feedback (B); positive feedback (C); overall feedback showed at the end of each level (D)

spatial position on the number line for all items. The items used comprised fractions with a numerator ranging from 1 to 25 and a denominator ranging from 2 to 30. That is, the items included one (e.g., $3/7$) and two-digit fractions (e.g., $7/27$). To investigate any potential improvements, we calculated the mean percentage absolute estimation error (PAE; Booth & Siegler, 2006).

Analyses

Data preprocessing and analyses were performed using R (R Version 4.0.5, R Core Team, 2021). Statistical requirements were checked and confirmed before running the planned analyses (i.e., linearity, homogeneity of variance, and normality of residuals) using the R package “performance” (Lüdtke et al., 2021). Prior to evaluating any potential training effects, possible performance differences between the two conditions (game-based vs. non-game-based training) at the pretest were examined using a Welch Two Sample *t*-test. A linear mixed-effects model was fitted to examine potential performance changes in PAE from pretest to posttest. Moreover, changes in PAE and response duration across the five training sessions were also examined by fitting separate linear mixed-effect models.

Linear mixed-effects models were fitted using “lmer” from the “lme4” R package (Bates et al., 2015). Linear mixed models provide various methodological and statistical advantages over mixed ANOVAs (Hilbert et al., 2019). Importantly, simulation studies show that linear mixed-effects models are very robust even if the distributional assumptions

are violated (Schielzeth et al., 2020). The p -values for linear mixed-effects models were calculated using Satterthwaite's degrees of freedom method available with the R package "lmerTest" (Kuznetsova et al., 2017). Marginal means were extracted using the "ggeffects" R package (Lüdtke, 2018). To support reporting, we used the "report" R package (Makowski et al., 2021). The R package ggplot2 (Wickham, 2016) was used for visualizing results.

Results

Pretest differences

The Welch Two Sample t -test evaluating the difference of PAE at pretest between the game ($mean=4.33$, $sd=1.56$) and non-game-based training condition ($mean=4.68$, $sd=1.20$) did not show a significant difference (difference=0.35%, 95% CI [-0.95, 0.26], $t(74.84)=-1.15$, $p=0.255$; Cohen's $d=0.25$, 95% CI [-0.68, 0.18]).

Pre-post differences—learning outcome

Differences in PAE from pretest to posttest were analyzed by fitting a linear mixed model (see also Fig. 3) using Restricted Maximum Likelihood [REML] to predict PAE with condition (i.e., game-based vs. non-game-based training) and time (pre vs. post) as fixed factors. The model also included a random intercept to account for participants' individual differences.

The model's total explanatory power was substantial (conditional $R^2=0.50$), and the part related to the fixed effects alone (marginal R^2) was 0.08. The effect of *time* was statistically significant ($beta=-0.54$, 95% CI [-0.91, -0.16], $t(164)=-2.81$, $p=0.005$; *Std.*

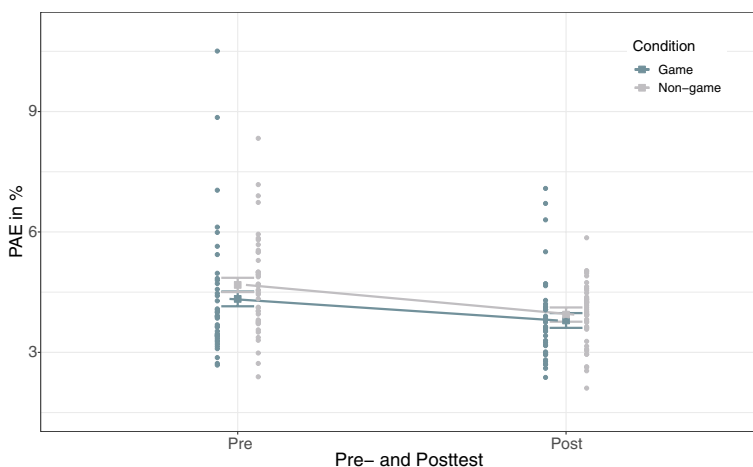


Fig. 3 PAE changes from pretest to posttest for game-based (petrol) and non-game-based conditions (grey). Points show the PAE of each individual. Squares represent the estimated marginal means. Error bars denote $1 \pm$ standard error

$\beta = -0.22$, 95% CI [-0.38, -0.07]) and indicated that PAE significantly decreased due to the training from pretest to posttest. Neither the effect of condition ($\beta = 0.35$, 95% CI [-0.15, 0.85]; $t(164) = 1.36$, $p = 0.173$; *Std. beta* = 0.20, 95% CI [-0.15, 0.55]) nor the interaction between time and condition was statistically significant ($\beta = -0.20$, 95% CI [-0.72, 0.32], $t(164) = -0.76$, $p = 0.449$; *Std. beta* = -0.08, 95% CI [-0.30, 0.13]).

Training session performance

Descriptive data on PAE and duration across all five training sessions are reported in Table 1.

PAE: Differences in PAE across the five training sessions were analyzed using a linear mixed-effect model (see Fig. 4). In particular, we fitted a linear mixed model using REML to predict PAE by condition (i.e., game-based vs. non-game-based training) and training sessions (i.e., five training sessions) as fixed factors. The model included a random intercept to account for participants' individual differences in baseline performance and a random slope to account for individual differences in learning rate.

