Effects of scripting on dialogues, motivation and learning outcomes in serious games

Hans van der Meij, Sabine Veldkamp and Henny Leemkuil

Hans van der Meij is senior researcher and lecturer. His research concentrates on technical communication, instructional video and ICT-use in education. Sabine Veldkamp is doctoral student. Her research focuses on factors that influence bullying and victimisation. Henny Leemkuil is assistant professor. His research mainly addresses the design of learner support in educational games. Address for correspondence: Hans van der Meij, Faculty of Behavioural, Management and Social Sciences, Department of Instructional Technology, University of Twente, P.O. Box 217, Enschede 7500 AE, the Netherlands. Email: H.vanderMeij@utwente.nl

Abstract
Research on serious games shows that complementary design measures are often needed to increase effectiveness. The present study investigated the role of scripted collaboration. In such a collaboration, players are assigned roles or given specific tasks to stimulate them to communicate more about essential game aspects (eg, arguing why a move should be made versus merely suggesting a move), and thereby raise learning. A risk of scripted collaboration is that it can constrain communication so much that motivation drops. An experiment with eighth-grade students is reported in which a scripted and non-scripted condition were compared for their effects on the level of dialogic acts, motivation and learning outcomes. In the scripted condition, players were assigned opposing roles in a game that revolved around making smart choices by considering each perspective. The findings supported the prediction that scripting would raise the level of the dialogues. Scripting led to significantly more dialogic acts about deeper level game features. In addition, players in the scripted condition achieved a significantly higher mean score on a knowledge posttest. An exploratory analysis showed that higher level dialogic acts correlated with higher posttest scores. Motivation was the same in both conditions.

Introduction
Computer games are competitive, situated, interactive (learning) environments based upon an underlying model, in which a challenging goal must be reached under certain constraints and uncertain circumstances. In games, players actively solve challenging situated problems, sometimes in cooperation with others (Leemkuil & de Jong, 2012). Serious games enable students to engage playfully in activities on meaningful topics. To achieve this, these games have features that can motivate and yield learning (Feng, González, Amor, Lovreglio, & Cabrera-Guerrero, 2018; Hummel et al., 2011; Tsekleves, Cosmas, & Aggoun, 2016). Just like computer games in general, serious games have entertaining qualities such as pleasing graphics, exciting plots and lively game characters that can keep students engaged for an extended period. In addition, serious games have features that can foster learning such as assignments, scaffolds, advice and external feedback (Wouters & Van Oostendorp, 2013).

Nevertheless, the amount of learning from a serious game can fall short. In such a case, a possible remedy is to equip the game with an instructional feature that can increase learning. Educators
are not usually capable of doing this, however. What they can do instead is create complementary instructional support before, during or after the game. Designs that can enhance learning before training such as pre-training (eg, Barzilai & Blau, 2014), or after training such as debriefing (eg, Bilgin, Baek, & Park, 2015), are beyond the scope of this study. Instead, we concentrate on a design option to enhance learning during the game, namely, collaboration.

Games often rely on a learning-by-doing mode that benefits from special measures to stimulate students to reflect on what they are doing (Wouters & Van Oostendorp, 2013). A prime candidate for achieving such a stimulating effect is collaboration. When a game is played in collaborative mode, this arrangement creates a natural context for players to communicate about the game. Empirical research by van der Meij, Albers and Leemkuil (2011) has already shown that collaborating partners frequently exchange ideas about what they are doing in a game. However, that study also showed that the exchanges mainly revolved around basic actions and superficial game features and rarely included predictions or arguments for game moves. In other words, the dialogues were low-level, and, possibly as a consequence, collaboration did not enhance learning more than solitary gameplay. Two other empirical studies that compared gameplay in solitary and collaborative mode likewise found no effect of conditions on learning outcomes (Chen, Wang, & Lin, 2015; Meluso, Zheng, Spires, & Lester, 2012) and this was reported in a recent meta-analysis as well (Clark, Tanner-Smith, & Killingsworth, 2016).

