Collaboration scripts for mastership skills

Hummel, Hans; Geerts, Walter; Slootmaker, Aad; Kuipers, Derek; Westera, Wim

Published in:
Interactive Learning Environments

DOI:
10.1080/10494820.2013.789063

Interactive Learning Environments, 2015
Vol. 23, No. 6, 670–682, http://dx.doi.org/10.1080/10494820.2013.789063

Collaboration scripts for mastership skills: online game about classroom dilemmas in teacher education

Hans Hummel[a,b]*, Walter Geerts[b], Aad Slootmaker[a], Derek Kuipers[b] and Wim Westera[a]

[a]*CELSTEC (Centre for Learning Sciences and Technologies), Open University, Valkenburgerweg 177, Heerlen, 6419 AT, The Netherlands;* [b]*IEC (Institute of Education and Communication), NHL University of Applied Science, Rengerslaan 10, Leeuwarden, 8917 DD, The Netherlands*

(*Received 3 September 2012; final version received 12 March 2013*)

Serious games are seen to hold potential to facilitate workplace learning in a more dynamic and flexible way. This article describes an empirical study into the feasibility of an online collaboration game that facilitates teachers-in-training to deal with classroom management dilemmas. A script to support these students in carrying out such practical tasks, independent of teacher intervention and in collaboration with peers, was designed and worked out in both a face-to-face and an online version of a 'mastership' game. After assigning and discussing practical dilemmas during a small group play session, solutions are worked out individually in the form of small advisory reports, and assessed by both teachers and peers (co-players in the group). Learning effects were measured and satisfaction was questioned for 9 players that played the online version and 10 players that played the face-to-face version of the game. Results show that the collaboration of students on classroom dilemmas can indeed be successfully facilitated by this script, and that learning results do not differ for both versions. The latter holds potential for offering online and more flexible ways of workplace learning. Especially, students playing the online version reported needs for simpler structure and clearer task instruction. The optimal level of structure in collaboration scripts, therefore, appears an issue for further study.

Keywords: serious games; scripted collaboration; classroom dilemmas; teacher education

1. Collaboration and serious games for workplace learning

Serious games are games that can educate, train or inform, either because they have been deliberately designed for learning or just happen to do so by coincidence. Educators call them 'serious' to denote that they are *not just* fun to play, but *also* hold potential as cognitive tools for learning and professional development (Michael & Chen, 2006; Prensky, 2006). Over the last decade it has been assumed that serious games offer many new learning opportunities and positive effects on learner motivation and learning outcomes, although empirical evidence has remained scarce so far (Kiili, 2007; Shaffer, 2006). Serious games not only support individual learning but can also mediate soft skills such as collaboration, reflection and social communication about wicked problems that are usually not addressed by other learning platforms (Gee, 2003).

*Corresponding author. Email: hans.hummel@ou.nl

This introductory section will describe how serious games may facilitate professional workplace learning, especially when collaboration on practical problems is involved. The next section then explains what is the role of scripting in such online collaboration games. Section 3 will further focus on mastership skills and describe the game that lies under study here (i.e. teachers learning to deal with classroom dilemmas). Sections that follow describe the exploratory study we carried out with a small group of higher vocational education students. Half the group played a face-to-face and half played an online version of the mastership game, discussing and collaborating on classroom dilemmas in small groups of four or five. Method (Section 4) and results of this study are presented (Section 5) and discussed, together with suggestions for future research (Section 6).

Workplace learning is no longer restricted to individuals acquiring or updating domain knowledge, but also has to deal with selecting and using this knowledge for certain problem situations in daily practice. Such learning is about acquiring competences like information skills and media literacy, problem solving, communication and collaboration skills, and above all critical reflection. Today's professionals will become lifelong learners that continuously have to face problem situations that are changing dynamically and rapidly. Furthermore, organizations' tacit knowledge plays a crucial role in solving their problems but such knowledge can only be expressed and accessed in direct collaboration on authentic tasks.