The model's total explanatory power was substantial (conditional $R^2 = 0.76$), and the part related to the fixed effects alone (marginal R^2) was 0.17. The effect of the *training session* was statistically significant and indicated that PAE decreased significantly over training sessions and, thus, participants improved across training sessions ($\beta = -0.24$, 95% CI [-0.31, -0.17], $t(417) = -6.82$, $p < 0.001$; *Std. beta* = -0.30, 95% CI [-0.39, -0.21]). Furthermore, the effect of the *condition* was statistically significant ($\beta = 0.66$, 95% CI [0.18, 1.14], $t(417) = 2.71$, $p = 0.007$; *Std. beta* = 0.59, 95% CI [0.26, 0.92]) reflecting that participants in the game-based condition committed smaller PAE during the training. However, the interaction effect of session and condition was not statistically significant ($\beta = 0.00742$, 95% CI [-0.09, 0.10], $t(417) = 0.15$, $p = 0.881$; *Std. beta* = 0.00915, 95% CI [-0.11, 0.13]).

Duration: Differences in response duration across the five training sessions were analyzed using a linear mixed-effect model (see Fig. 5). In particular, we fitted a linear mixed model using REML to predict response duration by condition (i.e., game-based vs. non-game-based training) and training sessions (i.e., five training sessions) as fixed factors. The model included a random intercept to account for participants' individual differences in baseline response duration and a random slope to account for individual differences in learning rate.

Table 1 Overview of the means of PAE and Duration for each training session for both conditions

	PAE				Duration			
	Game ($n = 41$)		Non-Game ($n = 44$)		Game ($n = 41$)		Non-Game ($n = 44$)	
	M	SD	M	SD	M	SD	M	SD
Session 1	4.28	1.21	4.99	1.22	4694.33	1020.17	4252.14	961.52
Session 2	3.69	1.07	4.37	1.11	4527.80	823.41	4155.33	932.72
Session 3	3.49	1.05	4.06	0.86	4314.24	881.75	3862.6	825.47
Session 4	3.27	0.87	3.96	1.09	4190.68	891.06	3730.64	806.76
Session 5	3.27	0.82	4.01	0.95	3930.14	840.21	3513.21	706.50

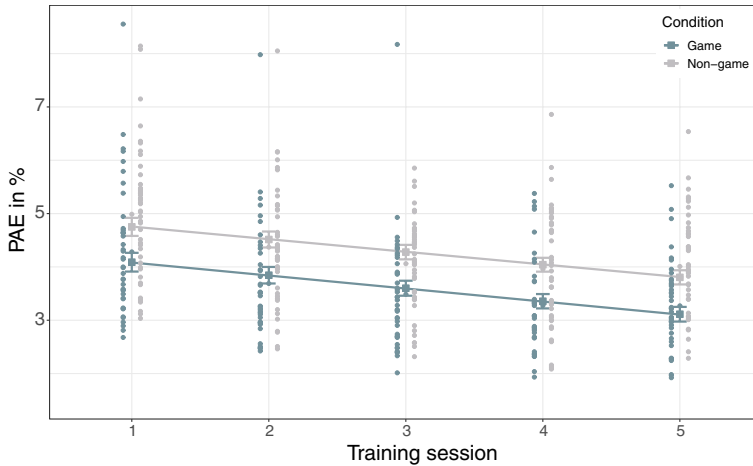


Fig. 4 PAE change across training sessions for game-based (petrol) and non-game-based condition (grey). Points show the PAE of each individual. Squares represent the estimated marginal means. Error bars denote $1 \pm$ standard error

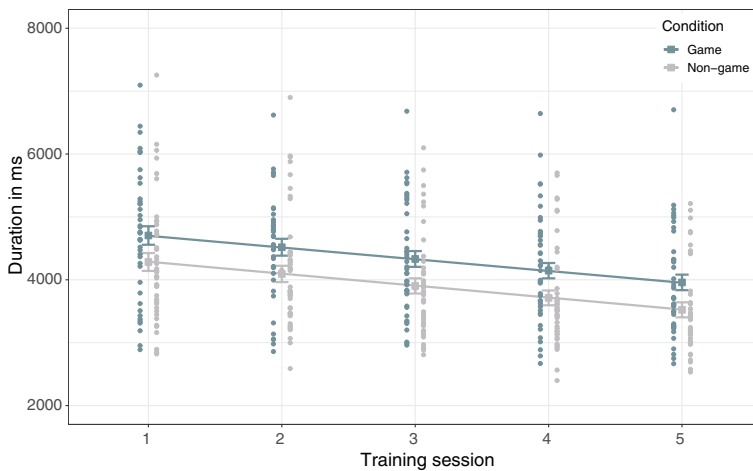


Fig. 5 Change of response duration across training sessions for game-based (petrol) and non-game-based condition (grey). Points show the response duration of each individual. Squares represent the estimated marginal means. Error bars denote $1 \pm$ standard error

The model's total explanatory power was substantial (conditional $R^2=0.89$) and the part related to the fixed effects alone (marginal R^2) was 0.13. The effect of training session was statistically significant indicating that participants became significantly faster across training sessions ($beta = -186.55$, 95% CI $[-237.79, -135.32]$, $t(417) = -7.14$, $p < 0.001$; *Std. beta* = -0.28 , 95% CI $[-0.36, -0.21]$). Furthermore, the effect of *condition* was statistically significant ($beta = -421.25$, 95% CI $[-822.39, -20.10]$, $t(417) = -2.06$, $p = 0.040$; *Std. beta* = -0.46 , 95% CI $[-0.83, -0.09]$) and indicated that participants in the game-based condition took more time to respond. Again, however, the interaction effect

of session and condition was not statistically significant ($\beta = -3.70$, 95% CI [-74.92, 67.51], $t(417) = -0.10$, $p = 0.919$; *Std. beta* = -0.00565, 95% CI [-0.11, 0.10]).