In the larger realm of education, similar findings for free or unstructured collaboration versus individual study have frequently been reported (eg, Barron, 2003; Cohen, 1994; Gegenfurtner, Veermans, & Vauras, 2013). This has led to the development of collaboration scripts in which pairs or larger groups of students are assigned roles, or given specific tasks (eg, Dillenbourg & Hong, 2008). Compared to unstructured collaboration, scripted collaboration has led to qualitatively better dialogues and higher learning outcomes in diverse learning arrangements (Weinberger, Stegmann, & Fischer, 2010).
The present study investigated whether similar beneficial effects from scripted collaboration could be obtained for serious games. A condition with scripted collaborative gameplay was compared with a non-scripted condition. Effects of condition on the quality of the dialogues, motivation, and learning outcomes were investigated.

Scripted collaboration

Research on collaborative learning shows that merely arranging for students to work together does not always result in rich, high-level interactions and subsequent learning (Johnson & Johnson, 1999). Certain boundary conditions must be satisfied to realise effective collaboration. An option that seems particularly well suited for enhancing the quality of the interactions during gameplay is scripted collaboration. Whereas free or unstructured collaboration merely arranges for people to play together with the possible outcome of low-level communications, scripted collaboration imposes constraints that can raise the level of interactions (Dillenbourg & Hong, 2008; Hummel et al., 2011).

The way this works is that the script partially disrupts the spontaneous, natural communications between players to make room for interactions that concern the core of what needs to be learned. The balance is delicate. Gameplay should not be overscripted, because that would diminish motivation and gives players too little room for grounding and otherwise having a normal interaction (see Clark et al., 2016; Dillenbourg & Jermann, 2007). At the same time, the script should increase the chances that the interactions will revolve around the key issues that players must learn. In other words, the script must give players a certain amount of freedom and simultaneously channel the communication so that the right kinds of dialogic acts emerge. One of the ways to achieve this is by creating a script that assigns a meaningful role to each player. The roles can have a beneficial effect on the motivation and actions of the player when he or she assumes the responsibilities befitting the prescribed identity.

There is a growing number of studies on scripted collaboration as part of game mechanics (Hämäläinen, Niilo-Rämä, Lainema, & Oksanen, 2018; Nebel et al., 2017; Oksanen & Hämäläinen, 2014). These studies are based on the idea that collaborative games do not automatically lead to successful collaboration (Zagal, Rick, & Hsi, 2006). The design focus in these studies lies on embedding game mechanics that structure collaboration in such a way that successful gameplay depends on successful collaboration. Features that are often used in designing for scripted collaboration include complementary actions, indirect actions and encrypted information (Hämäläinen et al., 2018; Oksanen & Hämäläinen, 2014). With complementary actions, players need to synchronise their individual tasks to perform a collaborative task. The mechanic of indirect actions is based on the idea that a player possesses information that another player needs in order to act. This forces players to exchange information. A similar situation is created with encrypted information.

To our knowledge, only a few empirical studies have investigated scripted collaboration with games that do not incorporate the mechanics for such collaboration just described. Hummel and his colleagues (2011) investigated the effect of a conflict script in a serious game on water management. The script assigned players roles belonging to opposing views that had to be factored in to achieve a satisfactory solution. That is, the players were asked to take on the role of a representative from an ecology or governance standpoint. The experimental setup was a within-subjects design. Players first engaged with the game in solitary mode and then wrote a draft report on a water management problem. Next, each player selected an ecology or governance perspective for another round of gameplay in solitary mode, after which they improved their draft report. This revised report then yielded written feedback from an unknown other player (a confederate) who
had taken on the opposing role. Finally, players reflected on the feedback, decided how to handle contradictory views and completed their report. Learning was assessed by comparing the scores for the first draft and final report. A statistically significant learning gain was found. A questionnaire showed that the players rated the collaboration as below average on aspects such as liking and helpfulness.

Chen and Law (2016) investigated the effect of scripted collaboration on motivation (ie, interest, competence and autonomy) and learning in a serious game on force and motion. The four conditions in the study were: solitary, scripted solitary, collaborative and scripted collaborative. Scripting consisted of a set of questions whose answers required the players to connect game features with real-world phenomena. These questions were presented after a first round of gameplay, and players could re-engage with the game to arrive at an answer. Learning was assessed with a knowledge test administered before and after gameplay. Participants in the scripted collaboration condition performed significantly better than those in the collaboration and the scripted solitary conditions, who, in turn, had a significantly higher mean learning outcome than those in the solitary condition. There was a significant, negative main effect of scripting on interest, competence and autonomy. Overall, the solitary condition yielded the highest scores on all three motivation measures.