Professional tasks, including collaboration, argumentation and negotiation are crucial for vocational education, especially when they aim to connect school knowledge to practical work. Serious collaboration games are considered to hold potential as more open, dynamic and flexible learning environments where such professional skills could be acquired through self-determined learning with little or none direct teacher intervention. We are interested to find out to which extent this holds true in educational practice, and what kind of learner support these learning environments then should contain to facilitate this kind of learning. Collaboration on practical problems or dilemmas can be *about* the game (and take place in a face-to-face context) or be an integral *part of* the online game (and take place in a virtual context) as well. In the latter case, such collaboration has to be enabled by the 'didactic script' or game play, which in this article will be denoted by 'scripted collaboration'. Collaboration (argumentation or negotiation) scripts have been scarcely implemented within educational games so far. They will use the situated context (or authentic case) to have learners' access tacit knowledge by sharing and co-creating new knowledge together (Bell, Kanar, & Kozlowski, 2008).

2. Collaboration scripts in serious games

Games are heavily inspired by learning-by-doing and experiential learning principles which hold tremendous potential for contextualized workplace learning. The way players move, progress and navigate within serious games to a large degree will depend on their self-explorative way-finding behaviour and will often have a substantial trial-and-error component. Without any structure or cueing, this is unlikely to lead to effective learning. Which moves are evoked, how much erroneous or meaningful learning takes place, will depend on the learner support that is provided, shared and distributed in the gaming environment. Learner support helps students select most useful information, compare and reflect on multiple perspectives of others, and monitor task progress and quality of learning output. This article examines the extent to which collaboration scripts can provide such learner support for workplace learning, and as a consequence improve learning efficiency and output. Collaboration scripts (Kobbe et al., 2007) are an instructional method that structures the collaboration by guiding the interacting partners through a sequence of interaction phases with designated activities and roles. Optimising the type and amount of structured collaboration are key issues of research here.

Scripting collaboration was examined in computer-supported collaborative learning (CSCL) (Gunawardena, Lowe, & Anderson, 1997; Van Bruggen, Kirschner, & Jochems, 2002). CSCL environments have shown to positively influence learning (Gunawardena, Carabajal, & Lowe, 2001). The interaction in CSCL between learners can lead to further elaboration and refinement of individually constructed schemata, since it (a) incites learners to make explicit the actual level of schema development and (b) demands them to explicitly compare their own schemata with schemata of others as to defend or criticize (Jeong & Chi, 2000). Dillenbourg and Jermann (2006) describe different core scripts, Hernandez Leo (2007) explored the potential of educational modeling languages to describe and implement particular collaboration scripts, and Kobbe et al. (2007) proposed a framework to describe collaboration scripts.

However, such collaboration scripts have hardly ever been implemented and tested in more open learning environments like serious games. No research has focused on the structural aspects nor has measured the learning effects of including such scripts in serious game play. We see these games as a platform for both playful and motivating learning methods. Collaboration scripts in serious games may provide adequate learner support by cueing social processes (elaboration, explanation, argumentation, and question asking) that might otherwise not occur. Students are more likely to explore relevant perspectives than in unstructured collaboration.

Dillenbourg (2002) mentions that a specification of how students should collaborate and solve the problem in an online environment requires five attributes: definition of the task, composition of the group, distribution of the task, the mode of interaction, and the timing of the phases. Especially, the distribution of activity is essential for these scripts. For instance, in a 'reciprocal teaching' script one player has to read and understand while another player has to monitor the other's understanding by asking questions and give feedback. Regarding the mode of interaction scripts may vary according to the degree of freedom learners are allowed in following the script. A low degree of 'coercion' in 'induced scripts' is elegant but often not sufficient to influence the collaborative processes. 'Prompted' or 'follow me' scripts have higher levels of 'coercion' and will steer the collaborative processes, at the risk of being perceived as too complex or rigid. Indeed, Dillenbourg reports a number of scripts (e.g. UniverSanté) that were perceived as too complex, and advises that scripts should be kept as simple as possible so that all actors are able to appropriate them.