Discussion

The current training study compared learning effects for a game-based and an equivalent non-game-based version of a number line estimation task. The results indicated that participants successfully improved their fraction magnitude estimation from pretest to posttest, but no significant differences were identified between training conditions. Furthermore, looking specifically at the training sessions, results revealed that participants estimated fraction magnitudes more accurately in the game-based condition as compared to the non-game-based condition. However, estimation also took more time in the game-based condition compared to the non-game-based condition. In the following, these results will be discussed in more detail.

Pre-Posttest comparisons

In line with our first hypothesis, participants in both conditions improved their fraction estimation accuracy from a paper-based pretest to a posttest. This demonstrates that a digital number line estimation task is an effective instructional approach to foster fraction magnitude estimation skills—even in educated university students who already showed a rather high performance at the pretest. This result is in line with previous studies examining the overall effectiveness of number line-based fraction instruction in general (e.g., Gersten et al., 2017; Gunderson et al., 2019; Opfer et al., 2016; Siegler et al., 2013) and game-based instruction in particular (e.g., Fazio et al., 2016; Kiili et al., 2018a, 2018b).

However, in contrast to our second hypothesis, improvements in estimation accuracy from the pretest to the posttest were not significantly more pronounced for the game-based as compared to the non-game-based training. The lack of a statistically significant learning difference between task versions seemed to suggest those game elements—even though not beneficial—did not significantly hamper learning either. Therefore, it seems that game elements cannot always be considered seductive, or at least the potential negative effects were smaller than could be expected in light of the cognitive theory of multimedia learning (Mayer, 2005). The current game design was guided by an intrinsic integration approach. That is, the game events were tied to Greek mythology; the number line was implemented as the walkable platform (trails) on the Mount Olympus, and the avatar was seeking hidden coins on the trails. In other words, the learner could experience fraction magnitudes as specific distances on the trails. In the non-game-based version, the learner also could experience the magnitudes, but in a more abstract way, as the mountain scenery and the walking character were not available. This might be a reason that game elements did not disturb learners, which would substantiate the results of Habgood and Ainsworth (2011), who demonstrated the beneficial effects of intrinsic integration. However, in the current study, we did not compare an intrinsic and extrinsic version of the game. Therefore, to better understand the effects of intrinsic integration, future studies are necessary.

In a similar vein, Schneider et al. (2016), investigated the effect of decorative pictures and found that not all types of decorative pictures were seductive. Given the right design choices, this means, pictures that evoke motivational and/or (positive) emotional states, as

used in emotional design (for a meta-analysis see Wong & Adesope, 2021), can also be conducive to learning. In their systematic review on the use of game elements in cognitive assessment, Lumsden and colleagues (2016a) also highlighted that the integration of game elements into tasks needs to be done very carefully. The concept of intrinsic integration (Habgood & Ainsworth, 2011; Kafai, 1996) might be one way of achieving a balance between the potential positive (i.e., generative processing) and negative (i.e., extraneous processing) effects of game elements (cf. Mayer, 2020).

Nevertheless, the current implementation of game elements into a number line estimation task did not yield a significantly larger training effect than a conventional, non-game-based version of the same task. We need to note here though, that the overall performance of participants at the beginning of the training can be considered rather high and thus room for improvement over the course of the training was limited. Consequently, it is also rather unlikely to identify differential effects as the skill level was already high. That is, a study using more difficult fractions or less skilled participants might reveal different results. The study should be replicated with pupils who are less experienced with the use of fractions. Moreover, as the pre-post measure was paper-based, no response duration was recorded, which could have acted as another indicator for learning (e.g. faster response times in the posttest).

The fact that the pre-post measure was paper-based and did not include any game elements, might also explain the lack of differential learning effects (i.e., higher estimation accuracy). Participants in the game-based condition as compared to participants in the non-game-based condition estimated more accurately and took more time per estimation in the training, which might be indicative of increased cognitive engagement (for a more detailed discussion see below). However, as the pre-post measure was done on paper (i.e., different modality) and without game elements, a potential positive effect of game elements on cognitive engagement and, thus, participants' willingness to invest mental effort might not have carried over to the posttest. At the same time, the potential positive effect of game elements on cognitive engagement during the training might not have been strong enough to be reflected in the learning outcome as assessed with the pre-post measures. Future studies should also consider using delayed posttests to study potential (differential) long-term effects and consolidation processes, which are hardly studied in this field of research. That is, one might find (more) positive or negative effects of game-based learning not immediately but only after longer retention intervals.

Performance in training sessions

Even though we were not able to identify any differential effects on the learning outcome between the two training conditions, our third hypothesis was partly confirmed. Participants indeed were significantly more accurate across training sessions in the game-based training as compared to participants in the non-game-based training. Sailer and Sailer (2021) reported similar results when comparing flipped classroom learning groups with and without game elements present. They also did not find any differences between the groups in a posttest but identified beneficial effects of using game elements during the learning/training phase on performance. This underscores the importance of studying the actual learning process when investigating the effects of game elements rather than analyzing learning outcomes alone (i.e., pretest–posttest differences).

In the current study, increased performance was only reflected in higher estimation accuracy but not, for instance, by faster response durations. In fact, participants in the

game-based training provided more accurate responses but also took longer to estimate their responses across training sessions. Consequently, we can only partly confirm our third hypothesis. Overall, in both conditions, accuracy increased and response duration decreased significantly across training sessions. While this general improvement in accuracy and speed is in line with previous studies on number line estimation training (e.g., Fazio et al., 2016; Kiili et al., 2018a, 2018b; Obersteiner et al., 2013; Wilson et al., 2009; Wortha et al., 2020), the higher accuracy but slower response duration in game-based vs. non-game-based training paint a more complex picture.

