Computers & Education 194 (2023) 104694

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

The strength and direction of the difficulty adaptation affect situational interest in game-based learning

Antti Koskinen[a,*], Jake McMullen[b], Minna Hannula-Sormunen[b], Manuel Ninaus[c], Kristian Kiili[a]

[a] Faculty of Education and Culture, Tampere University, P.O.Box 700, 33014, Finland
[b] Department of Teacher Education, University of Turku, 20014, Finland
[c] Institute of Psychology, University of Graz, Universitätsplatz 2, 8010, Graz, Austria

ARTICLE INFO

Keywords:
Game-based learning
Adaptive learning
Adaptation
Motivation
Situational interest

ABSTRACT

Research has shown that difficulty adaptation is a promising instructional technique in digital game-based learning. Nevertheless, although the strength and direction of difficulty adaptation can affect motivational outcomes, these effects remain insufficiently examined in game-based learning. This within-subject study examined how the strength and direction of difficulty adaptation affected motivational outcomes in digital game-based math learning. The participants were 167 Finnish fifth-graders who studied fractions with the Number Trace game. The game included 144 tasks, half were adapted according to participants' playing performance. Situational interest and perceived difficulty were measured several times with in-game self-report items during the intervention. The manipulation check confirmed that difficulty adaptation was implemented successfully as task correctness and perceived difficulty changed according to the strength and direction of adaptation. Regarding motivational outcomes, two-way repeated-measures ANOVAs showed that the difficulty adaptation increased situational interest, but only when the task difficulty was substantially adapted downwards. Contrary to our expectations, a substantial upwards adaptation of the task difficulty significantly decreased situational interest. Minor adaptation of difficulty did not affect situational interest. The current study contributes to the field of adaptive digital learning environments by highlighting the effects of the strength and direction of difficulty adaptation on motivational outcomes. Theoretical, practical, and methodological implications of the findings are discussed.

1. Introduction

Educational research has demonstrated that learning environments that adapt to learners' needs can enhance the effectiveness of instruction (e.g., Aleven, McLaughlin, Glenn, & Koedinger, 2017; Bloom, 1968; Plass & Pawar, 2020; Skinner, 1986). Advances in computer-supported learning systems, especially in real-time learning diagnostics, have facilitated the development of digital adaptive learning materials and environments (Aleven et al., 2017; Ninaus & Nebel, 2021; Bakkes, Tan, & Pisan, 2012; Plass & Pawar, 2020). Adaptation of digital learning environments refers to the capability of a system to modify its content and behavior according to learner

* Corresponding author.
E-mail addresses: antti.koskinen@tuni.fi (A. Koskinen), jamcmu@utu.fi (J. McMullen), mimarha@utu.fi (M. Hannula-Sormunen), manuel.ninaus@uni-graz.at (M. Ninaus), kristian.kiili@tuni.fi (K. Kiili).

https://doi.org/10.1016/j.compedu.2022.104694
Received 12 May 2022; Received in revised form 16 November 2022; Accepted 29 November 2022
Available online 30 November 2022

needs, preferences, and characteristics (Shute & Zapata-Riviera, 2007). The adaptation of digital instruction has been suggested to be a major emerging instructional advance that may shape the future of education (Zhang & Aslan, 2021). However, more research is needed to advance our understanding of how to implement and examine the effects of adaptation (Plass & Pawar, 2020).

In this paper, we focus on the adaptation of difficulty, a common adaptation approach in game-based learning (Liu, Moon, Kim, & Dai, 2020). Although previous research has shown that adaptation is a promising instructional technique in game-based learning (Wouters & Oostendorp, 2017), several studies have revealed that difficulty adaptation does not always have positive effects on motivational outcomes (e.g., Shute et al., 2021; Van Oostendorp, Van Der Spek, & Linssen, 2014; Vanbecelaere et al., 2020). These findings are inconsistent with the theoretical propositions of seminal motivational theories suggesting that balancing the difficulty of the game with the player's skill level should positively affect motivational outcomes (Csikszentmihalyi, 1990; Hattie, Hodis, & Kang, 2020; Ryan & Deci, 2020). We presume that measurements (e.g., post-game self-reports) that are not aligned with temporal changes in motivational outcomes may partly explain this disparity between the theoretical propositions and empirical findings. That is, dynamic changes in motivational outcomes, for instance, in response to the adaptation of difficulty of the learning activity, cannot be reflected by a single post-game self-report measurement but requires a more fine-grained assessment. Further, the strength and direction of the difficulty adaptation may have substantial effects on motivational outcomes, but these effects have rarely been examined. Thus, in the present study, we address these issues and advance the field by shedding light on the motivational effects of the strength and direction of difficulty adaptation during game-based math learning. In doing so, we aim to provide a better understanding of the motivational effects of difficulty adaptation and, therefore, facilitate the development of more effective and engaging learning environments.

2. Theoretical rationale

2.1. Foundations of motivational outcomes in game-based learning

Motivational states are unobservable psychological, neural, and biological processes that initiate, direct, and sustain individuals' behavior (e.g., Eccles & Wigfield, 2020; Hattie et al., 2020; Reeve, 2012; Ryan & Deci, 2020) and have an important role in students' learning (e.g., Howard, Bureau, Guay, Chong, & Ryan, 2021; Kriegbaum, Becker, & Spinath, 2018). In this study, we refer to motivational outcomes as observable manifestations of the motivational state of a student while interacting with a learning game (e.g., Ainley, 2012; Reeve, 2012). Various tasks (e.g., value, controllability, importance) and person-related dimensions (e.g., perceived competence, relatedness, goals, and costs) constantly affect the motivational state (e.g., Hattie et al., 2020). Learning activity that cultivates a positive motivational state is manifested in positive motivational outcomes (e.g., Reeve, 2012). Thus, for example, altering task attributes can result in improved motivational outcomes which can positively influence learning (Howard et al., 2021).

Game-based learning environments aim to foster positive motivational states by providing experiential learning activities that nurture several motivational dimensions and facilitate deep processing (i.e., generative processing; see Mayer, 2014) of essential learning content. In other words, game-based learning refers to redesigning the basic instructional task (i.e., by utilizing game mechanics beyond reward systems) to make it more interesting, meaningful, and ultimately more effective for learning (Plass, Homer, Mayer, & Kinzer, 2020).

Although several aspects of game-based learning environments can contribute to the players' motivational outcomes (e.g., Plass, Homer, & Kinzer, 2015), flow theory posits the balance of challenge and skill as the most important antecedent of optimal experience (Csikszentmihalyi, 1990). Thus, learning games should maximize positive motivational outcomes by matching the game difficulty to players' competence levels (Kiili, Lainema, de Freitas, & Arnab, 2014). The game should provide more challenging tasks to more skilled players and less challenging tasks to less skilled players (Kiili et al., 2014). According to the self-determination theory (e.g., Ryan & Deci, 2020), optimized challenges improve motivational states by nurturing players' perceived competence (Ryan, Rigby, & Przybylski, 2006), which is one of the main determinants of a positive motivational state in games (Przybylski, Rigby, & Ryan, 2006). For these reasons, balancing the difficulty of the game to provide players with a sufficient challenge can be regarded as one of the major determinants influencing positive motivational outcomes.

Learning games are usually designed so that the difficulty level increases as the game progresses (e.g., Ninaus, Kiili, McMullen, & Moeller, 2017; Kiili, Moeller, & Ninaus, 2018). This premise assumes that all students acquire skills needed in a game at the same rate and in the same way. However, such a one-size-fits-all design does not consider the possible large variation in students' skill levels, prior knowledge, and varying abilities to acquire skills while playing the game. Thus, some students may find the predefined difficulty curve either too easy or too hard. To overcome this design deficiency, scholars have emphasized the need to adapt the difficulty of the learning games to correspond with the student's skill level (e.g., Bakkes et al., 2012; Streicher & Smeddinck, 2016).

2.2. Adaptation of difficulty and motivational outcomes in game-based learning

The adaptation of digital learning environments is a process in which data gathered from the interaction with the learning environment is selected and analyzed to adapt aspect(s) of the learning environment to better respond to individuals' instructional needs (e.g., Plass & Pawar, 2020; Shute & Zapata-Rivera, 2007). The adaptation can be based on the assessment of cognitive, motivational, affective, and/or socio-cultural variables. Adaptation can be utilized to adjust aspects of the overall learning context where students study (macro-level) or aspects of the task (micro-level) (Aleven et al., 2017; Plass & Pawar, 2020). In game-based learning, the adaptation of difficulty is often implemented as a micro-level adaptation based on the assessment of students' cognitive variables (Liu et al., 2020). For example, Sampayo-Vargas, Cope, He, and Byrne (2013) found that micro-level adaptation of game difficulty produced significantly higher learning outcomes when compared to a non-adaptive version of the same game.

Previous research has shown that adaptation is a promising instructional technique in game-based learning (Wouters & Oostendorp, 2017). However, only a limited number of studies have examined the effects of difficulty adaptation on motivational outcomes (for a review, see Ninaus & Nebel, 2021; Liu et al., 2020; Sajjadi, Ewais, & De Troyer, 2022). Further, several studies have revealed that difficulty adaptation does not always have positive effects on motivational outcomes (Orvis, Horn, & Belanich, 2008; Sampayo-Vargas et al., 2013; Shute et al., 2021; Van Oostendorp et al., 2014; Vanbecelaere et al., 2020). The measured motivational outcomes of these studies included, for example, engagement (Orvis et al., 2008), training motivation (Van Oostendorp et al., 2014), and situational interest (Vanbecelaere et al., 2020). The non-significant results of these interventions might be partly explained by rather small sample sizes (Orvis et al., 2008; Van Oostendorp et al., 2014) or the relatively short duration of the interventions (Sampayo-Vargas et al., 2013). More importantly, Van Oostendorp et al. (2014) suggest that one possible explanation for their non-significant findings might have been that motivational outcomes were measured after the game, which may not adequately reflect the motivational state and its temporal changes *during* gameplay. In fact, measuring motivational outcomes during the interaction with a difficulty-adapted learning game is uncommon (Orvis et al., 2008; Sampayo-Vargas et al., 2013; Shute et al., 2021; Van Oostendorp et al., 2014; Vanbecelaere et al., 2020). Retrospective assessment of motivational outcomes can jeopardize the validity of the results, as participants do not always correctly remember their motivational states during the instruction (Krapp, 1999). On the other hand, measuring motivational outcomes during the learning activity can disturb the learning process and likewise jeopardize the validity of the results (Ainley & Patrick, 2006).