Ter Vrugte and colleagues (2015) investigated scripted collaboration in a math game on proportional reasoning. This experiment had the same four conditions as the study by Chen and Law (2016). Scripting consisted of the inclusion of a competitive element. That is, the players were informed that their game scores and knowledge progression would be compared with other players. During the game, the experimenter reported the highest game score after every 10 minutes, for all players. In addition, the experimenter gave the players an overview of all game scores after the first session of gameplay. Thereafter, the players could engage in a second round of gameplay. No overall effects of condition on learning were found, but prior knowledge played a moderating role. Scripted collaboration decreased learning for players with low prior-knowledge, whereas it enhanced learning for high prior-knowledge players.

Empirical studies on scripted collaboration (of the non-mechanic kind) in serious games are far and few between. In addition, there are differences in the kinds of scripts that are investigated, and the learning outcomes sometimes favour and sometimes do not favour scripted collaboration (Chen & Law, 2016; Hummel et al., 2011; ter Vrugte et al., 2015). The findings for motivation point to a negative effect of scripting (Chen & Law, 2016; Hummel et al., 2011). In view of these findings, the present study was set up as another exploratory investigation on scripted collaboration in game-based-learning.

**Experimental design and research questions**

This empirical study involved an experimental (scripted) condition in which pairs of students were given a script for playing a game. In the control condition pairs of students played the game without a script. The following research questions were investigated:

**Research question 1:** *What types of dialogues are stimulated by scripting?* Dialogic communications were coded in the same fashion as in the study by van der Meij, Albers, and Leemkuil (2011). The tested prediction is that participants in the scripted condition engage in relatively more higher level dialogic acts. That is, it is expected that the proportion of third-and fourth-level dialogic acts is higher in the scripted condition.

**Research question 2:** *What is the effect of scripting on motivation?* The study measured four motivational constructs (ie, interest, anxiety, confidence and challenge) that empirical studies have found to be related to performance and learning (Vollmeyer & Rheinberg, 2006). Arguments for
as well as against an effect of scripting on motivation can be given. For instance, scripting can be experienced as enhancing or reducing the challenge level of the task. Therefore, no specific hypothesis is tested for the effect of condition on motivation.

Research question 3: What is the effect of scripting on learning outcomes? The tested prediction is that the scripted condition enhances learning outcomes, due to higher quality of the dialogues. The association between the relative percentage of dialogic acts on a given level and the learning outcome is also explored.

Method
Participants
Participants in the study were 32 students, split into 16 pairs, from an eighth grade elementary school classroom. The study included an even number of male and female students with a mean age of 11.56 years ($SD = 0.62$). Students volunteered to participate in the study. Participants were randomly assigned to pairs, which were randomly assigned to conditions.

Instructional instruments
The experiment revolved around “Enercities,” a simulation game that challenges the players to build a sustainable city. The starting position is a town hall with acres of surrounding land that players can use for economic purposes (eg, agriculture, industry), or for building houses. The players must aim for the development of a city that has a sufficient supply of houses and industry to ensure an adequate income. In doing so, they should ensure that the change is durable. They should avoid allowing the city to exhaust its natural resources (such as energy). The players must find a way to balance the city’s development and its sustainability. In order to make the right choices for people, profit and planet, the players must carefully consider what to do where and when. The game has four levels, each increasing the available land and options to choose from. Change from one level to the next is automatic, and depends on the number of inhabitants in the city. The game displays a continuously updated game score to the players. The game teaches players about the relevance of renewable energy, energy savings, CO2 emission measures and minimising fossil fuel use (Knol & de Vries, 2011). This fits with Core Objective 39 of Dutch primary education: The pupils learn to handle the environment with care.

Script. There are numerous scripts to choose from. Important considerations for selecting the right kind of script for game-based-learning are usability and intended cognitive stimulus (eg, Dillenbourg & Jermann, 2007; Hummel, Geerts, Slootmaker, Kuipers, & Westera, 2015). Above all, the script should be easy to use. It should not require much advance training, nor should it impose too much additional cognitive load, if any, during gameplay. The script should also steer the conversations in the right direction without giving too much support. The script should stimulate the players to activate prior knowledge and invite reflection rather than involve much teaching or coaching, since that would foil the self-guided nature of gameplay.