3. Mastership game: collaboration script for classroom dilemmas

The cases under study here deal with (multiple perspectives on) classroom management dilemmas. What should a teacher do, for instance, when a pupil continues to disturb the lesson by insulting his peers. Should the problem be resolved during the lesson, even at the risk of loosing valuable time to the expense of the majority of students that is not involved in the conflict. Or should the problem be resolved after the class has been dismissed, even at the risk that disturbances will continue during the lesson. Teaching can be considered to be an exciting game. As a teacher without doubt you will have to face unexpected situations that demand you to find solutions on the fly. The NHL University of Applied Science offers a broad range of teacher training programs, and its Didiclass project set out to develop (video)cases for teachers-in-training to link the professional theory acquired to actual practice in the workplace (i.e. the actual classroom)

(Geerts, Mitzsche, & Van Laeken, 2009). In the context of this Didiclass project, some experienced didactics teachers developed the mastership game which helps students to find solutions to the some of the most prevailing practical classroom management dilemmas in a playful way that will help them become better teachers. It can be assumed that collaborating on problems first will later increase their ‘professional productivity’, simply because exchanging information and looking from various perspectives will increase the quality of the individual solutions, as shown by some CSCL studies mentioned in the previous section (Gunawardena et al., 2001; Jeong & Chi, 2000). The mastership game became available in a face-to-face version in 2010; an online version was developed to become operational in 2012, a version inspired by the same collaboration script (Figure 1).

The mastership game can be played in small groups of two till six students and does not require any moderation or other intervention by teachers. After selecting their avatars, they start group play both in the role of player (or problem owner) and of co-player (judging the way that players solve their problems). The game has a structure that consists of five consecutive phases, during which players discuss, elaborate and negotiate solutions to solve the problem. Communication is structured by various assignments and rules during these phases, but is possible by unstructured group chat as well. During the *first phase* players select 3 practical dilemmas, either out of a pile of 24, most prevailing practical classroom dilemmas (i.e. ‘How to maintain control in a good way’, ‘How to deal with negative colleagues’, or ‘How to deal with a pupil that does not want to get coached’), or by formulating

Figure 1. Screens of the online version of the mastership game: selecting three practical dilemmas in phase 1 (upper left hand), assigning and motivating themes in phase 3 (upper right hand), motivating and discussing declined themes in phase 4 (lower left hand), and peer assessment of elaborated assignments in phase 6 (lower right hand).

one of their own. Then, each player selects the problem that is considered most important. During the *second phase* players draw an exploratory assignment (e.g. 'Provide an exemplary experience that shows why this problem is important for you'). The elaboration is judged by the co-players until the group is satisfied. During the *third phase* players take turns in drawing theme cards (e.g. 'professional development', 'dealing with losses', or 'lesson preparation') that are placed at their co-players while motivating why this theme should be further explored in combination with the chosen dilemma, until every player has received three theme cards. In the *fourth phase*, players will negotiate and discuss which theme cards may be declined. Co-players may use jury cards and ask further questions to challenge players to further motivate their declined cards before the group agrees on the final selection. During the *fifth phase*, players select a practical assignment to further elaborate a solution for the problem in a short advisory report.

(Sections 4.1 and 4.3 will describe how game play was organized and assessed for both versions.)

The main hypotheses (research questions) to be answered in the next sections are fourfold: (1) it is feasible to implement collaboration scripts in small group play to support students in finding solutions to wicked professional problems, without further teacher intervention or guidance, while yielding sufficient learning outcomes; (2) it is possible to implement such a collaboration script in an online environment, without decreasing learning outcomes; (3) students will appreciate the collaboration set-up within such game play; and (4) it is possible to have students profit from online support and interaction, without decreasing user satisfaction. In other words, this article addresses the effectiveness of and satisfaction with collaboration within serious gaming.

4. Method

This section will consist of consecutive paragraphs describing the way we have setup our exploratory study. After describing the participants and the learning materials we used, we describe the procedure and assessment instruments we used to measure the effectiveness of and the satisfaction with the scripted collaboration in both conditions.

