Computers in Human Behavior 64 (2016) 21–33

Contents lists available at ScienceDirect

Computers in Human Behavior

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

The effectiveness of a math game: The impact of integrating conceptual clarification as support

Sylke Vandercruysse [a, b, *], Judith ter Vrugte [c], Ton de Jong [c], Pieter Wouters [d], Herre van Oostendorp [d], Lieven Verschaffel [a], Mariola Moeyaert [e], Jan Elen [a]

[a] CIP&T, Center for Instructional Psychology and Technology, KU Leuven, Dekenstraat 2, 3000 Leuven, Belgium
[b] Faculty of Psychology and Educational Sciences, Ku Leuven Kulak, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium
[c] Department of Instructional Technology, Faculty of Behavioral Sciences, University of Twente, 7500AE Enschede, Netherlands
[d] Department of Information and Computing Sciences, Utrecht University, 3508TB Utrecht, Netherlands
[e] Department of Educational Psychology and Methodology, State University of New York, 1400, Washington Avenue, 12222 Albany, NY, USA

ARTICLE INFO

Article history:
Received 12 November 2015
Received in revised form
31 May 2016
Accepted 6 June 2016
Available online 25 June 2016

Keywords:
Educational game
Instructional support
Conceptual clarification
Vocational education

ABSTRACT

This study investigates the impact of integrating conceptual clarifications as support in an educational math game, and explores the impact of adding this (internal vs. external) support on students' game and mathematical performance, intrinsic motivation, and game perception. Three conditions are established: a condition in which internal support is offered, a condition in which (identical) external support is offered, and a control condition in which no support is added to the game. One hundred twenty-two vocational secondary education students participated in this study. The results of this study indicate that students benefit from playing with an educational game in order to enhance their proportional reasoning skills. Adding conceptual clarifications as instructional support in an intrinsically integrated game is not recommended. If the support is given to the students anyhow, it is advised to offer it externally because internally integrating this support leads to a decrease in performance and motivation. Hence, not only support as such, but also the way it is integrated in the game-based learning process, might be decisive for its effectivity. Obviously, further research is warranted in order to replicate these findings also for other types of support, other game-based learning environments and other target groups.

1. Introduction

Despite the popularity and expectations of educational games, the spread and growth of digital games in educational contexts is still relatively in its infancy. Also, design issues concerning game-based learning environments (GBLEs) and the understanding of their impact on learning and instruction are still in early stages of advancement (de Freitas & Maharg, 2014; Koh, Kin, Wadhwa, & Lim, 2012). While time constraints, a lack of clear guidance, and game costs are some of the possible reasons for limited use (e.g., Charsky, 2011; Koh et al., 2012), a decisive barrier seems to be the lack of good, robust research on games and learning (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Girard, Ecalle, & Magnan, 2013; Van Eck, 2008). The evidence on the value of educational games is limited, inconsistent and inconclusive (de Freitas & Maharg, 2014; Perrotta, Featherstone, Aston, & Houghton, 2013). Claims stated in literature are often not empirically examined and consequently not confirmed or countered. One instance is the repeated claim in literature that instructional support features are a necessary part of GBLEs (de Freitas & Maharg, 2014). According to Ke (2009), without this additional support, the player would only learn to play the game instead of learning also the educational content embedded in the game. Unfortunately, the strong conviction of the need for instructional support in a game is not univocal empirically confirmed (Leemkuil & de Jong, 2011) and the results are inconsistent. Therefore, in this study the assumed benefits of integrating instructional support in games is investigated.

* Corresponding author. Ku Leuven Kulak, E. Sabbelaan 53 Mailbox 7654, 8500 Kortrijk, Belgium.
E-mail address: sylke.vandercruysse@kuleuven-kulak.be (S. Vandercruysse).

http://dx.doi.org/10.1016/j.chb.2016.06.004
0747-5632/

1.1. Instructional support in games

Instructional support, as defined by Tobias (2009) and Tobias, Fletcher, Dai, and Wind (2011), is any type of assistance, guidance or instruction to help students learn. Examples are scaffolds, explanations, directions, assignments, background information, monitoring tools, planning tools, etc. (Leemkuil & de Jong, 2011; Liu & Rojewski, 2013; Tobias et al., 2011). Adding instructional support is assumed to stimulate or facilitate students' game-based learning (GBL) processes. This assumption is confirmed by previous meta-analyses conducted by Ke (2009), Lee (1999) and Wouters and van Oostendorp (2013): simulation environments and games with instructional support can improve learning.

However, Wouters and van Oostendorp (2013) emphasize that adding instructional support to games is more complex since the effect is dependent on among others *the type of support* and *the cognitive activities* they target. The effect of the type of support becomes clear when looking at different studies. Mayer and Johnson (2010), for instance, provided evidence concerning the instructional effectiveness of reflection prompts in the form of feedback on conceptual learning (Mayer & Johnson, 2010). Another study however established that reflection promotes retention only in specific circumstances (i.e., non-interactive environments; Moreno & Mayer, 2005). Reflection prompts as support in yet another study were ascertained less promising as they did not affect students' mathematical performance and transfer (ter Vrugte et al., 2015a). Another type of support, i.e., supportive information and problem solving practice, was investigated by Darabi, Nelson, and Seel (2009). The results indicated a change in players' mental models after the supportive information, and only a little change in mental model after problem solving practice. Additional practice as support in combination with feedback was investigated by Liu and Rojewski (2013) and no effect of integrating these types of support in the game on participants' game enjoyment, academic achievement or motivation was found. Procedural information – which was intended to aid the reflection process – had no additional value either (ter Vrugte et al., 2015a). Also the integration of learning units, which provide explicit instruction of the mathematical thinking strategies used in the game, did not lead to better learning outcomes (Broza & Barzilai, 2011). Charsky and Ressler (2011) moreover found that using conceptual scaffolds (i.e., by using concept maps) decreased students' motivation to learn through gameplay. In addition to the type of support or the cognitive activities they target, *the format of support* also turned out to be decisive for its effectivity. When self-explanation as instructional support was added to a GBLE, students' transfer improved when they had to select an explanation from an onscreen list (Johnson & Mayer, 2010; Mayer & Johnson, 2010). However, when players had to type their explanations themselves into a text box, this effect disappeared. Finally, the *timing* for offering support in a GBLE also matters (Barzilai & Blau, 2014). Although no gain in problem solving was found for integrating formal knowledge representations in the form of learning units as support in a GBLE, learners with the support offered before the game performed better in the post-game assessment than students who were offered the same support afterwards or the students who were offered no support at all. Also Mayer, Mautone, and Prothero (2002) proved the effectiveness of presenting support to students before gameplay opportunity.

In short, the effectiveness of instructional support in games turns out to be unclear. These ambiguous findings might be a consequence of the diversity of the support (Leemkuil & de Jong, 2011; Tobias & Fletcher, 2012; Wouters & van Oostendorp, 2013). Developing a categorisation of different forms of instructional support could be helpful. Leemkuil and de Jong (2011) made such categorisation based on the aim of the support. They distinguish between support elements that (1) orient and focus attention on important elements or changes in the GBLE, (2) plan or choose actions and interventions, (3) evaluate the results of actions in the interventions, and (4) reflect or draw conclusions. Reiser (2004) made a distinction between software-based tools that support learners in complex and difficult tasks, which might also apply for GBL. This categorisation is based on the mechanisms the support activates: instructional support may (1) structure the task and reduce the complexity of it or (2) may problematize the subject matter. Finally, Wouters and van Oostendorp (2013) classified the instructional support based on the cognitive activities at which they target: (1) selecting relevant information by paying attention to the presented material and (2) mentally organizing the new information and the integration of this structure with prior knowledge. Using (one of) these three categorizations in studies may clear up the undetermined effectivity of instructional support in research on games since the diversity and complexity might get reduced. In this study, additional explanations or conceptual clarifications are added to a GBLE as support. This type of support can be categorised as (1) advisement support that focuses the attention of the players on important elements of the environment and the learning content that is integrated (Leemkuil & de Jong, 2011), (2) support that structures the task and reduces its complexity (Reiser, 2004) and (3) support that is mentally organizing the new information and link this with the prior knowledge (Wouters & van Oostendorp, 2013). In the 4C/ID model of van Merriënboer, this kind of support is labelled as supportive information and is typically called "the theory" (van Merriënboer, Clark, & de Croock, 2002). Additional to the supportive information, the 4C/ID model – a model focusing on the development of learning environments for complex cognitive skills – contains three other components: the learning tasks, procedural information, and part-task practice. Supportive information provides the bridge between the things students already know and what they need to know. Hence, prior knowledge gets activated, to which newly presented information gets connected (van Merriënboer et al., 2002). This support is optimally presented before learners start with their learning tasks. Doing so, prior knowledge can be activated and is available during task performance. It concerns information that is supportive to the learning and performance of non-recurrent aspects of the learning task. Ideally, the supportive information is connected to the different task classes and is evenly distributed across the whole learning activity (van Merriënboer et al., 2002). More specifically, supportive information reflects two types of knowledge (1) mental models that allow one to reason within the learning domain and (2) cognitive strategies that allow one to systematically approach problems in this domain (van Merriënboer & Kirschner, 2007; van Merriënboer et al., 2002).

