The Role of Surprise in Game-Based Learning for Mathematics

Pieter Wouters[1], Herre van Oostendorp[1(✉)], Judith ter Vrugte[2], Sylke vanderCruysse[3], Ton de Jong[2], and Jan Elen[3]

[1] Institute of Information and Computing Sciences, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands
{P.J.M.wouters,h.vanOostendorp}@uu.nl

[2] Department of Instructional Technology, University of Twente, Enschede, The Netherlands
{j.tervrugte,a.j.m.dejong}@utwente.nl

[3] Center for Instructional Psychology and Technology, KU Leuven, Louvain, Belgium
sylke.vandercruysse@kuleuven-kulak.be, jan.elen@ppw.kuleuven.be

Abstract. In this paper we investigate the potential of surprise on learning with prevocational students in the domain of proportional reasoning. Surprise involves an emotional reaction, but it also serves a cognitive goal as it directs attention to explain why the surprising event occurred and to learn for the future. Experiment 1 - comparing a surprise condition with a control condition - found no differences, but the results suggested that surprise may be beneficial for higher level students. In Experiment 2 we combined Expectancy strength (Strong vs. Weak) with Surprise (Present vs. Absent) using higher level students. We found a marginal effect of surprise on learning indicating that students who experienced surprises learned more than students who were not exposed to these surprises but we found a stronger positive effect of surprise when we included existing proportional reasoning skill as factor. These results provide evidence that narrative techniques such as surprise can be used for the purpose of learning.

1 Introduction

Despite the increasing popularity of serious games or game-based learning (GBL), recent meta-analytic reviews have shown that GBL is only moderately more effective and not more motivating than traditional instruction [11]. A potential problem with GBL is that the outcomes of players' actions in the game are directly reflected in the game world. This may lead to a kind of intuitive learning: players know how to apply knowledge, but they cannot explicate it [3]. The articulation of knowledge and underlying rules is important, because it triggers students to *organize* new information and *integrate* it with their prior knowledge [5] and thus construct a mental model that is more broadly applicable. This implies that genuine learning in GBL requires additional features in the game that will provoke the player to engage in the process of knowledge

A. De Gloria and R. Veltkamp (Eds.): GALA 2015, LNCS 9599, pp. 401–410, 2016.
DOI: 10.1007/978-3-319-40216-1_45

articulation. Typically in learning environments such knowledge articulation is often prompted by explicitly asking students to reflect on their actions and thoughts. In complex GBL environments this may compromise the motivating quality of the game or can be so cognitively demanding that learning will not take place [6].

The question raised in this paper is how we can stimulate players to engage in relevant cognitive processes that foster learning without jeopardizing the motivational appeal of the game [12]. A promising technique is the generation of manageable cognitive conflicts by introducing surprises [4]. Surprise involves an emotional reaction, but it also serves a cognitive goal as it directs attention to explain why the surprise occurred and to learn for the future [1]. In the context of learning a medical procedure with a serious game [9] demonstrated that surprise yielded superior knowledge structures, indicating that such events foster deep learning. They assume that in games players construct a mental model based on the story line, the events and the underlying rules of the game. The assumption of this study is that the effect of surprise also pertains to problem solving in serious games. Ideally, the mental model will enable the student to recognize specific characteristics of a problem and how to solve that problem. Because our aim is to integrate the instructional technique (i.e. the introduction of the surprise) with the learning content [2], the surprises have to be focused on what has to be learned, i.e. the mental model of proportional reasoning problems and methods to solve them. For this reason, the surprising events change some of the problem characteristics, and the solution method previously applied, is no longer applicable and the player has to re-evaluate the situation and decide which problem characteristics are relevant and which solution method is now most appropriate. We expect that surprise has a positive effect on learning because it stimulates relevant cognitive processes such as organizing and integrating information [5] without compromising the motivational appeal of computer games.

In this study we investigate the impact of surprise on learning and how this impact is moderated by the expectancy of the student (in the second study).

2 Experiment 1

In Experiment 1 a group of students playing the game with surprises occurring during the game was compared with a group without these surprises (control group). We expected that the group with surprises would learn more from the GBL than the control group.

2.1 Method

Participants and Design. The participants were 71 students from second year pre-vocational education. We adopted a pretest-posttest design with a control condition (N = 36) and a surprises condition (N = 35). Participants were randomly assigned to the conditions.