The conflict script satisfied both criteria. Essentially, it meant that one player in a team was asked to adopt one role (eg, representative of the ecology perspective), while the other team member had to take on the opposing role (eg, representative of the governance perspective). More specifically, one participant in a pair had to try to make the game’s environmental score as high as possible, while the partner was told to do the same for the economic score. The roleplay was expected to enhance the argumentation for and against each stance, and thus to yield better dialogues.

Research instruments
The dialogues were recorded, transcribed and coded in the same way as in the study by van der Meij, Albers and Leemkuil (2011). More specifically, we employed the same dialogue coding scheme.

© 2019 The Authors. British Journal of Educational Technology published by John Wiley & Sons Ltd on behalf of British Educational Research Association
that was used in the aforementioned study (see Table 1). The scheme is based on research literature on collaborative dialogues in education and distinguishes between four levels of verbalisations that represent an increasing depth of discussion of game features from one level to the next. First-level verbalisations refer to visible aspects of gameplay. Second-level verbalisations deal with actions that players are contemplating. Third-level verbalisations involve expectations of effects of game moves. Fourth-level verbalisations refer to game strategies. Each dialogic act was assigned to one of the four categories. A set of 20 randomly selected transcribed dialogic acts were coded by all three authors. This resulted in moderate to substantial interrater agreement scores (ie. Kappa scores of 0.59, 0.60 and 0.66). Inspection revealed that differences in coding mainly occurred between the two lower levels one and two, or between the two higher levels three and four. Subsequently, one of the authors coded all dialogic acts.

**Game experience questionnaire**

This questionnaire assessed the participants’ prior experience with games. It consisted of three closed-ended questions concerning time spent on gaming in general, experience with “Enercities,” and time spent on strategy games (ie. “In the last month, in an average week, how many hours did you spend on playing strategy games like The Sims, SimCity, Civilization?”). Answers were given in pre-determined categories for ranges of hours or experience.

**Learning style questionnaire**

This questionnaire assessed the participants’ preferences for a visual, auditory, reading/writing or kinaesthetic learning style. It was based on Fleming’s VARK model (see Dobson, 2009). The questionnaire consisted of 16 multiple-choice questions. The reported reliability (correlation) scores for the four styles were all above .75. Just as for game experience, this questionnaire was administered as a check on the random distribution of participants. Empirical research on the VARK

---

**Table 1: Coding scheme for the dialogic activities**

<table>
<thead>
<tr>
<th>Activity</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourth-level verbalisations</td>
<td>Indicating having seen or done something before Accounting for and reasoning out ideas</td>
<td>Just like in a real city Building a factory will bring us more money but it’s not good for the environment You know what I would have done. Put these energy things on the other side of the river. Then people will not be bothered</td>
</tr>
<tr>
<td>Third-level verbalisations</td>
<td>Forecasting future situations and contemplating effects.</td>
<td>If we do this, our environment is likely to become better I don’t think that they want to live there, if you build industry</td>
</tr>
<tr>
<td>Second-level verbalisations</td>
<td>Suggesting what actions to take, or responding to such a suggestion.</td>
<td>We are going to buy light industry We could build windmills at sea Eco-roofs are also possible. Shall we do that?</td>
</tr>
<tr>
<td>First-level verbalisations</td>
<td>Telling what happened in an event Inquiring about a feature of the interface</td>
<td>We have no money Look, this has increased What is this, forest?</td>
</tr>
</tbody>
</table>

---

© 2019 The Authors. British Journal of Educational Technology published by John Wiley & Sons Ltd on behalf of British Educational Research Association
questionnaire suggests that it has fair to good construct validity, but poor to moderate predictive validity (eg, Alzain, Clark, Ireson, & Jwaid, 2018; Fitkov-Norris & Yeghiazarian, 2015; Meyer, Stomski, Innes, & Armson, 2016). The present study did not aim to partake in the debate on the importance of learning styles for teaching and (adaptive) learning (see Newton & Miah, 2017). It merely assessed learning style as a control measure for random allocation of this trait to conditions.