Besides the diversity of the support, another possible explanation for the ambiguity in the results on the effect of support in GBLEs (e.g., the studies discussed above) might be that two forms of integration of instructional support can be distinguished (Honey & Hilton, 2011; Ke, 2009; Liu & Rojewski, 2013). In some studies the instructional support is internally integrated in the GBLE. This is the case when the support is integrated in the GBLE (e.g., Darabi et al., 2009; Johnson & Mayer, 2010; Liu & Rojewski, 2013; Mayer & Johnson, 2010; Moreno & Mayer, 2005; ter Vrugte et al., 2015a). In other studies external instructional support is investigated (e.g., Barzilai & Blau, 2014; Broza & Barzilai, 2011; Charsky & Ressler, 2011). Externally integrated support is support which is offered to the students apart from or in addition to the GBLE. There is no consensus about which type of support integration (i.e., internal or external) is most effective. Barzilai and Blau (2014) for instance concluded from their study that external support might help learners to form connections between game knowledge and formal school knowledge, and hence improve their knowledge. Offering external support, might however also focus students' attention too much on the difficulty of the learning content and make the gameplay less self-evident. Therefore, Charsky and Ressler (2011) and Liu and Rojewski (2013) propose to integrate this instructional support internally into the game so it becomes an ongoing part of the gameplay. This might enhance learning and motivation. However, an important consequence of internally integrating support in games – in turn – is that, depending on the format and type of the support, it might disrupt the game flow, and as in consequence the motivational nature of the game (Johnson & Mayer, 2010).

In sum, the literature is inconclusive about the effect of instructional support in GBL and whether this support should be internally and/or externally integrated into GBLEs. This study is further elaborating on this deadlock by investigating how to successfully design instructional support which facilitates GBL while maintaining enjoyment and flow. Comparing different ways of support integration might give insight on (1) the effectivity of the integration of support and (2) which of the two ways of integrating support (i.e., internal or external) yields the best results.

1.2. Learners' perception of the GBLE

In line with the mediational paradigm (Winne, 1987) and studies conducted with the same target group (i.e., vocational secondary education (VSE); e.g., Vandercruysse, van Weijnen, Vandewaetere, & Elen, 2015c), the importance of students' perception is acknowledged in this study. It is assumed that "learners are active actors in learning environments and not mere consumers of instructional designers' products" (Lowyck, Elen, & Clarebout, 2004, p. 429). The perception of instructional methods rather than the instructional methods themselves affect learning processes and products. Notwithstanding the intentions of designers and teachers, the ultimate effect of games seems to depend on the perception and use by students themselves (Lowyck et al., 2004; Winne, 1987). Students' perception, however, does not only relate to the instructional method and performance, but also to students' intrinsic motivation (Lowyck et al., 2004). When instruction is perceived as relevant or interesting, students will experience higher intrinsic motivation (Ryan & Deci, 2000). In this study, game perception is defined in line with the Game Perception Scale (GPS): (1) the expectation about the goals of the GBLE and, more specifically, whether players think of the GBLE as a leisure time or an instructional activity (perceived playfulness) and (2) the degree to which students expect that using the GBLE will stimulate their mathematical performance (perceived usefulness) (Vandercruysse et al., 2015b).

Studies investigating the impact of instructional support in games already endorsed the importance of students' perception. It appeared that students who are offered additional support had a higher perceived ease of the math skills learned in the game, although they enjoyed the learning less and had a more negative evaluation of the contribution of the game to learning than the students' playing with the GBLE alone (without additional support) (Broza & Barzilai, 2011). In another study, is was found that adding external support led to a decrease in learners' perception regarding how much they learned from the game but no reduced perception of flow or enjoyment was found (Barzilai & Blau, 2014). Hence, adding additional support to a game seems to influence the perception of students.

In sum, students' perception about the GBLE (with or without instructional support) is essential. Differences among the perception of participants may influence the effectiveness of the GBLE (Lowyck et al., 2004). How students perceive instructional interventions triggers their engagement in learning and influences the effect of the instructional interventions on their learning (Lowyck et al., 2004). Combining these findings with the abovementioned focus of the different possibilities to integrate instructional support in a GBLE, a difference in perception for the different ways of integrating support can be assumed.

1.3. Purposes of this study

In this clustered randomized design study, the focus is on the impact of integrating conceptual clarifications in a mathematical GBLE. Three conditions are constituted: a condition in which internal conceptual clarifications are integrated in the GBLE (i.e., ICC condition), a condition in which (identical) external conceptual clarifications integrated in the GBLE (i.e., ECC condition) and a control condition in which no additional support is offered. The current study explores the potential benefits to VSE students when a GBLE for mathematics is used, and more specifically, whether adding instructional support in a particular way(i.e., internal or external) produces different effects. .

Based on literature research, the following research questions are put forward:

1. Does playing a GBLE stimulates VSE students' proportional reasoning performance?
2. Does playing with an internally or externally support integrated GBLE differently influence students' performance than playing with a GBLE without additional support (Ke, 2009; Lee, 1999; Wouters & van Oostendorp, 2013)?
3. Does playing with an internally or externally support integrated GBLE differently influence students' intrinsic motivation than playing with a GBLE without additional support (Johnson & Mayer, 2010)?
4. Does playing with an internally or externally support integrated GBLE differently influence students' perceived playfulness than playing with a GBLE without additional support (Barzilai & Blau, 2014; Broza & Barzilai, 2011; Lowyck et al., 2004; Salomon, 1984)?
5. Does playing with an internally or externally support integrated GBLE differently influence students' perceived usefulness than playing with a GBLE without additional support (Barzilai & Blau, 2014; Broza & Barzilai, 2011; Lowyck et al., 2004; Salomon, 1984)?

Furthermore, this paper not only examines whether the intervention influences students' perception *after* using the GBLEs. In line with the mediational paradigm (Winne, 1987), it also investigates whether the effect of instructional support integration in a GBLE on students' performance, intrinsic motivation, and perception (as described in the abovementioned research questions) is influenced by the differences in students' *pre*-game perception.

6. Does students' perceived usefulness and playfulness before gameplay interact with the intervention for students' mathematical and game performance?
7. Does students' perceived usefulness and playfulness before gameplay interact with the intervention for students' intrinsic motivation?
8. Are the effects of the GBLE and the differences in instructional support integration influenced by students' computational skills and prior knowledge? This research question is the result of specific characteristics of the target group, i.e., VSE students. These students from less advanced levels of education show disengagement and demotivation which hinders growth in

numeracy and causes among other things passivity or limited investment and underachievement (Placklé et al., 2014). What is more, VSE students are characterized by a wide variety in cognitive abilities and potential (ter Vrugte et al., 2015b). Therefore it is investigated whether the expected learning effects of the instructional support integration are influenced by students' prior proportional reasoning knowledge and basic arithmetic skills. .

Fig. 1 offers an overview of the variables and the investigated relations between the variables.

2. Method

2.1. Participants

One hundred and twenty two students were recruited for this study. All participants were VSE students from the third and fourth year in Flanders (Belgium). Their age ranged between 15 and 19 year old ($M = 16.19$; $SD = 0.90$). Eleven classes from three schools were selected, resulting in five different specializations. Classes of students were randomly assigned to the conditions (see Table 1). An unbalanced gender division was found (i.e., 81 male students vs. 41 female students): there is an over-representation of boys in the ECC and control conditions. In the ICC condition the gender division is more balanced.

Fourteen students were excluded from the dataset because they did not attend for the complete intervention time. This resulted in 108 participants being used in the analyses.

2.2. Design

Three conditions (i.e., ICC, ECC and control condition) are used as between-subject variables in our pre-post cluster randomized subject design. The first two conditions contain conceptual clarifications as additional instructional support. This is done in order to facilitate players' sense making by helping them to make a link between their intuitive understanding of the game and more formal disciplinary representations of the learning content. The content of the instructional support is identical in the ICC and ECC conditions, but the way the support is implemented in the GBLE process differs. In the ICC condition, the conceptual clarifications are internally integrated in the GBLE and hence provided to the students in the GBLE while playing the game. The support elements are interwoven with the gameplay elements and presented to the students as an interactive tutorial. This type of support can be labelled as supportive information, also referred to as "the theory". Because the same body of knowledge underlies the ability to perform all proportional reasoning items (i.e., learning tasks) in the same subgame (i.e., task class), the supportive information is not coupled to individual items, but to subgames as a whole. The supportive information for each subsequent subgame is an addition to or an elaboration of the supportive information from previous subgames, since the composition of the subgames is based on different (non-recurrent) task-related and subject-related factors and hence reflects an increasing difficulty. Because the supportive information is relevant for all the items in the respective subgame, it is presented at the beginning of a new subgame. Furthermore, the information is kept available for the students during the whole subgame in form of a digital study book that can be activated.

Supportive information reflects two types of knowledge (van Merriënboer & Kirschner, 2007). The support offered in the interactive tutorial and study book concerns *the cognitive strategies* that allow the students to perform the tasks in the subgame and thus solve the problems (van Merriënboer et al., 2002; van Merriënboer & Kirschner, 2007). Therefore, the problem solving phases that an expert typically goes through while solving the problems as well as the rules-of-thumb that may be useful to successfully complete the different problem solving phases are presented to the learners. This was elaborated in collaboration with experts in the mathematical domain. Although the support especially focuses on these cognitive strategies, some information was also provided about *the mental models* that allow the students to reason within the task domain (i.e., proportional reasoning). Mental models are declarative representations of how the domain is organized and may contain general and abstract knowledge as well as specific cases that exemplify this knowledge (van Merriënboer & Kirschner, 2007; van Merriënboer et al., 2002). In the GBLE, for instance, the information provided to the students differs for the different problem types, presented in the different subgames which stimulates the students to pay attention to the differences between the types of proportional reasoning problems.