4.1 *Participants*

Nineteen teacher education trainees, third year students of the NHL University of Applied Science in the Netherlands, participated in this case study. The mastership game has a study load of about 10 hours and is awarded by half a European Credit point as part of workplace learning during their third year of curriculum. Participants are qualifying for a broad variety of first degree teaching positions, ranging from modern languages teaching, teaching didactics to science teaching. All had comparable prior knowledge since all were in the third year of their curriculum. Most students follow education in combination with work (as a dual or flexible learning trajectory) which explains the relatively high average age of 41.4 years (SD = 7.22), ranging between 29 and 52. Ten were male and nine were female students, equally divided over the four small groups. Gender and age showed to be neither related to learning outcomes nor satisfaction.

4.2 *Learning materials*

The face-to-face version of the 'mastership' game is delivered in an old-fashioned VHS box containing 66 cards: 24 cards containing practical dilemmas; 13 cards containing exploratory assignments; 10 cards containing themes, with 10 jury cards containing questions per theme; 8 cards containing the final assignments; and 1 card containing the instructions and rules for playing the game. The card version normally will be played by two till six players in a room at the university with a teacher being present, and take about two hours to play. The five phases involved and various types of cards were explained in the previous section. We developed the online version of the 'mastership' game using the ZK toolkit (http://www.zkoss.org). The collaboration script was implemented by applying the same phases, rules, and cards. However, we have to mention that for functional reasons an exact, one-on-one transformation of the face-to-face situation to an online situation was considered neither wise nor practical by the developers (the authors of this article), so some minor design adjustments were made without changing the initial structure or rules for collaboration.

The EMERGO toolkit for serious game development (Nadolski et al., 2008) was used for game run management and for storing and analyzing data. The toolkit is built in Java and the collaboration script described above was implemented as a separate Java component within the toolkit. The component is designed and built in such a way that it can later be reused and extended for other scripts and cases applying a similar collaboration pattern in their design. Game logic is neatly separated from the rest of the code in so-called GameScene classes. The game can be configured by a game author on several aspects like structure.

4.3 *Procedure*

Students participated in this study as part of their regular curriculum. They were approached by their teacher (being one of the authors of this article) and invited to be present at a certain place and time at the university for a two-hour meeting. Participants were notified in advance that this meeting would also be used for study purposes and randomly allocated to both versions of the game, with each condition containing two groups. The online participants received an e-mail before the meeting, containing the URL and their account. After the meeting all participants received a questionnaire about their appreciation of the game by e-mail. During the time of the meeting, face-to-face students went to a room to collaborate with their group members, and online students went to a computer room to independently work together online. A teacher was present in this computer room to control for any direct (non)verbal communication beyond the program.

During regular education, the fifth phase would be the final phase and outcome of the small group play. Students then elaborate and deliver their reports individually, and get graded by their teacher. For the purpose of this study, we included a *sixth* and final *phase* in which students had to grade the reports of their peers, in order to enable a comparison of the assessments by peers (co-players) and teachers. To be able to establish inter-rater reliability of the learning effect correction model developed for this study, student reports were independently assessed by two teachers. It was estimated that the elaboration of the reports would take about half a day. Because the study took place during an end-of-year exams period students were allowed two weeks to deliver their report and questionnaire, and to grade the reports of the peers in their group (by awarding one to five stars). Also, due to some technical problems, we faced in phase 6, online students were allowed to deliver and grade reports either online or by mail, like face-to-face students had to do. Other than in phase 6, students were able to pass through phases without technical problems. All data could be collected two weeks after the meeting took place.

4.4 *Learning effect correction model*

To measure individual learning output, the quality of the solutions provided for the classroom dilemmas was assessed by using a learning effect correction model, that was developed for this study by the teacher/topic expert (being one of the authors of this article). The elaborated reports were assessed on 'growth in professional productivity'. The five criteria to establish this growth were inspired by the development of 'design practice' (or practical theory) (Copeland & D'Emidio-Caston, 1998): (A) ownership (to what extent does student commit to solve this problem); (B) reflection (to what extent does student reflect on his own actions); (C) focus (to what extent does student attach the right amount of context to the problem); (D) nuance/complexity (to what extent is applying the solution feasible); and (E) richness/correctness (of the elaborated solution). Table 1 contains indications for the possible scores on these criteria, with total scores ranging from 0 to 10.

