



Educational theories and computer game design: lessons from an experiment in elementary mathematics education

Abdelghani Es-Sajjade¹ · Fred Paas^{2,3}

© Association for Educational Communications and Technology 2020

Abstract

Although educational games have been used for a considerable time, their true potential for enhancing achievement and motivation is still being explored. We argue in this paper that we may get closer to realizing this potential if the theoretical underpinning of educational games is improved. We developed a simple interactive math game based on insights from cognitive load and game design theory, in which players solve mathematical problems competitively and creatively. To investigate the effects of the game on achievement and motivation, we ran an experiment with fifth- and sixth-grade students. Our results show that students who played the game achieved better math results than those who did not. We did not find any significant differences in math motivation. Another promising finding, albeit beyond the scope of our hypotheses, relates to the potential effects of math games on student behavior.

Keywords Game-based learning · Cognitive load · Elementary education · Mathematics

Introduction

Research into the effectiveness of games as a learning vehicle has a long history, especially in the domain of mathematics (e.g., Kebritchi et al. 2010; Steiner and Kaufman 1969). In a review by Randel et al. (1992) of the effectiveness of games for educational purposes, math was the subject area where studies showed computer games to have the most positive impact. Similarly, in a recent meta-analysis by Wouters et al. (2013), computer-based games were found to have a small positive effect ($d=0.17$) on mathematics learning. However, the findings have not been conclusive. For example, in a recent review of trends in

✉ Abdelghani Es-Sajjade
aes.sajjade@gmail.com; abdelghani.es-sajjade@roehampton-online.ac.uk

Fred Paas
paas@essb.eur.nl

¹ University of Roehampton Online, Roehampton Ln, London SW15 5PU, UK

² Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands

³ Early Start Research Institute, University of Wollongong, Wollongong, Australia

serious gaming for education, Young et al. (2012) found little support for the educational value of video games in math.

One of the reasons for inconclusive evidence on the achievement effects of mathematics games is that the games are often not founded on established instructional theories (Kiili et al. 2015). Decisions to use one particular game rather than others are often based on intuitive assumptions by educators or parents. Games branded as “educational” or “serious” are assumed to have positive effects on motivation and achievement (Say and Bag 2017). However, such assumptions can mean that the full potential of educational games may not be realized.

Following a different approach, we used principles from cognitive load theory (CLT) (Paas and Sweller 2012; Sweller et al. 2011, 1998) to guide the design and development of a mathematics game so that it would minimize the ineffective (i.e., extraneous) load and maximize the effective (i.e., intrinsic) load on players’ working memory to enhance their math achievement. This recommendation caused us to leave out some of the more “fun” elements of games, such as advanced animation, which of course gave us a challenge in terms of how to keep the game entertaining for the players (see Sects. 3 and 4). To address this challenge, and to maintain a balance between achievement and intrinsic motivation, we also incorporated recommendations from game design theory (GDT) (Annetta 2010; Devlin 2011). While elements of the learning task that are designed to improve achievement are focused on stimulating learners’ cognitive processes, elements that seek to increase intrinsic motivation are focused on increasing learners’ willingness to invest more time and cognitive resources in the task (Garris et al. 2002; Wouters et al. 2013). The game we developed for this study includes clear instructions that provide not only descriptive information but also a step-by-step visual explanation. More information on the design of the game is given in Sect. 3. The central question we attempt to answer in our paper is: Can we improve elementary students’ math achievement and math motivation by using a simple computer game that we designed and built based on insights from CLT and GDT?

In an empirical study focused on students from the fifth and sixth grade, we hypothesized that students who played the computer game during their math classes would show higher achievement in math than those who followed the regular math lessons without the game-based intervention. Also, we hypothesized that students who played the game during lessons would demonstrate greater motivation to learn math than students who did not use the game. To test our hypotheses, we created an interactive mathematics game (“MATHE-RIAL”) for the last two grades of elementary education in the Netherlands.

Theory

Math games and elementary education

In a yearly survey by “Kennisnet” (2015), an IT services institute funded by the Ministry of Education, Culture and Science in the Netherlands, 65% of the teachers surveyed said they used simulations and games in class either “sometimes,” “often,” or “very often.” Although this finding shows that many teachers use games in class, it does not provide any information regarding the games’ effectiveness in improving achievement or motivation. Teachers, parents, and students are subjected to marketing by game developers, all of which claim to be the “game-changer” in digital education, but we argue there is a crucial role for researchers in helping teachers and parents distinguish between fad and fact.

Several studies have looked at the effects of using math games in elementary education. Rieber (1996) proposed that introducing play into elementary education benefits learners' individual achievement and their attitude to learning. Computer games, in particular, have been shown to have positive effects on learners' intellectual and social development (Dempsey et al. 2002; Ke and Grabowski 2007). Ke and Grabowski (2007), for example, undertook a study with 125 fifth graders using an experimental group and a control group (game-playing versus non-playing) and found that those in the experimental group made significantly better progress in math. An interesting nuance in this study is that learners who played the game at an appropriate level of difficulty also showed positive changes in attitude, while those who did not showed no change in attitude. This finding indicates that using computer games in class should be done carefully and deliberately, informed by intelligent design decisions, so as to provide a challenging gaming experience catering to different skill levels. Kebritchi (2008) consolidates this idea by demonstrating how the added value of computer games in elementary education is mediated by such factors as training for teachers, trainers, and administrators. The author adds that, for technology such as a computer game to have any positive results, a reliable and accessible IT infrastructure needs to be available, and this is linked to the notion of a school's IT provision. Games that involve advanced animations and complex game play can place a significant burden on the internet network and computer CPU, particularly for schools whose IT provision is less advanced. This does not mean that these schools should not use educational computer games. We argue that simple computer games based on recommendations from academic research can still deliver benefits in terms of attitude and achievement, so there is no need for students in schools with lower levels of IT provision to lose out.

