Education Tech Research Dev (2017) 65:1345–1368
DOI 10.1007/s11423-017-9530-5

DEVELOPMENT ARTICLE

Content integration as a factor in math-game effectiveness

Sylke Vandercruysse[1] · Judith ter Vrugte[2] · Ton de Jong[2] · Pieter Wouters[3] · Herre van Oostendorp[3] · Lieven Verschaffel[1] · Jan Elen[1]

Published online: 29 June 2017

Abstract In this study we focus on the integration of mathematical learning content (i.e., proportional reasoning) in game-based learning environments (GBLE). More specifically, two kinds of GBLEs are set up: an extrinsically integrated GBLE and an intrinsically integrated GBLE. In the former environment, the mathematical content is not part of the core mechanics and structure of the gaming world. In the latter environment, the mathematical content is delivered through the parts of the game that are the most fun to play and embodied within the structure of the gaming world and the players' interactions with it. Sixty-four vocational track students participated in the study, all of them working in either version of the self-developed GBLE "Zeldenrust". The results of this study suggest that the way the content is integrated in a GBLE (i.e., intrinsically or extrinsically) matters: contrary to our expectations, students who played the extrinsically integrated game showed higher learning gains, motivational gains and perceived usefulness than students who played the game in which the content was intrinsically integrated.

Keywords Educational game · Math game · Content integration · Intrinsic and extrinsic integration · Game perceptions

✉ Sylke Vandercruysse
sylke.vandercruysse@kuleuven-kulak.be

1 CIP&T, Center for Instructional Psychology and Technology, KU Leuven, Dekenstraat 2, 3000 Louvain, Belgium

2 Department of Instructional Technology, Faculty of Behavioral Sciences, University of Twente, 7500 AE Enschede, The Netherlands

3 Department of Information and Computing Sciences, Utrecht University, 3508 TB Utrecht, The Netherlands

Introduction

Educational games have attracted scrutiny from educational technology researchers and practitioners. Educational games are argued to have an effective and motivational effect on students' learning processes (e.g., O'Neil et al. 2005; Vogel et al. 2006) since they stimulate their understanding and performance (Hayes and Games 2008; Liu and Chu 2010). Games can draw players in, capture their interest and engagement (Barab et al. 2010).This means that, according to Barab et al. (2010), educational games are technological and methodological means for creating curricula that are immersive, interactive and experientially consequential. Nonetheless, firm evidence to back up all the positive claims and expectations remains limited (Connolly et al. 2012; Girard et al. 2013; Hays 2005; O'Neil et al. 2005; Randel et al. 1992; Sitzmann 2011; Vandercruysse et al. 2012). A more systematic and thorough research approach is recommended in which specific characteristics of game-based learning environments (GBLEs)—rather than games as such—are studied (Aldrich 2005; e.g., Warren et al. 2009). Furthermore, there is a need for far more research on the interaction between GBLE characteristics and student-related variables.

The current study examines the assumed benefits of games by focusing on the effects of one specific game characteristic—i.e., content integration—on students' motivation, math performance, and perception. The study is conducted with a specific group of students, being vocational secondary education (VSE) students. This target group has some specific characteristics as it contains a significant number of at-risk students. A lot of these students experienced a history of poor prior learning outcomes, have been exposed to numerous unsuccessful instructional intervention and, hence, show resistance to traditional educational materials (ter Vrugte et al. 2015). As these students are characterized by high disengagement levels, which in turn hinder numeracy progress and cause passivity or limited investment of effort (Placklé et al. 2014), the use of computer games might be particularly attractive because of their motivational and engaging nature and offer an alternative instructional method to keep them interested, motivated and engaged (ter Vrugte et al. 2015). This study examines whether the way the learning content is integrated into a GBLE matters for these students, and if recommendations can be made in order to design a GBLE in line with the specificity of this target group.

Content integration

Malone (1980, 1981) and Malone and Lepper (1987) were the first to consider the problem of integrating learning content into educational games. They proposed the concepts of *intrinsic and extrinsic fantasy* and assumed that the educational effectiveness of games depended on the way in which learning content was integrated into the fantasy context of the game. Building on these hypotheses, Habgood et al. (2005), and Habgood and Ainsworth (2011) made a similar distinction and categorized these games as *intrinsically and extrinsically integrated games*. Within this categorization, they shift the emphasis for the distinction between GBLEs from fantasy to the core game mechanics of digital games that embody the rule-systems and player interactions. Following their definition, intrinsically integrated games:

> (1) deliver learning material through the parts of the game that are the most fun to play, riding on the back of the flow experience produced by the game, and not interrupting or diminishing its impact and; (2) embody the learning material within the structure of the gaming world and the players' interactions with it, providing an

external representation for the learning content that is explored through the core mechanics of the gameplay. (Habgood et al. 2005, p. 494)

Extrinsically integrated games separate learning and playing components. After completing one part of the learning content, students are provided with a reward by having the chance to advance in the game without dealing with learning content (e.g., playing a subgame).

The integration of learning content into parts of the gameplay ensures game flow experiences. Because the flow experience is maintained continually, intrinsically integrated games are argued to motivate and engage players more than extrinsically integrated games (e.g., Garris et al. 2002). Clark et al. (2011) as well as Habgood and Ainsworth (2011) moreover report that intrinsically integrated games engage players with the learning content of the game over a longer period of time. In addition to students' increased motivation, playing with an intrinsically integrated game might also improve learning outcomes. For instance, Habgood and Ainsworth (2011) reported higher scores on a delayed mathematical post-test in the intrinsically integrated condition than in the extrinsically integrated condition. In the study of Clark et al. (2011), the learning progress was not as high as hoped for, but the learning experience during their intrinsically integrated condition seemed to have been supported.

In this study, our research focus is on the effect of extrinsically versus intrinsically integrating learning content (i.e., proportional reasoning) in games on students' performance and motivation.

Students' perception: mediating the influence of instructions

A direct effect of the instructional method on learning processes and products was assumed in the process–product paradigm. The cognitive turn and the growing awareness of the importance of the constructive nature of learning, however, resulted in the mediational paradigm (Winne 1982, 1987). This paradigm stresses the importance of students' active cognitive processes: *interpretations of* instructional methods rather than the instructional methods themselves affect learning processes and products because "learners are active actors in learning environments and not mere consumers of instructional designers' products" (Lowyck et al. 2004, p. 429). Different interpretations result in different processes and products (Winne 1987). Often, students' interpretations do not align with the intentions of teachers or designers, resulting in suboptimal outcomes (Lowyck et al. 2004). Salomon (1984) has demonstrated that *students' perception* of learning materials impacts learning. Students who perceive the material to be leisure time activities invest less mental effort compared to students who see them as more instructional (Salomon 1984). This interpretation appears to also apply to educational games.

Given the mediational paradigm, the role of students' perception is explicitly acknowledged in this study. Students' game perception is defined as (1) their expectations around the GBLE goals and more specifically whether players perceive the GBLE as leisure time (something fun or relaxing, not asking any effort) or as an educational activity (something more akin to work, asking some effort of the students) (perceived playfulness), and (2) the degree to which students believe that using a GBLE will enhance their performance (perceived usefulness) (Vandercruysse et al. 2015c).

Students' perception, however, not only relates to the instructional method (i.e., the way the GBLE is introduced to the students) and performance (i.e., how well they perform in the GBLE), but also to their intrinsic motivation (Lowyck et al. 2004). Students' perception of instruction as relevant and interesting is assumed to go along with a high intrinsic motivation (Kinzie 1990; Ryan and Deci 2000). The study of Herndon (1987) also suggests that intrinsic motivation might be stimulated when students are provided with relevant and interesting instructions.