Another possible explanation for the non-significant results is that past studies have not often examined the influence of the strength of the adaptation (i.e., how much the difficulty was adjusted). Based on seminal motivational theories (Csikszentmihalyi, 1990; Ryan & Deci, 2020), the strength of the adaptation can have a major influence on motivational outcomes. For example, if the game is too easy, a minor adaptation of the task difficulty may not provide a sufficient challenge for the student. In other words, the strength of the difficult adaptation is insufficient to influence the motivational state. Unfortunately, previous studies did not conduct manipulation checks that would provide evidence to assess this conjecture (Orvis et al., 2008; Sampayo-Vargas et al., 2013; Shute et al., 2021; Van Oostendorp et al., 2014; Vanbecelaere et al., 2020). Moreover, although task difficulty was adjusted to be higher or lower (i.e., the direction of the adaptation) in several studies, its effect on motivational outcomes was not examined (Orvis et al., 2008; Sampayo-Vargas et al., 2013; Van Oostendorp et al., 2014).

2.3. *Situational interest as a measurement of motivational outcomes*

Several different self-reporting questionnaires have been utilized to measure motivational outcomes in game-based learning environments (Eyupoglu & Nietfeld, 2019; Krath, Schürmann, & Von Korflesch, 2021). However, due to the increased emphasis on the temporal changes in motivational states (Hattie et al., 2020; Pekrun & Marsh, 2022; Wigfield & Koenka, 2020), scholars have emphasized the utilization of measurements that capture motivational outcomes during the learning activity (e.g., Pekrun & Marsh, 2022; Ryan & Deci, 2020). However, several self-reporting measurements of motivational outcomes require a lengthy scale that can considerably interrupt the learning activity (Ainley & Patrick, 2006). Thus, it is preferable to measure such motivational outcomes that can be assessed with a single-item (Ainley & Patrick, 2006; Allen, Iliescu, & Greiff, 2022).

In this study, situational interest is used to measure motivational outcomes as it can be measured with a single item (Schmidt & Rotgans, 2021). Situational interest is a transient and malleable motivational outcome associated with positive affect such as enjoyment (Ainley, 2012; Krapp, 2005; Knogler, 2017; Reeve, Lee, & Won, 2015). Situational interest arises from the interaction with the learning environment (Hidi & Renninger, 2006; Krapp, 2007; Reeve et al., 2015) and reflects how learning activity influences the motivational state (Ainley, 2012; Deci, 1992; Hidi & Renninger, 2006; Krapp, 2007; Reeve et al., 2015). Past studies have found that task attributes that influence motivational states such as scaffolds (Chen & Law, 2016), concreteness (Tapola, Veermans, & Niemivirta, 2013), autonomy (Renninger, Bachrach, & Hidi, 2019), and challenge (Chen, Darst, & Pangrazi. 2001) are reflected in situational interest measures. Regarding the personal attributes influencing situational interest, in particular, the role of individual interest, enduring trait-like interest (Hidi & Renninger, 2006), has been examined (e.g., Rotgans & Schmidt, 2018). However, previous research shows that individual math interest has a relatively small influence on situational interest in game-based math learning (Koskinen, McMullen, Halme, Hannula-Sormunen, Ninaus, & Kiili, 2022a).

Situational interest has been successfully measured with single-item approaches in several domains such as traditional reading (Alexander, Kulikowich, & Schulze, 1994), digital simulations (Tapola et al., 2013), and digital game-based learning (Ronimus, Kujala, Tolvanen, & Lyytinen, 2014). Single-item measurement of situational interest particularly emphasizes the emotional aspect of the motivational outcomes (Ainley, 2012; Fredrics & Wang, 2019) but still provides a sufficient reflection of the overall motivational state (e.g., Ainley & Patrick, 2006).

2.4. *The present study*

The present within-subject study examines the effects of difficulty adaptation on situational interest in digital game-based math learning. Seminal motivational theories hold that challenges balanced with learners' skills should enhance motivation. However, findings of past adaptive game-based learning studies do not support this proposition. We seek to clarify this contradiction by addressing several methodological weaknesses in previous adaptive game-based learning research. Firstly, we conducted a manipulation check to ensure that the implemented adaptation mechanism worked as expected. This is a crucial prerequisite for a valid investigation of the motivational outcomes in adaptive learning environments. Secondly, we measured situational interest multiple times during the intervention with in-game self-report items. Such a recurring measurement approach should grasp potential temporal motivational changes more accurately than the commonly employed post-game measures alone. Thirdly, we examined the effects of the strength and direction of difficulty adaptation, which are often neglected in game-based learning research.

Fig. 1 illustrates the research design, which is based on nine game worlds. Each game world had different instructional content. A game world consists of a basic-game level followed by an adapted-game level. Both of these levels had eight fraction estimation tasks and had similar mathematical content. The basic-game levels were designed to assess particular fraction competencies. Students' performance in the basic-game level determined the strength and direction of difficulty adaptation in the following adapted-game level (e.g., poor performance in the basic-game level led to a substantial decrease in the task difficulty in the following adapted-game level). The adaptation adjusted the difficulty of tasks by changing the required estimation accuracy of the fraction tasks (Fig. 2; see more details from chapter 3.2.3). The effects of difficulty adaptation were examined by comparing (a) task correctness, (b) perceived difficulty, and (c) situational interest across the nine pairs of basic- and adapted-game levels (Fig. 1). Task correctness and perceived difficulty were used in the manipulation check.

Research question 1 and hypotheses

1) *How did the adaptation of difficulty affect task correctness and perceived difficulty?* This research question was treated as a manipulation check to ensure that the implemented adaptation mechanism affected task correctness and perceived difficulty as designed. If the implementation of the difficulty adaptation was successful, we expected that:

H1.1. When task difficulty is decreased (i.e., less accurate answers are required) participants' have higher task correctness in adapted-game levels than in basic-game levels.

H1.2. When task difficulty is increased (i.e., more accurate answers are required) participants' have lower task correctness in adapted-game levels than in basic-game levels.

H1.3. The variance in task correctness in adapted-game levels is lower than in basic-game levels.

H1.4. When task difficulty is decreased (i.e., less accurate answers are required) participants' report lower perceived difficulty in adapted-game levels than in basic-game levels.

H1.5. When task difficulty is increased (i.e., more accurate answers are required) participants' report higher perceived difficulty in adapted-game levels than in basic-game levels.

H1.6. The variance in perceived difficulty in adapted-game levels is lower than in basic-game levels.

Research question 2 and hypotheses

2) *How does the adaptation of difficulty affect situational interest?* Only a few studies have investigated the effects of difficulty adaptation on motivational outcomes. However, motivational theories suggest that challenges balanced with learners' skills should enhance motivation (Csikszentmihalyi, 1990; Ryan & Deci, 2020). Consequently, we expect that:

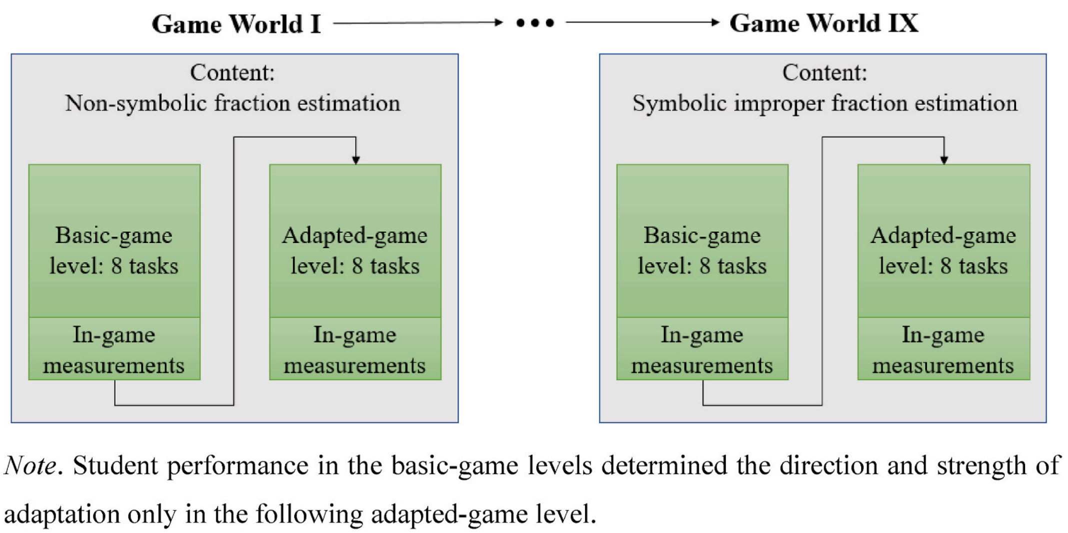


Fig. 1. Research design based on nine game worlds consisting of a basic-game level followed by an adapted-game level.
Note. Student performance in the basic-game levels determined the direction and strength of adaptation only in the following adapted-game level.