If a shift is made to educational games with a stronger theoretical foundation this indicates that academic experts have a vital role to play in game design and development. Intricate game designs, however, require substantial development budgets, which are often unavailable to elementary schools or educational researchers. We decided therefore to experiment with game design driven by established educational theories. We followed recommendations from cognitive load theory (Choi et al. 2014; Paas and Sweller 2012; Sweller et al. 1998, 2019) to examine whether a simple game, designed specifically for math development, could achieve the desired learning effect.

Cognitive load theory and math achievement

Cognitive load theory (CLT) revolves around optimizing learners' cognitive load in order to enhance learning performance (Sweller 1994). The central tenet of CLT is that individuals have a limited working memory (WM), which comprises components that process different types of information and which works in conjunction with an effectively unlimited long-term memory. The load on working memory can be managed through instructional design that uses different types of learning tasks and different types of information (text, audio, animation, etc.). With regard to learning tasks, for example, studies have found that worked examples impose a lower cognitive load and facilitates knowledge construction and transfer performance better than actually solving the problems (i.e., worked example effect; for a review see Renkl 2013). With regard to types of information, research into the modality effect has shown that effective working memory capacity can be increased by using both visual and auditory working memory rather than either of these two processors alone (for a meta-analysis see Ginns 2005). The modality effect is based on the assumption that working memory consists of two partially independent processors; one that is based

on auditory working memory that deals with verbal materials and another that uses visual working memory to deal with diagrammatic/pictorial information (e.g., Baddeley 1992). Splitting the processing in this way (i.e., using both visual and auditory working memory) means that if different types of information are used in combination—a picture plus narrative text, for example—a larger reservoir of cognitive capacity can be drawn upon to process that information, and this can have a positive effect on learning performance (Paas et al. 2003).

An intriguing finding in CLT is that using animations in instructional design has some negative effects, because dynamic visualizations such as animations are highly transitory. This means that if information disappears from the screen as the animation progresses, learners may be forced to process current information while also trying to remember previous information. Since working memory is limited in capacity (Cowan et al. 2015) and time (Peterson and Peterson 1959), if learners are not given any additional instructional support animations may therefore impose a high and unproductive load on their working memory, and this can have an adverse effect on learning (Ayres and Paas 2007a, b). Research has shown that animations become more effective when the accompanying text is spoken (e.g., Wouters et al. 2008a, b), when the animation is segmented (Spanjers et al. 2012), when cues are used to focus learners' attention (e.g., De Koning et al. 2011), and when learners can control the pace of the animation or replay certain sections (Hasler et al. 2007). These findings have important implications for contemporary games, which are hard to imagine without animations.

Another relevant finding in CLT research is the worked-example effect, which holds that, for novices, presenting full solutions to the problem (i.e., worked-out examples) that they must study carefully leads to better learning than the conventional strategy of requiring them to solve the problems (Sweller et al. 1998, 2019). CLT research has shown convincingly that the means–ends problem-solving strategy typically used by novices to solve conventional problems creates an unnecessary load on the working memory and interferes with learning (Paas 1992; Sweller and Cooper 1985). The means–ends strategy uses up considerable working memory capacity, because the learner must hold and process in working memory a series of different things: the current problem, the goal to be achieved, the relationships between them, any problem-solving operators that could reduce differences, and any sub-goals (Sweller et al. 2019). Using worked examples instead focuses learners' attention on the problem itself and the associated operators, thereby helping the learner to use existing knowledge to solve the problem. In addition to CLT, we also included suggestions from educational game design research (Annetta 2010; Annetta and Holmes 2006) to achieve our objective of enhancing motivation in math.

Game design theory and math motivation

We developed MATHERIAL as a collaborative initiative in which researchers and educators teamed up with game developers to build educational games in different domains. MATHERIAL is a two-player interactive math game in which players not only solve but also create mathematical equations while competing in several parallel games. The game was built in HTML5, using the JQuery Javascript library, and used a responsive website design. This meant that the game could be accessed by students using any of the popular smartphone, tablet, and desktop platforms.

The game starts with an invitation to watch a demo game, which explains the rules of the game and shows how an optimal equation can be created (Wouters et al. 2008a, b).



Fig. 1 Player in process of creating and solving a problem

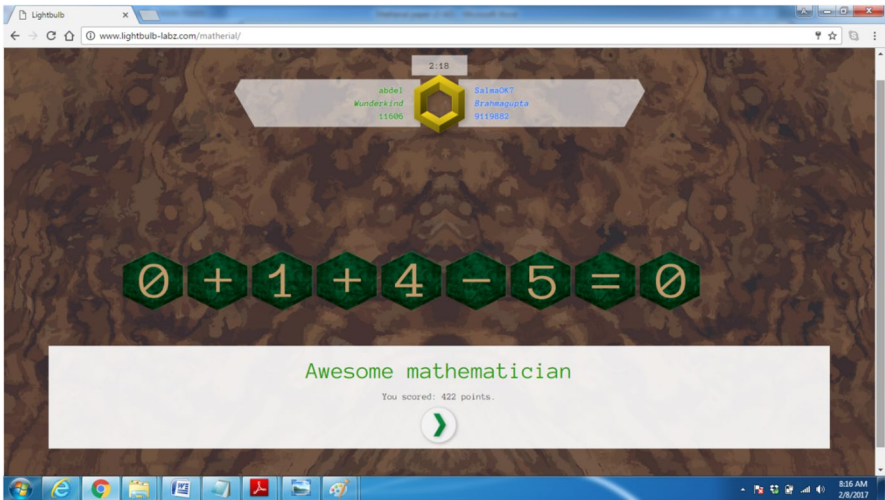


Fig. 2 Player has submitted a perfect equation

After the demo game, the player then either challenges another player or accepts a challenge. The player will also see a list of unfinished games if he/she has played before. Any two-player game ends as soon as one of the players submits an incorrect equation, which means the opponent wins the game. Players can invite classmates or peers from other schools to compete and can have several games running simultaneously. A screenshot of the player screen is shown in Figs. 1 and 2.