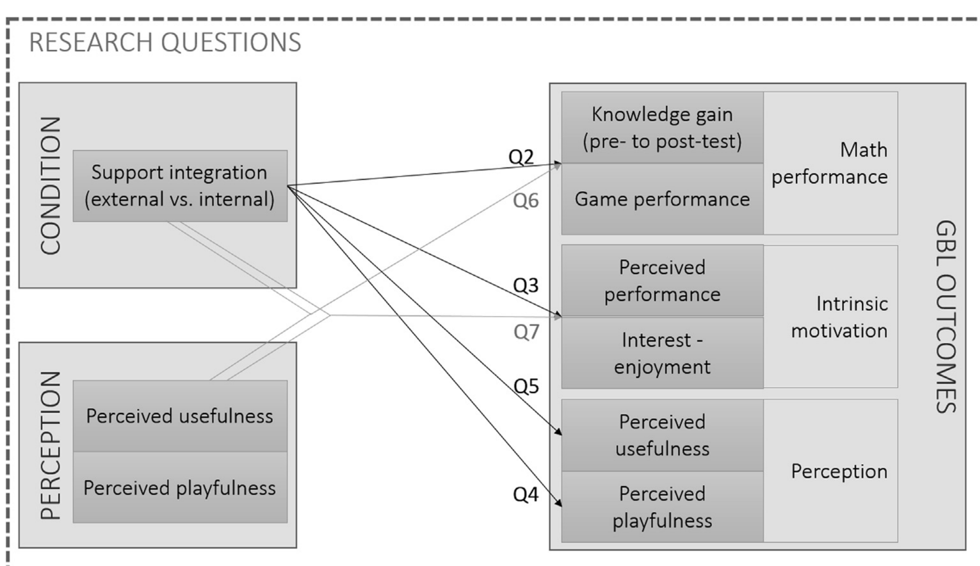

Fig. 1. Schematic overview of the research questions (indicated by "Q", followed by the research question mentioned in the section above.

Table 1
Conditions with number of participants (between brackets the number of students after drop-out).

	School	Specialization	Year	n_{Boys}	n_{Girls}	n_{Total}
Condition 1:	School 1	Electricity/Woodworking	4	18 (15)	1	
ICC	School 2	Hairdressing	3	0	14 (13)	
	School 3	Hairdressing	4	0	8 (7)	
						41 (36)
Condition 2:	School 1	Metalworking	4	8	0	
ECC	School 1	Construction industry	4	12 (11)	0	
	School 1	Construction industry	3	12 (9)	0	
	School 3	Hairdressing	4	1	8 (6)	
						41 (35)
Condition 3:	School 1	Metalworking	3	9 (8)	1	
Control	School 1	Metalworking	4	9 (8)	0	
	School 1	Woodworking	3	12 (11)	2	
	School 3	Hairdressing	4	0	7	
						40 (37)
n_{Total}				81 (71)	41 (37)	122 (108)

In the ECC condition, the amount of information that is provided, is identical to the conceptual clarifications provided in the ICC condition. However, the support is delivered in separate paper handouts external to the GBLE and thus independently from the flow-inducing gameplay. Students could easily consult these handouts during the different subgames. To align with the ICC condition, the information was also ordered based on the levels in the game and students were stimulated to consult the conceptual clarifications at the beginning of every new subgame because this is also the case in the ICC condition. Fig. 2 provides an example of a part of the conceptual clarifications in the ICC and ECC condition.

In the control condition no additional clarifications were given to the students.

2.3. Materials

2.3.1. GBLE: Zeldenrust

The self-developed GBLE “Zeldenrust” (Seldomrest) was used.

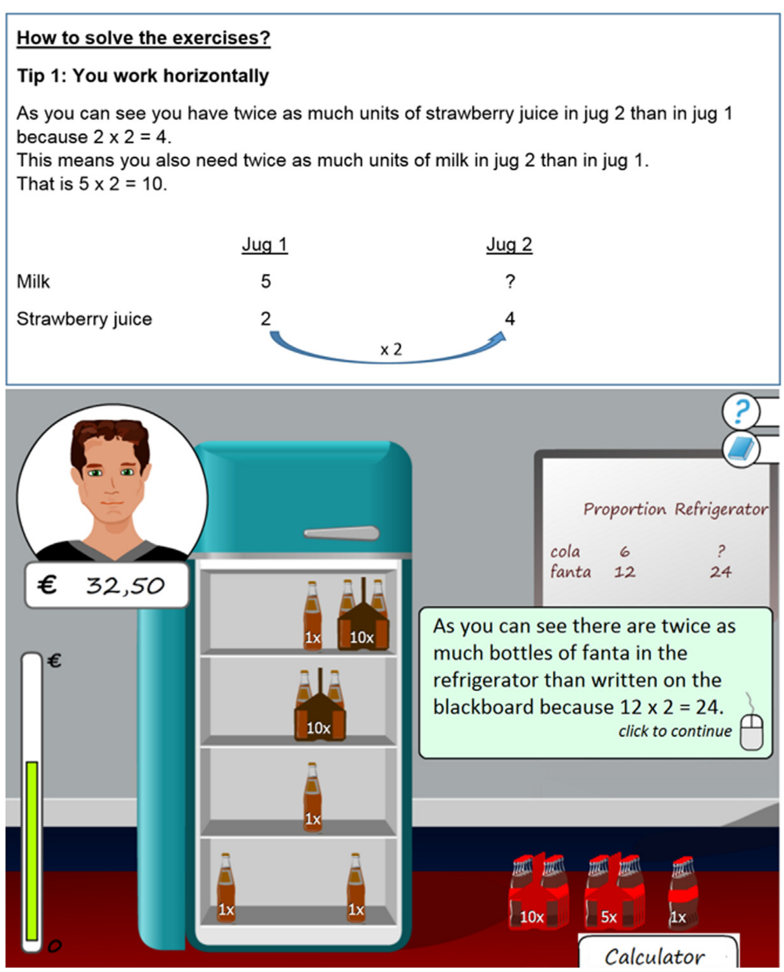

Fig. 2. Example of a translated part of the conceptual clarifications in the ICC condition and an extract from the handouts with conceptual clarifications in the ECC condition.

The GBLE is designed to be played individually. It offers a two-dimensional, cartoon-like environment with a play time of approximately 1.5–2 h (depending on the playing skills and mathematical ability of the players) and is meant for 14- to 16-year-old VSE students (for a comprehensive description see Vandercruysse, et al., 2015a). The original version of "Zeldenrust" was used in the ECC and control condition. In the ICC condition, changes were carried through and the conceptual clarifications were built in, in form of an interactive tutorial and study book (see procedure).

The game "Zeldenrust" starts with an introduction to the storyline of the game, which fits the social environment of the target group: he avatar, with which the player plays, wants to go on a holiday journey in the summer. To pay this journey, money needs to be earned. Therefore, players have to practice some jobs in a hotel (i.e., three in particular: fill the refrigerator, mix some cocktails and serve drinks to customers). Every job corresponds to a subgame (i.e., refrigerator game, blender game and serving game). Depending on how well these jobs are done, money for the journey can be earned.

The mathematical content in the game is proportional reasoning. Based on the literature (e.g., Harel & Behr, 1989; Vergnaud, 1983), three different types of proportional problems are integrated in the game. The missing value problems are problems in which one missing value in one of two ratios needs to be found (e.g., 5/10 = 15/?). The transformation problems contain two ratios, but one or two values needs to be adapted to obtain two equivalent ratios. In the example 7/18 and 21/50, 4 must be added to value 50 in the second ratio (i.e., 7/18 = 21/54) in order to obtain two equivalent ratios. For the comparison problems the relationship between two ratios needs to be determined, more specifically one ratio can be "equal", "smaller than" or "larger than" the other ratio (e.g., which ratio is the smallest: 3/9 or 10/35?). The three different problem types are presented to the players in a corresponding subgame. In the blender game, transformation items are included. The missing value problems are integrated in the refrigerator subgame and the comparison problems in the serving subgame. Depending on the problem type, a different number of attempts is built in. Due to the nature of the items of the comparison problems - they give either the correct answer or the wrong answer – only one attempt per item is possible. The missing value and transformation problems offer three attempts for each item. This is done to lessen frustration, to stimulate the players to rethink their calculations and to discourage guessing. All problem types have different difficulty levels based on several task-related and subject-related factors.

The game has four levels each containing the three subgames with a progressively difficulty across the levels. In each subgame, four proportional reasoning problem items are included, resulting in 48 proportional reasoning problem items in the entire game.

2.3.2. Measurements

The arithmetic tempo test *[TTR; Tempo Test Rekenen]* (De Vos, 1992) is offered to measure students' fluency in basic arithmetic skills. The test is an already validated test for Dutch students and standardized norms are available for fifth and sixth grade of VSE (see De Clerck et al., 2008) and is time restricted. The four basic mathematic operations (i.e., addition, subtraction, multiplication and division) are distributed over five sheets (i.e., one sheet for each arithmetic computation and one sheet with a mix of the arithmetic computations) with each page containing 40 problems with increasing difficulty. Students are stimulated to solve as many items as possible in the provided time (i.e., 1 min per sheet). The students' scores (i.e., the sum of all correct answers), reflect the level of automaticity and fluency with basic arithmetic computations. In the current study, the principle of automation was applied. This principle states that students have adequately mastered a fact or strategy when they are able to process the calculation within 3 s (van de Bosch, Jager, Langstraat, Versteeg, & de Vries, 2009). This means that, when applied to the TTR, students who score 100 or higher within 300 s are assumed to be computationally fluent.