Motivation questionnaire
This questionnaire was based on the FAM instrument (Rheinberg, Vollmeyer, & Burns, 2001; Vollmeyer & Rheinberg, 2006). First, students were informed that they would be engaged in playing a computer game in pairs. Next, each student was asked to answer 20 questions covering four motivational constructs: interest (eg, “I would work on this task even in my free time.”), probability of success (eg, “I think I am up to the difficulty of this task.”), anxiety (eg, “I am afraid I will make a fool out of myself.”), and challenge (eg, “If I can do this task, I will feel proud of myself.”). Answers were given on a 7-point Likert scale with values ranging from “not true” (1) to “true” (7). The same motivation questionnaire was administered before and after gameplay. Reliability analyses yielded satisfactory Cronbach alpha values only for interest (before = .83; after = .86) and challenge (before .70; after .90). Therefore, only the findings for these constructs will be reported.

Knowledge test
A knowledge test measured knowledge of the advantages and disadvantages of resources in certain situations, and knowledge of solutions for “problematic” situations. For instance, one item asked the students to mention three possibilities for increasing the city’s energy supply. Another item displayed a fictive game situation for which the students were asked to give arguments whether they would, or would not, consider building rooftop windmills. Another item presented a problematic situation in which the available energy supply was high but the ecological score and the score for well-being were low. Students were asked to analyse the situation and to come up with possible solutions. There were 2 closed-ended and 10 open-ended items. The closed-ended items first asked students to select one or more alternatives, and then to give arguments to support their choice(s). Scores that varied between 0 and 3 points could be obtained for the 12 items. The maximum score was 25. A codebook was used for scoring the answers. Test scores were converted to a percentage of possible points. Reliability analyses yielded a satisfactory Cronbach’s alpha value of .68.

Procedure
At the start of the experiment, participants were told that they would play a game in pairs. Then they completed the questionnaires about game experience, learning style and motivation. Thereafter, the pairs were formed and they were instructed to play the Enercities game. In the scripted collaboration, pairs were informed about the role they should take on during the game. The participants were not informed that after gameplay, a knowledge test would follow. Next, the game was played. During gameplay (25 minutes), the conversations were recorded. After gameplay, participants again completed the motivation questionnaire, and then took the knowledge test, which lasted 20 minutes. Participants were debriefed after completing this test.

Data analysis
Checks on random distribution of scores for game experience, learning style and initial motivation revealed no significant differences between conditions. There were also no homogeneity violations and therefore data testing was done with AN(C)OVAs. Comparisons between conditions for dialogic acts and knowledge used a one-tailed test with alpha set at 0.10. A Bonferroni correction for repeated testing of dialogic acts led to an alpha of 0.025 (0.10/4) for these scores. Comparisons for motivation were two-tailed with alpha set at 0.05. For effect sizes, the $d$-statistic
is reported (Field, 2013). This statistic tends to be qualified as small, medium and large for the values \(d = 0.20\), \(d = 0.50\) and \(d = 0.80\), respectively.

**Results**

**Dialogic acts**

Tables 2 and 3 show the absolute and relative numbers of dialogic acts per level and condition. The difference between conditions for the mean total of dialogic acts was not statistically significant, \(F(1,15) = 3.74, p = .074\). The overall findings indicate that the large majority of expressions involved first- or second-level dialogic acts. That is, they revolved around interface objects and descriptions of screen states or events. Dialogic acts on the third or fourth level were less frequent.

There was no difference between conditions for the proportion of first-level dialogic acts, \(F(1,15) = 0.01, p = .99\). In contrast, a statistically significant and substantial difference between conditions for the proportion of second-level dialogic acts was found, \(F(1,15) = 19.36, p = 0.001, d = 2.20\). Table 3 shows that these acts occurred relatively more frequently in the control condition. Exactly the opposite pattern was found for the third- and fourth-level dialogic acts. There was a statistically significant and substantial difference between conditions for the proportion of third-level dialogic acts, \(F(1,15) = 8.27, p = .012, d = 1.43\), as well as of fourth-level dialogic acts, \(F(1,15) = 28.82, p < .001, d = 2.68\). These dialogic acts occurred relatively more frequently in the scripted condition.