Figure 1 depicts the experimentation process during which players can move elements and place them in different slots as long as they do not click the “submit” icon. Players have a maximum of 180 s in which to do this. Figure 2 shows the screen that appears when

Table 1 Character progress in game

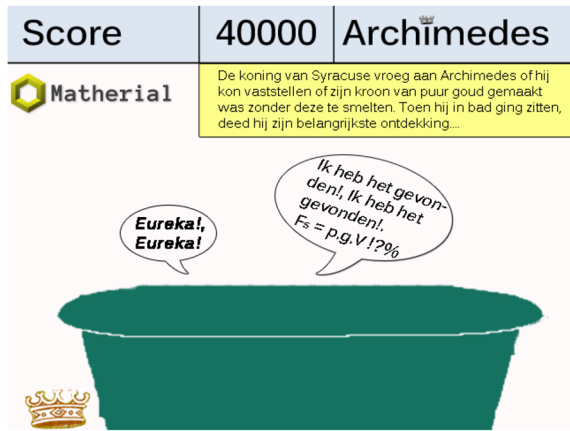
Points	Status
0	N00b de Newbie
1000	Abacus
2000	Solar-Powered Casio
5000	Texas Instrument
10,000	Wunderkind
20,000	Chang Tshang
40,000	Pythagoras
80,000	Archimedes
150,000	Fibonacci
300,000	Ramanujan
500,000	Omar Khayyám
750,000	Riemann
1,000,000	Andrew Wiles
2,000,000	Ibn Al-Haytham
3,000,000	Thales
5,000,000	Leibniz
7,000,000	Brahmagupta
10,000,000	Kolmogorov
13,000,000	John Von Neumann
16,000,000	Euler
20,000,000	Gauss
25,000,000	Euclid
30,000,000	Dr. Snuggles
40,000,000	El-Khwarizmi
50,000,000	Einstein

a player submits a perfect equation. Instead of presenting players with a traditional equation where they have to complete one side of the equation, we provide them with numbers and operators. The players can use as many or as few slots as they want to create and solve an equation. The more slots they use, the more points they earn, with bonus points being given for a perfect equation. A perfect equation is one in which the player uses all the available elements to cover all the slots.

Annetta (2010) proposes several guiding principles for educational game design: identity, immersion, interactivity, increased complexity, informed teaching, and instruction. *Identity* is the interface between the player and the game environment, and is often represented by an avatar representing the player's character. It maintains the player's interest in the game. The programming of sophisticated avatars with advanced abilities was beyond our financial capability, but we introduced *player status development* instead. Player status development is a feature that involves a journey in which players shift from one character to the next as their score increases (see Table 1).

As players accumulate more points, they reach different status levels, each associated with a significant event or influential figure in math history. For example, with 80,000 points players reach the Archimedes of Syracuse character level, named after the mathematician who invented ... hydrostatics, a significant development in the history of

Fig. 3 The status “Archimedes” is achieved when reaching a game score of 4000. The description in Dutch (yellow field) explains the interaction between the king of Syracuse and Archimedes before his famous discovery (Color figure online)



mathematics. This adds an additional educational dimension to the game. We assumed it would encourage competition and educate players in general math history. An example of Archimedes status is depicted in Fig. 3.

As previously mentioned, we avoided animations to ensure that the game would draw more on players' working memory capacity and would thus help to improve their math achievement. This made the concept of immersion for players difficult to achieve. Nonetheless, we expected that the competitive element of the game, particularly where players felt a sense of rivalry or had similar levels of skill, would lead to some degree of immersion.

We made good use of the principle of *interactivity*, as players solved equations in two-player battles while being invited by other challengers to take part in simultaneous games. We also incorporated some interactivity between the player and the game. The initial visual introduction of the game was a good example of this. Individuals could complete this introductory session at their own pace; i.e., they could click through pages and figures depending on whether they wanted to complete, repeat, or skip the instruction. The game also provided feedback after every solution submitted. Admittedly, there was not much variety in the range of feedback content, but there was at least a response from the game to indicate whether the player had submitted an incorrect, correct, or perfect solution.

We took the principle of *increased complexity* into consideration by building in different levels and increasing the number of elements at each level. At level 1, players were expected to build and solve an equation using four elements. For each subsequent level, we added two elements, which increased the challenge of the game. The highest level achieved by players during the experiment was level 12, which included equations with 26 elements.

Using a simple database structure (see Fig. 4), we were able to record events such as games won, games lost, character progression, level in current game, and whether the player had submitted a perfect solution. We downloaded performance reports, which we shared with the management of each participating school. The game's database had a feature that enabled reports to be downloaded, but this was a manual process. It required an administrator to log in, select the variables, and download the performance results. We could download the cumulative data at any given point. After downloading a report, we needed to organize it using spreadsheet software before sending it to each school. The ability to provide feedback through embedded assessment meets the principle of *informed teaching*.