Students' ability to solve proportional reasoning items is measured with two parallel versions of a self-developed *proportional reasoning test* (see ter Vrugte et al., 2015a; Vandercruysse et al., under revision). Both tests contain 16 items with comparable difficulty level. The items are tailored to the proportional reasoning items in the GBLE: four missing value items, four comparison items and four transformation items. Additionally, four transfer items which differ from the previous items in their use of context, are included in both tests. All the items are open-ended exercises with one possible correct answer. Because the pre- ($\alpha_{\text{pre}} = 0.68$) and post-test ($\alpha_{\text{post}} = 0.79$) are both parallel proportional reasoning tests, students' knowledge gain from pre- to post-test can be measured. Because the reliability of the pre-test – tested by calculating the internal consistency using Cronbach's alpha (Cronbach, 1951) – is unacceptably low, the Chronbach's alpha if items deleted are controlled. Hence, the comparison items are deleted from the test ($\alpha_{\text{pre}} = 0.74$). As for the pre-test, the four comparison items were removed from the post-test ($\alpha_{\text{post}} = 0.83$). The mean score without the comparison items is used in the analyses (i.e., number of correctly solved items divided by 12).

Students' *performance during gameplay* is analysed using two different indicators. The first indicator is the total game score which reflects an estimation of students' mathematical ability during gameplay in combination with their gaming skills. The second indicator, the proportion correct score, takes into account the number of items the students solved, the number of items solved correctly, and the number of attempts needed.

Students' *intrinsic motivation* before and after the playtime session was measured in order to calculate motivational gain scores. The same questions were used for pre- and post-measurement, albeit with adjusted tenses. The motivation questionnaire comprised two subscales of the Dutch version of the Intrinsic Motivation Inventory (IMI; McAuley, Duncan, & Tammen, 1989), using a 6-pint Likert scale. The first subscale is the interest/enjoyment subscale (7 items, e.g., "Playing this game was fun to do"). Reliability rates for this subscale were high ($\alpha_{\text{pre}} = 0.86$ and $\alpha_{\text{post}} = 0.90$). Second, the perceived competence subscale was used (6 items, "I was pretty skilled at playing this game"). Again, good reliability rates were found ($\alpha_{\text{pre}} = 0.76$ and $\alpha_{\text{post}} = 0.80$). Correlation between the two subscales was positive, large and statistically significant ($r_{pre} = 0.63$, $p_{pre} < 0.001$ and $r_{post} = 0.69$, $p_{post} < 0.001$).

Two subscales of the GPS were used to measure *students' perception of the GBLE* (Vandercruysse et al., 2015b). The perceived usefulness subscale measures the degree to which students believe that using a GBLE will enhance their performance (5 items, e.g., "I think playing this game could help me to learn proportional reasoning"). Good reliability rates were established ($\alpha_{\text{pre}} = 0.83$ and $\alpha_{\text{post}} = 0.90$). The perceived playfulness subscale measures whether players think of the game as something fun or something more akin to work (3 items, e.g., "I will characterize my experience with the game as "playing" rather than learning."). Again, reliability was acceptable ($\alpha_{\text{pre}} = 0.75$ and $\alpha_{\text{post}} = 0.85$) and the correlation between both subscales was positive and statistically significant ($r_{pre} = 0.25$, $p_{pre} = 0.009$ and $r_{post} = 0.26$, $p_{post} = 0.007$). The GPS was filled in before and after playtime, offering students the same questions, albeit with adjusted tenses.

2.4. Procedure

During a one-hour introduction and pre-test session, students were offered a short refresher lesson on proportional reasoning to activate their prior knowledge (Merrill, 2002). In line with the mathematical content integrated into the GBLE, general information on proportional reasoning was offered. During this explanation, as suggested by Vandercruysse et al., (under revision), students were also introduced to the GBLE with which they would play in the subsequent session. This introduction was followed by the questionnaires that measured students' pre-intrinsic motivation (IMI) and pre-perception of the GBLE (GPS). Also, the proportional reasoning test, which students had to individually complete, was administered together with the TTR that measured students' mastery of basic arithmetic skills.

After this pre-test session, students played the game for 2 h. Depending on the condition they were allocated to, they played with the GBLE with internally (i.e., ICC) or externally (i.e., ECC) integrated support, or with the GBLE without this support (i.e., control condition).

This playtime session was followed by the post-test session. Students again received the 16-item proportional reasoning test as well as the post-questionnaires that measured post-experimental motivation (IMI) and post-game perception (GPS). This final session lasted approximately 1 h.

3. Results

3.1. Initial differences between the conditions

A significance level of $\alpha = 0.05$ was set for all statistical significance tests. Table 2 lists the means per condition for all variables. No outliers were detected and standardized scores were used for all the analyses.

To identify possible initial differences between the three conditions, four (M)ANOVAs were conducted. In these four analyses, condition was the independent variable. The dependent variables were for the four separate analyses: (1) the score on the TTR, (2) the scores on the pre-test for the missing value items, transformation items and transfer items separately (MANOVA), (3) the subscales of the pre-motivation questionnaire (IMI; MANOVA) and (4) the score on the two pre-game perception subscales (GPS; MANOVA). The analyses revealed no significant differences between the three conditions for students' score on the TTR ($F(2, 105) = 1.90$, $p = 0.16$); the three different problem types on the pre-test (Wilk's $\lambda = 0.89$, $F(6, 206) = 2.00$, $p = 0.07$); the pre-intrinsic motivation (Wilk's $\lambda = 0.99$, $F(4, 208) = 0.40$, $p = 0.81$) and the pre-game perception (Wilk's $\lambda = 0.86$, $F(4, 208) = 0.98$, $p = 0.42$). Given these results, the conditions seem comparable with respect to prior proportional reasoning knowledge, basic arithmetic skills, pre-motivation, and pre-game perception.

As conditions were based on different stunt groups rather than individual students, multi-level analyses were conducted for all the dependent variables of the analyses to investigate potentially significant differences between the groups. Multi-level analyses were performed for these mixed model analyses (e.g., Seltman, 2015). The estimates of covariance parameters were inspected to decide whether the classes significantly differed from each other for the variables being tested. This gives a first indication that the variance between the classes is not determining for these variables, if the parameter of the intercept is not significant. The analyses revealed no significant differences. Furthermore, the variance between the classes was compared with the variance within the classes (i.e., estimated within-class residual variance). Although this remains an interpretation of the researchers, the variance between the classes was in most cases negligible small in comparison to the variance within the classes (e.g., for the pre-perceived playfulness $\beta_0 = 0.94$; $\sigma_e^2 = 47.05$). Hence, the differences between classes were not considered in the analyses.

3.2. The effect of integrating conceptual clarification in a GBLE on proportional reasoning skills and game performance

A repeated-measures ANOVA with time (proportional reasoning performance before to after gameplay) as within-subject factor and condition and problem types as between-subject factors was conducted. Results revealed a main effect of time on students' proportional reasoning test score (Wilk's $\lambda = 0.79$, $F(3, 103) = 9.13$, $p < 0.001$, $\eta^2 = 0.21$). Further analysis revealed that for all three problem types (i.e., missing value items ($F(1, 105) = 8.48$, $p = 0.004$, $\eta^2 = 0.08$), see Fig. 3; transformation items ($F(1, 105) = 21.80$, $p < 0.001$; $\eta^2 = 0.17$), see Fig. 4 and transfer items ($F(1, 105) = 6.73$,

Table 2
Means (and Standard Deviations) per condition for all variables.

	ICC condition ($n = 36$)	ECC condition ($n = 35$)	Control condition ($n = 37$)
Performance (pre)			
TTR - Basic arithmetic skills (max. 200)	106.94 (19.99)	111.80 (25.69)	117.17 (21.77)
Proportional reasoning pre-test (in %)	42.82 (26.96)	60.48 (21.42)	58.10 (19.26)
Motivation (pre)			
IMI - Interest/Enjoyment (max. 42)	24.14 (6.36)	25.37 (7.73)	23.47 (6.52)
IMI - Perceived Competence (max. 36)	20.58 (4.69)	21.69 (5.82)	20.67 (4.43)
Perception (pre)			
GPS - Pre-perceived usefulness (max. 30)	16.69 (4.71)	18.40 (6.19)	16.58 (4.80)
GPS - Pre-perceived playfulness (max. 18)	9.31 (3.56)	10.40 (3.62)	9.77 (2.91)
Game Performance			
Total Game score	877.92 (365.30)	1119.81 (399.41)	1068.76 (300.37)
Proportion Correct Score - Missing Value	0.55 (0.29)	0.64 (0.26)	0.63 (0.24)
Proportion Correct Score - Comparison	0.61 (0.15)	0.67 (0.15)	0.64 (0.15)
Proportion Correct Score - Transformation	0.41 (0.25)	0.55 (0.22)	0.50 (0.23)
Proportional reasoning post-test (in %)	56.94 (31.09)	76.67 (22.67)	66.67 (24.07)
Motivation (post)			
IMI - Interest/Enjoyment (max. 42)	22.22 (6.51)	26.29 (8.39)	26.44 (7.83)
IMI - Perceived Competence (max. 36)	20.67 (4.55)	21.29 (6.22)	22.61 (5.01)
Perception (post)			
GPS - Post-perceived usefulness (max. 30)	17.22 (5.35)	19.06 (6.48)	19.06 (5.61)
GPS - Post-perceived playfulness (max. 18)	9.22 (3.41)	10.14 (3.57)	10.53 (3.96)

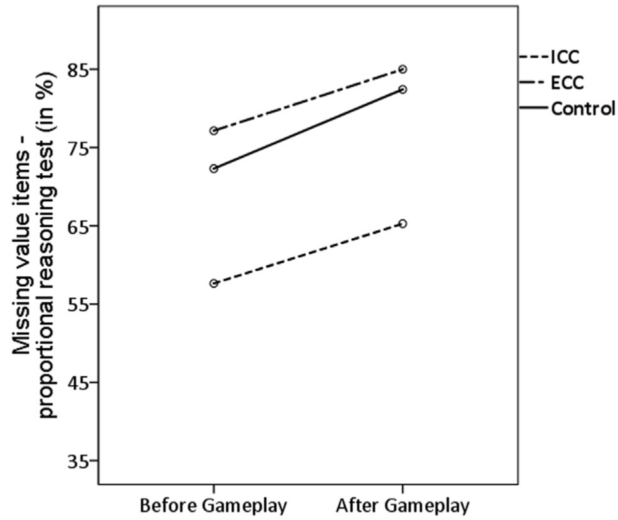

Fig. 3. Main effect of time and of condition on proportional reasoning gain score for missing value items (in %).