4.5 *Student satisfaction questionnaire*

The student satisfaction questionnaire was developed for this study by a learning technology expert (being one of the authors of this article). It contains 19 items to establish the students' appreciation of various game aspects, pertaining to the structure (S, five questions), user-friendliness and clarity (U, five questions), the timing of the phases (T, two questions), the quality of the dilemmas and assignments (Q, five questions), and the interaction during collaboration (I, two questions). The focus on structure and clarity of instruction was inspired by studies (Dillenbourg, 2002; Dillenbourg & Hong, 2008) showing this often to be problematic. Even when structure and clarity of the script (the logistics) are perfect, game play will lead to nowhere when the quality of assignments, players, information exchanged (the content) is of poor quality; this is why we added some items to check for this. (Clearly, the letters referring to these five aspects and +/− signs referring to the positive/negative formulation in Table 2 were not listed in the original questionnaire.) All these items used a Likert scale with 5 values, ranging from (1) fully do not agree to (5) fully do agree. The median value (neutral), therefore, is 3.0. Depending on the positive (+)/negative (−) formulation of the items, values below can be interpreted as (slightly) negative/positive and all values above as (slightly) positive/negative appreciations. Item

Table 1. Sub-scales and scoring categories of the learning effect correction model.

Sub-scales	Insufficient (0 points)	Sufficient (one point)	Good (two points)	Score
A. Ownership	Refers to others: 'They will solve the problem'	'I will take action'	The answer shows real commitment.	0–2
B. Reflection	No reflection	Some reflection, partly rich	Rich reflection	0–2
C. Focus	The problem has not been framed/ focused	The problem has partly been focused	The problem is rich and has been correctly focused	0–2
D. Nuance/ complexity	The answer does not contain nuance	The answer is correctly linked to one design pattern	The answer is correctly linked to (a network of) more design patterns	0–2
E. Richness/ correctness	The elaboration is not correct	The elaboration is partly rich and correct	The elaboration is rich and correct	0–2
Total score				0–10

Table 2. Items of the satisfaction questionnaire.

Item	Aspect	+/–	Statement
1	U	+	The way to play the game is clear, playing rules are clear
2	Q	+	The elaborations (of practical assignments) by co-players were of sufficient quality
3	Q	+	The composition of the group was good (regarding interest and level of expertise)
4	U	+	The user-interface of the game is clear and user-friendly
5	S	+	Group play was possible without teacher intervention, the collaboration process has been determined well in advance
6	T	–	The time allowed to play was too low
7	S	–	The amount of game structure is too low
8	U	+	The time allowed for each phase was too low
9	S	–	The amount of structure in each phase is too high
10	T	–	The time allowed for each phase was too high
11	U	–	The way to collaborate during each phase was too complex
12	I	+	Mutual interaction and collaboration proceeded well and were useful
13	Q	+	Feedback (assigning cards, peer assessment, etc.) from co-players was useful (in further elaborating my assignment)
14	Q	+	The elaborations of the exploratory assignments by co-players were of sufficient quality
15	S	+	Using jury cards was useful and proceeded well
16	U	+	Collaboration rules (for peer assessment, taking turns, when to proceed to next phase, etc.) were clear
17	S	–	Mutual dependency during collaboration (awaiting feedback, taking turns, etc.) was too high
18	Q	+	The elaborations of the final assignments by co-players were of sufficient quality
19	I	+	It was a fun and effective way to play the mastership game

20 was an open question, allowing room for comments and suggestions. Table 2 contains the list of items.

5. Results

This section provides answers to the fourfold research question we posed at the end of the third section: (1) is it feasible to implement collaboration scripts in small group play to support students in finding solutions to wicked professional problems, without further teacher intervention, while yielding sufficient learning outcomes? (2) Is it is possible to implement such a collaboration script in an online environment (i.e. serious game) without decreasing learning outcomes? (3) Will students appreciate the collaboration set-up within such game play? And (4) is it possible to have students profit from online support and interaction without decreasing user satisfaction? We will now, respectively, present the objective learning effect measures (answering the first two sub-questions) and the subjective questionnaire measures (answering the last two sub-questions). Finally, we will present some more qualitative impressions that shed more light on these appreciations.