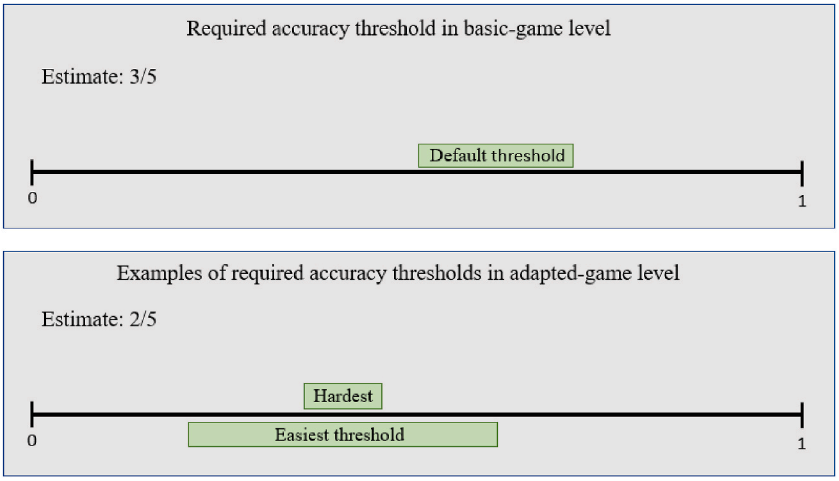


Fig. 2. Example of the adaptation of the estimation requirements of the fraction tasks in basic- and adapted-game levels.

H2.1. When task difficulty is decreased (i.e., less accurate answers are required) participants' will report higher situational interest in adapted-game levels than in basic-game levels.

H2.2. When task difficulty is increased (i.e., more accurate answers are required) participants' will report higher situational interest in adapted-game levels than in basic-game levels.

3. Method

3.1. Participants

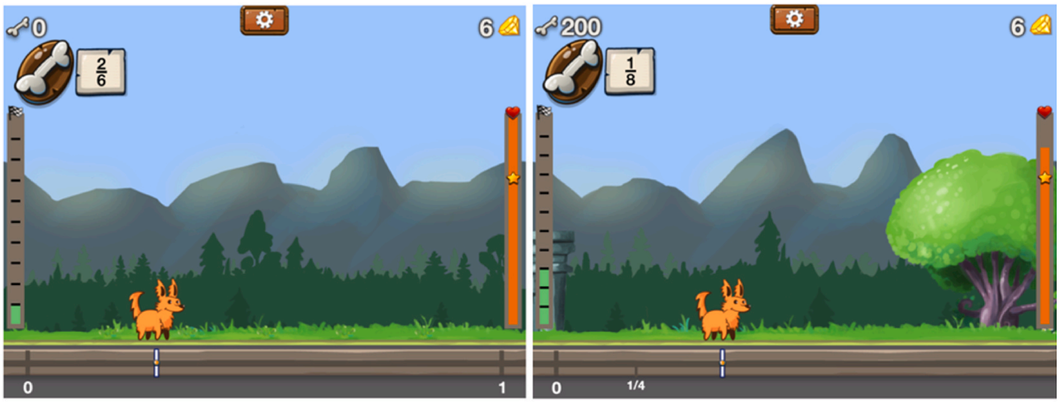
The sample comprised 167 Finnish fifth-graders (82 boys; 85 girls; age $M = 11.61$ years; $SD = 0.97$; six participants did not report their age) from eight classes in four schools. The schools were from varying socioeconomic status areas in a city located in Finland. All participants had written parental permission to participate in the study, and their verbal consent was given before beginning the study. Ethical board and municipality approval were granted for this study.

3.2. Description of the Number Trace learning game

The Number Trace learning game (Koskinen, McMullen, Ninaus, & Kiili, 2022b; Kiili, Lindstedt, Koskinen, Halme, Ninaus, & McMullen, 2021; Greipl, Klein, Lindstedt, Kiili, Moeller, Karnath et al., 2021) was used in this study. The Number Trace game environment is a configurable research and learning environment developed for rational number instruction. This environment and its predecessor, Semideus, has been successfully utilized in several studies focusing on enhancing understanding of rational number learning and game-based learning (e.g., Ninaus et al., 2017; Kiili et al., 2018; Koskinen et al., 2022b). The design of the Number Trace game is based on several game-based learning design principles, such as the utilization of emotional design (Plass & Kaplan, 2016), intrinsic integration of instructional content (Walkington, 2021), as well as goals and feedback (Kiili et al., 2014). For this study, the Number Trace game was configured to serve two purposes: (a) to support the development of fifth-graders' conceptual fraction knowledge according to the guidelines of the Finnish national core curriculum and (b) to examine the effects of the difficulty adaptation.

To support conceptual fraction knowledge development, the game is built around estimating fraction magnitudes on a number line. Previous research has demonstrated that the use of number line estimation is an effective and engaging instructional approach in game-based fraction learning (e.g., Kiili et al., 2018; Braithwaite & Siegler, 2021). In the game, the player estimates the magnitudes of given target fractions by moving a dog along a number line. The game features two types of number line estimation tasks (1) the basic number line estimation task (e.g., Siegler & Opfer, 2003), in which the player estimates the position of a target number on a number line based on the start and the endpoint of the number line (e.g., estimate 2/6 on a number line ranging from 0 to 1; left column of Fig. 3): and (2) the unbounded number line estimation task (e.g., Cohen & Blanc-Goldhammer, 2011), where the player estimates the position of a target number on a number line based on the start point and a given number on the number line (e.g., estimate 1/8 on a number line, where the starting point 0 and the location of 1/4 are shown, but the endpoint of the number line is not shown; right column of Fig. 3).

Fig. 3 Various feedback mechanics are implemented in the Number Trace game. After a correct answer, the player receives bones as points, and the dog looks happy (Fig. 4, left). In contrast, after a wrong answer, the dog is sad; the player loses energy and does not get

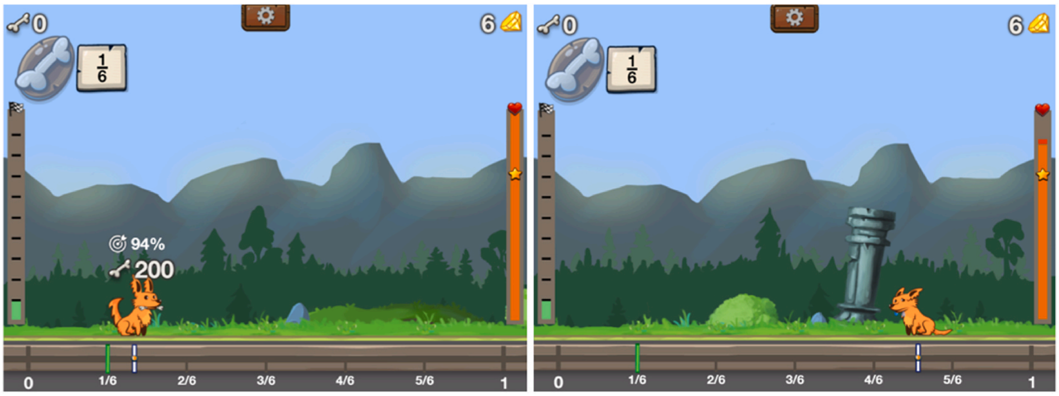


Note. On the left side is a basic number line estimation task, and on the right side is an unbounded number line estimation task. The top left corner in both pictures shows the fraction being estimated, and above it, the points (bones) acquired by the player. The unbounded task on the right has a hatch mark pointing where 1/4 is on the number line. The bar on the left shows progress in the level, and the bar on the right shows the remaining energy.

Fig. 3. Utilized task types.
Note. On the left side is a basic number line estimation task, and on the right side is an unbounded number line estimation task. The top left corner in both pictures shows the fraction being estimated, and above it, the points (bones) acquired by the player. The unbounded task on the right has a hatch mark pointing where 1/4 is on the number line. The bar on the left shows progress in the level, and the bar on the right shows the remaining energy.

points (Fig. 4, right). After a correct answer, or after two wrong answers, players are shown the position of the correct answer and instructional partitioning of the number line (Fig. 4). After finishing the game level, players receive delayed feedback in the form of 1–3 stars, depending on their overall performance in the level just played.

The game also features enemies to enhance players' engagement and foster the development of conceptual fraction number understanding. The enemies' jump pattern was designed to aid players' fraction estimation. Fig. 5 (left) shows the enemy making a pattern of jumps between 1, 4/5, and 3/5. The enemy leaves a trace behind and hence creates temporary hatch marks on the number



Note. Immediate positive feedback after a correct answer (left), immediate negative feedback after a second incorrect answer (right). The figure also shows the position of the correct answer, partitioning of the number line, and estimation accuracy in the case of a correct answer.

Fig. 4. Examples of feedback.
Note. Immediate positive feedback after a correct answer (left), immediate negative feedback after a second incorrect answer (right). The figure also shows the position of the correct answer, partitioning of the number line, and estimation accuracy in the case of a correct answer.

line. To benefit from enemies, the player must interpret how the jump pattern partitions the number line and how the partitioning relates to the current estimation task. If the enemy touches the player's character, the character loses energy. The player can avoid enemies by moving the dog or destroy the enemies by jumping on them. If the player did not like the enemies, he or she could press the 'destroy enemy' button (Fig. 5) to get rid of the enemy instantly. Furthermore, some tasks included graphical dividers embedded in the game world (see added flowers in Fig. 5, right). The dividers were not designed to be a key pedagogical feature but could be used by those players who recognized their values to help partition the number line into units.

3.2.2. Level structure and progression

The game was comprised of nine game worlds (Table 1). Each game world had different mathematical content. A game world consisted of two content-matched levels: a basic-game level followed by an adapted-game level (Table 2). Each level contained eight estimation tasks. The levels could be completed only once to ensure that each participant received the same number of tasks. The second and sixth tasks of the level featured an enemy, and the first and fourth tasks featured dividers (Table 2). The levels of unbounded estimation tasks did not feature enemies or dividers. The player had two attempts in every task to get the correct answer before the game automatically advanced to the next task, level, or game world. The level progression was designed so that the difficulty of the content gradually increased. Altogether, the game consisted of 144 fraction estimation tasks.