Taken together, knowledge of how students perceive a given instructional intervention—in this case a GBLE—seems essential because inter-individual differences in perception may affect the effectiveness of the intervention (Lowyck et al. 2004; Struyven et al. 2008). Furthermore, the way in which students perceive instructional interventions triggers learning engagement, and thus influences learning (Elen and Lowyck 2000; Entwistle 1991; Lowyck et al. 2004; Shuell and Farber 2001). Combining these findings with the abovementioned focus on the different possibilities to integrate mathematical content in a GBLE, a difference in perception might be expected for the two types of GBLEs developed for this study. Additionally, it can be hypothesized that the effect of content integration (extrinsic vs. intrinsic) on mathematical performance and students' motivation will be influenced by students' perception of the GBLE.

Purposes of this study

In this clustered randomized design study, the focus is on the integration of proportional reasoning as mathematical learning content. More specifically, two kinds of GBLEs are studied: an intrinsically and an extrinsically integrated GBLE. Based on the definition of Habgood et al. (2005), the mathematical content in the intrinsically integrated game is delivered through those parts of the game that are the most fun to play and embodied within the structure of the game and the players' interactions with it. In the extrinsically integrated environment, the mathematical content is not part of the core mechanics and structure of the game, but is only introduced at the beginning of every subgame as a separate mathematical exercise. After solving these questions, students are rewarded with the opportunity to play a subgame without being presented any learning content. The current study focuses on the potential benefit to VSE students when a GBLE for mathematics is used and, more specifically, whether integrating this mathematical content in a particular way (intrinsic vs. extrinsic) produces different effects.

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

1. Does playing a GBLE influence VSE students' proportional reasoning performance?
2. Does playing with an intrinsically integrated GBLE differently influence VSE students' performance than playing with an extrinsically integrated GBLE does (Clark et al. 2011; Habgood and Ainsworth 2011)?
3. Does playing with an intrinsically integrated GBLE differently influence VSE students' intrinsic motivation than playing with an extrinsically integrated GBLE does (Clark et al. 2011; Habgood and Ainsworth 2011)?
4. Does playing with the intrinsically integrated GBLE differently influence VSE students' perceived playfulness than playing with the extrinsically integrated GBLE does (Lowyck et al. 2004; Salomon 1984)?
5. Does playing with an intrinsically integrated GBLE differently influence VSE students' perceived usefulness of the GBLE than playing with an extrinsically integrated GBLE does?

Furthermore, this paper not only examines whether the intervention influences VSE students' perception *after* using the GBLEs. In line with the mediational paradigm (Winne 1987), it also investigates whether the effect of content 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 VSE students' perceived usefulness and playfulness before gameplay interact with the intervention for students' mathematical and game performance?
7. Does VSE students' perceived usefulness and playfulness before gameplay interact with the intervention for students' intrinsic motivation?

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

Methods

Participants

Participants in the study were 64 students from VSE schools (ages ranged between 14 and 17 years old; $M = 14.79$; $SD = .74$) in Flanders (Belgium). Six classes with three different specializations (woodworking/mechanics, hairdressing, and cookery/care) from two schools were selected. Student classes were randomly assigned to the conditions. The conditions showed an unbalanced gender division (see Table 1), which is further discussed in the limitation section. The participants form a homogeneous group in terms of educational background as they live in the same region and have similar educational backgrounds, access to computers and levels of ICT knowledge.

Only data from students that completed the entire intervention (four 50-min course hours) were included for analysis in the dataset. Accordingly, the data from six students were discarded, resulting in data from 58 participants on all measured variables.

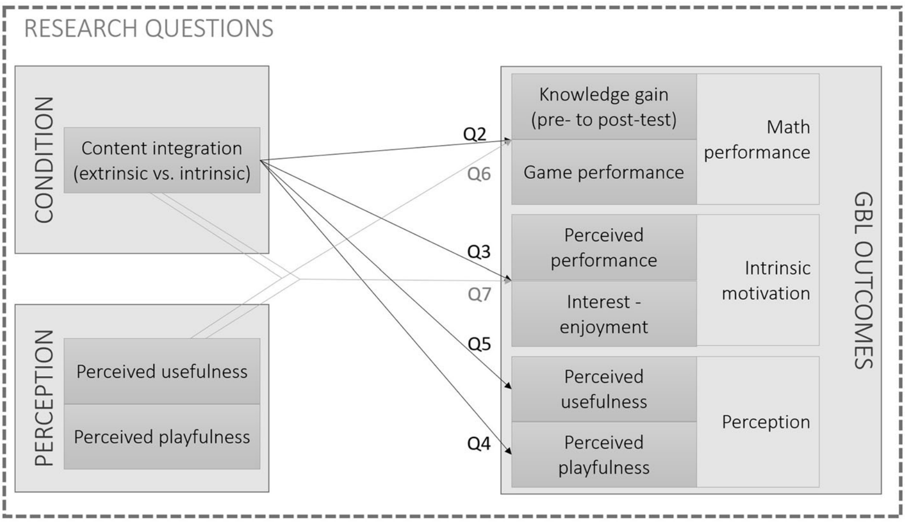

Fig. 1 Schematic overview of the research questions (indicated by "Q", followed by the research question mentioned in the section above

Table 1 Conditions with number of participants in the different classes (between brackets the number of students after drop-out and used in the analyses)

	School	Specialization	Grade	n_{Boys}	n_{Girls}	n_{Total}
Condition 1: Intrinsic integration	School 1	Woodworking/ mechanics	3	14 (13)	3 (2)	
	School 2	Hairdressing	3	0	7	
	School 2	Hairdressing	4	0	7 (6)	
						31 (28)
Condition 2: Extrinsic integration	School 1	Cookery/Care	3	1	12	
	School 2	Hairdressing	3	2	8 (7)	
	School 2	Hairdressing	4	0	10 (8)	
						33 (30)
n_{Total}				17 (16)	47 (42)	64 (58)

Grade 3 and 4 in a secondary school in Flanders equal grade 9 and 10 in an American High School

Design

In this quasi-experimental study, a pre-post clustered randomized subject design with two conditions (intrinsic integration vs. extrinsic integration) as between-subjects variables was adopted. In the intrinsic integration condition, students played with a GBLE in which the mathematical content (i.e., proportional reasoning, see below) was delivered through parts of the game (i.e., the subgames). In the extrinsic integration condition, students played with another version of the same GBLE in which the mathematical learning content was presented as a series of separate mathematical exercises. In this version, the mathematical content was not integrated into the structure of the game but presented to the students prior to each subgame in the form of traditional mathematical exercises. After students have answered these items, the game continues by entering a subgame but without mathematical items (as is the case in the intrinsically integrated version of the GBLE). This implies that the students playing with the extrinsically integrated version, advance in the game by solving some tasks in the subgames (e.g., fill the refrigerator as fast as possible), without the need to calculate the amount of bottles that need to be placed into it (i.e., the amount is given to the students). Only the time restriction is a prerequisite for solving the task appropriately.

Materials

GBLE: Zeldenrust

Two versions of the self-developed mathematical game called "Zeldenrust" (Seldomrest) were created for the purpose of this study. The GBLE was 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 of the game design and the elements integrated in the GBLE such as the goal, content, tools, feedback and scoring, see Vandercruysse et al. 2015b). The storyline fits the social environment of the target group: the players arrive in a hotel where they will work during the holidays to earn money for a summer trip. To earn money, they have to perform a number of jobs. Every job (subgame) contains a number of problems, and depending on how well the jobs are executed, money for the journey can be earned (lead game).

In the first version of the GBLE, *the intrinsically integrated version,* the mathematical content is interwoven with the game mechanics. Figure 2 shows an example of a challenge in one of the three subgames, i.e., the refrigerator subgame. In this subgame, the player has to fill the refrigerator with the correct number of cola bottles. In the example, the refrigerator contains 12 bottles of orange soda that cannot be removed. Outside the refrigerator are removable cola bottles and crates with a fixed number of cola bottles in them. The player has to drag (crates with) bottles of cola to the refrigerator so that the ratio between cola and orange soda is 16 to 4. In other words, the player has to drag 48 bottles of cola to the refrigerator. In the other two subgames also, (i.e., the serving subgame and the blender subgame), the player can only advance in the game after having simultaneously mastered both the game mechanics and the mathematical content. Both aspects cannot be separated from each other in the intrinsically integrated version of the GBLE. This means that the game-play is not interrupted by the mathematical learning content because it is completely interwoven with the game mechanics and storyline.