Materials.

Domain. The domain of proportional reasoning comprises three problem types: comparison problems, missing value problems, and transformation problems. In comparison problems students have to find out whether one proportion is "more than", "lesser-than" or "equal to" another proportion. In missing value problems one value in one of two proportions is missing. Students have to find this "missing value" in order to ensure that both proportions are equal. Transformation problems involve two proportions as well and all values are known, but the proportions are not equal. Students have to find out how much has to be added to one or more of the proportions in order to make both proportions equal (for a more extensive description see [7]).

Game Environment. In the 2D game (Flash/ActionScript 3) called *Zeldenrust* players have a summer job in a hotel (http://www.projects.science.uu.nl/mathgame/zeldenrust/index.html). By doing different tasks the players can earn money that can be used to select a holiday destination during the game: the more money they earn, the further they can travel. During the game the player is accompanied by the manager, a non-playing character, who provides information about the task and gives feedback regarding the performance on the task. The game comprises a base game and several subgames. The base game provides the structure from which the subgames can be started. After selecting an avatar, the players receive an introduction animation in which the context of the game is presented and finally enter the "Student room" from which the player can control the game (e.g., for example by choosing a specific subgame). Each task is implemented as a subgame and covers a specific problem type in the domain of proportional reasoning. The tasks are directly related to proportional reasoning (e.g., mixing two drinks to make a cocktail according to a particular ratio directly involves proportional reasoning skills). The actual assignments are described on a *whiteboard*. With drag-and-drop or clicking, the player can accomplish the assignment, but the specific action depends on the subgame. To further motivate the player, a *"geldmeter" (money meter)* is implemented which visualizes the amount of money that the player will receive after an assignment. Correct and incorrect actions during an assignment are directly reflected in the money meter. For example, if the player breaks a bottle, the money meter will decrease (and the color becomes redder); if the player places bottles in the refrigerator the money meter will increase (and becomes greener). The money meter also shows the (accumulated) amount of money that the player has earned. The player can use a built-in *calculator*, but using it will cost some money. Depending on the subgame the player has to perform a typical action (e.g., closing the door of the refrigerator) to receive *verbal feedback* from the manager of the hotel who tells whether the answer is correct or not (e.g., 'Excellent' or 'You have too much Cola in relation to Fanta'). If the answer is correct the money meter will be increased.

The *surprise* condition comprised a non-playing niece character in the introduction animation telling that she sometimes will make it difficult to carry out the task. When a surprise occurred the niece character had popped up and told that she had changed something. This change involved specific characteristics of the task whereby the solution method of the player doesn't apply anymore and the player has to reconsider the original solution method. Figure 1 gives an example of the occurrence of a surprise.

Figure 1a depicts the starting situation. The player can solve the problem by looking at the ratio within: the number of Fanta in the refrigerator is twice as much as the number of Fanta in the desired proportion (12 Fanta) since 12 * 2 = 24, so the number of Cola also has to be doubled (9 * 2 = 18 Cola). When the player is implementing the solution, the surprise occurs (Fig. 1b). When the niece character has disappeared the characteristics of the task are modified (Fig. 1c); that is, the desired proportion is now 5 Cola per 10 Fanta. The ratio "within" is not applicable anymore and the player can better use a method based on the ratio "between" (the desired proportion is 5 Cola/10 Fanta, so the number of Cola in the refrigerator should also be half the number of Fanta, 12/24). In total the players received 8 surprises (four in both the missing value and the transformation subgames).

Tests/Measures. The arithmetic tempo test, the *TTR (Tempo Test Rekenen),* measures fluency in basic arithmetic operations i.e., addition, subtraction, multiplication, and division. The TTR score is calculated as the sum of correct answers. The range of possible scores is 0–200. Proportional reasoning skill was measured with a test consisting of 12 open questions: four questions for each problem type. There were two versions of the test. The comparability of both versions was tested in pilot study. Each answer of the pre- and posttest was coded as 0 (wrong answer or no answer) or 1 (correct answer). As game performance indicators the number of correct game tasks and the time spent on game tasks were used.

Procedure. The experiment run on school computers and comprised three sessions of 50 min. In session 1 the experiment was introduced and the pretest was administered. In the second session, a week later, the participants played the game (40 min). All actions of the players during playing the game were logged. The posttest was administered in the third session (a week after playing the game).