3.2.3. Implementation of adaptation

The implementation of the adaptation was based on pairs of content-matched game levels (see Tables 1 and 2). In these pairs, the difficulty of the first level was fixed (basic-game level), and the difficulty of the following level was adapted (adapted-game level) based on the participants' estimation accuracy in the preceding content-matched basic-game level. For example, participants' estimation accuracy on Level 1 B was used to adapt Level 1 A. After the participants completed the tasks of the adapted-game level, they proceeded to the next pair of levels (basic followed by adapted), with respective adaptation procedures. The adaptation did not change the difficulty of the next basic-difficulty level, as the mathematical content was different in each level pair. The adaptation of difficulty was implemented by adjusting the fraction magnitude estimation accuracy threshold required for an answer to be considered as correct (Table 3). Fraction estimation accuracy was chosen as an adapted instructional variable as it is a valid indicator of conceptual fraction competence (for a meta-analysis, see Schneider et al., 2018).

In the basic-game levels, the required estimation accuracy[1] for answers being considered correct was always >90%. In other words, a correct answer could not deviate from the correct location by more than 10% of the overall length of the number line. This threshold was chosen because results of previous studies have indicated that the average fraction magnitude estimation accuracy of fifth-graders is around 90–92% (e.g., Tian, Bartek, Rahman, & Gunderson, 2021). Participants' average estimation accuracy on tasks 3–8 in the basic-game levels was used to adapt the required estimation accuracy threshold in the following adapted-game level. Tasks 1 and 2 were not taken into account in the adaptation, as these tasks were considered as an introduction to different mathematical content. The adapted estimation accuracy requirements were applied to all the tasks in the adapted-game level.

The adaptation featured five different levels (Table 4). The downward adaptation of difficulty featured three levels. The upwards adaptation had only two levels, as the requirements of fraction estimation accuracies over 95% were considered arbitrary and not pedagogically meaningful. Table 4 shows how the average estimation accuracy of the basic-game level was used to adapt the required estimation accuracy of the tasks in the following adapted-game level. For example, if the participant's mean estimation accuracy in the basic-game level was 93.5%, the required estimation accuracy in the following adapted-game level was over 92% (level of adaptation: +1). On the other hand, if the participant's mean estimation accuracy in the basic-game level was 83%, the required estimation accuracy in the following adapted-game level was over 80% (level of adaptation: 3). If the participant's average estimation accuracy in the basic-game level was 90–92%, they did not receive any adaptation, as the default difficulty was determined as appropriate (level of adaptation: 0). Adaptation was defined as major if the task's correctness requirements were adapted over ± 5% from the required 90% estimation accuracy of the basic-game level (Table 4).

If the participants received an adaptation of difficulty, they were informed about the change at the beginning of the level. These messages were designed not to explicitly inform participants about the adaptation of the required estimation accuracy to eliminate possible misuses of the adaptation mechanics. Instead, participants whose estimation accuracy requirement was increased were informed that the ground in this level where the bones are hidden is very hard and they must be very accurate to find the hidden bones. If the estimation accuracy requirements were decreased, the participants were informed that the ground in this level is soft, so the bones are easier to find, but they must try to be as accurate as possible.

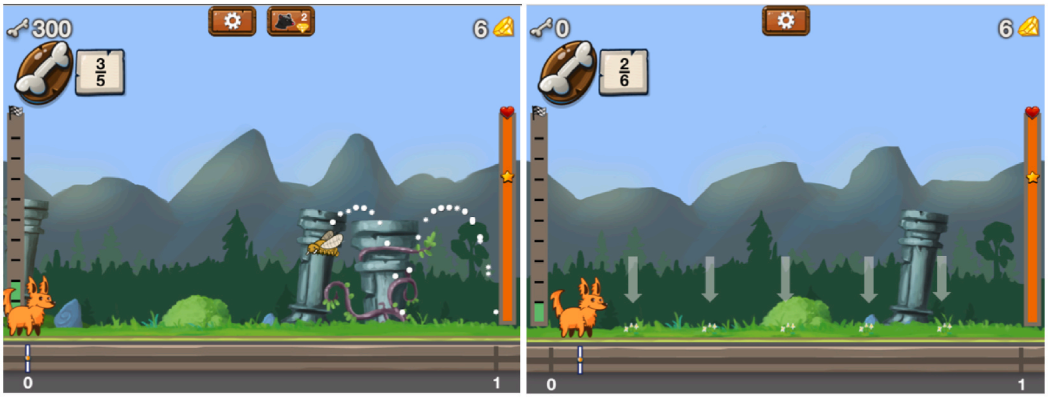
3.3. Measures

The present study used two types of measurements: (a) in-game measurements of motivational factors and (b) in-game metrics.

3.3.1. In-game measurements

In-game self-report measures were used to assess the participants' situational interest and perceived difficulty during the gameplay. The in-game measurements utilized a similar interface that was used in the instructional part of the game to minimize the risk of

[1] Estimation accuracy was calculated based on percent absolute accuracy (100 - ((|Participant's Answer–Correct Answer|)/Numerical Range * 100)) (e.g., Fazio, Kennedy, & Siegler, 2016).



Note. Enemy making a mathematically instructive jump pattern to aid fraction estimation (left). The button to destroy the enemy is at the top center of the screen. Embedded graphical dividers (highlighted with arrows) (right).

Fig. 5. Additional instructional features.
Note. Enemy making a mathematically instructive jump pattern to aid fraction estimation (left). The button to destroy the enemy is at the top center of the screen. Embedded graphical dividers (highlighted with arrows) (right).

Table 1
The level design of the game and measurement points of in-game measures.

	Game World I		Game World II		Game World III	
Level	1 B	1 A	2 B	2 A	3 B	3 A
Content	Non-symbolic fraction estimation		Symbolic fraction estimation		Symbolic fraction estimation; Denominator >10	
Measurements	SI; PD	SI; PD	SI	SI	SI; PD	SI; PD
	Game World IV		Game World V		Game World VI	
Level	4 B	4 A	5 B	5 A	6 B	6 A
Content	Symbolic fraction estimation		Unbounded symbolic fraction estimation		Unbounded symbolic fraction estimation	
Measurements	SI; PD	SI; PD	SI	SI	SI; PD	SI; PD
	Game World VII		Game World VIII		Game World IX	
Level	7 B	7 A	8 B	8 A	9 B	9 A
Content	Symbolic mixed number estimation		Symbolic improper fraction estimation		Symbolic improper fraction estimation	
Measurements	SI; PD	SI; PD	SI	SI	SI; PD	SI; PD

Note. On the level row: B = basic-game level, A = adapted-game level. On the measurements row, SI = situational interest; PD = perceived difficulty.

Table 2
Example of content-matched estimation tasks of the second game world and instructional features.

Task	2 B	2 A	Instructional feature
1.	2/6	1/6	Divider
2.	3/5	2/5	Enemy
3.	7/8	6/7	
4.	1/3	5/8	Divider
5.	4/7	3/7	Enemy
6.	1/4	3/4	
7.	3/8	2/3	
8.	5/9	4/9	
9.	Situational interest measurement		
10.	Possible perceived difficulty measurement		

breaking the game flow (Fig. 6). Instead of estimating fraction magnitudes, participants answered questions using the same number-line-based mechanic (Fig. 6). Previously, the utilization of similar measurements did not encourage careless responding and provided an effective way to capture students' motivational outcomes (Koskinen et al., 2022b; Kiili et al., 2021). Situational interest was measured after each of the game levels, and perceived difficulty was measured after the game levels: 1 B, 1 A, 3 B, 3 A, 4 B, 4 A, 6 B, 6 A, 7 B, 7 A, 9 B, and 9 A (see Table 2).

Table 3
Overview of the utilized adaptation.

Learner variable	Instructional variable	How to adapt	Time scale	Outcome
Cognitive ability: conceptual fraction knowledge	Task difficulty: fraction estimation accuracy	Correct answer threshold	Eight fraction estimation tasks	Motivational: situational interest

Note. Framework synthesized from Aleven et al. (2017), Sajjadi et al. (2022), and Liu et al. (2020).

Table 4
Adaptation of required estimation accuracy for the correct answer.

Average estimation accuracy in the basic-game level	Required estimation accuracy in the adapted-game level for the correct answer	Level of adaptation
>95%	>95%	+2; major
>92–95%	>92%	+1; minor
>90–92%	>90%	0
>88–90%	>88%	−1; minor
>85–88%	>85%	−2; major
<85%	>80%	−3; major

Situational interest was measured with one statement: "Tasks on this game level were really interesting" on a continuous scale from one ("totally disagree") to five ("totally agree") (Fig. 6). Test–retest approach (Spearman's rank correlation coefficient) was used to evaluate the reliability of the repeated one-item situational interest measure. The reliability of the situational interest measure was evaluated by comparing the basic-game level to the adapted-game level situational interest of the participants who did not receive difficulty adaptation (level of adaptation: 0; see Table 4). We used this group of students to assess the reliability of the measurement as theoretically their situational interest should not change much between the two levels that had similar content and difficulty. Test–retest reliability rating indicated decent reliability (Spearman's rho = .60, $p < .001$) (Matheson, 2019). The reliability was similar to what has been reported in previous studies that utilized a very similar measurement of situational interest (Koskinen et al., 2022a; Kiili et al., 2021).

The *perceived difficulty* was measured with one statement: "How difficult the tasks of this game level were" on a continuous scale from one ("extremely easy") to nine ("extremely difficult"). This measurement approach was adopted from DeLeeuw and Mayer (2008).

3.3.2. *In-game metrics*

The game logged players' answers in fraction estimation tasks which were used to calculate estimation accuracy. In the basic-game level tasks, a participant's answer was considered correct if the estimation accuracy was over 90%. For example, if the participant's estimation accuracy was 85% in the basic-game level, the answer was considered to be incorrect. At the adapted-game level tasks, the correctness of a participant's answer was based on the requirements of the adapted difficulty (see Table 3).

Note. Players indicated their level of agreement with the statement by marking on the number line.

Fig. 6. In-game measurement of situational interest.
Note. Players indicated their level of agreement with the statement by marking on the number line.