The (mathematical) learning content integrated into the GBLE is proportional reasoning. This topic was chosen because it is a well-defined mathematical domain with concrete applications. In addition, proportional reasoning is part of the Flemish vocational curriculum and VSE students often experience difficulties with proportional reasoning (Vlaamse overheid 2009). The GBLE can hence be considered both a practice environment and an instructional environment. The GBLE is a practical environment as it offers exercises that test content that students should master. At the same time, it is an instructional environment as it offers additional information on strategies to solve the different kinds of proportional reasoning problems. This instructional information, i.e., the learning content, is provided to the students in form of a handbook in the GBLE. This content-related tutorial is permanently accessible for the players during the game and gives players information about the different types of proportional reasoning problems and the strategies

Fig. 2 Example of refrigerator subgame task: how many bottles cola do you need?

they can use to solve these problems. This information is supportive to the learning of solving different proportional reasoning problems and provides a bridge between students' prior knowledge and the learning tasks. With this information, players should be able to handle the problems presented in the subgames. This tutorial is not automatically activated but can be activated by the players whenever they need it.

Three different types of proportional problems (e.g., Harel and Behr 1989; Kaput and West 1994; Vergnaud 1983) were included in the GBLE: (1) missing value problems, (2) transformation problems and (3) comparison problems. Each problem type corresponds to a subgame. Missing value problems (i.e., the refrigerator subgame) are problems in which a missing value in one of two ratios needs to be found. A schematic presentation of such a problem is for instance a/b = ?/d or a/b = c/? (e.g., ½ = 6/?). The second type are transformational problems (i.e., the blender subgame), in which two ratios are offered but the values need to be adapted to create two equivalent ratios. For instance, in the 4/16 and 14/64 example, 2 must be added to value 14 in the second ratio to make this ratio equivalent to the first one (4/16 = 16/64). For the third problem type, the comparison problems (i.e., serving subgame), 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: 4/8 or 12/20?). All three problem types can be subdivided into different difficulty levels. This division is based on several task-related and subject-related factors, which influence students' performance on proportional reasoning problems (Kaput and West 1994; Karplus et al. 1983; Tourniaire and Pulos 1985; Vergnaud 1983).

The game consists of four levels. At each game level, all three problem types (which progressively become more difficult with every level) are presented to the players in the corresponding subgame. Every subgame contains four proportional reasoning problems. So when players have finished the game, they have completed 48 proportional reasoning problems. Depending on the problem type, players either have one or three attempts to solve the items. For the missing value and transformation problems, three attempts for each item are allowed. Multiple attempts are allowed to lessen frustration, to stimulate the players to rethink their calculations and to discourage guessing. Due to the nature of the comparison problem items—students either give the correct or wrong answer—only one attempt is possible for every item. If the players cannot find the correct answer after having used up the number of allotted attempts, they automatically move on to the next task. The students always earn the same score after giving a correct answer, regardless of their number of attempts.

In the *extrinsically integrated version*, the learning content and game mechanics are separated from each other. The proportional reasoning problems are presented to players as classical mathematical problems before the subgames are activated. While solving these mathematical exercises, there is no link to the storyline of the game. This might disrupt the flow-inducing gameplay and the game-flow of the students since the exercises and the game itself are alternately offered to students. The number and type of exercises are, however, identical to the exercises provided in the intrinsically integrated version. When students solved the items offered to them, the subgames get activated. In advancing through the subgame, no mathematical operations have to be conducted. Figure 3 provides an example of an exercise in both versions. In the intrinsically integrated version (pictured at the top), the mathematical assignment is integrated into the storyline. Players have to fill the refrigerator according to the proportions they are given. The game mechanics and storyline are intertwined with the learning content. In the extrinsically integrated version (pictured below), the learning content is provided to the students the same way it is offered

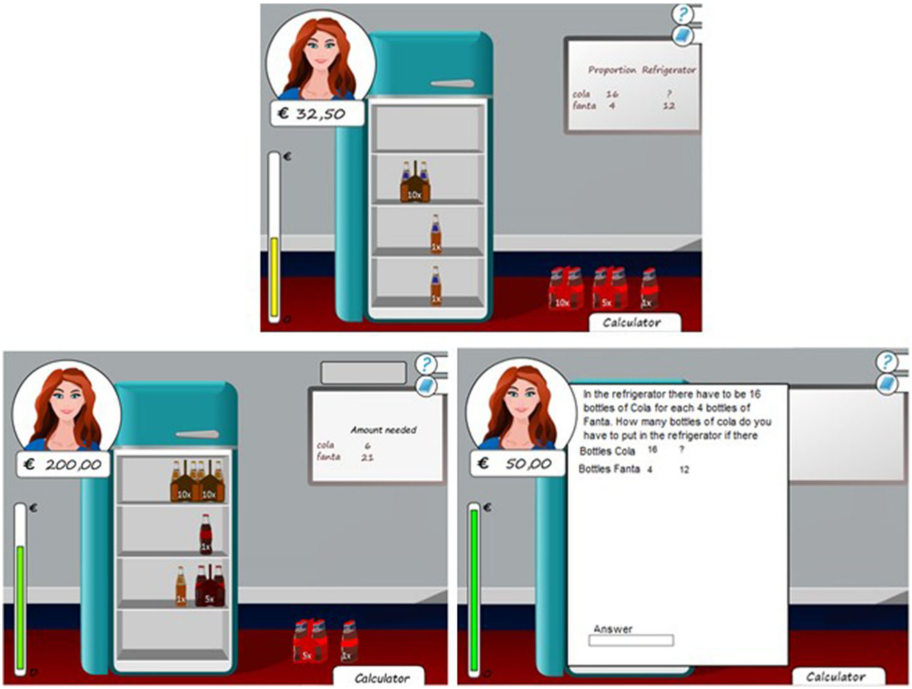

Fig. 3 Example of a missing value problem in the intrinsic (*top*) and extrinsic (*below*) integration versions of Zeldenrust

in pencil-and-paper tests before the subgame is activated (Fig. 3 below—right). No link with the game mechanics (i.e., the correct answer needs to be completed in the answer box at the bottom of the screen) is made. After solving the exercises, the players are asked to fill the refrigerator with a specific given amount of bottles within a given time span (Fig. 3 below—left). This implies the same drag-and-drop activity has to be done, but no mathematical calculations need to be executed.

Measurements

Basic arithmetic skills

Students' basic arithmetic skills were tested through administration of a timed arithmetic test (i.e., the TTR; Tempo Test Rekenen; De Vos 1992). The TTR, an already validated test for Dutch students (see De Clerck et al. 2008), was designed to measure students' fluency in basic arithmetic computations (addition, subtraction, multiplication and division). The four operations are divided over five pages: one page for each arithmetic computation and one page that offers a mix of the four operations. Every page contains 40 arithmetic problems of increasing difficulty. The test is time-restricted. For every page, one minute is provided, during which students have to solve as many items as possible. The scores students obtain reflect their level of automaticity and thus their mastery of basic arithmetic skills. In line with ter Vrugte et al. (2015), the sum of all the correct answers on the TTR is used (with a minimum of zero and a maximum of 200). The TTR has developed standardized norms for Flemish VSE students (5th and 6th year of secondary education). The principle of automation states that students have adequately mastered a fact or strategy when they are able to process the calculation within three seconds (van de Bosch et al. 2009). When applied to the TTR, this means that students who score 100 or higher within 300 s are assumed to be computationally fluent. This last principle is applied in the present study.