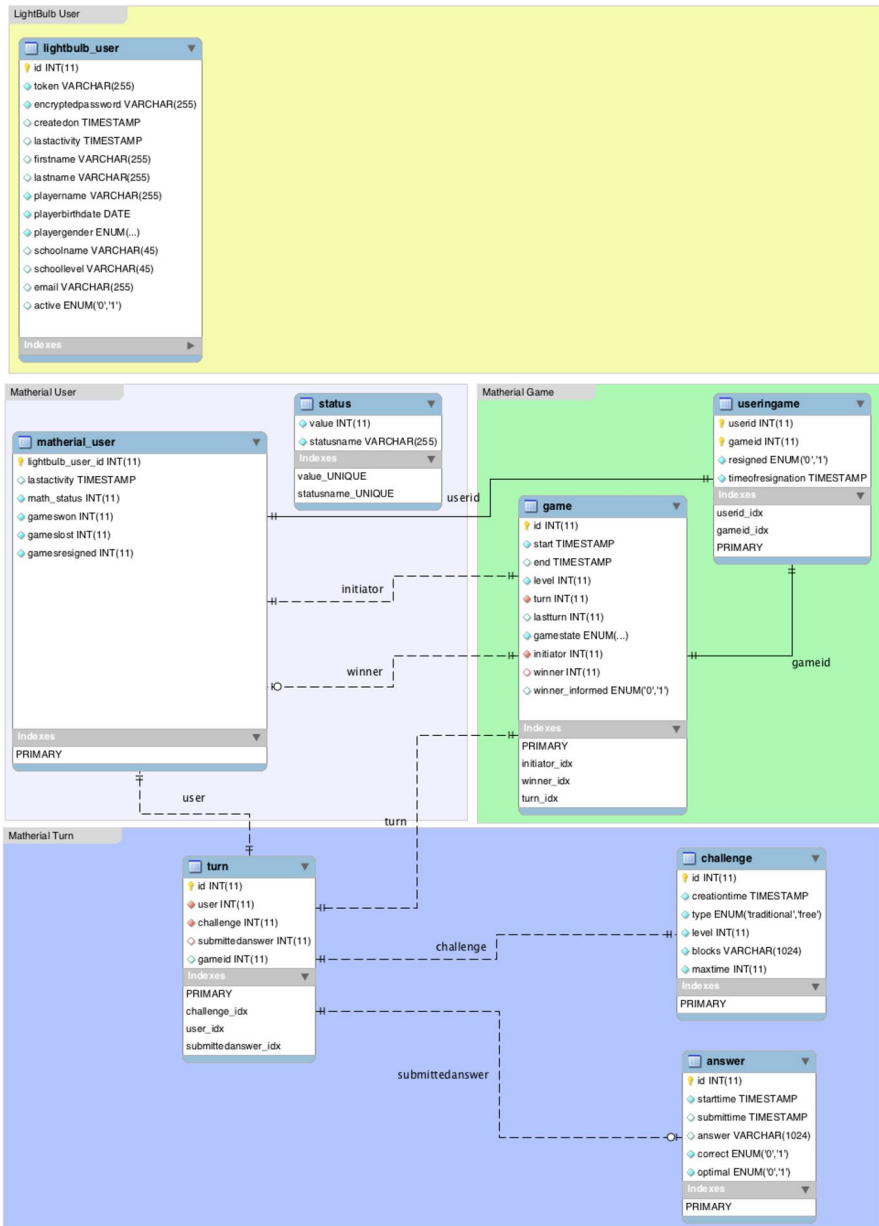


Fig. 4 Database structure for recording events and reporting results

Finally, the game also provided visual *instruction* to players by demonstrating the rules (e.g., operator preference) and steps needed to create a perfect solution. Instead of presenting this instruction as textual information, we presented it in the form of a worked example (i.e., demo game); this is as recommended by CLT research, which suggests that instruction in the form of solved problems by experts is more effective than a trial-and-error approach.

We did not assume players had not seen an operator preference before as we were dealing with fifth- and sixth-grade students. Our goal was to improve the extent to which students could apply this rule creatively using computer game play.

Overall, the general research question that guided our investigation was whether we could improve achievement and motivation to learn math using a simple computer game that we had designed and built specifically for this purpose, based on insights from CLT and GDT. Because we had included CLT, we expected there to be a positive effect on students' math achievement. We initially asked schools to allow students from the experimental group to play for 30 min each school day during math class. Schools could not commit to that, but they could commit to four days per week. We were allowed to introduce the game in math class for 4 weeks, as the intervention required considerable resources from the schools. Therefore,

Hypothesis 1 was: Students who have played the game for 2 h per week over a 4-week period will achieve better math results than students who have not played the game.

Furthermore, although CLT research on animations and motivation is somewhat limited, we expected that avoiding complex animations would not necessarily decrease students' math motivation. Research has shown that in many cases animations are not seen as more effective than static graphics (e.g., Hoeffler and Leutner 2007), nor as more motivating (Kim et al. 2007). Also, we expected that the recommendations we drew on from GDT would enable us to make the learning experience more enjoyable.

Hypothesis 2 Students who have played the game for 2 h per week over a 4-week period will report a higher level of motivation to learn math than students who have not played the game.

Method

Participants and experiment implementation

We ran the experiment with five schools in different cities in the Netherlands (i.e., Breda, Dordrecht, Maastricht, Rotterdam, and Schiedam). We had a total of 227 participants (101 males, 126 females), all from the fifth and sixth grade. The average age of the participants was 11.28 years ($SD=0.81$). The experimental group comprised 115 students and the control group 112 students. To have a similar number of fifth and sixth graders in both the experimental and the control group, we asked each school to provide a control and an experimental group from either the fifth or the sixth grade. We used a quasi-experimental design in which the two groups within each school were randomly assigned to the treatments. The ratio of males to females was 51:64 in the experimental group and 49:63 in the control group. The ratio of fifth to sixth grade students was 72:43 in the experimental condition and 38:74 in the control condition.

Before the experiment started, we gave presentations to teachers and students in each school explaining the game play and the experimental design. Students in the experimental group were instructed to play MATHERIAL for 2 h per week, spread over four days, over a period of 4 weeks. Students in the control group followed the regular math lessons without the intervention. After 4 weeks all students took a written test, with one part in which they

had to solve mathematical equations, and a second part in which they had to answer questions about their motivation to learn math. During the experiment, we visited the schools several times and kept participants and teachers up to date regarding the scoreboard and intermediate results. A Facebook page was created to provide information to teachers and students from the participating schools. On this page, students were invited to share screenshots of their scores in the game. The Facebook page added an element of interactivity at the level of the (student) player community. This was extremely useful as it allowed students to learn from each other by sharing perfect solutions. Helping students to learn from each other is one of the four meta-principles for fostering knowledge integration in information and communications technology-mediated science (Annetta 2010; Linn 2004).

Data collection

We obtained demographical data on the participants before the start of the experiment in order to create user accounts for the game. After participants had played the game for 4 weeks, we organized an assessment session with all of them. The first part of this assessment was a written math test where we recorded the start and end time for each participant. This test was designed to reflect the achievement hypothesis and required students to solve fifteen traditional equations of the kind taught in math class, of varying lengths and levels of difficulty. For each of the equations, between 1 and 8 points could be awarded, depending on the difficulty of the equation. The maximum score for this part of the test was 60 points.