Tables 4 and 5 show the spread of the dialogic acts across the four communication levels for the pairs in each condition. These data indicate that third- and fourth-level communications occurred relatively more frequently in all pairs in the scripted condition. They also reveal that, with one exception, all pairs in the scripted condition had a higher percentage of fourth-level dialogic acts than the pairs in the control condition.

**Motivation**

Table 6 shows the findings for interest and challenge. At both measurement points, and for both constructs, the scores were slightly above the neutral scale value. This indicates that the participants started and ended with neutral to positive appraisals for these motivational constructs. An ANCOVA with prior interest as covariate showed no difference between conditions for interest.

| Table 2: Mean totals (standard deviations) of dialogic acts per level and condition |
|----------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| First level                      | Second level    | Third level     | Fourth level    | Total           |
| M (SD)                           | M (SD)          | M (SD)          | M (SD)          | M (SD)          |
| Scripted (n = 8)                 |                 |                 |                 |                 |
| 396 (18.1)                       | 481 (28.0)      | 131 (5.6)       | 139 (10.0)      | 1147 (52.0)     |
| Control (n = 8)                  |                 |                 |                 |                 |
| 403 (20.7)                       | 669 (42.3)      | 87 (7.2)        | 46 (5.8)        | 1205 (70.9)     |

| Table 3: Proportions (standard deviations) of dialogic acts per level and condition |
|----------------------------------|-----------------|-----------------|----------------|
| First level                      | Second level    | Third level     | Fourth level   |
| M (SD)                           | M (SD)          | M (SD)          | M (SD)         |
| Scripted (n = 8)                 |                 |                 |                 |
| 35.1 (6.7)                       | 40.6 (7.6)      | 12.2 (3.7)      | 12.1 (4.1)     |
| Control (n = 8)                  |                 |                 |                 |
| 35.1 (6.1)                       | 55.1 (5.3)      | 6.6 (3.9)       | 3.2 (2.4)      |
| Total (n = 16)                   |                 |                 |                 |
| 35.1 (6.2)                       | 47.9 (9.8)      | 9.4 (4.7)       | 7.6 (5.6)      |
after gameplay, $F(1,31) = 0.08, p = .78$. Likewise, an ANCOVA with prior challenge as covariate showed no difference between conditions for challenge after gameplay, $F(1,31) = 2.33, p = .14$.

### Learning outcomes

The control condition achieved a mean score of 49.5% ($SD = 15.1$) on the knowledge test, while the scripted condition had a mean score of 61.3% ($SD = 20.0$). An ANOVA showed that this difference was statistically significant, $F(1,31) = 3.52, p = .035$. The effect size statistic pointed to a moderate effect, $d = 0.66$.

An exploratory analysis of the relationships between dialogic acts and knowledge test score yielded low and slightly negative correlations with first-level dialogic acts ($r = -.26, p = .15$) and second-level dialogic acts ($r = -.20, p = .27$). There was a non-significant positive relationship with third-level dialogic acts ($r = .20, p = .28$). The relationship between knowledge test score and fourth-level dialogic acts was positive and statistically significant ($r = .47, p = .006$).

© 2019 The Authors. British Journal of Educational Technology published by John Wiley & Sons Ltd on behalf of British Educational Research Association
Discussion and conclusion
The aim of the study was to investigate whether scripting a collaboration between players of a serious game would affect the quality of the dialogues, motivation and learning outcomes.

Dialogic acts
A very large majority, 90.2%, of the dialogic acts in the control condition consisted of first- and second-level verbalisations. This outcome is comparable to the score of 88% found for these verbalisations by van der Meij, Albers and Leemkuil (2011). Likewise, the proportions of third- and fourth-level verbalisations in the control group were comparable across studies.

The issue can be raised whether the communications revolved more around game strategies and tactics than learning content (see Frank, 2012). When we looked into this matter, we discovered that it was surprisingly hard to find unequivocal examples of “gaming the game” expressions. Only occasionally did talk about (potential) points gained or lost for certain actions provide clear tell-tales of gaming-the-game dialogic acts (eg, “It costs 100 and we have only 108, so we’re not going to do that,” “Victory points are important, they help you raise your score” and “Gee nearly 200”). In many more instances, gaming-the-game expressions coincided with relevant observations about game content. That is, there were many instances in which a game feature, or score, seemed to prompt a content-related reflection (eg, “See how expensive setting up a solar central is,” “We overspent, now we cannot do much” and “This decreases extremely. Actually we should move this further away from the city”). In short, the question about frequency of gaming-the-game expressions is complex but certainly worthwhile further consideration. However, the findings for learning outcomes, especially in the experimental condition, speak against a preponderance of gaming-the-game communications.