Proportional reasoning

Before and after the playtime session, students' mathematical performance at the level of proportional reasoning was measured using a *self-developed proportional reasoning pre- and post-test* (see ter Vrugte et al. 2015). This test was developed in collaboration with two experts in this topic to ensure the (content)validity of the test. The parallel test versions with comparable difficulty levels contain 16 questions with only one correct answer and no time limit. The questions in the tests show a considerable overlap with the items in the subgames. More specifically, this overlap was manifest in the order of increasing difficulty, the context and structure of the problems and proportional reasoning problem types (i.e., missing value problems (four items), transformational problems (four items) and comparison problems (four items)). This equal amount of items for each problem type further improves the (content)validity of the test. Additionally, the students were presented with four transfer questions, two missing value and two transformation items. These transfer items differ from the previous items in their use of context. Because the pre- and post-test are both parallel proportional reasoning tests, students' knowledge gain from pre- to post-test can be measured. At both intervals, the reliability—which was tested by calculating the internal consistency using Cronbach's alpha (Cronbach 1951)—appeared appropriate ($\alpha_{pre} = .74$ and $\alpha_{post} = .76$). For the analysis, the mean scores on these tests were used (i.e., number of correctly solved items divided by 16).

Game performance

Finally, students' performance during gameplay was also analyzed. For this performance, two different indicators were used. The first indicator is the total game score that represents an estimation of students' combined mathematical ability during gameplay and their gaming skills (e.g., deftness at dragging and dropping the necessary number of cola units in the refrigerator game). Because the total game score assesses students' math *and* gaming skills, a second, more objective indicator of students' proportional reasoning skills during or (in case of the extrinsically integrated condition) before gameplay was also used: the proportion correct score. This indicator reflects students' ability to solve the proportional reasoning problems, while also taking into account the number of attempts needed. More specifically, the percentage of items correctly solved (i.e., the number of items correctly solved divided by the number of items solved) is divided by the percentage attempts per item (i.e., the total number of attempts divided by the number of items solved). For example, if a student solved 14 missing value items, 10 of which were correctly answered while 30 attempts were needed, the student would have a proportion correct score of (10/14)/(30/14) = .33. For each problem type, this proportion correct score was calculated separately since the number of attempts differs for the different problem types (one vs. three).

Motivation

Students' intrinsic motivation was measured through the Dutch version of the intrinsic motivation inventory (IMI; McAuley et al. 1989; Plant and Ryan 1985). This questionnaire was presented to the students before and after the playtime session, which made it possible to measure students' motivational gains (by subtracting the motivation score measured after the playtime session with the motivation score previously measured). The same questions were used pre- and post-measurement, albeit with adjusted tenses. Two relevant IMI subscales were selected using a 6-point Likert scale: the interest/enjoyment subscale (seven items, e.g., "I enjoyed playing this game very much") and the perceived competence subscale (6 items, "I think I am pretty good at playing this game"). The correlation between the two subscales was positively significant ($r_{pre} = .30$, $p_{pre} = .02$ and $r_{post} = .67$, $p_{post} < .001$).

Game perception

Students' perception of the GBLE as defined in the introduction was measured through the game perception scale (GPS; Vandercruysse et al. 2015c). This questionnaire contains two subscales and uses a 6-point Likert scale. The first subscale measures the degree to which students believe that using a GBLE will enhance their performance (perceived usefulness; five items, e.g., "I think that playing this game is useful to learn fractions"). The second subscale measures students' expectations about the GBLE goals, and more specifically whether players view the game as something fun or something more akin to work (perceived playfulness; three items, e.g., "I was playing the game rather than working/learning."). The same questions were used pre- and post-measurement, but with adjusted tenses. The reliabilities for both subscales for both time periods were appropriate (perceived usefulness subscale: $\alpha_{pre} = .84$ and $\alpha_{post} = .92$ and the perceived playfulness subscale: $\alpha_{post} = .83$), that is, except for the pre-perceived playfulness. This reliability rate was unacceptable ($\alpha_{pre} = .53$). This subscale was consequently not used in the analyses. Again, students completed the questionnaire before and after their playtime session.

Procedure

The intervention was scheduled during regular 'project general subjects' (PGS) lessons and lasted four course hours for each class. The learning content in the GBLE is learning content that is offered to the students during these PGS courses, and hence the intervention was part of the regular school program of the students. Because the study was planned during and as part of their regular school schedule and curriculum, the teacher and principals of the secondary schools were asked for permission to participate in the study. The students were offered detailed information about the goals and set-up of the study and had the opportunity at the start of the intervention to refuse to participate. All students however, participated in the study.

During a one-hour introduction and pre-test session, students were offered a short refresher lesson on proportional reasoning to activate their prior knowledge (Merril 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. (2015a, d), 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 two hours. Students from the intrinsic integrated condition played with the intrinsically integrated version of the game, and the extrinsic integration condition with the extrinsically integrated version of the game. As the intervention took place during regular course hours, this playtime session took place at the classroom level, of which the classes were randomly assigned classes to the different versions of the game (see participants for this allocation of classrooms to conditions). Within this allotted time frame of 2 course hours, 72% of the students finished the game. 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 one hour.

Results

Initial differences

A significance level of $\alpha = .05$ was set for all statistical significance tests. No outliers were detected and the data were approximately normally distributed for all variables. The assumption of equal variances was tested with Levene's test for each analyses and the results show that we can feel confident that the assumption of equal variances is met. Table 2 lists the means (and standard deviations) per condition for all the variables. To facilitate interpretation of the analysis results, standardized scores of the variables are used.

To identify possible initial differences between both conditions, three ANOVAs were conducted with condition as the independent variable, and the score on the TTR, the score on the pre-test and the score on the perceived usefulness subscale (GPS) as separate dependent variables. These ANOVAs revealed no significant differences between both conditions with respect to students' score on the pre-test (F (1, 56) = 1.57, $p = .22$), on the TTR (F (1, 56) = 2.05, $p = .16$) and the pre-perceived usefulness (F (1, 56) = 2.53, $p = .12$). A MANOVA was conducted with the two subscales of the pre-motivation questionnaire (IMI) as dependent variables. For the MANOVA (pre-intrinsic motivation) also, no differences between both conditions were found (Wilk's $\lambda = .97$, F (2, 55) = 1.01, $p = .37$). It can consequently be concluded that both participant groups were comparable in terms of prior proportional reasoning problem-solving abilities, basic arithmetic skills, pre-game motivation and pre-game perception.

The subsequent analyses assessed prior knowledge and motivation by using gain scores for performance and motivation measurements, so that participants' pre-measurements were also considered in the analyses. Because it was assumed that students' pre-game perception would influence their motivation, performance and post-game perception, the interaction effects between the conditions and the pre-game perception were also examined.

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

Table 2 Means (and standard deviations) per condition for all variables

	Intrinsic integration ($n = 28$)	Extrinsic integration ($n = 30$)
Pre-performance		
TTR—basic arithmetic skills (max. 200)	113.71 (23.91)	105.50 (19.73)
Proportional reasoning pre-test (in %)	60.71 (25.68)	52.08 (26.69)
Pre-motivation		
IMI—interest/enjoyment (max. 42)	24.00 (6.91)	25.67 (7.02)
IMI—perceived competence (max. 36)	21.43 (4.76)	18.40 (4.80)
IMI—total score (max. 78)	45.43 (9.47)	43.97 (10.02)
Pre-perception		
GPS—pre-perceived usefulness (max. 30)	17.75 (4.61)	19.70 (4.71)
Game performance		
Total game score	1155.55 (429.38)	1926.20 (578.91)
Proportion correct score—missing value	.71 (.37)	.67 (.25)
Proportion correct score—comparison	.65 (.17)	.53 (.16)
Proportion correct score—transformation	.56 (.21)	.55 (.25)
Proportional reasoning post-test (total in %)	68.09 (17.95)	73.75 (19.93)
Post-motivation		
IMI—interest/enjoyment (max. 42)	24.64 (8.67)	30.40 (7.81)
IMI—perceived competence (max. 36)	21.64 (6.74)	23.20 (5.23)
IMI—total score (max. 78)	46.29 (14.00)	53.60 (12.05)
Post-perception		
GPS—post-perceived usefulness (max. 30)	18.79 (5.73)	22.80 (6.04)
GPS—post-perceived playfulness (max. 18)	8.50 (3.18)	9.00 (3.85)
GPS—post-total score (max. 48)	27.29 (7.17)	31.80 (5.65)

covariance parameters were scrutinized to determine whether the classes significantly differed from each other on the variables being tested. When the parameter of the intercept was not significant, this offered a first indication that the variance between the groups was not determining for this variable. However, this variance between the groups was also compared to the variance within the groups (residual parameter). In most cases, the variance between the classes was negligibly small in comparison to the variance within the classes (e.g., for the pre-IMI enjoyment subscale intercept estimate = .80, residual estimate = 99.43), although this might be subject to interpretation. The analyses revealed no significant differences. The differences between the groups were consequently not considered in the analyses.