In hindsight, the resulting pattern of increased accuracy and response duration makes sense from a cognitive engagement perspective. That is, participants were willing to invest more effort into accurate estimations, which also required more time. In other words, cognitively engaged participants might have been more thoughtful about their estimation strategies to optimize their accuracy leading to longer response durations. Schwartz and Plass (2020), for instance, consider the amount of time spent on a task an important indicator for cognitive engagement in games. Consequently, the game narrative/fantasy, visual appearance, and virtual incentives increased participants' willingness to invest more mental effort as compared to participants in the non-game-based training. This is in line with the RETAIN model (Gunter et al., 2008) as it emphasizes the importance of properly integrating the educational content into the game story/fantasy, which can influence the perception of relevance in a learning task, consequently impacting the learners' level of engagement and their behavior during the learning process.

However, the pattern might also indicate that participants in the game-based training might have prioritized accuracy over response duration across training sessions. The opposite appears to be true for participants in the non-game-based training—at least compared to participants in the game-based condition. It seems unlikely that participants in the game-based condition had an overall better understanding of fraction magnitude, as neither in the pretest nor the posttest participants in the game-based condition were more accurate than participants in the non-game-based condition (see also Fig. 3). That is, only in the training participants in the game-based condition were more accurate than participants in the non-game condition.

In general, the higher accuracy in game-based training might be a result of increased cognitive engagement, which is defined as the level of mental investment of a learner in a learning activity (Fredricks et al., 2004; Schwartz & Plass, 2020). While many studies measure cognitive engagement using questionnaires (Greene, 2015), the current study needs to rely on behavioral indicators, i.e., performance in the training sessions. Accordingly, the interpretation that the higher accuracy in game-based training is the result of higher cognitive engagement needs to be treated with caution. Nevertheless, previous studies have demonstrated that game elements can increase participants' level of cognitive engagement so that they invest more mental effort leading to deeper essential and generative processing (e.g., Bernecker & Ninaus, 2021; Chang et al., 2016; Ninaus et al., 2015; Plass et al., 2015). Bernecker and Ninaus (2021), for instance, showed that using game elements in a working memory task reduced task disengagement, which indirectly affected task accuracy via subjective effort. Similarly, Lumsden et al., (2016a, 2016b) showed that participants indicated investing less effort in a cognitive task when no game elements were present compared to when game elements were used in the task. Mekler et al. (2017) demonstrated that users invested more effort in a task (i.e., quantity and quality of tags in an image annotation task) when game elements were present.

At the same time, increased cognitive engagement might have also led to the observed increased response durations in the game-based training. Wiley et al., (2020) showed that

the use of points—a popular and often extrinsically integrated game element—in a cognitive task increased response durations but also increased error rates as compared to a basic version without game elements. In a different condition, the authors included a narrative to the cognitive task, which did not affect performance metrics. The current study used multiple game elements (i.e., visual aesthetics, game narrative, and a virtual incentive system) and realized principles of intrinsic integration, which might explain the different findings. As such, specific constellations of game elements might shape participants' behavior in diverse ways. This interpretation, however, goes beyond the current study and needs to be studied more comprehensively and in a systematic fashion in the future. That is, future studies might systematically modulate the presence or absence of single vs. multiple game elements as well as their degree of intrinsic integration.

The fact that the participants spent more time on each task/fraction in the game-based training, however, is in sharp contrast to the extensive literature on the so-called time-on-task hypothesis, i.e., a positive association between learning outcomes and time-on-task (Carroll, 1963). Recent evidence suggests, though, that the correlation between time-on-task and learning is weak and that time-on-task is necessary for learning to occur but not sufficient (Godwin et al., 2021). In fact, the authors argued that measurements used for time-on-task and learning vary considerably across studies, and so do the associations found. In the current study, we only measured the response duration for each item/fraction, which is only one part of the overall interaction time with the training. Accordingly, it might not be surprising that we did not find results that align with the time-on-task hypothesis (Carroll, 1963). An alternative interpretation related to the missing relation between time-on-task and learning outcome might be related to the used pre/posttest. Paper–pencil versions of a number line estimation task were used with no game elements present. In case game elements lead to increased cognitive engagement, participants' willingness to invest mental effort in the pre/posttest might not have been the same as during the game-based training.

The longer response duration in the game-based learning condition might, in fact, also indicate that participants had to process the provided narrative (i.e., novelty, extraneous processing of visuals), which was not necessary for the non-game-based condition. However, response duration only reflected the time from item onset to the first answer/response of the participants. The actual narrative used in the game-based condition was already provided before starting the actual training levels. Furthermore, the more extensive delayed feedback (i.e., 3-star rating) was shown after each level only, and the coins for correct responses were awarded directly after participants' responses, and therefore the processing of these features should not have affected response duration directly. Only the visual appearance of the task (game-based vs. non-game based), which remained constant across all training sessions, differed during task execution. Thus, the potential additional processing necessary for the visually more extensive game-based version, if indeed the case, should be limited to the first training or even just the first few trials. The response duration differences across training sessions between game-based and non-game-based conditions, however, remained rather stable (see Table 1). Accordingly, the difference in response duration should mainly indicate that participants invested more time estimating the fractions in the game-based compared to the non-game-based training. Nevertheless, future studies will need to systematically investigate the interaction between intrinsic/extrinsic integration of game elements, time-on-task, and different kinds of cognitive processing (i.e., extraneous, essential, and generative processing). The use of eye-tracking measures, for instance, might be particularly helpful to determine which game elements are processed and for how long. In the current study, we could only rely on the performance metrics (i.e., accuracy and

response duration) acquired during the training. Thus, our interpretation as to why the used game elements increased accuracy and response duration remains speculative.