The second part of the assessment focused on the motivation hypothesis. In this part, we asked participants to answer 17 questions using a four-point Likert scale ranging from 1 (Never) to 4 (Always). A final score was calculated by summing the scores (maximum of 68) for each item. We adapted the instrument from the Dutch 'Attitude towards Mathematics' scale, designed by Martinot et al. (1988) (see also Vandecandelaere et al. 2012). This scale measures the extent to which individuals enjoy learning mathematics. The instrument proved sufficiently reliable, with a Cronbach's alpha of 0.73.

To determine whether there were any differences in mathematical proficiency between groups at the start of the study, we used the scores achieved by students in the mathematics section of a national test of educational achievement (CITO) administered in the fifth grade. We converted scores to a scale of 0–100. In the Netherlands, this test is generally used in the last phase of elementary school (when students are around the age of 12) to determine what type of high school individual students should enter.

Results

Quantitative results

After 4 weeks of play, we collected the following score data which we summarize in Table 2.

We removed students who did not play the game ($N=25$) and those who did not attend the post-game test ($N=8$), leaving us with a final total of 115 students in the

Table 2 Score range per school

School indicator	Grade	Number of students	Min score	Max score	Score range
A	5	34	118	793,018	792,900
B	5	22	4468	243,053	238,585
C	5	17	0	1,245,399	1,245,399
C	6	24	0	163,894	163,894
D	6	15	13,188	470,784	457,596
E	6	11	75,429	401,734	326,305

experimental group. An independent samples t-test on the CITO test scores¹ revealed that there was no difference in mathematical proficiency between those in the experimental group ($M=62.06$, $SD=16.77$) and those in the control group ($M=63.08$, $SD=18.36$) ($t(68) = -0.238$, $p=0.81$).

A one-way analysis of variance (ANOVA) was used to determine whether our intervention had an effect on the math test results (achievement hypothesis) of those in the experimental condition, on the time taken to complete the test (achievement hypothesis), and on their motivation. Partial eta-squared was used to measure effect size, with 0.01, 0.06, and 0.14 being considered small, medium, and large respectively (Cohen 1988). To investigate the effect on math achievement, we looked mainly at the number of points students had accumulated after solving the 15 math equations. The lowest score achieved was 7, while the highest was 59. In addition to looking at the test result, we were also interested in the time students needed to complete the test as a secondary measure of achievement.

With regard to the test result, an ANOVA revealed that our intervention had worked, ($F(1, 226)=15.975$, $p<0.0005$), and that the effect size was medium to large (partial eta squared=0.066). Consistent with the first hypothesis, students who played the MATHE-RIAL game scored higher on the math test ($M=29.3$, $SD=11.6$) than those who did not ($M=23.9$, $SD=9.3$). With regard to the time (in seconds) needed to complete the math test, an ANOVA did not reveal any effect from the intervention ($F(1, 226)=3.175$, $p>0.05$ (experimental condition, $M=1119$, $SD=446$, control condition, $M=1015$, $SD=428$)), and the effect size was small (partial eta squared=0.014). With regard to motivation to learn math, an ANOVA revealed no effect of the intervention ($F(1, 226)=0.006$, $p>0.93$ (experimental condition, $M=38.75$, $SD=8.18$, control condition, $M=38.65$, $SD=9.76$)), and the effect size was very small (partial eta squared=0.0001).

Using games in math education can differentiate learning approaches

During our final presentations to participating schools, one of the schools shared with us the case of a sixth-grade student whom they referred to as having behavioral challenges. He had been in the experimental group of that particular school. The school's management was preparing to transfer him to a special needs education institute. During the experiment, however, he was the most active player of the game in his school and ended up as the

¹ Due to privacy issues we were only able to obtain mathematical proficiency scores for a limited number of children (i.e., experimental condition $N=31$, control condition $N=39$).

second-best player across all five schools, which was recognized on the game's Facebook page. In the Facebook competition we compiled a top 10 of best players across the schools and celebrated the winners by giving them special recognition and prizes. Positive changes in this particular student's behavior accompanied his outstanding performance in the game. This development caused the school's management to invite the education inspection authority to demonstrate to them this student's case. In an e-mail, the school's Information and Communication Technology Coordinator told us that:

His behavior changed from negative to very positive. He was smiling again, helpful, and communicative with his classmates and teachers. His grades became better; he was encouraged by his classmates to keep playing because he had achieved such a high score in the game. We could see him shining and growing.

Although this finding was particularly exciting, a perhaps more important issue is whether this student's outstanding game performance should have come as a surprise. The fact that the school invited the inspection authority to reconsider his case, indicates that the event must have been unexpected. Although it is promising that our intervention brought out the skills of this student, because it indicates that games can offer different approaches to learning, this school should already have had mechanisms in place to bring out those skills. The student's different approach to learning may have been wrongly diagnosed as "behavioral challenges," leading to a perhaps unjust transfer to a school for special educational needs. We therefore encourage future research into other positive effects of games. Additional studies could provide a better understanding of how educational games can help schools to work with a broader range of learning approaches, which could help reveal the potential in more students.

Discussion

We started this paper with the question: Can we improve elementary students' math achievement and math motivation by using a simple, educational theory-based computer game? With limited resources, we built MATHERIAL, an interactive game in which fifth- and sixth-grade students were invited to creatively solve math problems in an interactive setting. As our guiding principles for game development, we drew on recommendations from cognitive load theory (Paas and Sweller 2012; Sweller et al. 1998, 2011) and game design theory (Annetta 2010; Devlin 2011).

The results of our study suggest that MATHERIAL players demonstrated higher math achievement than their peers who did not play. This adds to the findings of other studies that reported that computer games can have a positive effect on students' performance in math (Annetta et al. 2009; Harris et al. 2008; Ke and Grabowski 2007; Van Den Heuvel et al. 2013). Our contribution to this research is that educators do not always have to resort to using expensive commercial games. Instead, we propose that simple computer games, which are theoretically underpinned by principles, might also provide positive effects in terms of math achievement. Future research using simple, purpose-built games in other learning domains, e.g., literacy, could test whether these may have similar effects on learner achievement.