2.2 Results and Conclusion

In order to test the effect of surprise on learning we used the combined score of the items of the two problem types in which surprise was applied (missing value – Refrigerator subgame; transformation – Blender subgame). Table 1 shows the results for each condition on proportional reasoning skill. A paired samples T-test reveals that playing the game ($t(70) = 2.73$, $p = .008$, $d = 0.29$) improves learning. We also conducted a hierarchical regression analysis to investigate whether game performance indeed is predictive for learning. The first block comprised pretest score and TTR; the second block consisted of correct game tasks and time spent on game tasks. The number of correct game tasks explains an additional 6 % extra of the variance in posttest performance ($B = .11$, $SE\ B = .04$, $\beta = .39$, $p = .007$). An independent T-test with overall learning gain (posttest score – pretest score) as dependent variable shows no difference between control and surprise group ($t(69) = .07$, $p > .05$, $d = .00$).

The surprise group did not perform better than the control group. There are two plausible explanations for the failure to find a beneficial effect of surprise. To start with, the processing of surprise requires a certain level of cognitive flexibility and metacognitive skills. Students must perceive and understand that the changes in the

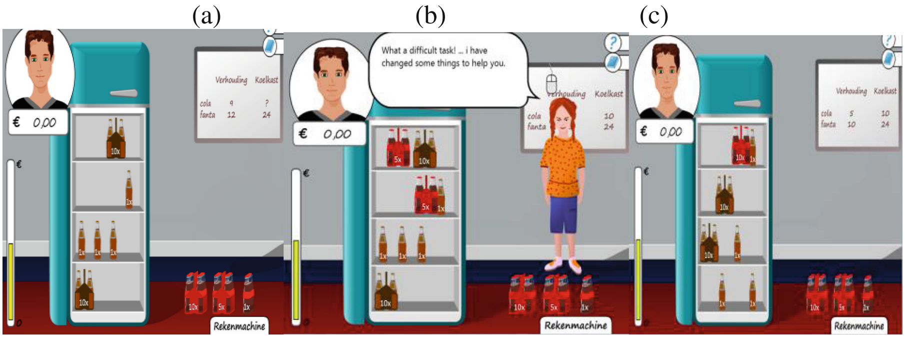

Fig. 1. (a) Starting situation in a task with a surprise, (b) surprise (c) modification of task characteristics in the game *Zeldenrust*.

Table 1. Mean scores and standard deviations on the dependent variables of Experiment 1

TTR	Control		Surprise	
	76 (12)		75 (20)	
	Pre	Post	Pre	Post
	M (SD)	M (SD)	M (SD)	M (SD)
All items	4.32 (2.33)	5.07 (2.65)	4.38 (2.01)	5.02 (2.56)
Surprise items	2.30 (2.87)	2.88 (2.34)	2.06 (1.71)	2.66 (2.31)
Comparison	2.02 (.99)	2.20 (.91)	2.32 (1.83)	2.36 (1.18)

Note: Range of All items is 0 to 12. All items means all proportional reasoning skill items. Surprise items (are missing value and transformation items) range is 0–8. Range comparison items is 0–4.

problem situation of the game are not superficial but that some deeper characteristics of the problem have been altered, see that the changes may have consequences for the chosen solution method and consider whether another method is more appropriate. For students who do not possess these skills sufficiently, the surprise can be confusing or even frustrating because their solution method is thwarted. The students in this experiment can be classified in one of three educational levels that may give an indication regarding cognitive flexibility and metacognitive skills: (1) basic with a practical orientation, (2) moderate with a larger theoretical component and (3) high with a large theoretical (cognitive) component. Probably students in the third level have higher metacognitive skills than those in the first and second level. Figure 2 shows the posttest scores on the surprise items for the three groups in each condition. Although ANCOVAs (posttest score surprise items as dependent variable; condition and educational level as independent variables and TTR and pretest score surprise items as covariate) showed no significant effects (main effect condition and level $F < 1$; interaction of condition and level $F(1, 67) = 1.36$, $p > .05$), theoretical level students seem to benefit more from surprises.