A speculative but more likely explanation of the observed pattern of results might be that the used game elements changed participants' priorities or even strategies in performing the game-based or non-game-based tasks, which might explain the observed pattern in accuracy and response duration. In particular, the provided feedback during the training might have been perceived to be more salient and/or relevant in the game-based as compared to the non-game-based condition. In particular, in the game-based condition, the avatar communicated success with jubilant gestures, and participants were awarded virtual coins, while in the non-game-based condition, only a green check mark was shown. As such, the fantasy in the game-based condition made the feedback more meaningful to the participants and thereby affected participants' way of interacting with the learning content, which is in line with the RETAIN model proposed by Gunter et al. (2008).

This effect might have been even magnified because of the overall high performance of the participants or the rather seemingly simple task. That is, participants in the non-game-based condition might have not seen practical value in optimizing their already high estimation accuracy. In contrast, the game elements in the game-based condition might have provided additional value to participants to further optimize their accuracy. That is, choosing a more challenging task for the participants might reveal a different pattern. Furthermore, in our study, correct/incorrect feedback was only determined by estimation accuracy and not response duration, participants in the game-based condition might have been more focused on receiving positive feedback than participants in the non-game-based condition. However, future studies are needed to substantiate this interpretation empirically. For instance, future studies might wish to consider utilizing think-aloud protocols or self-explanations when solving tasks in the game-based vs. non-game-based conditions to better understand the strategies employed (e.g., Kiili et al., 2019a) and whether they differ between the game-based and non-game-based condition.

Conclusions

The current study showed that a game-based, as well as a non-game-based version of the number line estimation training, improved participants' fraction magnitude understanding, which confirmed our first hypothesis. The results substantiated and extended earlier findings that games relying on intrinsic integration can be successfully used in fraction instruction. However, contrary to our second hypothesis, participants did not learn significantly more from completing the game-based as compared to the non-game-based training. Nevertheless, participants performed better in the game-based as compared to the non-game-based training as reflected by higher accuracy. At the same time, participants in the game-based training took longer to respond as compared to participants in the non-game-based training. It seems that the processing of the task was altered by the use of intrinsically integrated game elements, which might have increased both essential and generative processing. Accordingly, participants in the game-based training might have invested more cognitive effort in their estimations, hence, were more cognitively engaged. Further, in the current task, game elements might have altered participants' priorities. That is, participants might have prioritized accuracy over speed in the game-based condition leading to more careful selection and use of estimation strategies. Furthermore, the current study

highlighted the relevance of studying the actual learning/training process rather than relying on learning outcomes alone (i.e., pretest–posttest differences).

Acknowledgements The authors acknowledge the financial support by the University of Graz.

Funding Open access funding provided by University of Graz.

Data availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. <https://doi.org/10.18637/jss.v067.i01>
- Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E. (2012). Toward a taxonomy linking game attributes to learning. *Simulation & Gaming*, 43(6), 729–760. <https://doi.org/10.1177/1046878112439444>
- Benbow, C. P., & Faulkner, L. R. (2008). Rejoinder to the Critiques of the National Mathematics Advisory Panel Final Report. *Educational Researcher*, 37(9), 645–648. <https://doi.org/10.3102/0013189X08329195>
- Bernecker, K., & Ninaus, M. (2021). No pain, no gain? Investigating motivational mechanisms of game elements in cognitive tasks. *Computers in Human Behavior*, 114, 106542. <https://doi.org/10.1016/j.chb.2020.106542>
- Booth, J. L., & Newton, K. J. (2012). Fractions: Could they really be the gatekeeper's doorman? *Contemporary Educational Psychology*, 37(4), 247–253. <https://doi.org/10.1016/j.cedpsych.2012.07.001>
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, 42(1), 189–201. <https://doi.org/10.1037/0012-1649.41.6.189>
- Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., & Pereira, J. (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. <https://doi.org/10.1016/j.compedu.2015.11.003>
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record: The Voice of Scholarship in Education*, 64(8), 1–9. <https://doi.org/10.1177/016146816306400801>
- Chang, M., Evans, M. A., Kim, S., Norton, A., Deater-Deckard, K., & Samur, Y. (2016). The effects of an educational video game on mathematical engagement. *Education and Information Technologies*, 21(5), 1283–1297. <https://doi.org/10.1007/s10639-015-9382-8>
- Chen, Y. C., Lu, Y. L., & Lien, C. J. (2021). Learning environments with different levels of technological engagement: A comparison of game-based, video-based, and traditional instruction on students' learning. *Interactive Learning Environments*, 29(8), 1363–1379. <https://doi.org/10.1080/10494820.2019.1628781>
- Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, 86(1), 79–122. <https://doi.org/10.3102/0034654315582065>
- de Marcos, L., Domínguez, A., de Saenz Navarrete, J., & Pagés, C. (2014). An empirical study comparing gamification and social networking on e-learning. *Computers & Education*, 75, 82–91. <https://doi.org/10.1016/j.compedu.2014.01.012>
- Fazio, L. K., Kennedy, C. A., & Siegler, R. S. (2016). Improving children's knowledge of fraction magnitudes. *Plos One*, 11(10), e0165243. <https://doi.org/10.1371/journal.pone.0165243>

- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Gallagher-Mitchell, T., Simms, V., & Litchfield, D. (2018). Learning from where 'eye' remotely look or point: Impact on number line estimation error in adults. *Quarterly Journal of Experimental Psychology*, 71(7), 1526–1534. <https://doi.org/10.1080/17470218.2017.1335335>
- Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, Motivation, and learning: A research and practice model. *Simulation & Gaming*, 33(4), 441–467. <https://doi.org/10.1177/1046878102238607>
- Gersten, R., Schumacher, R. F., & Jordan, N. C. (2017). Life on the number line: Routes to understanding fraction magnitude for students with difficulties learning mathematics. *Journal of Learning Disabilities*, 50(6), 655–657. <https://doi.org/10.1177/0022219416662625>
- Godwin, K. E., Seltman, H., Almeda, M., Davis Skerbetz, M., Kai, S., Baker, R. S., & Fisher, A. V. (2021). The elusive relationship between time on-task and learning: Not simply an issue of measurement. *Educational Psychology*, 41(4), 502–519. <https://doi.org/10.1080/01443410.2021.1894324>
- Greene, A. (2015). Measuring cognitive engagement with self-report scales: Reflections from over 20 years of research. *Educational Psychologist*, 50(1), 14–30. <https://doi.org/10.1080/00461520.2014.989230>
- Greipl, S., Klein, E., Lindstedt, A., Kiili, K., Moeller, K., Karnath, H.-O., Bahnmüller, J., Bloechle, J., & Ninaus, M. (2021). When the brain comes into play: Neurofunctional correlates of emotions and reward in game-based learning. *Computers in Human Behavior*, 125, 106946. <https://doi.org/10.1016/j.chb.2021.106946>
- Greipl, S., Moeller, K., Kiili, K., & Ninaus, M. (2020). Different performance, full experience: A learning game applied throughout adulthood. *International Journal of Serious Games*, 7(3), 79–98. <https://doi.org/10.17083/ijsg.v7i3.359>
- Gross, S. I., Gross, C. A., Kim, D., Lukowski, S. L., Thompson, L. A., & Petrill, S. A. (2018). A comparison of methods for assessing performance on the number line estimation task. *Journal of Numerical Cognition*, 4(3), 554–571. <https://doi.org/10.5964/jnc.v4i3.120>
- Gunderson, E. A., Hamdan, N., Hildebrand, L., & Bartek, V. (2019). Number line unidimensionality is a critical feature for promoting fraction magnitude concepts. *Journal of Experimental Child Psychology*, 187, 104657. <https://doi.org/10.1016/j.jecp.2019.06.010>
- Gunter, G. A., Kenny, R. F., & Vick, E. H. (2008). Taking educational games seriously: Using the RETAIN model to design endogenous fantasy into standalone educational games. *Educational Technology Research and Development*, 56(5–6), 511–537. <https://doi.org/10.1007/s11423-007-9073-2>
- Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences*, 20(2), 169–206. <https://doi.org/10.1080/10508406.2010.508029>
- Hamdan, N., & Gunderson, E. A. (2017). The number line is a critical spatial-numerical representation: Evidence from a fraction intervention. *Developmental Psychology*, 53(3), 587–596. <https://doi.org/10.1037/dev0000252>
- Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, 80, 152–161. <https://doi.org/10.1016/j.compedu.2014.08.019>
- Hilbert, S., Stadler, M., Lindl, A., Naumann, F., & Bühner, M. (2019). Analyzing longitudinal intervention studies with linear mixed models. *TPM Testing, Psychometrics, Methodology in Applied Psychology*, 26, 101–119. <https://doi.org/10.4473/TPM26.1.6>
- Hwang, G.-J., & Chen, P.-Y. (2022). Interweaving gaming and educational technologies: Clustering and forecasting the trends of game-based learning research by bibliometric and visual analysis. *Entertainment Computing*, 40(43), 100459. <https://doi.org/10.1016/j.entcom.2021.100459>
- Kafai, Y. B. (1996). Learning design by making games: children's development of design strategies in the creation of a complex computational artifact. In Y. B. Kafai & M. Resnick (Eds.), *Constructionism in practice designing, thinking, and learning in a digital world* (1st ed., pp. 71–96). Routledge. <https://doi.org/10.4324/9780203053492>
- Kiili, K., Koskinen, A., Lindstedt, A., & Ninaus, M. (2019a). Extending a Digital Fraction Game Piece by Piece with Physical Manipulatives. In M. Gentile, M. Allegra, & H. Sobke (Eds.), *Games and Learning Alliance, GALA 2018*: Vol. 11385 LNCS (Issue 7th Games and Learning Alliance (GALA) Conference), pp. 157–166. https://doi.org/10.1007/978-3-030-11548-7_15
- Kiili, K., Koskinen, A., & Ninaus, M. (2019b). Intrinsic integration in rational number games—A systematic literature review. *CEUR Workshop Proceedings*, 2359(January), 35–46.