Effect of content integration on students' proportional reasoning and game performance

A repeated-measures ANOVA with time (*proportional reasoning test* before to after gameplay in order to show the evolution of the scores of the students) as within-subject factor and content integration (intrinsic vs. extrinsic) as between-subject factor revealed a main effect of time on students' proportional reasoning test score (*Wilk's* $\lambda = .66$, F (1,

56) = 28.39, $p < .001$; $\eta^2 = .34$; see Fig. 4). The mean score on the proportional reasoning test significantly differed between the pre- and post-gameplay session (see research question 1). Students performed significantly better on the post-test than on the pre-test.

This effect was qualified by an interaction between time and content integration (*Wilk's* $\lambda = .92$, F (1, 56) = 5.14, $p = .03$, $\eta^2 = .08$; see Fig. 4). Although all students improved from pre- to post-test, students in the extrinsic integration condition (from pre-test 52.08% to posttest 73.75%) improved significantly more than students in the intrinsic integration condition (from pre-test 60.71% to post-test 68.09%).

The effect of content integration on students' *game performance* was also analyzed. First, an ANOVA was conducted (since there is only one dependent variable measured) with the total game score as the dependent variable and condition as factor. A significant effect for condition on students' total game score was found (F (1, 54) = 35.25, $p < .001$, $\eta^2 = .40$). Students in the extrinsically integrated condition ($M = 1926.20$; $SD = 578.91$) scored significantly higher than students in the intrinsically integrated condition ($M = 1155.55$; $SD = 429.38$). Because the total game score tallies the gaming and puzzle-solving skills of the students, an additional analysis was performed with the proportion correct scores for each subgame/problem type (see measurements). The proportion correct scores (see method section) are a more objective indicator of students' proportional reasoning ability during gameplay. A MANOVA with the three proportion correct scores (which significantly correlated with each other) for each problem type as dependent variables and condition as factor revealed a significant effect (*Wilk's* $\lambda = .82$, F (3, 52) = 3.89, $p = .01$, $\eta^2 = .18$). Univariate testing showed the effect to be significant for the comparison items (F (1, 54) = 11.75, $p < .001$, $\eta^2 = .18$). Post-hoc testing indicated that the intrinsically integrated condition ($M = .65$; $SD = .17$) outperformed the extrinsically integrated condition ($M = .53$; $SD = .16$). For the missing value and transformation problems, no significant differences between both conditions were found for efficiency scores.

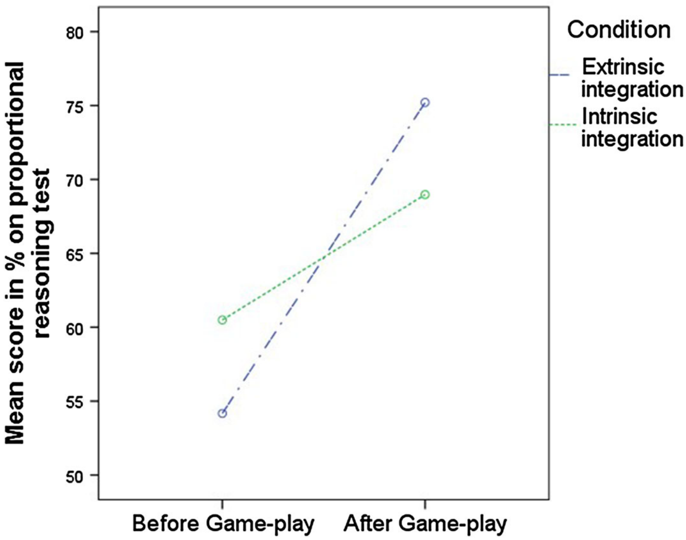

Fig. 4 Main effect of time (before to after gameplay) and interaction effect of time and condition on students' proportional reasoning gain scores (in %)

In sum, the results demonstrate an effect of content integration on students' math performance (see research question 2). During gameplay, the intrinsically integrated condition performed better on the comparison items than the extrinsically integrated condition. However, for the gain score on the proportional reasoning tests and the total game score, the effect was the opposite: the extrinsic condition outperformed the students in the intrinsically integrated condition. These findings seemingly contradict each other. In the below discussion part, these somewhat surprising results are discussed at length.

Effect of content integration on students' motivation

To investigate the main effect of content integration in the game on students' intrinsic motivation, a MANOVA was conducted. Condition was used as factor and the gain scores (in order to take the evolution of the scores into account) of the two subscales of the IMI (post-IMI subscale score minus pre-IMI subscale score for the perceived interest/enjoyment and perceived competence subscales)—which significantly correlated with each other ($r = .71, p < .001$)—as dependent variables. The results showed that the integration of the content in a GBLE (condition) had a significant main effect on students' intrinsic motivational gain scores (Wilk's $\lambda = .85$, F (2, 53) = 4.78, $p = .012$, $\eta^2 = .15$). Further analyses demonstrated a significant difference between the conditions for the score on the perceived interest/enjoyment subscale of the IMI (F (1, 54) = 7.22, $p = .007$, $\eta^2 = .13$), as well as for the perceived competence subscale of the IMI (F (1,54) = 8.39, $p = .005$, $\eta^2 = .14$). The extrinsic integration condition ($M = 4.83$; $SD = 5.48$) showed a significantly higher increase in interest and enjoyment than the intrinsically integrated condition ($M = .64$; $SD = 6.23$). For perceived competence also, the extrinsic condition ($M = 4.80$; $SD = 5.89$) increased more than the intrinsic integration condition did ($M = .21$; $SD = 5.36$). These results show that content integration influences students' intrinsic motivation (see research question 3) and, more concretely, that students who play in the extrinsic conditions are more intrinsically motivated than the other group.

Effect of content integration on students' game perception

The effect of the content integration (condition) on students' post-game perception of the GBLE was investigated through two ANOVAs (because of a non-significant correlation between the two subscales) that used condition as factor and the two GPS subscales (perceived playfulness and perceived usefulness; post-measurements; $r = .11$, $p = .403$) as dependent variables. The first ANOVA with post-perceived playfulness as the dependent variable showed no significant main effect of condition (F (1, 54) = .30; $p = .59$; $\eta^2 = .006$). Students' perceived playfulness was not significantly different after having playing with either an intrinsically or extrinsically integrated GBLE (see research question 4). The second ANOVA with post-perceived usefulness as the dependent variable, showed a small significant main effect of condition (F (1, 54) = 3.96; $p = .05$; $\eta^2 = .07$). After having played with an extrinsically integrated game, students saw the GBLE as more useful than students who had played with a GBLE in which the content was intrinsically integrated (see research question 5).