3.4. Procedure

This study was conducted in regular school classes during regular school days and was administered by regular classroom teachers during mathematics lessons. The study protocol did not rigidly regulate the activities of the participants and participating teachers. During the intervention, participants continued their other school activities. With this naturalistic study design and instructional content corresponding to the requirements of the national core curriculum, we aimed for the learning situation and task demands to resemble characteristics of the authentic use of game-based learning in the schools (i.e., high ecological validity) (Holleman, Hooge, Kemner, & Hessels, 2020).

Fig. 7 shows that before the intervention, the teachers were provided with clear guidance about the research and procedures of the study. The intervention consisted of four 45 min sessions of gameplay during a two-week period. Administrating classroom teachers were given the freedom to integrate these four sessions within the regular school timetable. Giving additional fraction instruction during the intervention was not allowed. Before the instructional content of the game (144 fraction tasks), participants completed 14 onboarding tasks (i.e., step-by-step instruction) to become familiar with the game mechanics and interface and two tasks ensuring that the participants had mastered the core game mechanics. During playing, the participants could proceed at their own pace and discuss with each other if they had technical problems with the game.

3.5. Analysis

We utilized a within-subject design to determine how the adaptation of difficulty affected participants' task correctness (RQ1), perceived difficulty (RQ1), and situational interest (RQ2). We ran three two-way repeated measures ANOVAs with each variable separately treated as a dependent variable. Since individuals had different levels of adaptation at different points in the game, the unit of analysis was the content-matched levels, with repeated-measures being the basic and adapted versions of these pairs of levels. Hence, each participant had nine pairs of observed variables for each dependent variable. *Levels of adaptation* (i.e., the strenght and direction of adaptation; see Table 4) were used as a between-subject factor to determine the effects of each of the six levels of adaptation on the three dependent variables. The *difficulty condition* (basic- or adaptive-game level) was used as a within-subject variable. The F-test of equality of variances was conducted to test whether variances in the fixed- and adaptive-difficulty levels were significantly different. IBM SPSS statistics 26 was used to analyze the data.

4. Results and discussion

4.1. Implementation of the adaptation

First, we examined how the adaptation of difficulty affected task correctness and perceived difficulty (RQ1).

To determine whether adaptation of difficulty affected participants' task correctness (RQ1), we conducted a two-way repeated-measures ANOVA with difficulty condition (basic- or adaptive-game level) as a within-subject factor, level of adaptation (-3, -2, -1, 0, 1, 2) as a between-subject factor, and task correctness (correct or incorrect) as the dependent variable. The analysis shows that there was a significant main effect of difficulty condition on task correctness, $F\,(1, 1428) = 92.0$, $p < .001$, $\eta 2p = .06$. Moreover, there was a significant interaction effect between the level of adaptation and difficulty condition, $F\,(5, 1428) = 184.24$, $p < .001$, $\eta 2p = .39$. As shown in Table 5, when task difficulty was increased in levels of adaptation 1 and 2, participants' task correctness decreased, confirming H1.1; when task difficulty was decreased or remained the same, in levels of adaptation -3 through 0, participants' task correctness increased, confirming H1.2. Furthermore, there was a significant difference in mean variances between difficulty conditions in task correctness, $F = 14 > F_{0.01} = 1.15$. The mean variance of adapted-game level task correctness (var = 0.05) across all levels of adaptation was significantly smaller than the mean variance of basic-game level task correctness (var = 0.7), confirming H1.3.

To determine whether adaptation of difficulty affected participants' perceived difficulty (RQ1), a two-way repeated-measures ANOVA was conducted with difficulty condition as a within-subject factor, level of adaptation as a between-subject factor, and perceived difficulty as the dependent variable. The analysis shows that there was a non-significant main effect between the difficulty conditions on perceived difficulty, $F\,(1, 948) = 2.55$, $p = .11$, $\eta 2p = .00$. However, as expected, there was a significant interaction effect between level of adaptation and difficulty condition, $F\,(5, 948) = 11.22$, $p < .001$, $\eta 2p = .06$. Paired samples t-tests revealed that participants who received the most adaptation of difficulty (levels of adaptation -3 and 2) reported significant changes of perceived difficulty between difficulty conditions (Table 6). Participants reported lower perceived difficulty when the task difficulty was decreased, confirming H1.4. The participants who received a major increase in task difficulty reported higher perceived difficulty. However, the participants who received a minor increase in task difficulty reported lower perceived task difficulty, thus H1.5 was

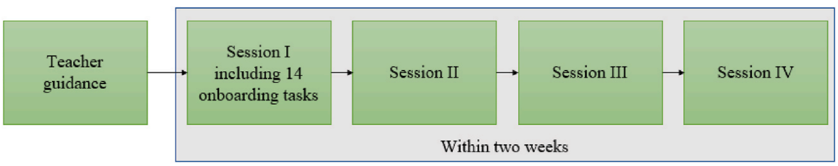


Fig. 7. Flow chart of the procedures.

Table 5
Estimated marginal means of the task correctness of basic-game level, adapted-game level, t-tests, and effect sizes.

Level of adaptation	*N*	Basic-game level		Adapted-game level		T-test	Cohen's *d*
		M	SD	M	SD		
−3; Major	322	0.30	0.19	0.58	0.26	t (321) = **19.76, p < .001**	1.23
−2; Major	99	0.56	0.17	0.74	0.23	t (98) = **8.02, p < .001**	0.89
−1; Minor	82	0.65	0.16	0.77	0.21	t (81) = **5.21, p < .001**	0.64
0	150	0.72	0.13	0.76	0.18	t (149) = **3.07, p = .003**	0.25
1; Minor	269	0.81	0.11	0.75	0.18	t (268) = **−4.72, p < .001**	0.40
2; Major	512	0.94	0.08	0.78	0.20	t (511) = **−17.11, p < .001**	1.05

Note. Significant *t*-test values are shown in bold.

partly confirmed. Moreover, the F-tests showed that there was a significant difference in mean variances between difficulty conditions in perceived difficulty, $F = 1.20 > F_{0.01} = 1.18$. The mean variance of adapted-game level perceived difficulty (var = 4.56) across all levels of adaptation was significantly smaller than the mean variance of basic-game level perceived difficulty (var = 5.46), confirming H1.6.

4.1.3. Discussion of the implementation of the adaptation

As hypotheses 1.1–1.6 were confirmed or partly confirmed, the implementation of adaptation was determined to be successful; adaptation of difficulty affected participants' task correctness and perceived difficulty almost as expected. Fig. 8 shows that participants' task correctness changed with the adaptation level (strength and direction of adaptation). The effect sizes of the task correctness of the participants who received either a major decrease or a major increase in their task difficulty were interpreted to indicate a large effect (Kraft, 2020). Further, participants' perceived difficulty changed almost in total conjunction with the adaptation level. In particular, only participants who received a major adaptation of difficulty reported significant changes in their perceived difficulty. The effect sizes of the perceived difficulty of the participants who received either a major decrease or a major increase in their task difficulty were interpreted to indicate a small to medium effect (Kraft, 2020). Fig. 9 shows that the adaptation significantly balanced the large initial differences in skill levels toward a better ratio of correct–incorrect answers for learning (Wilson, Shenhav, Straccia, & Cohen, 2019). The perceived task difficulty was balanced from extreme values to more appropriate difficulty levels (e.g., Mills et al., 2013).

Our results show that by utilizing a well-established fraction learning task, we were able to design a relatively simple but effective adaptation mechanism. The implementation of the adaptation was based on manipulating the required fraction magnitude estimation accuracy. Fraction estimation accuracy is considered a valid indicator of conceptual fraction knowledge (e.g., Schneider et al., 2018). Thus, it provided a strong tool to assess learners' competencies in basic-game levels and determine the appropriate level of task difficulty for adapted-game levels. Thereby, we encourage designers to ground adaptation mechanisms on well-established principles of the instructional domain. Previous research can guide the selection of the appropriate learner variables (i.e., a particular skill, competence, etc.), ensure valid assessment of these variables, and guide the implementation of adaptation (e.g., difficulty, feedback, scaffolding, etc.) to develop efficient and effective adaptive learning environments.

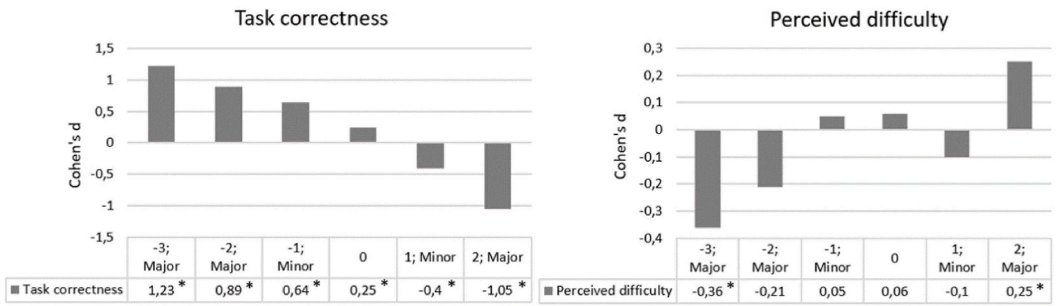
4.2. Effects of adaptation on situational interest

To determine whether the adaptation of difficulty affected participants' situational interest (RQ2), a two-way repeated-measures ANOVA with difficulty condition as a within-subject factor, level of adaptation as a between-subject factor, and situational interest as the dependent variable was conducted. The analysis shows that there was a non-significant main effect between the difficulty conditions on situational interest, $F(1, 1428) = 0.67, p = .415, \eta 2p = .00$. However, as expected, there was a significant interaction effect of the level of adaptation and difficulty condition, $F(5, 1428) = 6.73, p < .001, \eta 2p = .02$. In particular, paired samples t-tests revealed that there were significant differences in situational interest between the difficulty conditions when participants received major adaptations of difficulty, but not for minor adaptations of difficulty (Table 7). That is, the adaptation of difficulty significantly enhanced situational interest for those participants who received a major decrease in task difficulty, confirming H2.1. Contrary to our

Table 6
Estimated marginal means of the perceived difficulty of basic-game level, adapted-game level, t-tests, and effect sizes.