5.1 *Learning effect measures*

First of all, we found that most participants' reports could be graded as sufficient, which can be considered to be an overall success of playing the 'mastership' game in both conditions.

Grades below 6.0 were considered not sufficient, and only four students received either a 5.0 or 5.5, with these unsufficient grades being equally divided over both conditions. The average grade (also combined for both raters) for all participants was M = 7.00, SD = 1.32. We, therefore, *accept our first hypothesis* with a positive answer to the first sub-question: yes, it is possible to implement collaboration in small group play without teacher intervention but with sufficient learning effect.

When we compare the grades for both conditions, we do not find much difference in learning effect between the online condition (M = 6.94, SD = 1.33) and the face-to-face group (M = 7.05, SD = 1.38). One-factor analysis of variance (ANOVA) shows this difference to be only slightly in favour of the face-to-face condition, but being by far not significant (F (1,17) = 0.29, p = 0.868, $\eta_p^2 = 0.002$), with values of the partial-eta-squared below 0.01 showing neglectable effect size according to Cohen (1988). Furthermore, the average and normalized ratings awarded by peers (co-players) appear to be highly correlated to the grades awarded by teachers (r = 0.715, p = 0.001, two-tailed). Average ratings by peers overall are somewhat higher then teacher grades for both the online condition (M = 7.42, p = 0.87) and face-to-face condition (M = 7.30, p = 1.29), and in contrast slightly in favour of the online condition, but this difference again by far is not significant (F (1,17) = 0.57, p = 0.814, $\eta_p^2 = 0.003$). We, therefore, *accept our second hypothesis* with a positive answer to the second sub-question: yes, it is possible to implement collaboration in an online environment without decreasing learning outcomes.

The reliability of the learning effect correction model was determined by applying Cohen's Kappa for inter-rater reliability (with k = 2). For the total instrument, the standard (and rather strict) Kappa measure appeared poor (K = 0.16, p = 0.124, δ = 0.13). Closer inspection on the level of the five sub-scales of the instrument revealed that inter-rater reliability was moderate to almost perfect for four out of five sub-scales, but that there was no agreement on sub-scale B. Reflection (K = −0.01), and that excluding this sub-scale would immediately increase the overall Kappa to moderate (K = 0.47) and acceptable. Closer inspection of the differences in scores between both raters revealed that rater one consistently awarded higher grades (M = 7.42, p = 1.22) then rater two (M = 6.63, p = 1.53), which appeared to be done in the same manner over all sub-scales. Since Kappa does not take into account degrees of disagreement between observers (all disagreements are considered as total disagreement) when having ordered categories, we decided it would be preferable to use a weighted Kappa. Since the difference between the first and second categories had the same importance as the difference between the second and third categories (0, 1, and 2 were the scoring categories in each sub-scale), we used linear weighting. The weighted Kappa for the total instrument appeared moderate (K_w = 0.46, δ = 0.09), and when excluding sub-scale B appeared even good (K_w = 0.66, δ = 0.09). We, therefore, decided it was very acceptable to use the original instrument, even without correction for the lack of agreement on one sub-scale. Kappa's between 0 and 20 are considered 'poor' or 'light', between 20 and 40 as 'fair', between 40 and 60 as 'moderate', between 60 and 80 as 'substantial' or 'good', and between 80 and 100 as 'almost perfect' or 'very good' (Heuvelmans & Sanders, 1993, p. 450).

5.2 *Satisfaction measures*

Table 3 presents the average scores on all items of the questionnaire for both conditions. Considering the formulation of items (listed in Table 2), we can observe that over both conditions most aspects have been valued as slightly positive (above neutral), with the exception of items 4, 8, 11, and 17. This means that the interface (item 4), information about

Table 3. Average scores on the satisfaction questionnaire items for both versions.