Interaction effect of condition and perception on students' (mathematical and game) performance and motivation

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

The ANOVA with learning gain scores (proportional reasoning pre-test to post-test in order to take the evolution of the scores into account) as the dependent variable and condition as factor showed no significant interaction effect between condition and perceived usefulness ($F(2, 54) = .41$, $p = .67$; $\eta^2 = .02$). The moderating effect of students' pre-game perception on proportional reasoning performance could not be shown. Additionally, students' game performance was investigated for the interaction effect between condition and pre-perceived usefulness. No significant interaction effect was found between condition and students' pre-perceived usefulness for their total game scores ($F(2, 54) = 1.20$, $p = .31$; $\eta^2 = .04$). For students' efficiency in solving the proportional reasoning items in the GBLE (measured with the three proportion correct scores that significantly correlate with each other; MANOVA) also, no interaction between condition and perception was found (Wilk's $\lambda = .82$, $F(6, 104) = 1.83$, $p = .10$; $\eta^2 = .10$; see research question 6).

Moreover, no interaction effect between integration of the content into the game (condition) and pre-perceived usefulness for students' *intrinsic motivation* (measured with the two subscale scores of the IMI that significantly correlate with each other; MANOVA) was found (Wilk's $\lambda = .93$, $F(4, 106) = .97$, $p = .43$; $\eta^2 = .04$; see research question 7).

Table 3 offers an overview of the results.

Table 3 Overview of the results of the study

Number	Research question	Results
Q 1	Does playing the GBLE influence students' proportional reasoning performance?	All students improved from pre- to post-test
Q 2	Has content integration an influence on	
	Proportional reasoning?	Extrinsic > Intrinsic
	Game performance?	Game score: Extrinsic > Intrinsic Prop. correct: Intrinsic > Extrinsic
Q 3	Has content integration an influence on intrinsic motivation?	Extrinsic > Intrinsic
Q 4	Has content integration an influence on perceived playfulness?	No difference between condition
Q 5	Has content integration an influence on perceived usefulness?	Extrinsic > Intrinsic
Q 6	Does condition interact with perception for	
	Proportional reasoning?	No interaction effect
	Game performance?	No interaction effect
Q 7	Does condition interact with perception for intrinsic motivation?	No interaction effect

Discussion

In this study, VSE students were presented with a GBLE. The first analysis revealed that the GBLE stimulated students' *proportional reasoning abilities*, which was measured through their progress from pre- to post-test, regardless of the content integration manner. Other studies that had students play with this GBLE (e.g., ter Vrugte et al. 2015) demonstrated similar learning gains, which increases the robustness of the finding that the GBLE is powerful.

The focus of this study was on the effect of different *ways of integrating* learning content into a GBLE for VSE students. The results provide evidence for this effect. More specifically, participants who played with the extrinsically integrated GBLE showed higher *learning gains*, higher *motivational gains* and higher *perceived usefulness* than the participants who played with a GBLE with intrinsically integrated content. The generalizability of these findings is discussed below.

This *effect of content integration on motivation and performance* was not completely in line with our expectations based on literature data. At the outset, it was for instance assumed that intrinsically integrating the content would stimulate students and make them outperform those students who played in the extrinsically integrated condition. Instead, the results of the study indicated the opposite. For VSE students' total game score, the extrinsically integrated condition outperformed the intrinsically integrated condition. However, when efficiency in solving the proportional reasoning items during gameplay was taken into account, the intrinsically integrated condition performed better, although only for one problem type. The type of proportional reasoning problem for which a significant effect was found (i.e., the comparison problems) partly explains this finding. A distinctive aspect of these comparison items was that students had to choose between two possible answers rather than calculating or changing a missing or given value. Also, only one attempt was allowed. Hence, the proportion correct of these items did not assess students' as well and it may have been attained based on chance (as there is always a one in two chance that the answer is correct). For the other two problem types, students were permitted three attempts so that their skill level was more accurately captured by the proportion correct score. However, no significant difference between the conditions was found for those problem types. What is more, the effect of the comparison items disappeared when students solved proportional reasoning problems outside the GBLE. In this case, students who played with the extrinsically integrated version performed better.

If these findings are linked to students' results at the level of *learning gain scores*, *several explanations* arise. *First*, integrating the learning content into the game mechanics proved a complex and difficult process, despite the ensured flow-experience, for our particular target group—i.e., VSE students with a significant number of at-risk youths (ter Vrugte et al. 2015). Students who play with this kind of GBLE experience more difficulties in learning the content because they simultaneously have to cope with two competing demands: the educational game and the gameplay elements (Shaffer 2004). This distracts their attention from the learning content. The proportional reasoning items were presented separately from the gameplay elements to students in the extrinsically integrated condition. They, in contrast, did not have to cope with both demands simultaneously: either they solved the exercises or they played the game. These VSE students consequently devoted more attention to the learning contents, were probably more stimulated to find the correct answer and completed more attempts to solve the items correctly. This particularity of the target group might also explain why the findings from this study deviate from findings in other previous conducted studies. As for instance the results of the study of Clark et al. (2011) was conducted with students from 7th to 9th grade science classes, these students might not struggle with the simultaneously asked demands occurring while playing an intrinsically integrated game. Also the study of Habgood and Ainsworth (2011) implied a different target group, being 7–8 year primary school students, and hence representing a more heterogeneous and less at-risk group of students than the VSE students used in the current study. The fact that the participants in the study of Habgood and Ainsworth (2011) had high prior computer experiences as they are members of an after-school computer club, they might have higher acquaintance with computer games, which might have a less detrimental impact on the higher demands an intrinsically integrated game poses to the players and hence not disturb the game-flow, while this is the case for VSE students.

A *second* possible explanation is the similarity between the exercise formats in the extrinsically integrated GBLE and the pre-and post-test. This similarity was higher in the extrinsically integrated condition. This might explain why students in the intrinsically integrated condition experienced more difficulties in transferring their mathematical knowledge from one context (the game) to the next (the paper-and-pencil test) (Habgood and Ainsworth 2011). The intrinsically integrated GBLE does not appear to help VSE students make the leap from tacit understanding during gameplay to more formalized knowledge in the classroom (Clark et al. 2011) while the extrinsically integrated GBLE does. The specific characteristics of the target group (i.e., their below average cognitive capability and potential which hinders the less evident transfer of learning content) seems of influence in explaining the different findings in comparison with previous studies (i.e., Clark et al. 2011). However, also the way the intrinsically integrated version of the game was designed might be an explanation for this. In intrinsically integrated versions of a game, the mathematical content is delivered through the parts of the game that are the most fun to play and embodied within the structure of the gaming world and the players' interactions with it. However, it might be that the characteristics of the GBLE used in this study (i.e., 2D-cartoon like game) make the intrinsic integration of the learning content less intense than for instance is the case in more sophisticated GBLEs (e.g., a 3D-adventure game as 'Zombie division in Habgood and Ainsworth 2011). This difference in degree of intrinsic integration, due to the difference in the parts of the game that embody the structure of the gaming world, might explain the difference in findings across studies. Hence further investigation whether this difference in degree in intrinsic integration seems necessary.

Third, playing in the extrinsically integrated condition seems to be more *intrinsically motivating* for VSE students. The difficulties students experienced in the intrinsically integrated condition (see above) frustrated them to such a degree that it reduced their motivation. Inherent to the design of the extrinsically integrated condition, students in this condition received a reward each time a particular learning content was tackled: they were given the opportunity to play a subgame in which no learning content was integrated (and hence some leisure time activity). This extrinsic reward may have stimulated students' intrinsic motivation (Hoffmann et al. 2009). For this specific target group, providing variation in educational games in the form of learning content that is alternated with playtime opportunities (without learning content) seems to improve learning gains for proportional reasoning items and also seems to be more motivating in terms of perceived enjoyment.