Level of adaptation	*N*	Basic-game level		Adapted-game level		T-test	Cohen's *d*
		M	SD	M	SD		
−3; Major	222	5.81	2.26	5.00	2.18	t (221) = **−5.52, p < .001**	0.36
−2; Major	87	5.38	2.18	4.94	1.97	t (86) = −1.95, p = .054	0.21
−1; Minor	67	4.72	2.08	4.82	1.99	t (66) = 0.49, p = .63	0.05
0	116	4.74	2.07	4.87	2.10	t (115) = 0.64, p = .53	0.06
1; Minor	177	4.49	2.35	4.26	2.15	t (176) = −1.37, p = .17	0.10
2; Major	285	3.80	2.16	4.33	2.13	t (284) = **4.60, p < .001**	0.25

Note. Significant *t*-test values are shown in bold.



Note. Significant values based on t-tests are marked with an asterisk (*), indicating p<.05 (see Tables 5 and 6 for more detail).

Fig. 8. Distribution of effect sizes of levels of adaptation when comparing task correctness and perceived difficulty on adapted-game levels to basic-game levels.
Note. Significant values based on t-tests are marked with an asterisk (*), indicating p < .05 (see Tables 5 and 6 for more detail).

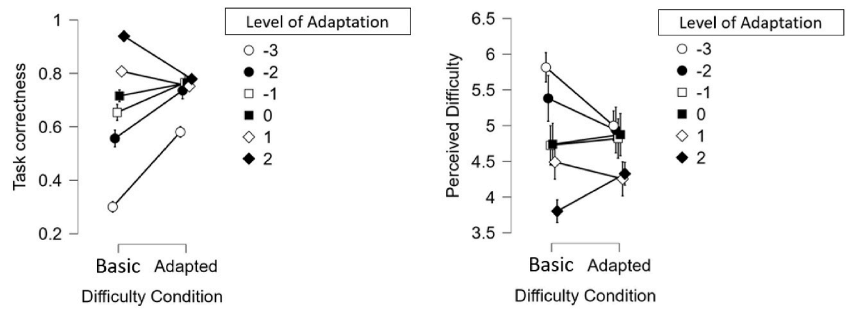


Fig. 9. Change of task correctness (left) and perceived difficulty (right) between difficulty conditions grouped by level of adaptation.

expectations, participants who received a major increase in task difficulty reported significantly lower situational interest. Thus, H2.2 was rejected.

4.2.2. Discussion of the effects of adaptation on situational interest

These results partially align with propositions of seminal motivational theories, which suggest that the balancing task difficulty with learner's skills should enhance motivational outcomes (Csikszentmihalyi, 1990; Ryan & Deci, 2020). The adaptation of difficulty positively impacted situational interest, but only for struggling participants for whom task difficulty was decreased in a major way (Fig. 10). Surprisingly, participants who received a major increase in task difficulty reported significantly lower situational interest. Regarding the adopted study design (only the required estimation accuracy threshold changed between the measurement points), the effect sizes of the participants who received a major decrease in their task difficulty were interpreted to indicate a medium positive effect (Kraft, 2020). Similarly, the effect sizes of the participants who received major increases in their task difficulty were interpreted to indicate a medium negative effect (Kraft, 2020). Nonetheless, participants who received a major increase in their task difficulty still reported the highest situational interest among the different adaptation groups.

The adaptation resulted in corresponding changes in participants' perceived difficulty and situational interest. Only participants

Table 7
Estimated marginal means of situational interest of basic-game level, adapted-game level, t-tests, and effect sizes.

Level of adaptation	*N*	Basic-game level		Adapted-game level		T-test	Cohen's *d*
		M	SD	M	SD		
−3; Major	322	2.51	1.17	2.73	1.11	$t\,(321) = \mathbf{3.54}, \boldsymbol{p < .001}$	0.19
−2; Major	99	2.78	1.22	2.99	1.23	$t\,(98) = \mathbf{2.65}, \boldsymbol{p < .01}$	0.17
−1; Minor	82	2.94	1.20	2.96	1.19	$t\,(81) = 0.20, p = .84$	0.01
0	150	3.17	1.12	3.13	1.08	$t\,(149) = -0.44, p = .66$	0.03
1; Minor	269	3.12	1.18	3.04	1.20	$t\,(268) = -1.29, p = .20$	0.06
2; Major	512	3.41	1.09	3.25	1.13	$t\,(511) = \mathbf{-3.76}, \boldsymbol{p < .001}$	0.14

Note. Significant *t*-test values are shown in bold.

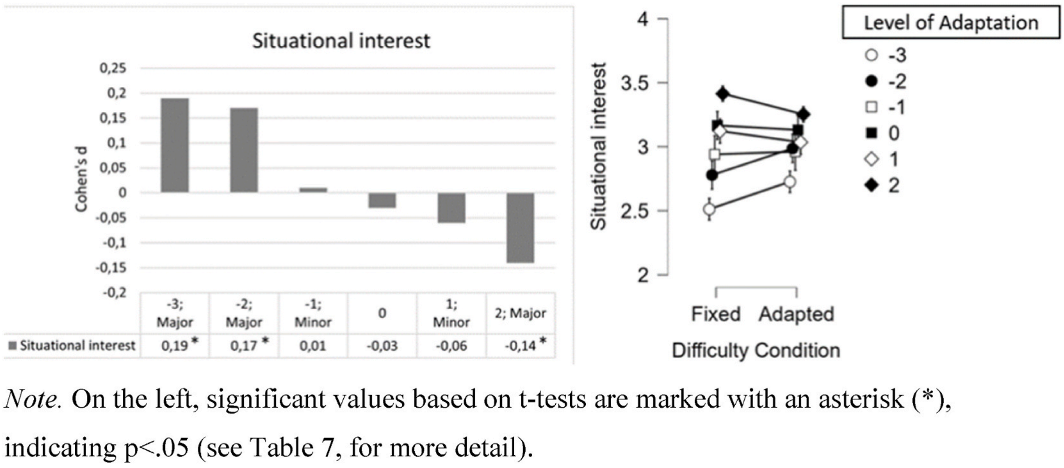


Fig. 10. On the left, the distribution of effect sizes of levels of adaptation when comparing situational interest on adapted-game levels to basic-game levels. On the right, situational interest changes between difficulty conditions, grouped by the level of adaptation.
Note. On the left, significant values based on t-tests are marked with an asterisk (*), indicating $p < .05$ (see Table 7, for more detail).

who received a major adaptation of difficulty reported significant changes in perceived difficulty and situational interest; minor adaptations to the difficulty did not have a significant effect on perceived difficulty or situational interest. This indicates that the instructional design did affect motivational outcomes within the learning activity, but only if the adaptation of the task difficulty was substantial. Importantly, the non-significant main effect between difficulty conditions on situational interest indicates that without examining the effects of strength and direction of adaptation, our results would have indicated that adaptation of difficulty does not affect motivational outcomes. These findings can partly explain why previous research in digital game-based learning has not found unequivocal evidence of the effects of difficulty adaptation on motivational outcomes, as previous studies have rarely examined the effects of strength and direction of adaptation.

5. Implications and limitations

5.1. Theoretical implications

The results of this study contribute to the development of models of adaptive learning environments by highlighting the effects of the strength and direction of difficulty adaptation on motivational outcomes (e.g., Plass & Pawar, 2020). Thus, this work can be seen as a preliminary step in establishing a basic understanding of the dimensions affecting motivational outcomes in adaptive digital learning environments. The results support previous propositions stating that adaptation can enhance the educational value of digital learning environments (e.g., Aleven et al., 2017; Plass & Pawar, 2020; Shute & Zapata-Rivera, 2007). By incorporating adaptation of difficulty into learning games, we can provide more appropriate challenges for students, which can lead to improved motivational outcomes. Crucially, our results show that there is a boundary condition for the effective utilization of the adaptation of difficulty. The strength of difficulty adaptation needs to surpass a certain threshold to influence motivational outcomes.

However, the results of this study cannot be fully explained from the perspective of enhanced difficulty–skill balance. In particular, in our study, the decrease in situational interest of participants who received a major increase in task difficulty may have resulted from (i) the changes in the amount of positive (and negative) feedback (e.g., Ryan & Deci, 2020), (ii) varying performance demands (unclear goals) between the difficulty conditions decreasing control over the requirements of the tasks and the outcome of their answers (e.g., Csikszentmihalyi, 1990), and (iii) the increases in task difficulty decreasing the expectations of success (e.g., Eccles & Wigfield, 2020). This suggests that scholars should apply a broad theoretical framework to better understand the factors that affect motivational outcomes when digital instruction is adapted.

5.2. Practical implications

As online teaching continuously increases, the importance of adaptive learning environments will increase in the future. Therefore, it is important to understand how to design adaptive game-based learning environments that are as engaging and efficacious. This study offers practical implications for designing engaging learning environments. The designers of the game-based learning environments should consider the capabilities of the difficulty adaptation mechanism in a way that it can balance the challenges of the games for individuals with a great variety of skill levels. Furthermore, the designers should consider techniques to balance the negative effects of possibly decreased amounts of positive feedback when the task difficulty is adapted upwards. The use of such techniques could increase satisfaction with digital learning materials.

5.3. Methodological implications

One of the main methodological implications of this work is that motivational measures should be administered several times in adaptive learning games to increase sensitivity to the temporal changes in the motivational state that the real-time adaptation aims to induce. Moreover, we suggest using in-game measures that are integrated into the core game mechanics to minimize the possible negative effects that repeated out-game self-report measures can have (e.g., disrupting the learning experience and flow of the game). Following these suggestions might improve the validity of game-based learning research and thus advance our understanding of how instructional design affects motivational outcomes.