Regarding our second hypothesis, even though we used insights from GDT, playing the game did not seem to have any discernible effect on students' motivation to learn math. In commercial games, vivid animations and sophisticated scenarios are considered crucial

to provide an entertaining experience. Although our game had neither of these, we still expected it to have a positive effect on motivation because of the increasing levels of complexity, the player's journey through math history, and the social interaction between the competitors (Annetta 2010; Devlin 2011). These factors did not, however, lead us to observe any positive effects on students' motivation. This could be because we incorporated into the game features that we believed would be entertaining but that proved not to be so to students. Another reason could be the design of the questionnaire; i.e., perhaps the questions were not appropriately framed to allow us to detect any effects on motivation. The participants ranged in age between 9 and 13, and hence it is possible that some of them did not understand the questions correctly.

We would welcome further research on how simple, scientifically validated games can affect learner motivation. The availability of lower-cost and perhaps simpler computer games might give more schools scope and incentive to introduce these learning tools, which could reduce their dependence on the commercial computer game industry. This creates a win-win situation in which schools benefit by having access to effective educational games, and researchers benefit by acquiring data for further research in this important area. Independent game development by educators and researchers could also serve broader societal interests, in that it could help to provide access to educational technology to as many social groups as possible and could also seek to ensure that, where possible, students are not excluded because their guardians cannot afford expensive game subscriptions.

We have experimented with various strategies for directly and indirectly regulating cognitive load. Future studies could validate these design features intended to regulate cognitive load, and explore possible interaction effects with other relevant variables such as students' prior math knowledge or language proficiency. It is challenging to regulate cognitive load in practice and we therefore expect several game design iterations before an effective design will become available. We therefore strongly encourage future research using MATHERIAL or similar games to build on the findings of this study.

Finally, although this was not within the scope of our hypotheses, our discovery of positive changes in the behavior of one sixth-grade student who played MATHERIAL was inspiring and intriguing. Research has shown that instructors play a critical role in facilitating and optimizing the experience of educational technology for different approaches to learning (Bunninga et al. 2010; Marco et al. 2013). If the student in our experiment was indeed a special needs learner, then our finding would indicate a promising avenue for future research on how educational games might affect the achievement and motivation of special needs learners. Overall, this intriguing outcome gave additional meaning to this research project, and we feel it is worth exploring on whether this finding is more widely applicable.

We acknowledge that this study has some limitations. First, there was a clear imbalance in composition between the experimental and control groups, as there were 72 fifth graders in the experimental group and 74 sixth graders in the control group. In future studies, it will be important to ensure that, if the game is being used with students from more than one grade, the experimental and control groups are carefully balanced in terms of the grades. Second, our study was based on an ecologically valid intervention as the game was played in a classroom setting. However, we could not control how much time participants spent on playing the game or on any other type of relevant practice, or how much time their teachers spent discussing the game. As can be seen from Table 2, school C included students with a score of 0; the fact that these students did not score any points seems to indicate they spent little or no time on the game. Although these students were excluded from the experimental group, we cannot rule out other causal mechanisms for the current results. Concerning

better control for time-on-task effects, in future research students or their guardians could be asked to keep a diary in which they record how much time throughout the day was exactly spent on playing the computer game.

A final limitation relates to the mathematical proficiency scores, which we could not get for all participants, and the fact that, for some of the participants (i.e., sixth-grade), these scores were already one year old. Although classes within schools were randomly assigned to the experimental and control condition, we cannot say with confidence that the groups started at the same level. Future research should use a pretest to determine the mathematical proficiency level of all participants at the start of the study.

Practical implications

Our findings have several implications for those involved in educational game development. We categorize these implications into: (a) design decisions and consumer acceptance, (b) administrative issues, and (c) fit with the school curriculum.

First, regarding design decisions and consumer acceptance, we made a number of assumptions, based on the theories we referred to in this study. The fact that we did not find a positive effect on motivation could be due to our interpretation of how these theories would translate into educational game design. Academics should not base their game design on theoretical principles alone. Collaborating more closely with educators, gaming professionals, and students might result in a better outcome and a more entertaining experience for end-users. For example, our use of references to historical mathematical thinkers may have been more exciting for us than for the end-users. We believe that we could have prevented this and other issues by involving the end-users and their teachers much earlier in the development of the game. Nonetheless, after the experiment, we received suggestions from students and teachers on how to improve the game. Future versions of the game that include these features may be possible if we are able to attract funding for R&D. Table 3 summarizes some of the features that students and teachers wanted to see in a new version of the game. The original table is more extensive than the one presented here, and this extract from it serves merely as illustrative evidence. The columns of Table 3 are explained as follows:

- Number: a unique reference to the desired game feature.
- Requirement: a short label referring to the desired game feature.
- Description: a detailed description of the desired game feature.
- C/I/E: the area of the game that would include this feature. We identified three main areas: *Content* relates to learning content, *interface* pertains to visual presentation, and *experience* revolves around ease of use.
- MoSCoW prioritization: How important is it to add a particular feature? Must have, Should have, Nice to have, or Will not have.
- Justification: an explanation for the chosen MoSCoW prioritization.

A second implication is that user administration needs to be thought through very carefully. Registering and maintaining user accounts in the game's system was labor-intensive. We used a specific format of.csv files that were filled in by teachers who sent them to us before we uploaded the new accounts to the system. It frequently