Another important perspective to consider in extending the coding of the dialogues is the ICAP framework. According to Chi, Kang, and Yaghmourian (2017), dialogues can be distinguished as interactive, constructive, active or passive acts of which the first yields the most learning. In interactive activities both dialogic partners make substantive contributions on the same issue, and react to each other’s contribution. The coding scheme in the present study merely recorded the participants’ constructive processes (eg, self-explanations). It did not measure whether the dialogues were interactive. Future studies on collaborative gameplay should probably also address this feature of the communication between pairs.

The script did not affect the amount of talking. The pairs in the two conditions were talking about the same amount, but about different things. It was not that the pairs in the scripted condition had to work through the same first- and second-level stuff as the non-scripted condition and then add in the argumentation that came with having opposing viewpoints being represented. If anything, the pairs in the scripted condition talked less than those in the control condition. However, the distribution pattern for the dialogic acts in the scripted group sharply contrasted with that in the control group. As expected, the conflict script significantly and substantially raised the level of the dialogues. Third- and fourth-level verbalisations were more frequent when students took on opposing roles during gameplay. In short, the script helped raise the quality of the dialogues.

Motivation
An important design challenge for scripted collaboration is that it should not compromise motivation during gameplay. The script must give players the freedom to communicate in a manner that is not severely restricted. It should avoid overscripting, in which case the communications become too constrained by imposed rules (Clark et al., 2016; Dillenbourg & Jermann, 2007).
The data showed that the participants’ appraisals for game interest and challenge was virtually unaffected by their gameplay during the study, and, more importantly, there was no effect of condition. Slightly positive outcomes for interest and challenge were found before and after gameplay, in the control as well as the scripted conditions.

Learning outcomes
The findings on the knowledge test showed that there was a significant, moderate effect of condition. As predicted, participants in the scripted group scored higher on this test. One possible reason is that the conflict script externalised the trade-off between ecology and economy that is key to being successful in the Enercities game and for being successful in the knowledge test as well. In other words, the script did not add another task to the playing of the game, but merely highlighted a challenge that was already part of the game.

In a recent anthology, Arnseth, Hanghøj and Silseth (2018) qualified scripting as a promising intervention for raising the level of the discussions in collaborative gameplay and thereby enhancing learning. The present study provides tentative support for this line of reasoning. Not only was there an effect of scripting on the knowledge test administered after gameplay, but the correlational findings also pointed to a connection between the level of the dialogic acts and learning outcomes.

It also seems valuable to create and investigate collaboration scripts that stimulate co-construction of knowledge. Just as was intended in this study, such a script should stimulate partners to engage in communications that are vital to game understanding and knowledge development, without being overscripted. The jigsaw method can be a possible basis for constructing such a script for game-based-learning. The main tenet of jigsaw is that a topic, usually a problem, is divided into parts that are worked on by different students whose ideas must be combined to complete the view (Aronson & Patnoe, 2011). Thus, students are given a responsibility for, and develop expertise in, a subtopic. To successfully solve the problem, they must then share and communicate their acquired knowledge with the other team members. The jigsaw method has been found beneficial for students’ actions, and for learning (eg, Karacop & Doymus, 2013; Walker & Crogan, 1998). But it has also been found that complementary measures are needed to strengthen positive interdependence (eg, Deiglmayr & Schalk, 2015; Roseth, Lee, & Saltarelli, 2019). In other words, the design of a suitable jigsaw-like, or other, co-construction script for game-based learning is not an easy matter. The present study is a small step towards creating such a script for effective collaborative gameplay.

Acknowledgement
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Statements on open data, ethics, and conflict of interest
The data can be requested from the corresponding author.

The study complied with the Ethical Standards of the University of Twente. A research proposal describing the study was approved by the ethics committee of the University. Participants gave informed consent.

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

© 2019 The Authors. British Journal of Educational Technology published by John Wiley & Sons Ltd on behalf of British Educational Research Association