Moreover, our study provides methodological insights for the researchers examining issues outlined in the educational technology and addictions research agenda proposed by Melo et al. (2020). We recommend that when examining the effects of adaptive advanced educational technology and game-based learning on students' well-being, manipulation checks should be conducted, and the effects of the strength and direction of the adaptation should be examined. Manipulation checks and comprehensive analyses are crucial to avoid misinterpretation of the effects of adaptation. Furthermore, we encourage scholars to develop a consistent practice to disseminate how adaptation was implemented to allow for systematic comparisons between studies. Table 3 can serve as a starting point for developing such practices.

5.4. Limitations and future directions

This study has limitations that should be considered when interpreting current results and planning future studies. First, only situational interest was used to measure motivational outcomes. It is possible that adaptation of difficulty may influence differently on other motivational outcomes. Future studies should consider measuring a wider range of motivational outcomes to explore why upward and downward adaptation of difficulty had differential effects on situational interest. For example, examining the changed ratio between positive and negative feedback within the situated expectancy-value theory (e.g., Eccles & Wigfield, 2020) could be a starting point for such research. Further, even though previous research has indicated that individual math interest has only a relatively small influence on situational interest in a game-based learning setting (Koskinen et al., 2022a), longitudinal studies should be conducted to examine how difficulty adaptation affects individual interest over longer periods of time.

Second, repeated measures of motivational outcomes may have affected participants' self-reporting behavior. As participants were asked to focus on their experience, they may have become more self-aware of their motivational states over time. However, this is the case for any study employing the same measure several times (in a row). In contrast to most other studies, the assessment of motivational outcomes in the current study was implemented directly within the game environment using game mechanics. Such approaches and their differences from conventional approaches need to be studied systematically in the future. Still, one might assume that the direct integration of subjective measures into the learning environment better reflects the actual learning experience than externally administered measures (Koskinen et al., 2022b).

Third, the used difficulty adaptation technique differed from typically used techniques. That is, an adaptation that varied estimation accuracy threshold is an exemption among the common difficulty adaptation techniques that often rely, for example, on adapting the content (e.g., providing easier or harder fractions to estimate). However, we selected this technique because it is well-aligned with the nature of the number line estimation task and allows each player to engage with the same math content. It is noteworthy that specific adaptation techniques may be more feasible and effective for certain tasks, subject domains, or even game types. Thus, future studies should examine whether the present findings on motivational outcomes hold when different difficulty adaptation techniques are used. These studies could also consider inter-individual differences, for example, the meaning of prior knowledge, and how motivational outcomes are related to learning outcomes.

Fourth, because the game included different kinds of math content, the difficulty adaptation only changed the difficulty of eight tasks in a row. This led to the constant fluctuation of the task difficulty (required estimation accuracy threshold), which may have influenced participants' situational interest. Therefore, future studies should examine the effects of the difficulty adaptation that continuously assesses skills and adapts the difficulty of the tasks accordingly (e.g., task or step-loop adaptation; see Aleven et al., 2017 for details). However, these studies should also examine the effects of the strength and direction of difficulty adaptation by administering repeated measurements of motivational outcomes during the learning activity. Also, longitudinal studies could be conducted to examine whether adaptive learning environments produce negative or positive effects on students' well-being, an important consideration in digital learning environments (Melo et al., 2020). In fact, the results of our study suggest that difficulty adaptation increased struggling students' intrinsic motivation by nurturing their experience of competence, which is shown to be positively related to students' well-being outcomes (Howard et al., 2021). Moreover, a recent study demonstrated that an adaptive math game increased students' learning efficiency compared to a non-adaptive version of the game (Debeer, Vanbecelaere, Van Den Noortgate, Reynvoet, & Depaepe, 2021). Together these results suggest that adaptive digital learning environments may provide means to mitigate the potential health risks that the increasing use of educational technology can pose. Nevertheless, these issues need further clarification. Regardless of these limitations, we believe this work has made an important contribution to our understanding of the effects of the difficulty adaptation on motivational outcomes.

6. Conclusion

Developing digital learning environments to respond to individual needs can substantially enhance the effectiveness of education (e.g., Aleven et al., 2017; Plass & Pawar, 2020). This study contributed to the field of adaptive digital learning environments by highlighting the effects of the strength and direction of difficulty adaptation on motivational outcomes. The results revealed that the adaptation of difficulty had a positive impact on students' situational interest in digital game-based learning, but only when students received a major decrease in task difficulty. Moreover, students who received a major increase in their task difficulty reported significantly lower situational interest. We demonstrated the usefulness of repeated in-game measurements in capturing such temporal changes in motivational outcomes that the real-time adaptation induced. However, because we used a domain-specific difficulty adaptation technique, the generalization of these results should be exercised with caution. Overall, our research suggests that further work is needed to determine how the strength and direction of difficulty adaptation affect motivational outcomes when different difficulty adaptation techniques are used.

Author statement

Antti Koskinen: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Jake McMullen**: Methodology; Writing - Review & Editing, Supervision, Funding acquisition. **Manuel Ninaus**: Conceptualization, Writing- Reviewing and Editing. **Minna Hannula-Sormunen** Methodology, Writing-Reviewing, Supervision, Funding acquisition, Project administration. **Kristian Kiili**: Conceptualization, Writing - Review & Editing, Funding acquisition, Project administration, Supervision, Methodology, Software,

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We gratefully acknowledge the support of the Academy of Finland (Grants 310338, and 326618), and the Strategic Research Council (SRC) established within Academy of Finland (Grant 336068).

References

Ainley, M. (2012). Students' interest and engagement in classroom activities. In L. S. Christenson, L. A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 283–302). Boston, MA: Springer Science + Business Media.

Ainley, M., & Patrick, L. (2006). Measuring self-regulated learning processes through tracking patterns of student interaction with achievement activities. *Educational Psychology Review, 18*(3), 267–286. https://doi.org/10.1007/s10648-006-9018-z

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2017). Instruction based on adaptive learning technologies. In R. E. Mayer, & P. Alexander (Eds.), *Handbook of research on learning and instruction* (2nd ed., pp. 522–560). Routledge.

Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). The influence of topic knowledge, domain knowledge, and interest on the comprehension of scientific exposition. *Learning Ind Diff., 6*(4), 379–397. https://doi.org/10.3102/00028312031002313

Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single item measures in psychological science: A call to action. *European J Psychol Ass., 38*(1), 1–5. https://doi.org/10.1027/1015-5759/a000699

Bakkes, S., Tan, C. T., & Pisan, Y. (2012). Personalised gaming: A motivation and overview of literature. In *Proceedings of the 8th australasian conference on interactive entertainment: Playing the system* (pp. 1–10). New York: Association for Computing Machinery.

Bloom, B. S. (1968). Learning for mastery. Instruction and curriculum. Regional education laboratory for the carolinas and Virginia, topical papers and reprints, number 1. *Evaluation comment, 1*(2), n2.

Braithwaite, D. W., & Siegler, R. S. (2021). Putting fractions together. *J Edu Psychol., 113*(3), 556. https://doi.org/10.1037/edu0000477

Chen, A., Darst, P. W., & Pangrazi, R. P. (2001). An examination of situational interest and its sources. *British J Edu Psychol., 71*(3), 383–400. https://doi.org/10.1348/000709901158578

Chen, C. H., & Law, V. (2016). Scaffolding individual and collaborative game-based learning in learning performance and intrinsic motivation. *Comput Human Behav., 55*, 1201–1212. https://doi.org/10.1016/j.chb.2015.03.010

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bull Rev, 18*(2), 331–338. https://doi.org/10.3758/s13423-011-0059-z

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.

Debeer, D., Vanbecelaere, S., Van Den Noortgate, W., Reynvoet, B., & Depaepe, F. (2021). The effect of adaptivity in digital learning technologies. Modelling learning efficiency using data from an educational game. *British J Edu Technol., 52*(5), 1881–1897. https://doi.org/10.1111/bjet.13103

Deci, E. L. (1992). The relation of interest to the motivation of behavior: A self-determination theory perspective. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 43–70). Lawrence Erlbaum Associates, Inc.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Edu Psychol., 100*(1), 223. https://doi.org/10.1037/0022-0663.100.1.223

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Edu Psychol., 61*, Article 101859.

Eyupoglu, T., & Nietfeld, J. (2019). Intrinsic motivation in game-based learning environments. In D. Ifenthaler, & Y. Kim (Eds.), *Game-based assessment revisited* (pp. 85–102). Cham: Springer. https://doi.org/10.1007/978-3-030-15569-8_5.

Fazio, L. K., Kennedy, C. A., & Siegler, R. S. (2016). Improving children's knowledge of fraction magnitudes. *PLoS One, 11*(10). https://doi.org/10.1371/journal.pone.0165243

Fredrics, H., & Wang. (2019). Addressing the challenge of measuring student engagement. In K. A. Renninger, & S. E. Hidi (Eds.), *Cambridge handbook on motivation and learning* (pp. 689–712). Cambridge: Cambridge University Press.

Greipl, S., Klein, E., Lindstedt, A., Kiili, K., Moeller, K., Karnath, H. O., & 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

Hattie, J., Hodis, F. A., & Kang, S. H. (2020). Theories of motivation: Integration and ways forward. *Contemporary Edu Psychol, 61*, Article 101865. https://doi.org/10.1016/j.cedpsych.2020.101865

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Edu Psychol, 41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Holleman, G. A., Hooge, I. T., Kemner, C., & Hessels, R. S. (2020). The 'real-world approach'and its problems: A critique of the term ecological validity. *Frontiers Psychol., 11*, 721. https://doi.org/10.3389/fpsyg.2020.00721

Howard, J. L., Bureau, J., Guay, F., Chong, J. X., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives Psychol Sci., 16*(6), 1300–1323. https://doi.org/10.1177/1745691620966789

Kiili, K., Lainema, T., de Freitas, S., & Arnab, S. (2014). Flow framework for analyzing the quality of educational games. *Entertainment Computing, 5*(4), 367–377. https://doi.org/10.1016/j.entcom.2014.08.002

Kiili, K., Moeller, K., & Ninaus, M. (2018). 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. J. M., Lindstedt, A., Koskinen, A., Halme, H., Ninaus, M., & McMullen, J. (2021). Flow experience and situational interest in game-based learning: Cousins or identical twins. *International Journal of Serious Games, 8*(3), 93–114. https://doi.org/10.17083/ijsg.v8i3.462

Knogler, M. (2017). Situational interest: A proposal to enhance conceptual clarity. In P. A. O'Keefe, & J. M. Harackiewicz (Eds.), *The science of interest* (pp. 109–124). Cham: Springer.