Table 3 New features for next version

No	Requirement	Description	C/I/E	MoSCow prioritization	Justification
1.1.1	Mnemonic	Add a mnemonic in the next version to help students apply the operator preference rules	C	S	Students struggle to remember the operator preference rule which affects their performance. For some students this was not yet introduced in class
1.1.2	Background theme	Present a visual progression of different background themes as the game level increases	I	S	Feedback received from students across schools. The interface is too monotonous
1.1.3	Search engine	Include a search engine that allows searching for particular opponents	C	S	Students regularly asked for games against friends
1.1.4	Tips	For more advanced levels (3+) the game should give tips to help students complete more challenging calculations. For example, if half the time has passed then game fills two slots	C	S	Suggested by a teacher responsible for a class with learning difficulties
1.1.6	Math stories	Introduce story-based math problems in the game	C	N	Suggested by students
1.1.7	Sound	Introduce sound effects to make the game more fun. For example, a happy sound if the answer is correct and a buzzer if the answer is incorrect	E	M	Regularly asked by students. Could make the game more enjoyable
1.1.8	Avatars	Allow students to personalize the game by choosing an avatar that represents them in the game. This avatar can also be seen by other players	E	N	The option to add a personal touch to the game would make the game more enjoyable

happened that users lost their password or wanted to make changes to their username. This required manual effort from the research team as students or teachers did not have access to the user admin section of the system. Care should be taken to devise an efficient process for user administration and it is important to set out clearly at the start who is to be responsible for this. One suggestion might be to allocate an administrative user account to each school. The school can then appoint one of its staff, either a teacher or an administrator, to manage the user accounts for that school. Alternatively, the researchers involved could ask one of their PhD students to take on this role in exchange for data that they could use for a thesis.

Third, at times, we struggled with communication between team members. We mainly worked virtually, with team members based in different geographical locations. We used e-mail and online collaboration software to coordinate tasks and share ideas. These platforms proved particularly useful for discussing suggestions on how to solve technical issues or for making changes in the game's graphical user interface. A disadvantage of working virtually was the slow pace of action. We often had to wait for several days to get a response to an urgent question, which in turn affected subsequent steps and actions. This was particularly disadvantageous in the case of technical issues. An example was that, in the first week, some students received the system message "we think you're trying to cheat," although that was not the case. This feedback mechanism was built in to flag up attempts to refresh the page and reset the timer, which would allow players more than the maximum time to finish a game task. It took several days to solve this problem, which may have adversely impacted learner motivation.

Another reason for the slow pace of action within the team was that the software developers were involved in several projects simultaneously. This research project was not always their first priority as they also had to run a business, and we thus had to compete for their attention. Working virtually was not the best way of developing this type of game. Although it offered us flexibility in planning, dividing, and executing tasks, it reduced efficiency, particularly in terms of communication and action.

A final implication centers on the fit with the school curriculum. Developing a game that can be used across different school grades creates a challenge regarding how to achieve such a fit. In our experiment we targeted two grades, where the students were focusing on different mathematical topics at that particular time. An important question is: Should an educational math game focus on fit with the school curriculum at all? We argue this does not have to be a priority as a game should be complementary to activities at school, not supplementary. If future projects were more concerned with fit with the school curriculum, then we would recommend building a game that has a narrower focus in terms of the targeted end-users and the point at which it is to be used. A game that is designed to be used with one particular grade of students, and is introduced in the semester in which similar topics are covered in the curriculum, could indeed offer an exciting area for future research.

Acknowledgements This paper would not have been possible without the Grace and Mercy of God, the Most Gracious, the Most Merciful. Furthermore, we would like to thank Younes Hida, Theo Lagendijk, Coen de Koning, Hubert Knulst, and Andre Limburg for contributing to this project in various ways. We are also grateful to the participating schools and their managers for allowing us to work closely with their students throughout this research project. Finally, we would like to thank Professor Lin Lin and the four anonymous reviewers for their excellent comments and constructive criticism.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Aldrich, C. (2004). *Simulation and the future of learning*. San Francisco: Wiley.
- Annetta, L. A. (2010). The "I's" have it: A framework for serious educational game design. *Review of General Psychology, 14*(2), 105–112.
- Annetta, L. A., & Holmes, S. (2006). Creating presence and community in a synchronous virtual learning environment using avatars. *International Journal of Instructional Technology and Distance Learning, 3*(8), 27–43.
- Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M. T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education, 53*(1), 74–85.
- Ayres, P., & Paas, F. (2007a). Making instructional animations more effective: A cognitive load approach. *Applied Cognitive Psychology, 21*(6), 695–700.
- Ayres, P., & Paas, F. (2007b). Can the cognitive load approach make instructional animations more effective? *Applied Cognitive Psychology, 21*(6), 811–820.
- Baddeley, A. (1992). Working memory. *Science, 255*, 556–559.
- Bunning, K., Heath, B., & Minnion, A. (2010). Interaction between teachers and students with intellectual disability during computer-based activities: The role of human mediation. *Technology and Disability, 22*(1,2), 61–71.
- Choi, H. H., Van Merriënboer, J. J., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: Towards a new model of cognitive load. *Educational Psychology Review, 26*(2), 225–244.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cowan, N., Ricker, T. J., Clark, K. M., Hinrichs, G. A., & Glass, B. A. (2015). Knowledge cannot explain the developmental growth of working memory capacity. *Developmental Science, 18*(1), 132–145.
- Dempsey, J. V., Haynes, L. L., Lucassen, B. A., & Casey, M. S. (2002). Forty simple computer games and what they could mean to educators. *Simulation & Gaming, 33*(2), 157–168.
- De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2011). Attention cueing in an instructional animation: The role of presentation speed. *Computers in Human Behavior, 27*, 41–45.
- Devlin, K. (2011). *Mathematics education for a new era: Video games as a medium for learning*. Boca Raton, FL: CRC Press.
- Fouze, A. Q., & Amit, M. (2017). Development of mathematical thinking through integration of ethnomathematic folklore game in math instruction. *Eurasia Journal of Mathematics, Science and Technology Education, 14*(2), 617–630.
- Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*, 441–467.
- Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE), 1*(1), 1–4.
- Gee, J. P. (2005). What would a state of the art instructional video game look like? *Innovate: Journal of Online Education, 1*(6).
- Ginn, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*, 313–331.
- Harris, A., Yuill, N., & Luckin, R. (2008). The influence of context-specific and dispositional achievement goals on students' paired collaborative interaction. *British Journal of Educational Psychology, 78*, 355–374.
- Hasler, B. S., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology, 21*, 713–729.
- Hoeffler, T., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*, 722–738.
- Ke, F. (2008a). A case study of computer gaming for math: Engaged learning from gameplay? *Computers & Education, 51*, 1609–1620. <https://doi.org/10.1016/j.compedu.2008.03.003>.
- Ke, F. (2008b). Computer games application within alternative classroom goal structures: Cognitive, meta-cognitive, and affective evaluation. *Education Technology Research and Development, 56*, 539–556. <https://doi.org/10.1007/s11423-008-9086-5>.
- Ke, F., & Grabowski, B. (2007). Gameplaying for maths learning: cooperative or not? *British Journal of Educational Technology, 38*(2), 249–259.
- Kebritchi, M. (2008). Effects of a computer game on mathematics achievement and class motivation: An experimental study. *Dissertation Abstracts International Section: A Humanities and Social Sciences, 69*(6-A), 2121.
- Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education, 55*(2), 427–443.