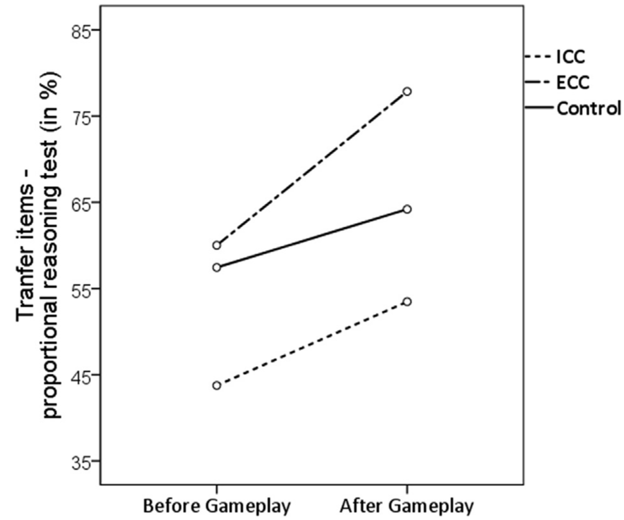

Fig. 5. Main effect of time and of condition on proportional reasoning gain score for transfer items (in %).

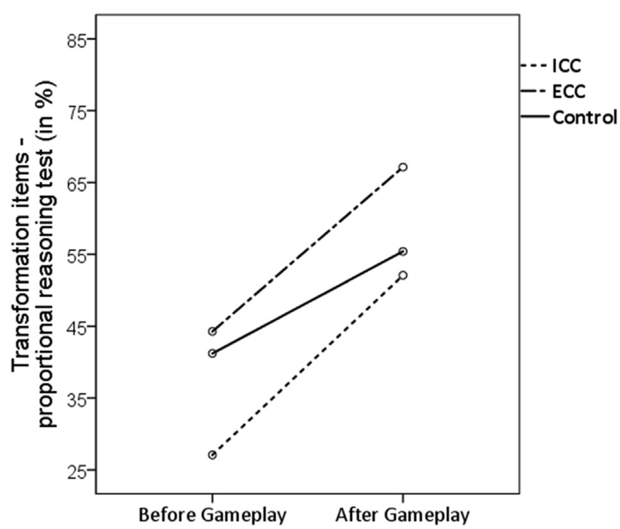

Fig. 4. Main effect of time on proportional reasoning gain score for transformation items (in %).

$p = 0.011$, $\eta^2 = 0.06$), see Fig. 5), students significantly improved from pre-to post-test after playing "Zeldenrust" (see research question 1).

This effect was not qualified by an interaction between time and condition (Wilk's $\lambda = 0.98$ $F(6, 206) = 0.38$, $p = 0.88$). However, a significant main effect of condition was found (Wilk's $\lambda = 0.85$, $F(6, 206) = 2.94$, $p = 0.009$, $\eta^2 = 0.08$). The ICC condition performed significantly lower than the two other conditions after gameplay. Further analysis show that this difference is significant for the missing value items ($F(2, 105) = 5.82$, $p = 0.004$, $\eta^2 = 0.10$) and for the transfer items ($F(2, 105) = 6.27$, $p = 0.003$, $\eta^2 = 0.11$), but not for the transformation items ($F(2, 105) = 2.95$, $p = 0.06$).

In order to analyse the effect of the integration of conceptual clarification on students' *game performance*, an ANOVA was conducted. The total game score was used as dependent variable and condition as factor. A significant effect of condition on students' total game score was found ($F(2, 105) = 4.73$, $p = 0.011$, $\eta^2 = 0.08$). Students in the ICC condition ($M = 877.92$; $SD = 365.30$) scored significantly lower than the students in the ECC ($M = 1119.81$, $SD = 399.41$) and control condition ($M = 1083.46$, $SD = 309.36$). An additional MANOVA was conducted with the proportion correct scores for each problem type as dependent variables and condition as factor. No significant effect was found (Wilk's $\lambda = 0.92$, $F(6, 194) = 1.31$, $p = 0.25$).

In sum, the results demonstrate an effect of adding conceptual clarifications to a GBLE on students' mathematical performance on the proportional reasoning test, but not on their game performance, analysed with the proportion correct score (see research question 2). These findings are discussed at length in the below discussion part.

3.3. The effect of integrating conceptual clarifications in a GBLE on students' motivation

A MANOVA was conducted with condition as factor and the gain scores of the two subscales of the IMI – which correlated significantly positive with each other ($r = 0.57$, $p < 0.001$) – as dependent variables. Results showed a main effect of condition (Wilk's $\lambda = 0.85$, $F(4, 194) = 4.27$, $p = 0.002$, $\eta^2 = 0.08$). Further analyses showed a significant difference between the conditions for the score on the perceived interest/enjoyment subscale ($F(2, 98) = 7.32$, $p = 0.001$, $\eta^2 = 0.13$), but not for perceived competence ($F(2,98) = 2.10$, $p = 0.13$). Post-hoc Bonferroni comparisons showed that the ICC condition had less enjoyment and interest in the game ($M = -1.92$, $SD = 5.86$), while the students in the ECC ($M = 0.91$, $SD = 4.80$) and control condition ($M = 2.97$, $SD = 6.49$) showed an increase in enjoyment and interest in the game (see research question 3). As can be derived from Fig. 6, the control condition showed the highest increase.

3.4. The effect of integrating conceptual clarifications in a GBLE on students' game perception

To investigate whether conceptual clarifications had an effect on students' post-game perception of the GBLE, a MANOVA was conducted. Condition was taken as factor and the two GPS subscales as dependent variables. No main effect of condition was found on students' post-game perception (Wilk's $\lambda = 0.95$, $F(4, 196) = 1.38$, $p = 0.24$; see research questions 4 and 5).

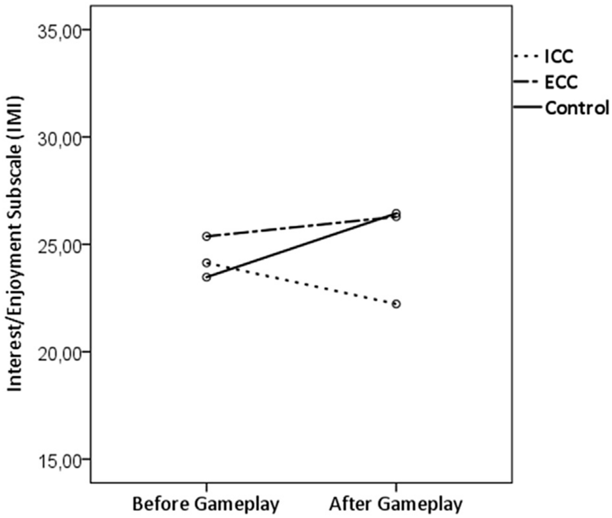

Fig. 6. Main effect of time and condition on students' interest/enjoyment gain score (subscale IMI).

3.5. The interaction effect of condition and perception on students' performance and motivation

Because the expectation was that the effect of support integration might be influenced by students' pre-game perception (i.e., perceived usefulness and playfulness of the GBLE), the interaction effects of condition and students' pre-perceived usefulness and playfulness were also analysed.

The MANOVA with learning gain scores as dependent variable and condition as factor showed no significant interaction effect between condition and perceived usefulness (Wilk's $\lambda = 0.92$, $F(9, 236) = 0.92$, $p = 0.51$) or between condition and perceived playfulness (Wilk's $\lambda = 0.98$, $F(9, 236) = 0.21$, $p = 0.99$). Additionally, no significant interaction effect was found between condition and pre-perceived usefulness ($F(3, 99) = 0.89$, $p = 0.45$) or pre-perceived playfulness ($F(3, 99) = 0.34$, $p = 0.79$) for the total game score of students. Also for the efficiency in solving the proportional reasoning items in the GBLE (measured with the three proportion correct scores; MANOVA), no interaction between condition and perception was found (pre-perceived usefulness; Wilk's $\lambda = 0.96$, $F(9, 236) = 0.41$, $p = 0.93$ and pre-perceived playfulness; Wilk's $\lambda = 0.97$, $F(9, 236) = 0.37$, $p = 0.95$; see research question 6).