Koskinen, A., McMullen, J., Halme, H., Hannula-Sormunen, M., Ninaus, M., & Kiili, K. (2022a). The role of situational interest in game-based learning. In M. Bujik, J. Koivisto, & J. Hamari (Eds.), *Proceedings of the 6th international GamiFIN conference* (pp. 54–63).

Koskinen, A., McMullen, J., Ninaus, M., & Kiili, K. (2022b). Journal of Computer Assisted Learning. *Journal of Computer Assisted Learning.* https://doi.org/10.1111/jcal.12728

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Edu Res., 49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Krapp, A. (1999). Interest, motivation and learning: An educational-psychological perspective. *European J Psychol Edu., 14*(1), 23–40. https://doi.org/10.1007/BF03173109

Krapp, A. (2005). Basic needs and the development of interest and intrinsic motivational orientations. *Learning Inst., 15*(5), 381–395. https://doi.org/10.1016/j.learninstruc.2005.07.007

Krapp, A. (2007). An educational–psychological conceptualisation of interest. *Int J Edu Vocational Guid., 7*(1), 5–21. https://doi.org/10.1007/s10775-007-9113-9

Krath, J., Schürmann, L., & Von Korflesch, H. F. (2021). Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Comp Human Beha., 125*, Article 106963. https://doi.org/10.1016/j.chb.2021.106963

Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Edu Res Rev, 25*, 120–148. https://doi.org/10.1016/j.edurev.2018.10.001

Liu, Z., Moon, J., Kim, B., & Dai, C. P. (2020). Integrating adaptivity in educational games: A combined bibliometric analysis and meta-analysis review. *Edu Technol Res Develop., 68*(4), 1931–1959. https://doi.org/10.1007/s11423-020-09791-4

Matheson, G. J. (2019). We need to talk about reliability: Making better use of test-retest studies for study design and interpretation. *PeerJ, 7*, Article e6918. https://doi.org/10.7717/peerj.6918

Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learning Inst., 29*, 171–173. https://doi.org/10.1016/j.learninstruc.2013.04.003

Melo, C., Madariaga, L., Nussbaum, M., Heller, R., Bennett, S., Tsai, C. C., et al. (2020). Educational technology and addictions. *Comp Edu., 145*, Article 103730. https://doi.org/10.1016/j.compedu.2019.103730

Mills, C., D'Mello, S., Lehman, B., Bosch, N., Strain, A., & Graesser, A. (2013). What makes learning fun? Exploring the influence of choice and difficulty on mind wandering and engagement during learning. In *International conference on artificial intelligence in education* (pp. 71–80). Berlin, Heidelberg: Springer.

Ninaus, M., Kiili, K., McMullen, J., & Moeller, K. (2017). Assessing fraction knowledge by a digital game. *Computers in Human Behavior, 70*, 197–206. https://doi.org/10.1016/j.chb.2017.01.004

Ninaus & Nebel. (2021). A Systematic Literature Review of Analytics for Adaptivity Within Educational Video Games. In *Frontiers in Education* (p. 308). Frontiers. https://doi.org/10.3389/feduc.2020.611072.

Orvis, K. A., Horn, D. B., & Belanich, J. (2008). The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Comp Human Beha., 24*(5), 2415–2433. https://doi.org/10.1016/j.chb.2008.02.016

Pekrun, R., & Marsh, H. W. (2022). *Research on situated motivation and emotion: Progress and open problems*. Learning and Instruction, Article 101664. https://doi.org/10.1016/j.learninstruc.2022.101664

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Edu Psychol., 50*(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533

Plass, J., Homer, B., Mayer, R. E., & Kinzer, C. K. (2020). Theoretical foundations of game-based and playful learning. In J. Plass, R. E. Mayer, & B. Homer (Eds.), *Handbook of game-based training* (pp. 3–24). Cambridge, MA: MIT Press.

Plass, J. L., & Kaplan, U. (2016). Emotional design in digital media for learning. In S. Y. Tettegah, & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 131–161). New York: Elsevier.

Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Res Technol Edu., 52*(3), 275–300. https://doi.org/10.1080/15391523.2020.1719943

Reeve, J. (2012). A self-determination theory perspective on student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 149–172). New York: Springer.

Reeve, J., Lee, W., & Won, S. (2015). Interest as emotion, as affect, and as schema. In K. A. Renninger, M. Nieswandt, & S. Hidi (Eds.), *Interest in mathematics and science learning* (pp. 79–92). Washington: American Educational Research Association.

Renninger, K. A., Bachrach, J. E., & Hidi, S. E. (2019). Triggering and maintaining interest in early phases of interest development. *Learning, Cul Soc Int., 23*, Article 100260. https://doi.org/10.1016/j.lcsi.2018.11.007

Ronimus, M., Kujala, J., Tolvanen, A., & Lyytinen, H. (2014). Children's engagement during digital game-based learning of reading: The effects of time, rewards, and challenge. *Comp Edu., 71*, 237–246. https://doi.org/10.1016/j.compedu.2013.10.008

Rotgans, J. I., & Schmidt, H. G. (2018). How individual interest influences situational interest and how both are related to knowledge acquisition: A microanalytical investigation. *J Edu Res., 111*(5), 530–540. https://doi.org/10.1080/00220671.2017.1310710

Ryan, R. M., & Deci, E. L. (2020a). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Edu Psychol., 61*, Article 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Ryan, R. M., & Rigby, C. S. (2020b). Motivational foundations of game-based learning. In J. Plass, R. E. Mayer, & B. Homer (Eds.), *Handbook of game-based training* (pp. 153–177). Cambridge, MA: MIT Press.*Handbook of Game-Based Learning*.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion, 30*(4), 344–360. https://doi.org/10.1007/s11031-006-9051-8

Sajjadi, P., Ewais, A., & De Troyer, O. (2022). Individualization in serious games: A systematic review of the literature on the aspects of the players to adapt to. *Enter Computing, 41*, Article 100468. https://doi.org/10.1016/j.entcom.2021.100468

Sampayo-Vargas, S., Cope, C. J., He, Z., & Byrne, G. J. (2013). The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game. *Computers & Edu, 69*, 452–462. https://doi.org/10.1016/j.compedu.2013.07.004

Schmidt, H. G., & Rotgans, J. I. (2021). Epistemic curiosity and situational interest: Distant cousins or identical twins? *Edu Psychol Rev., 33*(1), 325–352. https://doi.org/10.1007/s10648-020-09539-9

Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., et al. (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

Shute, V., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C. P., … Sun, C. (2021). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *J Computer Assisted Learning, 37*(1), 127–141. https://doi.org/10.1111/jcal.12473

Shute, V. J., & Zapata-Rivera, D. (2007). Adaptive technologies. *ETS Research Report Series, 2007*(1), i–34. https://doi.org/10.1002/j.2333-8504.2007.tb02047.x

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*(3), 237–250. https://doi.org/10.1111/1467-9280.02438

Skinner, B. F. (1986). Programmed instruction revisited. *Phi Delta Kappan, 68*, 103–110.

Streicher, A., & Smeddinck, J. D. (2016). Personalized and adaptive serious games. In R. Dörner, S. Göbel, M. Kickmeier-Rust, M. Masuch, & K. Zweig (Eds.), *Entertainment computing and serious games. Lecture notes in computer science* (pp. 332–377). Cham: Springer. https://doi.org/10.1007/978-3-319-46152-6_14.

Tapola, A., Veermans, M., & Niemivirta, M. (2013). Predictors and outcomes of situational interest during a science learning task. *Instructional Sci., 41*(6), 1047–1064. https://doi.org/10.1007/s11251-013-9273-6

Tian, J., Bartek, V., Rahman, M. Z., & Gunderson, E. A. (2021). Learning improper fractions with the number line and the area model. *J Cognition and Develop., 22*(2), 305–327. https://doi.org/10.1080/15248372.2021.1890603

Van Oostendorp, H., Van Der Spek, E. D., & Linssen, J. (2014). Adapting the complexity level of a serious game to the proficiency of players. *EAI Endorsed Trans Game-Based Learning, 1*(2), 8–15. https://doi.org/10.4108/sg.1.2.e5

Vanbecelaere, S., Van den Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2020). The effectiveness of adaptive versus non-adaptive learning with digital educational games. *J Computer Assisted Learning, 36*(4), 502–513. https://doi.org/10.1111/jcal.12416

Walkington, C. (2021). Intrinsic integration in learning games and virtual instruction. *Edu Technol Res Develop., 69*(1), 157–160. https://doi.org/10.1007/s11423-020-09886-y

Wigfield, A., & Koenka, A. C. (2020). Where do we go from here in academic motivation theory and research? Some reflections and recommendations for future work. Contemporary. *Educational Psychology, 61*, Article 101872. https://doi.org/10.1016/j.cedpsych.2020.101872

Wilson, R. C., Shenhav, A., Straccia, M., & Cohen, J. D. (2019). The eighty five percent rule for optimal learning. *Nature Comm., 10*(1), 1–9. https://doi.org/10.1038/s41467-019-12552-4

Wouters, P., & Oostendorp, H. V. (2017). Overview of instructional techniques to facilitate learning and motivation of serious games. In P. Wouters, & H. van Oostendorp (Eds.), *Instructional techniques to facilitate learning and motivation of serious games* (pp. 1–16). Cham: Springer.

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers in Education: Artificial Intelligence, 2*, 100025. https://doi.org/10.1016/j.caeai.2021.100025