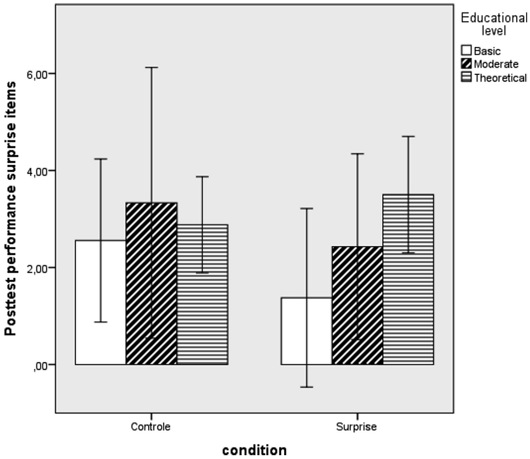

Fig. 2. Posttest performance score surprise items (missing value and transformation items) for basic, moderate and theoretical level students in Experiment 1.

The second explanation concerns the order in which problems with different characteristics were presented to the player. Often the characteristics of a new problem were different from the preceding one which may have thwarted the emergence of a strong expectation. If this is true, the potential beneficial effect of surprise may not have been fully realized. In that case surprise triggering the necessity to retrieve and update the mental model will be weak.

3 Experiment 2

In Experiment 2 we investigated the two possible explanations discussed in Experiment 1. First, participants were recruited from the theoretical educational levels instead of all three educational levels. Second, we introduced a second independent variable in which expectancy was manipulated.

3.1 Method

Participants and Design. The participants were 94 students from second year pre-vocational education. We adopted a pretest-posttest design with two independent variables *Surprises* (Yes or No) and *Expectancy strength* (Strong or Weak) that were factorial combined resulting in 4 conditions: Surprises and Strong expectancy ($N = 22$), Surprises and Weak expectancy ($N = 23$), No surprises and Strong expectancy ($N = 26$) and No surprises and Weak Expectancy ($N = 23$). Participants were randomly assigned to conditions. We tested three hypotheses: (1) Playing the game will improve learning; (2) We expect a main effect of surprise indicating that surprise will increase learning more because it triggers students to interpret the changes in the problem characteristics caused by the surprise and the consequences for the solution process; (3) In addition, we hypothesize an interaction between surprise and expectancy strength indicating that a surprise after multiple problems with the same characteristics (Strong expectancy) will have the largest learning effect because the strong unexpectedness of the surprise will incite students more (or deeper) to think about the changes in the characteristics of the problem characteristics caused by the surprise and the consequences for the solution process.

Materials.

Domain. The same domain was used as in Experiment 1.

Game Environment. The game environment was the same as in Experiment 1 with a modification in the surprise conditions. In Experiment 2 the screen first became brighter (to make the surprise more salient) and then dimmed again before the niece character appeared. The implementation of the factor *Expectancy strength* implied that the characteristics of the problems changed. Strong expectancy was defined as a series of problems with the same characteristics. Weak expectancy was defined as a series of problems in which the characteristics varied. The tests, procedure and scoring were the same as in Experiment 1.

3.2 Results and Conclusion

As in Experiment 1 we used the combined score of the surprise. Table 2 shows the results for each condition on proportional reasoning skill. To test hypothesis 1 we conducted a paired-samples T-test. The results show that playing the game improves learning ($t(93) = 2.54$, $p = .013$, $d = .25$). This is partly corroborated by the results of the hierarchical regression analysis: the number of correct tasks explains 10 % extra of the variance in posttest performance ($B = .10$, $SE\ B = .03$, $\beta = .34$, $p = .002$). Hypotheses 2 and 3 were tested with 2*2 ANCOVA with Surprise and Expectancy strength as independent variables, posttest score as dependent variable and TTR and pretest scores as covariate.

For the surprise items we found a marginally significant main effect for Surprise ($F(1, 90) = 3.16$, $p = .079$). The main effect for Expectancy strength and the Surprise*Expectancy strength interaction were not significant (both $F(1, 90) < 1$). For the comparison items we did not find main or interaction effects (all $F < 1$). The participants in this study represent a heterogeneous population which is reflected in the large SD. We assumed that better performing students would possess the (meta) cognitive skills to deal with the surprises and benefit from the cognitive processes that they trigger. We divided the sample in low and high level students based on the median score of 6 on the pretest. Figure 3 shows the results.