- Kennisnet (2015). Vier in balans-monitor 2015. De belangrijkste resultaten voor het primair onderwijs [Four in balance-monitor 2015: The most important results for elementary education.]: Retrieved from https://www.kennisnet.nl/fileadmin/kennisnet/leren_ict/visie_implementatie/bijlagen/Kennisnet_Vier_in_balans-monitor_2015_De_belangrijkste_resultaten_voor_het_primair_onderwijs.pdf
- Kiili, K., Devlin, K., & Multisilta, J. (2015). Editorial: Is game-based math learning finally coming of age? *International Journal of Serious Games*, 2(4), 1–4.
- Kim, S. I., Yoon, M., Whang, S. M., Tversky, B., & Morrison, J. B. (2007). The effect of animation on comprehension and interest. *Journal of Computer Assisted Learning*, 23(3), 260–270.
- Linn, M. C. (2004). Using ICT to teach and learn science. In R. Holliman & E. Scanlon (Eds.), *Mediating science learning through information and communication technology* (pp. 9–26). London: Routledge Falmer.
- Marco, J., Cerezo, E., & Baldassarri, S. (2013). Bringing tabletop technology to all: Evaluating a tangible farm game with kindergarten and special needs students. *Personal and Ubiquitous Computing*, 17(8), 1577–1591.
- Martinot, M. J., Kuhlemeier, H. B., & Feenstra, H. J. M. (1988). Het meten van affectieve doelen: De validering en normering van de belevingsschaal voor wiskunde (BSW) [Measuring affective goals: Validation and standardization of the Attitude Scale towards Mathematics (ASM)]. *Tijdschrift voor Onderwijsresearch*, 13, 65–76.
- Mohamed, H., & Jaafar, A. (2010, June). Challenges in the evaluation of educational computer games. In *Information Technology (ITSim), 2010 International Symposium in* (Vol. 1, pp. 1–6). IEEE.
- Okolo, C. M. (1992). The effect of computer-assisted instruction format and initial attitude on the arithmetic facts proficiency and continuing motivation of students with learning disabilities. *Exceptionality*, 3, 195–211. <https://doi.org/10.1080/09362839209524815>.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429–434.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38(1), 1–4.
- Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review*, 24(1), 27–45.
- Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58(3), 193–198.
- Randel, J. M., Morris, B. A., Wetzel, C. D., & Whitehill, B. V. (1992). The effectiveness of games for educational purposes: A review of recent research. *Simulation & Gaming*, 23(3), 261–276.
- Renkl, A. (2013). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38, 1–37.
- Rieber, L. P. (1996). Seriously considering play: designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology, Research, and Development*, 44(1), 43–58.
- Say, S., & Bag, H. (2017). The evaluation of the effect of a newly designed computer game on 7th grade students' motivation towards science and aggression. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(8), 5379–5393.
- Spanjers, I. A., van Gog, T., Wouters, P., & van Merriënboer, J. J. (2012). Explaining the segmentation effect in learning from animations: The role of pausing and temporal cueing. *Computers & Education*, 59(2), 274–280.
- Steiner, H. G., & Kaufman, B. A. (1969). Checker games in operational systems as media for an inductive approach to teaching algebra. *Educational Studies in Mathematics*, 1(4), 445–483.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory* (Vol. 1). New York: Springer.
- Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, 2(1), 59–89.
- Sweller, J., Van Merriënboer, J. J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31, 261–292.
- Vandecandelaere, M., Speybroeck, S., Vanlaar, G., De Fraine, B., & Van Damme, J. (2012). Learning environment and students' mathematics attitude. *Studies in Educational Evaluation*, 38(3–4), 107–120.

- Van den Heuvel-Panhuizen, M., Kolovou, A., & Robitzsch, A. (2013). Primary school students' strategies in early algebra problem solving supported by an online game. *Educational Studies in Mathematics*, 84(3), 281–307.
- Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, 34(3), 229–243.
- Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Students*, 23(3), 34–41.
- Wouters, P., Paas, F., & van Merriënboer, J. J. (2008a). How to optimize learning from animated models: A review of guidelines based on cognitive load. *Review of Educational Research*, 78(3), 645–675.
- Wouters, P., Paas, F., & Van Merriënboer, J. J. G. (2008b). Observational learning from animated models: Effects of modality and reflection on transfer. *Contemporary Educational Psychology*, 34, 1–8.
- 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.
- Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., et al. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research*, 82(1), 61–89.

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

Abdelghani Es-Sajjide works as a lecturer at the University of Roehampton Online. He is a Marie Curie Fellow and member of the MANETEI (EU FP7) group in which examined the management of emerging technologies. In addition to his research and teaching, Abdelghani has advised technology startups in the UK, the Netherlands and Saudi-Arabia in the roles of mentor and coach.

Fred Paas is Professor and chair of the research group Educational and Developmental Psychology at the Department of Psychology, Education and Child Studies of Erasmus University Rotterdam in the Netherlands. He is also a Professorial Fellow at the School of Education/Early Start of the University of Wollongong in Australia. He has been using the theoretical framework of cognitive load theory to investigate the instructional control of cognitive load in the training of complex cognitive tasks. His recent research interests include the role of the physical environment in cognition and learning, evolutionary approaches to cognitive load and learning, and the effects of physical activities (e.g., gestures and exercise) on cognition and learning.