The interaction effect was also analysed for intrinsic motivation. Results show a significant interaction effect between condition and the pre-perceived usefulness for intrinsic motivation (Wilk's $\lambda = 0.81$, $F(6, 194) = 3.61$, $p = 0.002$, $\eta^2 = 0.10$; see Fig. 7). Further analyses show that this interaction effect is significant for the perceived competence subscale ($F(3, 98) = 5.58$, $p = 0.001$, $\eta^2 = 0.15$) and not for the interest/enjoyment subscale ($F(3, 98) = 1.38$, $p = 0.25$). Particularly, based on Bonferroni post-hoc comparisons, the effect is significant for the ICC condition. Students with a low pre-perceived usefulness of the game, show an increase in perceived competence after gameplay. Students with a high pre-perceived usefulness of the GBLE show a (high) decrease in perceived competence after gameplay. For pre-perceived playfulness, no significant interaction effect was found (Wilk's $\lambda = 0.88$, $F(6, 194) = 1.98$, $p = 0.07$). Hence, evidence was found for the moderating effect of students' perceived usefulness on their intrinsic motivation (see research question 7).

Table 3 offers an overview of the results.

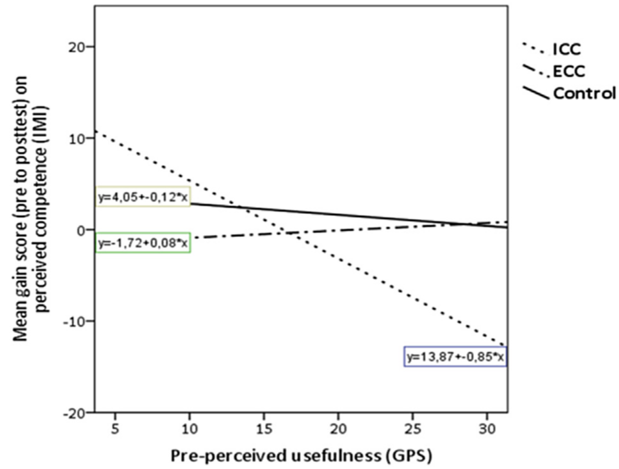

Fig. 7. Significant interaction effect of pre-perceived usefulness and condition for students' perceived competence (subscale IMI).

3.6. The effect of proportional reasoning prior knowledge and basic arithmetic skills

It was investigated whether students' prior mathematical ability affected students' ability to learn from the game and the impact of condition. Students' prior math ability was measured with the TTR and proportional reasoning pre-test. Students either mastered basic arithmetic skills (score above 100) or not (score below 100), and students' either mastered prior proportional reasoning skills (score above 50%) or not (score below 50%). This resulted in three groups of students: a group without or with low basic arithmetic skills, a group with basic arithmetic skills but below proportional reasoning skills, and a group with basic arithmetic and proportional reasoning skills. An ANOVA with time as within-subject factor and condition and students' prior mathematical ability as between-subject factors was conducted. The analysis revealed a significant main effect of time (Wilk's $\lambda = 0.69$, $F(1, 99) = 45.23$, $p < 0.001$, $\eta^2 = 0.31$) and a significant interaction between time and prior mathematical ability (Wilk's $\lambda = 0.79$, $F(2, 99) = 13.03$, $p < 0.001$, $\eta^2 = 0.21$). Fig. 8 shows that the students with less basic arithmetic skills performed better on the pre-test than the students with basic arithmetic skills but without or with low proportional reasoning skills. However, this last group of students caught up remarkably for the post-test, and although also the other students profited from playing the game, the group of students without or with low basic arithmetic skills improved less. These results indicate that the GBLE is a learning tool for all the students, nonetheless, students with basic arithmetic skills but without or with low proportional reasoning skills profit most from playing with the GBLE. Hence, students with different levels of mathematic ability are differently affected by playing the game. In addition it was investigated to which degree these prior mathematical abilities predict the post-test score. A regression analysis was conducted in which students' prior arithmetic abilities (score on the TTR) and students' prior proportional reasoning abilities (score on the pre-test) were used as predictors for students' post-test performance. Both predictors were entered simultaneously. The tests to see if the data met the assumption of collinearity indicated that multicollinearity was not a concern. The results of the regression indicated that students' prior mathematical ability was a significant predictor for students' proportional reasoning ability after gameplay. More specifically, it explains 32.00% of the variance in post-test performance, $R^2 = 0.103$, $F(2,105) = 6.00$, $p = 0.003$. The pre-test score

Table 3
Overview of the results of the study.

N°	Research question	Direction of effect
Q 1	Does playing the GBLE influence students' proportional reasoning performance?	All students improved for all 3 problem types
Q 2	Has support integration (i.e., CC) an influence on	ICC < ECC and control
	• Proportional reasoning?	Game score: ICC < ECC and control
	• Game performance?	
Q 3	Has support integration (i.e., CC) an influence on intrinsic motivation?	Enjoyment: ICC < ECC and control
Q 4	Has support integration (i.e., CC) an influence on perceived playfulness?	No differences between conditions
Q 5	Has support integration (i.e., CC) an influence on perceived usefulness?	No differences between conditions
Q 6	Does condition interact with perception for	No differences between conditions
	• Proportional reasoning?	
	• Game performance?	
Q 7	Does condition interact with perception for intrinsic motivation?	ICC: the higher the perceived usefulness, the lower the perceived competence.

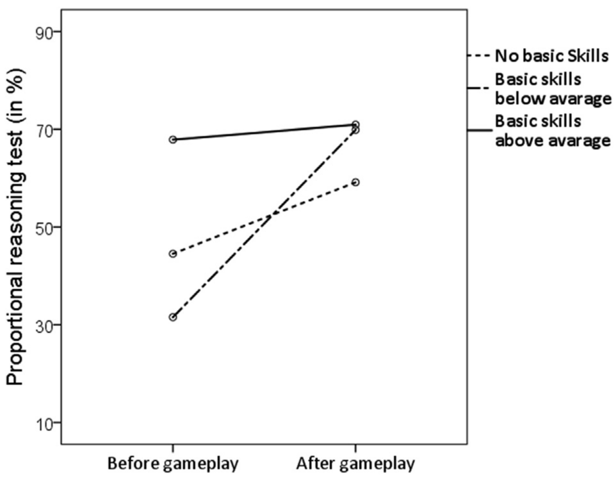

Fig. 8. Main effect of time and an interaction between time and prior mathematical ability on students' proportional reasoning score (in %).

significantly predicts the post-test score (β = 0.35, $p = 0.002$). The prior arithmetic ability however did not contribute to the prediction of the post-test performance ($p = 0.46$).

The interaction between condition and prior mathematical ability was not statistically significant ($F(4, 99) = 1.27$, $p = 0.29$). Finally, no significant interaction between time, condition, and prior mathematical ability was found (Wilk's $\lambda = 0.97$, $F(4, 99) = 0.77$, $p = 0.55$).

4. Discussion

In this study, a math GBLE was introduced to VSE students. The results show that, after using the GBLE, VSE students perform significantly better in solving proportional reasoning problems. These results are in line with previous studies in which the same GBLE was used for the same target group (e.g., ter Vrugte et al., 2015a; Vandercruysse et al., under revision). This emphasizes the learning effect of the GBLE, although the integrated learning content in the GBLE is rather complex and VSE students often experience difficulties with it (Vlaamse Overheid, 2009). What is more, the GBLE seems especially effective for students with basic arithmetic skills (i.e., who are computational fluent) but with limited proportional reasoning skills.

This study addressed the effectiveness of two ways of integration of conceptual clarifications (internal versus external) as instructional support for playing with the GBLE. The ICC condition showed lower performance during and after gameplay than the other two conditions. No difference between the ECC and control condition was found concerning performance. Although no assumptions were made about which way of integration would be most effective, the difference between both conditions needs some explanation. Also, the better results of the control condition compared to the ICC condition is surprising as it was assumed that the conditions with support would outperform. Different issues can be considered for these results. A first lesson learned from the current study is that the way the supportive information is provided may play a role. The supportive information in the ICC condition was integrated as an interactive tutorial. The support was an embedded and programmed set of information given to all the students at the start of every new subgame; irrespective whether the players needed this information or not. Hence, the students in the ICC condition had to simultaneously cope with two forms of competing demands: the educational game with the integrated support and the gameplay elements (Shaffer, 2004). This might have been too overwhelming and have resulted in information overload. Students in the ECC condition, in contrast, could process the handouts at their own pace and could choose themselves when to consult the support. Moreover, the addition of support in the GBLE before they got the chance to play and discover the environment by themselves might have directed the students' attention too much to the difficult learning content. As a result, gameplay became less autonomous for the ICC condition (Charsky & Ressler, 2011), which may explain their lower performance. Another issue is the GBLE "Zeldenrust" which is an intrinsically integrated GBLE. A previous study (Vandercruysse et al., under revision) showed that this GBLE forms a challenge for this target group, because intertwining gameplay and learning content is highly complex for them. Although an attempt was made to reduce this complexity by integrating additional support, the results indicate that internally adding information in the already complex GBLE is detrimental for VSE students' performance and motivation. What is more, integrating the support internally is more unfavourable for students' performance than offering no support at all. Hence, when integrating conceptual clarifications in a GBLE, the GBLE itselves needs to be considered as well. A final issue is the impact of the behaviour of the learners. It is possible that the students in the ECC condition did not read or used the support. Although they were stimulated to do so during the intervention, they could not be forced to read the information. Previous studies already showed that instructional support remains frequently unused (Nelson, 2007) and help offered in computer environments (not necessarily GBLEs) is seldom used by learners (Aleven, Stahl, Schworm, Fischer, & Wallace, 2003). The support in the ICC condition was embedded in the GBLE so students could not ignore it. For the ECC condition, students might not have used the supportive information at all and hence resemble the control condition. This may explain the lack of difference between the ECC and the control condition. Future studies could provide a control mechanism in order to be able to know if students read the external support or not and hence be able to decide whether this type of support is beneficial for this target group. If students in the ECC condition actually read the additional support and the same results are found, this would indicate that it is not the support as such, but the way it is offered to the students (i.e., internal or external) that evokes an effect. Up to now, no studies compared both ways of integration yet. This has led to an impasse without a clear view on the effect of instructional support and more specifically on the effect of the different ways the support can be integrated in GBLEs. This study slightly indicates that the way support is integrated matters, even though there is no certainty about the actual use of the external support. Consequently, future research could investigate whether the findings in this study concerning students' performance are reproducible and generalizable to other forms of support or to other GBLEs. Additionally, generalizability for other target groups can also be the topic of future research since it was demonstrated by ter Vrugte et al. (2015a) that VSE students deviate from other target groups concerning the effect of instructional support in GBLEs.