Related to the *influence of students' perception*, the extrinsically integrated condition moreover proved to have a higher perceived usefulness than the intrinsically integrated condition. This shows that students are indeed active actors in learning environments

(Lowyck et al. 2004). Nevertheless, no evidence was found in this study that supported the hypothesis that perception of instructional methods (i.e., the pre-perceived usefulness of the GBLE) affects students' proportional reasoning ability, game performance and motivation. These findings are in contrast to statements indicating that "learners are active actors in learning environments and not mere consumers of instructional designers' products" (Lowyck et al. 2004, p. 429). Reformulated, we found no evidence for the assumption that way in which students perceive instructional interventions triggers their engagement in learning and their learning results (Elen and Lowyck 2000; Entwistle 1991; Lowyck et al. 2004; Salomon 1984; Shuell and Farber 2001).

Some other issues in this study require further explanation. Firstly, the participants and more specifically the composition of the conditions, resulted in two problems: (1) the number of participants, and (2) the composition of the condition based on the complete groups of students. First, the number of participants was rather low. Although 64 students were recruited for this study, only 58 complete data sets were obtained. This may make it difficult to generalize the findings of this study. Secondly, the composition of the conditions is likely to have produced other generalization caveats. The conditions were composed based on complete groups rather than separate individuals. This was a conscious choice and it was made based on several reasons. Because the experiment was scheduled during regular PGS lessons, which did not take place during the same time periods for the different groups, and because it lasted four course hours for each class, it was impossible to place students from different groups together to form one condition, or to make a combination of individual students over groups in one condition. Because of the study design and specificity of both conditions (e.g., students' individual computer screens were visible to their neighbors so that the different integrations of the learning content in the GBLE might have distracted students), offering both conditions to one group would have been an ill-advised move. In order to guarantee the equality of procedure between the classes/conditions, the researcher took on the responsibility of guiding the experiments and giving students in different groups the same instructions. Nonetheless, the groups still could differ from each other as a consequence of the groups' different specificity in VSE. First of all, the classes were gender-specific (partly due to specific study orientations, see below), which resulted in an over-representation of either boys (e.g., woodworking, mechanics) or girls (e.g., hairdressing courses, care). For this study, almost half the intrinsic condition group was comprised of "woodworking" boys, while over a third of the extrinsic group consisted of "cookery" girls. We have previously argued that girls have less initial computer and game knowledge, which may result in greater difficulty in using a game application (Vandercruysse et al. 2012). However, the extrinsic integration condition outperformed the intrinsic integration condition for the total game score. This score assessed students' math ability together with their gaming skills. This might indicate that the influence of gender on GBL processes may not be as great as has previously been assumed. What is more, there seems to be no consistent gender effect on learning from games (Ke 2009) and gender does not seem to influence the effects of games on motivation and learning (Huizenga et al. 2010). Aside from any gender differences, the groups' performance may also have been influenced by the different study orientations. In order to examine whether there was a significant difference between the groups for the dependent variables used in the analyses, additional multi-level analyses were conducted. Not one multi-level analysis showed a significant difference between the groups for the measured variables before they participated in the study. The differences between the groups were therefore disregarded in the analyses. Even though the different groups started the treatment with comparable scores, this did not mean that they experienced and interacted with the two versions of the GBLE in the same way. Taken together, this may point to generalization problems for VSE in general and can be seen a limitation of the study. Future research should hence focus on (1) the possible differences between study orientations in VSE. This could be done by investigating whether the findings are stable for all study orientations within VSE, or exploring whether a particular type of VSE student gains more from intrinsically integrated games than extrinsically integrated games. Furthermore, generalizability to (2) other types of education may be limited since ter Vrugte et al. (2015) have demonstrated that VSE students differ from other target groups for the effect of instructional support in GBLEs. Instructional support that has been proved to be successful does not guarantee success for this target group. Further extending this reasoning, this may also be true for content integration. Extrinsic content integration was found to be effective for VSE students. However, whether this can be generalized to other target groups remains to be studied. These specific differences between different target groups should be investigated, more specifically whether students from other education levels react similarly on the same interventions or not. This would enable us to pinpoint the decisive aspects of a target group for the effect of GBL and whether it is students' prior knowledge, their motivation, their previous gaming or school experiences, etc. that influences the GBL outcomes.

Furthermore, a number of other *limitations* remain. For instance, the provided learning time was relatively short to attain deep learning since the mean playtime was 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, however, might have provided more information about the quality of the acquired knowledge after the students' played an intrinsically or extrinsically integrated mathematical game (i.e., long-term effects). In addition, the study design did not include a control condition, i.e., a group of learners who were not engaged in game-based learning, which prevents the conclusion that the two interventions were the primary or only cause for students' improved performance and motivation. However, the inclusion of such a condition was not necessary to answer the research questions and, what is more, the inclusion of a control condition would also imply several difficulties (e.g., the media-debat becomes prominent) and rise some questions (e.g., 'How does the ideal control condition look like?'). Therefore no control condition was included in the study. In addition, subsequent studies should implement a short debriefing phase after gameplay. During this debriefing, students and teachers might interact with each other and students would be able to share their game experiences, discuss problems they encountered during game-play and reflect on their performance. As encouraged by Felicia (2011), Watson et al. (2011) and Charsky and Mims (2008), such a debriefing phase helps students understand their mistakes and stimulates reflection, which heightens the potential for transfer.

This study investigated the impact of one instructional element of a GBLE, i.e., the learning content, and more specific whether the way this element is integrated, matters. Obviously, further research is warranted in order to replicate the findings of this study. First, the replication of these finding with other target groups can researched (see above). Second, the different degrees of intrinsic integration (and its impact on students learning processes; see above) can be investigated. Furthermore, it should be investigated whether the way the learning content is integrated is for instance dependent on the learning content. It might be the case that several learning domains might profit from intrinsically integrated GBLEs while other domains do not and are better presented in extrinsically integrated GBLEs. Another suggestion for follow-up research is the investigation of the impact of content integration on the context in which the GBLE is used. Additional research is also needed to investigate the effect of distinct game characteristics (i.e., the content integration in this study) on students' motivation, performance and perception since this type of research might support the development of effective educational games. When conducting such research, we should however also be aware of the fact that—although isolating one specific element in order to investigate its impact on students' performance, motivation and perception—changing one element in a larger GBLE, always affects the GBLE design as a whole because of having an impact on the way the remaining elements are related to each other or interact with each other (Salomon 1990). This makes the call for replication studies even bigger in order to gain insight in the potential impact of GBLE.

In sum, the results of this study indicate that VSE students benefit from playing with an extrinsically integrated GBLE. When the learning content is explicitly and separately presented in a GBLE, students in this target group seem to obtain higher learning and motivational gains, compared to a GBLE in which the learning content is integrated in the game story and game mechanics. These findings are surprising and contrast with the results that would be expected based on previous literature.

Acknowledgements This study was funded by The Netherlands Organisation for Scientific Research/ Programme Council for Fundamental Scientific Education Research (NWO/PROO—Project No. 411-00-003) and the Fund of Scientific Research (FWO—Project No. G.O.516.11.N.10).

Compliance with ethical standards

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

References

Aldrich, C. (2005). *Learning by doing: A comprehensive guide to simulations, computer games, and pedagogy in e-learning and other educational experiences*. San Fransisco, CA: Pfeiffer.

Barab, S. A., Gresalfi, M., & Ingram-Goble, A. (2010). Transformational play: Using games to position person, content, and context. *Educational Researcher, 39,* 525–536.

Charsky, D., & Mims, C. (2008). Integrating commercial off-the-shelf video games into school curriculums. *TechTrends, 52*(5), 38–44. doi:10.1007/s11528-008-0195-0.

Clark, D. B., Nelson, B. C., Chang, H.-Y., Martinez-Garza, M., Slack, K., & D'Angelo, C. M. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States. *Computers & Education, 57,* 2178–2195. doi:10.1016/j.compedu.2011.05.007.

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. doi:10.1016/j.compedu.2012.03.004.

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

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.