We ran an ANCOVA with the posttest score on the surprise items as dependent variable; surprise, expectancy strength and level as fixed factors and TTR as covariate. We expected to see an interaction between surprise and expectancy and between surprise and level, indicating that high level students would benefit more from surprises than low level students. The results, however, only show significant main effects for

Table 2. Mean scores and standard deviations on the dependent variables of Experiment 2

TTR	Surprise			
	Strong expectancy		Weak expectancy	
	Pre	Post	Pre	Post
	M (SD)	M (SD)	M (SD)	M (SD)
All items	5.71 (2.34)	6.90 (3.18)	6.00 (3.05)	6.95 (3.29)
Surprise items	3.04 (1.88)	4.28 (2.23)	4.00 (2.41)	4.79 (2.53)
Comparison	2.66 (1.06)	2.61 (.97)	2.00 (1.06)	2.17 (1.20)
TTR	No surprise			
	Strong expectancy		Weak expectancy	
	Pre	Post	Pre	Post
	M (SD)	M (SD)	M (SD)	M (SD)
All items	5.07 (2.41)	5.53 (3.46)	5.56 (1.61)	5.86 (3.16)
Surprise items	3.11 (2.41)	3.57 (2.14)	3.21 (1.44)	3.47 (2.44)
Comparison	1.96 (.95)	1.96 (1.18)	2.35 (1.07)	2.39 (1.40)

Note: Range of All items is 0 to 12. All items means all proportional reasoning skill items. Surprise items (are missing value and transformation items) range is 0–8. Range comparison items is 0–4.

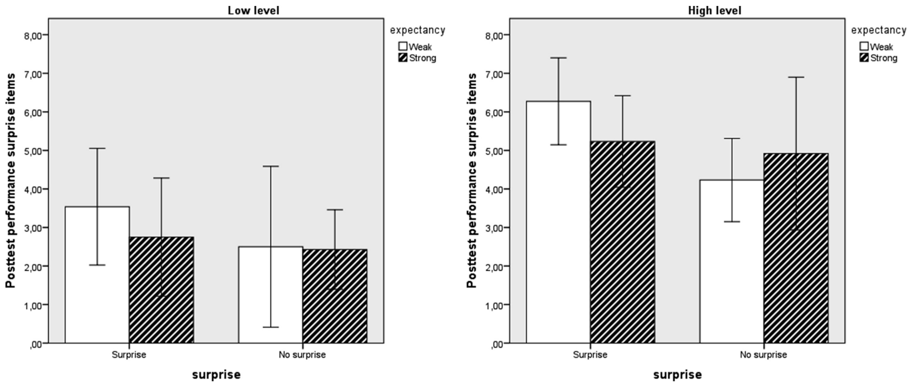

Fig. 3. Posttest performance on surprise items (missing value and transformation items) for low (left) and high (right) level students.

Surprise ($F(1, 85) = 4.12$, $p = .046$) and Level ($F(1, 85) = 18.98$, $p = .000$), but all other main or interaction effects were not significant (Expectancy*Surprise: $F(1, 90) = 1.06$, $p > .05$; all other effects $F < 1$).

4 General Discussion

In two experiments we found that learning improves by playing the game. This corroborates earlier findings regarding serious games in general (cf. [11, 12] and other studies with the game *Zeldenrust* [6]). In both experiments we also found that effective game play (the number of correct game tasks) is predictive for posttest performance. In Experiment 1 we failed to find a beneficial effect of surprise, though high educational level students did benefit. We provided two arguments for this finding. The participants did not possess sufficient (meta)cognitive skills and the expectancy factor which is important for surprise was not optimal utilized. In Experiment 2 we operationalized these new demands by focusing on higher cognitive level students and by manipulating the strength of expectancy. We found a marginal effect of surprise on learning indicating that students who experienced surprises learned more than students who were not exposed to these surprises but we found a stronger effect of surprise when we included existing proportional reasoning skill as factor. These results connect with other studies that find cognitive effects of narrative techniques like surprise in games [9, 13]. These results also imply that instructional techniques such as surprises should be applied with care. An important precondition for effective surprise is that players have sufficient cognitive flexibility and metacognitive skills to orientate on the task, to re-evaluate the results at the moment when the surprise occurs and to reflect on the performed actions. Students who lack these competencies can be overwhelmed by the additional cognitive demands that are introduced by these surprises. Possibly, the effect of surprise can be increased by offering students additional instructional support during the problems before the surprise intervention occurs which may help them to select an appropriate method for a problem. One could think of exercises that help them to automatize part-tasks such as multiplication tables so that they can more easily identify intern or extern ratios and/or worked examples in which strategies for specific types of problems are modelled.