- Kiili, K., Moeller, K., & Ninaus, M. (2018a). Evaluating the effectiveness of a game-based rational number training—In-game metrics as learning indicators. *Computers & Education*, *120*, 13–28. <https://doi.org/10.1016/j.compedu.2018.01.012>
- Kiili, K., Ojansuu, K., Lindstedt, A., & Ninaus, M. (2018b). Exploring the educational potential of a game-based math competition. *International Journal of Game-Based Learning*, *8*(2), 14–28. <https://doi.org/10.4018/IJGBL.2018040102>
- Kocadere, S. A., & Çağlar, Ş. (2015). The design and implementation of a gamified assessment. *Journal of E-Learning and Knowledge Society*, *11*(3), 85–99. <https://doi.org/10.20368/1971-8829/1070>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*. <https://doi.org/10.18637/jss.v082.i13>
- Liao, Y., Chang, L. E., & Chang, C.-C. (2010). Game-based Learning vs. Traditional Instruction: A Meta-Analysis of Thirty-Eight Studies from Taiwan. *Technology and Teacher Education Annual Vo—Conf 21, March 2010*, 1491.
- 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. <https://doi.org/10.1016/j.tine.2013.06.005>
- Lüdecke, D. (2018).ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), 772. <https://doi.org/10.21105/joss.00772>
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60), 3139. <https://doi.org/10.21105/joss.03139>
- Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. *JMIR Serious Games*, *4*(2), e11. <https://doi.org/10.2196/games.5888>
- Lumsden, J., Skinner, A., Woods, A. T., Lawrence, N. S., & Munafò, M. (2016). The effects of gamelike features and test location on cognitive test performance and participant enjoyment. *PeerJ*, *4*, e2184. <https://doi.org/10.7717/peerj.2184>
- Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdecke, D. (2021). Automated Results Reporting as a Practical Tool to Improve Reproducibility and Methodological Best Practices Adoption. *CRAN*. Retrieved March 17, 2022, from <https://github.com/easystats/report>
- Matthews, N., Mattingley, J. B., & Dux, P. E. (2022). Media-multitasking and cognitive control across the lifespan. *Scientific Reports*, *12*(1), 4349. <https://doi.org/10.1038/s41598-022-07777-1>
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816819.004>
- Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology*, *70*(1), 531–549. <https://doi.org/10.1146/annurev-psych-010418-102744>
- Mayer, R. E. (2020). Cognitive foundations of game-based learning Richard E. Mayer Introduction. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (1st ed., pp. 83–110). The MIT Press.
- Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 279–315). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.015>
- Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1998). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology*, *80*, 514–523.
- Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, *71*, 525–534. <https://doi.org/10.1016/j.chb.2015.08.048>
- Mihulowicz, U., Klein, E., Nuerk, H.-C., Willmes, K., & Karnath, H.-O. (2015). Spatial displacement of numbers on a vertical number line in spatial neglect. *Frontiers in Human Neuroscience*, *9*, 1–10. <https://doi.org/10.3389/fnhum.2015.00240>
- Moeller, K., Fischer, U., Nuerk, H.-C., & Cress, U. (2015). Computers in mathematics education—Training the mental number line. *Computers in Human Behavior*, *48*, 597–607. <https://doi.org/10.1016/j.chb.2015.01.048>
- Newmann, F. M. (1992). *Student engagement and achievement in American secondary schools*. New York: Teachers College Press.
- Ninaus, M., Kiili, K., Wortha, S. M., & Moeller, K. (2021). Empirische Arbeit: Motivationsprofile bei Verwendung eines Lernspiels zur Messung des Bruchverständnisses in der Schule—Eine latente Profilanalyse. *Psychologie in Erziehung Und Unterricht*, *68*(1), 42–57. <https://doi.org/10.2378/peu2021.art03d>

- Ninaus, M., Pereira, G., Stefitz, R., Prada, R., Paiva, A., Neuper, C., & Wood, G. (2015). Game elements improve performance in a working memory training task. *International Journal of Serious Games*, 2(1), 3–16. <https://doi.org/10.17083/ijsg.v2i1.60>
- Nuraydin, S., Stricker, J., Ugen, S., Martin, R., & Schneider, M. (2023). The number line estimation task is a valid tool for assessing mathematical achievement: A population-level study with 6484 Luxembourgish ninth-graders. *Journal of Experimental Child Psychology*, 225, 105521. <https://doi.org/10.1016/j.jecp.2022.105521>
- Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction*, 23(1), 125–135. <https://doi.org/10.1016/j.learninstruc.2012.08.004>
- Opfer, J. E., Thompson, C. A., & Kim, D. (2016). Free versus anchored numerical estimation: A unified approach. *Cognition*, 149, 11–17. <https://doi.org/10.1016/j.cognition.2015.11.015>
- Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, 50(4), 258–283. <https://doi.org/10.1080/00461520.2015.1122533>
- Plass, J. L., Homer, B. D., Mayer, R. E., & Kinzer, C. K. (2020). Theoretical foundations of game-based and playful learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (1st ed., pp. 3–24). The MIT Press.
- Praet, M., & Desoete, A. (2014). Number line estimation from kindergarten to grade 2: A longitudinal study. *Learning and Instruction*, 33, 19–28. <https://doi.org/10.1016/j.learninstruc.2014.02.003>
- R Core Team. (2021). *R: A language and environment for statistical computing*. Retrieved March 17, 2022, from <https://www.r-project.org/>
- Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review*, 7(3), 216–237. <https://doi.org/10.1016/j.edurev.2012.05.003>
- Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32(1), 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Sailer, M., & Sailer, M. (2021). Gamification of in-class activities in flipped classroom lectures. *British Journal of Educational Technology*, 52(1), 75–90. <https://doi.org/10.1111/bjet.12948>
- Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allogue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, 11(9), 1141–1152. <https://doi.org/10.1111/2041-210X.13434>
- Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of number line estimation with mathematical competence: A meta-analysis. *Child Development*, 89(5), 1467–1484. <https://doi.org/10.1111/cdev.13068>
- Schneider, S., Nebel, S., & Rey, G. D. (2016). Decorative pictures and emotional design in multimedia learning. *Learning and Instruction*, 44, 65–73. <https://doi.org/10.1016/j.learninstruc.2016.03.002>
- Schwartz, R. N., & Plass, J. L. (2020). Types of engagement in learning with games. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (1st ed., pp. 53–80). The MIT Press.
- Sidney, P. G., Thalluri, R., Buerke, M. L., & Thompson, C. A. (2019a). Who uses more strategies? Linking mathematics anxiety to adults' strategy variability and performance on fraction magnitude tasks. *Thinking & Reasoning*, 25(1), 94–131. <https://doi.org/10.1080/13546783.2018.1475303>
- Sidney, P. G., Thompson, C. A., & Rivera, F. D. (2019b). Number lines, but not area models, support children's accuracy and conceptual models of fraction division. *Contemporary Educational Psychology*, 58(March), 288–298. <https://doi.org/10.1016/j.cedpsych.2019.03.011>
- 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. *Psychological Science*, 14(3), 237–250. <https://doi.org/10.1111/1467-9280.02438>
- Stigler, J. W., Givvin, K. B., & Thompson, B. J. (2010). What community college developmental mathematics students understand about mathematics. *Mathematics Teacher*, 1(3), 4–16.
- Sullivan, J. L., Juhasz, B. J., Slatery, T. J., & Barth, H. C. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review*, 18(3), 557–563. <https://doi.org/10.3758/s13423-011-0081-1>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)
- Toda, A. M., Valle, P. H. D., & Isotani, S. (2018). *The dark side of gamification: an overview of negative effects of gamification in education* (pp. 143–156). Springer. https://doi.org/10.1007/978-3-319-97934-2_9