Elen, J., & Lowyck, J. (2000). Instructional metacognitive knowledge: A qualitative study on conceptions of Frenchmen about instruction. *Journal of Curriculum Studies, 32,* 421–444. doi:10.1080/002202700182637.

Entwistle, N. (1991). Approaches to learning and perceptions of the learning environments: Introduction to the special issue. *Higher Education, 22,* 201–204. doi:10.1007/BF00132287.

Felicia, P. (2011). How can digital games be used to teach the school curriculum. Retrieved from http://linked.eun.org/c/document_library/get_file?p_l_id=22779&folderId=24664&name=DLFE-783.pdf

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33,* 441–467. doi:10.1177/1046878102238607.

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. doi:10.1111/j.1365-2729.2012.00489.x.

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences, 20,* 169–206. doi:10.1080/10508406.2010.508029.

Habgood, M. P. J., Ainsworth, S. E., & Benford, S. (2005). Endogenous fantasy and learning in digital games. *Simulation & Gaming, 36,* 483–498. doi:10.1177/1046878105282276.

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

Hayes, E. R., & Games, I. A. (2008). Making computer games and design thinking: A review of current software and strategies. *Games and Culture, 3–4,* 309–332. doi:10.1177/1555412008317312.

Hays, R. T. (2005). *The effectiveness of instructional games: A literature review and discussion*. Technical Report No. 2005-004, Naval Air Warfare Center Training Systems Division, Orlando, FL.

Herndon, J. N. (1987). Learner interests, achievement, and continuing motivation in instruction. *Journal of Instructional Development, 10*(3), 11–14. doi:10.1007/BF02905781.

Hoffmann, K. F., Huff, J. D., Patterson, A. S., & Nietfeld, J. L. (2009). Elementary teachers' use and perception of rewards in the classroom. *Teaching and Teacher Education, 25,* 843–849. doi:10.1016/j.tate.2008.12.004.

Huizenga, J., Admiraal, W., & ten Dam, G. (2010). A literature review of claims on learning and motivation in game-based learning. In B. Meyer (Ed.), *Proceedings of the 4th European conference on games based learning*. Reading: Academic Publishing.

Kaput, J., & West, M. M. (1994). Missing-value proportional reasoning problems: Factors affecting informal reasoning patterns. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 235–287). Albany, NY: State University of New York Press.

Karplus, R., Pulos, S., & Stage, E. K. (1983). Early adolescents' proportional reasoning on 'rate' problems. *Educational Studies in Mathematics, 14,* 219–233. doi:10.1007/BF00410539.

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. doi:10.4018/978-1-59904-808-6.ch001.

Kinzie, M. B. (1990). Requirements and benefits of effective interactive instruction: Learner control, self-regulation, and continuing motivation. *Educational Technology Research and Development, 38,* 5–21. doi:10.1007/BF02298244.

Liu, T. Y., & Chu, Y. L. (2010). Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers & Education, 55,* 630–643. doi:10.1016/j.compedu.2010.02.023.

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

Malone, T. W. (1980). What makes things fun to learn? Heuristics for designing instructional computer games. In *Proceedings of the 3rd ACM SIGSMALL symposium and the 1st SIGPC symposium* (pp. 162–169). doi:10.1145/800088.802839

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 5,* 333–369. doi:10.1016/S0364-0213(81)80017-1.

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivation for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning and instruction Volume 3: Conative and affective process analysis* (Vol. 3, pp. 223–253). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

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. doi:10.1080/02701367.1989.10607413.

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

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal, 16,* 455–474. doi:10.1080/09585170500384529.

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. doi:10.13152/IJRVET.1.2.2.

Plant, R. W., & Ryan, R. M. (1985). Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: An investigation of internally controlling styles. *Journal of Personality, 53,* 435–449. doi:10.1111/j.1467-6494.1985.tb00375.x.

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 and Gaming, 23,* 261–276. doi:10.1177/1046878192233001.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25,* 54–67. doi: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. doi:10.1037/0022-0663.76.4.647.

Salomon, G. (1990). Studying the flute and the orchestra: Controlled vs. classroom research on computers. *International Journal of Educational Research, 14,* 521–531.

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

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

Shuell, T. J., & Farber, S. L. (2001). Students' perceptions of technology use in college courses. *Journal of Educational Computing Research, 24,* 119–138. doi:10.2190/YWPN-H3DP-15LQ-QNK8.

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64,* 489–528. doi:10.1111/j.1744-6570.2011.01190.x.

Struyven, K., Dochy, F., Janssens, S., & Gielen, S. (2008). Students' experiences with contrasting learning environments: The added value of students' perceptions. *Learning Environments Research, 11,* 83–109. doi:10.1007/s10984-008-9041-8.

ter Vrugte, J., de Jong, T., Wouters, P., Vandercruysse, S., Elen, J., & van Oostendorp, H. (2015). When a game supports prevocational math education but integrated reflection does not. *Journal of Computer Assisted Learning, 31*(5), 462–480. doi:10.1111/jcal.12104.

Tourniaire, F., & Pulos, S. (1985). Proportional reasoning: A review of the literature. *Educational Studies in Mathematics, 16,* 181–204. doi:10.1007/BF02400937.

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.

Vandercruysse, S., Desmet, E., Vandewaetere, M., & Elen J. (2015a). Integration in the curriculum as a factor in math-game effectiveness. In J. Torbeyns, E. Lehtinen & J. Elen (Eds.), *Describing and studying domain-specific serious games. Advances in game-based learning* (pp. 133–153). Cham: Springer International Publishing AG.

Vandercruysse, S., ter Vrugte, J., de Jong, T., Wouters, P., van Oostendorp, H., & Verschaffel, L., et al. (2015b). "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. Advances in game-based learning* (pp. 63–81). Cham: Springer International Publishing AG.

Vandercruysse, S., Vandewaetere, M., & Clarebout, G. (2012). Game based learning: A review on the effectiveness of educational games. In M. M. Cruz-Cunda (Ed.), *Handbook of research on serious games as educational, business, and research tools* (pp. 628–647). Hershey, PA: IGI Global.

Vandercruysse, S., Vandewaetere, M., Maertens, M., ter Vrugte, J., Wouters, P., & de Jong, T., et al. (2015c). Development and validation of the Game Perceptions Scale (GPS). *Journal of Educational Multimedia and Hypermedia, 24*(1), 43–74.

Vandercruysse, S., van Weijnen, S., Vandewaetere, M., & Elen J. (2015d). Competitie als game-element integreren in de BSO-klaspraktijk. *Pedagogische Studiën, 92*(3), 179–201.

Vergnaud, G. (1983). Multiplicative structures. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 127–174). Orlando, FL: Academic Press.

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.

Vogel, J. J., Greenwood-Ericksen, A., Cannon-Bowers, J., & Bowers, C. A. (2006). Using virtual reality with and without gaming attributes for academic achievement. *Journal of Research on Technology in Education, 39,* 105–119. doi:10.1080/15391523.2006.10782475.

Warren, S. J., Stein, R. A., Dondlinger, M. J., & Barab, S. A. (2009). A look inside a MUVE design process: Blending instructional design and game principles to target writing skills. *Educational Computing Research, 40,* 295–321.

Watson, W. R., Mong, C. J., & Harris, C. A. (2011). A case study of the in-class use of a video game for teaching high school history. *Computers & Education, 56,* 466–474. doi:10.1016/j.compedu.2010.09.007.

Winne, P. H. (1982). Minimizing the black box problem to enhance the validity of theories about instructional effects. *Instructional Science, 11,* 13–28. doi:10.1007/BF00120978.

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. doi:10.1016/0742-051X(87)90025-.

Sylke Vandercruysse (PhD, 2015) works at the Center for Instructional Psychology and Technology (KU Leuven). Her research is on game-based learning and the interaction with learner-related variables.

Judith ter Vrugte (PhD, 2016) works at the University of Twente in the Department of Instructional Technology (Faculty of Behavioural 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.

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.