Two other lines of research can also be interesting. First, there is some evidence that metacognitive skills in math improve with small differences in age [8]. The students in this study had a mean age of 13.9 years (second year class) and the metacognitive skills of some may have been insufficiently developed. Another point is that the students come from the least advanced of three Dutch educational tracks. It would be interesting to replicate this study with older students in the same educational level or students from a higher educational track. A second research avenue pertains to the characteristics of the game. The game *Zeldenrust* has a repetitive character, students engage in the same type of tasks which require similar actions. It is not unlikely that students finally will expect that the niece character will reappear and modify the nature of the task. In that case they may anticipate these events and thus undermine the potential effect of surprise. If that is the case more variation in surprise can perhaps further increase its effectiveness.

It may seem that the introduction of surprise adds more difficulty to the problems presented. Two comments to this suggestion can be made. Firstly, making a problem in first instance somewhat harder, e.g. by omitting particular information, can improve learning, particularly of the underlying rules of the problem [10]. Secondly, we want to point out that surprise did have a positive effect on learning, particularly of students with sufficient (meta)cognitive capabilities.

Acknowledgments. This research is funded by the Netherlands Organization for Scientific Research (project Number No. 411-00-003).

References

1. Foster, M.I., Keane, M.T.: Predicting surprise judgments from explanation graphs. In: International Conference on Cognitive Modeling (ICCM). Groningen University, Groningen, The Netherlands (2015)
2. Habgood, M.P.J., Ainsworth, S.E.: Motivating children to learn effectively: exploring the value of intrinsic integration in educational games. J. Learn. Sci. **20**, 169–206 (2011)
3. Leemkuil, H., de Jong, T.: Instructional support in games. In: Tobias, S., Fletcher, D. (eds.) Computer games and instruction, pp. 353–369. IAP, Charlotte, NC (2011)
4. Loewenstein, G.: The psychology of curiosity: a review and reinterpretation. Psychol. Bull. **116**, 75–98 (1994)
5. Mayer, R.E.: Multimedia learning and games. In: Tobias, S., Fletcher, J.D. (eds.) Computer Games and Instruction, pp. 281–305. Information Age Publishing, Greenwich, CT (2011)
6. ter Vrugte, J., de Jong, T., Wouters, P., Vandercruysse, S., Elen, J., van Oostendorp, H.: When a game supports prevocational math education but integrated reflection does not. J. Comput. Assist. Learn. (2015). doi:10.1111/jcal.12104
7. Vandercruysse, S., ter Vrugte, J., de Jong, T., Wouters, P., van Oostendorp, H., Elen, J.: 'Zeldenrust': a mathematical game-based learning environment for vocational students. In: Proceedings of Research on Domain-Specific Serious Games: State-of-the-Art and Prospects Conference. University of Leuven, Belgium (2014)
8. van der Stel, M., Veenman, M.V., Deelen, K., Haenen, J.: The increasing role of metacognitive skills in math: a cross-sectional study from a developmental perspective. ZDM Int. J. Math. Educ. **42**(2), 219–229 (2010)
9. van der Spek, E.D., van Oostendorp, H., Meyer, J-JCh.: Introducing surprising events can stimulate deep learning in a serious game. Br. J. Educ. Technol. **44**, 156–169 (2013)
10. van Nimwegen, C., van Oostendorp, H., Tabachneck-Schijf, H.J.M.: The role of interface style in planning during problem solving. In: Bara, B., Barsalou, L., Bucciarelli, M. (eds.) Proceedings of the 27th Annual Cognitive Science Conference, pp. 2271–2276. Lawrence Erlbaum, Mahwah, NJ (2005)
11. Wouters, P., van Nimwegen, C., van Oostendorp, H., van der Spek, E.D.: A meta-analysis of the cognitive and motivational effects of serious games. J. Educ. Psychol. **105**, 249–265 (2013)
12. Wouters, P., van Oostendorp, H.: A meta-analytic review of the role of instructional support in game-based learning. Comput. Educ. **60**, 412–425 (2013)
13. Wouters, P., van Oostendorp, H., van der Spek, E.D., Boonekamp, R.: The role of game discourse analysis and curiosity in creating engaging and effective serious games by implementing a back story and foreshadowing. Interact. Comput. **23**, 329–336 (2011)