Concerning the results for the effect of condition on game performance an additional remark should be made. As was shown by the results, a significant difference for game performance was found only for the total game score. The proportion correct score, which is a more objective reflection of students' ability to solve proportional reasoning problems, revealed no significant difference. The students in the ICC condition, however, had to spend a considerable amount of time during gameplay to the conceptual clarifications. As it is not clear whether the ECC condition spent the same amount of time using this support, and the control condition had no additional information to process, students in the ECC and control conditions were able to spend more time to the gameplay and the learning content than the ICC condition. This might explain the higher gamescore. As the proportion correct score takes the number of items solved into account, this might be a reason why the differences between the conditions disappeared when using this score. Furthermore, the fact that support had an effect on the proportional reasoning performance but not on the game performance confirmed the findings of Leemkuil and de Jong (2011). They also concluded that the relationship between instructional support and game performance needs not necessarily be the same as the relationship between instructional support and test performance.

Playing with a GBLE in which conceptual clarifications are internally integrated, in addition to the negative effect on students' performance, also had negative effects on students' motivation. This decrease in motivation was not found for students in the ECC and control condition. On the contrary, both conditions showed an increase in enjoyment. The integrated support might have disrupted the game flow in the ICC condition (Johnson & Mayer, 2010). The ECC and control conditions had no interruptions and participants could play the game at their own pace. Further, the experienced difficulties as just discussed might have frustrated the students in the ICC condition. Finally, as the ECC and control condition performed better during gameplay (i.e., higher total game score), this might have been an extra stimulation and fun factor, while the students in the ICC condition performed less and experienced more difficulties with the proportional reasoning problems.

No significant difference between the three conditions concerning their perceived usefulness and perceived playfulness was established and no interaction between the conditions and students' pre-game perception was found for their performance. However, evidence was provided for the assumption that the perception of instructional methods (i.e., the pre-perceived usefulness of the GBLE) affects students' intrinsic motivation. For the ICC condition, students with a low pre-perceived usefulness of the game, showed an increase in perceived competence after gameplay. Surprisingly, students with a high pre-perceived usefulness of the GBLE showed a (high) decrease in perceived competence after gameplay. The complexity of the ICC condition might explain this. Students with high perceived usefulness, might been overwhelmed by the flood of information offered in the ICC version of the GBLE. Students with low expectations about the usefulness of the GBLE, might been positively surprised by the possibilities of the GBLE and their own experience with the GBLE. These findings are in contrast with the assumption that students who perceived the GBLE as a less useful tool to learn proportional reasoning were less motivated and invested less mental effort than students who perceived the GBLE as a more useful tool to learn (Salomon, 1984). In sum, the evidence for the importance of students' perception in GBL processes was less clear as expected.

This study contains some limitations. The first relates to the sample of participants which shows an over-representation of boys in the ECC and control conditions. In the ICC condition the gender division is more balanced. It is argued that girls have limited initial computer and game knowledge, possibly resulting in a greater difficulty in using a game (Vandercruysse, Vandewaetere, & Clarebout, 2012). The unbalanced division of gender over the conditions, and especially the under-representation of girls in the ECC and control conditions, might have played a role in the actual ease of use of the GBLE in this study. The allocation of classes instead of individuals to conditions is another limitation of the study, although this was a conscious choice and a pragmatic decision. The interventions were planned during regular lessons and students from different classes could not be regrouped to form a condition since the PGS lessons took not place on the same moments for the different classes. Furthermore, the interventions took four lessons for each class, which made a combination of individual students over classes in one condition practically impossible. Offering two or three conditions in one class (that participated at the same moment) was also impossible considering the study design (and the specificity of the three conditions, i.e., offering additional handouts or not). However, to examine if there was a significant difference between the classes for the dependent variables, additional multilevel analyses were conducted. None of the multilevel analyses showed a significant difference between the classes before they participated in the study. Therefore, the differences between classes were no longer considered or taken into account in the analyses. Another limitation was the lack of control of the use of support in the ECC condition, as discussed above. Future studies can further elaborate on this. A final limitation concerns the short-term intervention (i.e., four course hours for the complete intervention). This implies that the provided learning time was relatively short to attain deep learning since mean playtime was approximately 80 min. However, the game focused on a very specific and well-defined domain of mathematics (i.e., proportional reasoning) and the provided playtime seemed sufficient to be able to produce a learning effect. A delayed post-test or a long-term intervention, however, might provide more information about the quality of the acquired knowledge after the students' played with the (different versions of the) GBLE "Zeldenrust" (i.e., long-term effects). Hence, future research might investigate whether the findings of this study also apply for long-term intervention.

To conclude, the results of this study indicate that VSE students benefit from playing with a GBLE in order to enhance their proportional reasoning skills. Adding conceptual clarifications as instructional support in an intrinsically integrated GBLE is not recommended. If the support is given to the students anyhow, it is advised to offer it externally because internally integrating this support leads to a decrease in performance and motivation. The results in this study indicate that support as such, but also the way it is integrated in the GBL process, might be decisive for its effectivity. Obviously, further research is warranted in order to replicate these findings also for other types of support, other GBLEs and other target groups.

Funding

This study was funded by The Netherlands Organisation for Scientific Research/Programme Council for Fundamental Scientific Education Research (NWO/PROO – project number 411-00-003) and the Fund of Scientific Research (FWO – project number G.O.516.11.N.10).

Conflict of interest

The authors declare that they have no conflict of interest.

References

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research, 73*, 277–320. http://dx.doi.org/10.3102/00346543073003277.

Barzilai, S., & Blau, I. (2014). Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences. *Computers & Education, 70*, 65–79. http://dx.doi.org/10.1016/j.compedu.2013.08.003.

van de Bosch, E., Jager, J., Langstraat, H., Versteeg, B., & de Vries, M. (2009). *Remedierend Rekenprogramma Automatiseren de Zuidvallei [remediating mathprogramme automation de Zuidvallei]*. Ede: Giralis.

Broza, O., & Barzilai, S. (2011). When the mathematics of life meets school mathematics: Playing and learning on the "my money" website. In Y. Eshet-Alkalai, A. Caspi, S. Eden, N. Geri, & Y. Yair (Eds.), *Learning in the technological era: Proceedings of the sixth chais conference on instructional technologies research 2011* (pp. 92–100). Ra'anana, Israel: The Open University of Israel.

Charsky, D. (2011). Making a connection: Game genres, game characteristics, and teaching structures. In M. Khosrow-Pour (Ed.), *Gaming and simulations: Concepts, methodologies, tools and applications* (pp. 1823–1846). Hershey, PA: IGI Global. http://dx.doi.org/10.4018/978-1-60960-195-9.ch710.

Charsky, D., & Ressler, W. (2011). "Games are made for fun": Lessons on the effects of concept maps in the classroom use of computer games. *Computers and Education, 56*, 604–615. http://dx.doi.org/10.1016/j.compedu.2010.10.001.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*, 661–686. http://dx.doi.org/10.1016/j.compedu.2012.03.004.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. http://dx.doi.org/10.1007/BF02310555.

Darabi, A. A., Nelson, D. W., & Seel, N. M. (2009). Progression of mental models throughout the phases of a computer-based instructional simulation: Supportive information, practice, and performance. *Computers in Human Behavior, 25*, 723–730. http://dx.doi.org/10.1016/j.chb.2009.01.009.

De Clerck, D., Lahou, S., Marrannes, J., Milleville, M., Van Hul, K., & Vonckx, C. (2008). *Traject bij vermoeden dyscalculie. [Track when presumption of dyscalculia]*. Brussel: VCLB-service.

De Vos, T. (1992). *Tempo test rekenen handleiding en verantwoording [tempo test arithmetic manual and justification]*. Amsterdam: Pearson.

de Freitas, S., & Maharg, P. (2014). Series editors introduction. In N. Whitton (Ed.), *Digital games and learning: research and theory* (pp. xiii–xiv). New York: Routledge.

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: How effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning, 29*, 207–219. http://dx.doi.org/10.1111/j.1365-2729.2012.00489.x.

Harel, G., & Behr, M. (1989). Structure and hierarchy of missing-value proportion problems and their representations. *Journal of Mathematical Behavior, 8*, 77–119.