- Tokac, U., Novak, E., & Thompson, C. G. (2019). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning*, 35(3), 407–420. <https://doi.org/10.1111/jcal.12347>
- Van't Noordende, J. E., van Hoogmoed, A. H., Schot, W. D., & Kroesbergen, E. H. (2016). Number line estimation strategies in children with mathematical learning difficulties measured by eye tracking. *Psychological Research*, 80(3), 368–378. <https://doi.org/10.1007/s00426-015-0736-z>
- Wickham, H. (2016). *ggplot2. Media Issue July* (Vol. 35). Springer International Publishing. <https://doi.org/10.1007/978-3-319-24277-4>
- Wiley, K., Vedress, S., & Mandryk, R. L. (2020). How Points and Theme Affect Performance and Experience in a Gamified Cognitive Task. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. <https://doi.org/10.1145/3313831.3376697>
- Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain, and Education*, 3(4), 224–234. <https://doi.org/10.1111/j.1751-228X.2009.01075.x>
- Wong, R. M., & Adesope, O. O. (2021). Meta-analysis of emotional designs in multimedia learning: A replication and extension study. *Educational Psychology Review*, 33(2), 357–385. <https://doi.org/10.1007/s10648-020-09545-x>
- Wortha, S. M., Bloechle, J., Ninaus, M., Kiili, K., Lindstedt, A., Bahnmueller, J., Moeller, K., & Klein, E. (2020). Neurofunctional plasticity in fraction learning: An fMRI training study. *Trends in Neuroscience and Education*, 21, 100141. <https://doi.org/10.1016/j.tine.2020.100141>
- Wouters, P., van Nimwegen, C., van Oostendorp, H., & van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105(2), 249–265. <https://doi.org/10.1037/a0031311>
- Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review*, 30, 100326. <https://doi.org/10.1016/j.edurev.2020.100326>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Manuel Ninaus is Tenure-Track Professor of Digital Technologies and Psychology at University of Graz, Austria. His research interests include educational technologies and learning analytics. In his research projects he utilizes (neuro-)physiological measures.

Rodolpho Cortez is a PhD student at Federal University of Rio Grande do Norte, Brazil. His research interests encompass learning technologies, game-based learning, math learning and neuropsychological development.

Izabel Hazin is Associate Professor at Federal University of Rio Grande do Norte, Brazil. Her research interests include neuropsychological development, numerical cognition, math learning and information technology to giftedness children.

Kristian Kiili is Professor of Game-based Learning at Tampere University, Finland. His current research interests are related to learning sciences, game-based learning and assessment, and mathematics learning games.

Silke M. Wortha is a Research Fellow at Loughborough University, United Kingdom. Her research focuses on proportion processing across the lifespan. Furthermore, she is interested on cognitive and physical interventions to promote healthy brain aging.

Elise Klein is Associate Research Professor (CNRS) at University of Paris, France. She investigates mathematical competences and learning processes as well as their neurocognitive correlates in children and adults with and without mathematical difficulties. Furthermore, she focuses on promoting cognitive activation in learning.

Elisabeth M. Weiss is Professor of Clinical Psychology at University of Innsbruck, Austria. Her research interests include the biopsychological correlates of various psychiatric diseases and evaluates the impact of different interventions on affective and cognitive processes.

Korbinian Moeller is Professor of Mathematical Cognition at Loughborough University, United Kingdom. His research interests include neurocognitive foundations of mathematical cognition and developmental psychology in the context of educational games and learning analytics.

Authors and Affiliations

Manuel Ninaus^{1,2,3}  · **Rodolpho Cortez**^{1,4} · **Izabel Hazin**⁴ · **Kristian Kiili**⁵ · **Silke M. Wortha**⁶ · **Elise Klein**^{7,8} · **Elisabeth M. Weiss**² · **Korbinian Moeller**^{3,6,9}

✉ Manuel Ninaus
manuel.ninaus@uni-graz.at

¹ Institute of Psychology, University of Graz, Universitätsplatz 2, 8010 Graz, Austria

² Department of Psychology, University of Innsbruck, Innsbruck, Austria

³ LEAD Graduate School and Research Network, University of Tübingen, Tübingen, Germany

⁴ Federal University of Rio Grande Do Norte, Natal, Brazil

⁵ Faculty of Education and Culture, Tampere University, Tampere, Finland

⁶ Centre for Mathematical Cognition, School of Science, Loughborough University, Loughborough, UK

⁷ Université de Paris, LaPsyDÉ, CNRS, Sorbonne Paris Cité, Paris, France

⁸ Leibniz-Institut für Wissensmedien, Tübingen, Germany

⁹ Individual Development and Adaptive Education for Children at Risk Centre, Frankfurt Am Main, Germany