Honey, M. A., & Hilton, M. (2011). *Learning science through computer games and simulations*. Washington, DC: The National Academies Press.

Johnson, C. I., & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. *Computers in Human Behavior, 26*, 1246–1252. http://dx.doi.org/10.1016/j.chb.2010.03.025.

Ke, F. (2009). A qualitative meta-analysis of computer games as learning tools. In R. E. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education* (pp. 1–32). Hershey, PA: IGI Global. http://dx.doi.org/10.4018/978-1-59904-808-6.ch001.

Koh, E., Kin, Y. G., Wadhwa, B., & Lim, J. (2012). Teacher perceptions of games in Singapore schools. *Simulation & Gaming, 43*, 51–66. http://dx.doi.org/10.1177/1046878111401839.

Lee, J. (1999). Effectiveness of computer-based instructional simulation: A meta analysis. *International Journal of Instructional Media, 26*(1), 71–85.

Leemkuil, H., & de Jong, T. (2011). Instructional support in games. In S. Tobias, & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 353–369). Charlotte, NC: Information Age Publishing Inc.

Liu, Y., & Rojewski, J. W. (2013). Effects of instructional support in game-based learning: An analysis of educational games from design and application perspectives. In R. McBride, & M. Searson (Eds.), *Proceedings of society for information Technology & teacher education international conference 2013* (pp. 43–50). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

Lowyck, J., Elen, J., & Clarebout, G. (2004). Instructional conceptions: Analyses from an instructional design perspective. *International Journal of Educational Research, 41*, 429–444. http://dx.doi.org/10.1016/j.ijer.2005.08.010.

Mayer, R. E., & Johnson, C. I. (2010). Adding instructional features that promote learning in a game-like environment. *Journal of Educational Computing Research, 42*, 241–265. http://dx.doi.org/10.2190/EC.42.3.a.

Mayer, R. E., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology, 94*, 171–185. http://dx.doi.org/10.1037/0022-0663.94.1.171.

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport, 60*, 48–58. http://dx.doi.org/10.1080/02701367.1989.10607413.

van Merriënboer, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development, 50*, 39–64. http://dx.doi.org/10.1007/BF02504993.

van Merriënboer, J. J. G., & Kirschner, P. A. K. (2007). *Ten steps to complex learning: A systematic approach to four-component instructional design* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development, 50*(3), 43–59. http://dx.doi.org/10.1007/BF02505024.

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology, 97*, 117–128. http://dx.doi.org/10.1037/0022-0663.97.1.117.

Nelson, B. C. (2007). Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment. *Journal of Science Education and Technology, 16*, 83–97. http://dx.doi.org/10.1007/s10956-006-9039-x.

Vlaamse Overheid. (2009). *Peiling wiskunde in de eerste graad secundair onderwijs (B-stroom). [Mathematical survey in the first and second year of secondary education.]*. Brussel: Vlaams Ministerie van Onderwijs en Vorming.

Perrotta, C., Featherstone, G., Aston, H., & Houghton, E. (2013). *Game-based Learning: Latest evidence and future directions (NFER research Programme: Innovation in education)*. Slough: NFER.

Placklé, I., Könings, K. D., Jacquet, W., Struyven, K., Libotton, A., van Merriënboer, J. J. G., et al. (2014). Students' preferred characteristics of learning environments in vocational secondary education. *International Journal for Research in Vocational Education and Training (IJRVET), 1*, 107–124. http://dx.doi.org/10.13152/IJRVET.1.2.2.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences, 13*, 273–304. http://dx.doi.org/10.1207/s15327809jls1303_2.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54–67. http://dx.doi.org/10.1006/ceps.1999.1020.

Salomon, G. (1984). Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology, 76*, 647–658. http://dx.doi.org/10.1037/0022-0663.76.4.647.

Seltman, H. J. (2015). *Experimental design and analyses*. Retrieved from http://www.stat.cmu.edu/%7Ehseltman/309/Book/Book.pdf.

Shaffer, D. W. (2004). Pedagogical praxis: The professions as models for post-industrial education. *Teachers College Record, 106*, 1401–1421.

Tobias, S. (2009). An eclectic appraisal of the success or failure of constructivist instruction. In S. Tobias, & T. D. Duffy (Eds.), *Constructivist theory applied to education: Success or failure?* (pp. 335–350). New York: Routledge, Taylor and Francis.

Tobias, S., & Fletcher, J. D. (2012). Learning from computer games: A research review. In S. De Wannemacker, S. Vandercruysse, & G. Clarebout (Eds.), *Serious games: The challenge* (vol. CCIS 280, pp. 6–18). Berlin, Germany: Springer-Verlag. http://dx.doi.org/10.1007/978-3-642-33814-4_2.

Tobias, S., Fletcher, J. D., Dai, D. Y., & Wind, A. P. (2011). Review of research on computer games. In S. Tobias, & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 127–221). Charlotte, NC: Information Age Publishing Inc.

Van Eck, R. (2008). A guide to integrating commercial off-the-shelf (COTS) games into your classroom. In R. E. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education* (pp. 179–199). Hershey, PA: IGI Global. http://dx.doi.org/10.4018/978-1-59904-808-6.ch011.

Vandercruysse, S., ter Vrugte, J., de Jong, T., Wouters, P., van Oostendorp, H., Verschaffel, L., et al. (2015a). "Zeldenrust": A mathematical game-based learning environment for prevocational students. In J. Torbeyns, E. Lehtinen, & J. Elen (Eds.), *Describing and studying domain-specific serious games* (pp. 63–81). Cham, Switzerland: Springer International Publishing AG.. http://dx.doi.org/10.1007/978-3-319-20276-1_5

Vandercruysse, S., ter Vrugte, J., de Jong, T., Wouters, P., van Oostendorp, H., Verschaffel, L., et al. (2016). Content integration as a factor in math game effectiveness. *Educational Technology Research & Development* (under revision).
Vandercruysse, S., van Weijnen, S., Vandewaetere, M., & Elen, J. (2015c). Competitie als game element integreren in de BSO-klaspraktijk. [Integrating competition as game element in the vocational secondary classroom]. *Pedagogische Studiën, 92*, 179–201.
Vandercruysse, S., Vandewaetere, M., & Clarebout, G. (2012). Game-based learning: A review on the effectiveness of educational games. In M. Cruz-Cunha (Ed.), *Handbook of research on serious games as educational, business, and research tools* (pp. 628–647). Hershey, PA: IGI Global. http://dx.doi.org/10.4018/978-1-4666-0149-9.ch032.
Vandercruysse, S., Vandewaetere, M., Maertens, M., ter Vrugte, J., Wouters, P., de Jong, T., et al. (2015b). Development and validation of the game perception scale (GPS). *Journal of Educational Multimedia and Hypermedia, 24*, 43–74.
Vergnaud, G. (1983). Multiplicative structures. In R. Lesh, & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 127–174). Orlando, FL: Academic Press.
ter Vrugte, J., de Jong, T., Vandercruysse, S., Wouters, P., van Oostendorp, H., & Elen, J. (2015b). How competition and heterogeneous collaboration interact in prevocational game-based mathematics education. *Computers & Education, 89*, 42–52. http://dx.doi.org/10.1016/j.compedu.2015.08.010.
ter Vrugte, J., de Jong, T., Wouters, P., Vandercruysse, S., Elen, J., & van Oostendorp, H. (2015a). When a game supports prevocational math education but integrated reflection does not. *Journal of Computer Assisted Learning, 31*, 462–480. http://dx.doi.org/10.1111/jcal.12104.
Winne, P. H. (1987). Why process-product research cannot explain process-product finding and a proposed remedy: The cognitive mediational paradigm. *Teaching and Teacher Education, 3*, 333–356. http://dx.doi.org/10.1016/0742-051X(87)90025-4.
Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education, 60*, 412–425. http://dx.doi.org/10.1016/j.compedu.2012.07.018.

Sylke Vandercruysse, PhD, is a post-doctoral researcher at the Centre for Instructional Psychology and Technology (KU Leuven). Her research is on (the instructional design of) game-based learning and blended learning

Judith ter Vrugte is a doctoral student at the University of Twente in the Department of Instructional Technology (Faculty of Behavioral and Management Sciences). Her research focusses on serious games and the identification of variables that can foster game-based learning.

Ton de Jong is full professor of Instructional Technology at the University of Twente, Faculty of Behavioral Sciences where he acts as department head of the department Instructional Technology.

Pieter Wouters holds a PhD (2007) in Instructional Design (How to optimize cognitive load for learning from animated models). His current research focuses on cognitive and motivational processes in learning with serious games.

Herre van Oostendorp is associated professor Human Media Interaction, at Utrecht University, Department of Information and Computing Sciences (Interaction Technology group). His research activities are on the areas of cognitive modeling of web-navigation and knowledge acquisition, and cognitive learning with serious games.

Lieven Verschaffel is full professor at the KU Leuven. He works at the Centre for Instructional Psychology and Technology and his research interest pertain to teaching and learning of (elementary) mathematics.

Mariola Moeyaert, PhD, is Assistant professor Statistics and Research design at the Department of Educational Psychology and Methodology at the State University of New York. Her Major research interests and publications are in the field of multilevel analysis, meta-analysis and single-case experimental data.

Jan Elen is full professor at the KU Leuven. He works at the Centre for Instructional Psychology and Technology and his research interests pertain to instructional design in higher education and especially the use of instructional interventions by students.