	Online (n = 9)		Face-to-face (n = 10)		All (N = 19)		
Item	*M*	SD	*M*	SD	*M*	SD	$p\Delta$
1	2.44	1.33	4.20	0.92	3.37	1.42	0.004
2	3.11	0.93	4.30	1.06	3.74	1.15	0.019
3	3.89	1.17	4.60	0.97	4.26	1.09	0.163
4	1.78	0.97	4.00	0.82	2.95	1.43	0.000
5	2.00	1.00	4.10	0.99	3.11	1.45	0.000
6	3.78	1.56	2.00	1.15	2.84	1.61	0.011
7	2.67	1.22	1.90	0.88	2.26	1.09	0.132
8	2.67	1.58	2.80	1.62	2.74	1.56	0.858
9	2.33	1.00	2.30	1.06	2.32	1.00	0.945
10	2.78	0.83	1.60	0.70	2.16	0.96	0.004
11	4.22	1.30	2.30	1.49	3.21	1.68	0.009
12	2.67	1.41	4.60	0.96	3.68	1.53	0.003
13	3.00	1.12	4.10	1.19	3.58	1.26	0.054
14	3.33	0.86	4.50	0.97	3.95	1.08	0.014
15	2.89	0.33	3.10	1.45	3.00	1.05	0.676
16	1.78	0.44	4.40	0.51	3.16	1.42	0.000
17	4.44	1.01	2.00	1.05	3.16	1.61	0.000
18	2.78	0.44	3.80	1.13	3.32	1.00	0.022
19	2.56	1.51	4.70	0.48	3.68	1.53	0.001

available time to play (item 8), the complexity of playing together (item 8), and the interdependencies on each other to proceed (item 17) were negatively valued for both conditions. Differences in satisfaction between conditions appear significant. The last colon presents the significances of the difference ($p\Delta$) between both group means on each item after running an ANOVA. With the exception of items 3, 7, 8, 9, and 15, most differences appear significant <0.05 (with items 4, 5, 16, 17, and 19 yielding very high significances <0.001). More importantly, all these significant differences appear to be in favour of the face-to-face condition. All participants report problems with the structure and dependencies being too complex, and the user-interface and task instructions not being clear in advance for the online version, and consider it more effective to play the game face-to-face (item 19). Online players report that when these issues are resolved, playing the game online would be a more flexible and promising way. Most problematic in the current online version appeared to be the user-friendliness of the interface (item 4), unclarity about the interactions required (item 5), and unclarity about the rules for interaction (mutual assessment, awaiting turns, and completion rules) (items 16 and 17).

5.3 *Qualitative impressions and remarks*

A closer inspection of the comments and suggestions provided to the open item 20 shows that participants in both conditions report that it was fun to play the game. They especially liked to not only elaborate their own case, but also be able to provide input to and learn from others, and feel this way of collaboration is a useful way of increasing their professional teaching skills. Players in both conditions but mostly in the online condition report that instructions were unclear and players therefore were waiting much time for each other to proceed. They suggest that the interface instructions and performance in the online environment need much better elaboration. This was confirmed by teacher observations and students' communication that was logged during the online game chat.

Overall, based on these findings on satisfaction, students do appreciate to collaborate on these problems in such a playful way, and we therefore *accept our third hypothesis* with a positive answer to the third sub-question: yes, students do appreciate the collaboration set-up within such game play. It also became very clear that participants in the online group reported significantly less satisfaction with this game play on most aspects questioned. We, therefore, have to *decline our fourth hypothesis* with a negative answer to the fourth sub-question: no, it was not possible to have students profit from online support and interaction without decreasing user satisfaction.

6. Conclusion and discussion

Results from this study show that collaboration of students on classroom management dilemmas can indeed be successfully facilitated by this script, and that the learning outcomes do not differ for both versions (as objectively measured). The latter holds strong potential for offering online and more flexible ways of workplace learning. All students feel this is a useful way to collaborate, and appreciate the collaboration game. However, students playing the online version report significantly stronger needs for simpler structure and clearer task instruction. It is interesting to note that, however, online play was valued less; learning outcomes were equal for both conditions. To some extent, this can be attributed to unfamiliarity with online learning, but we also acknowledge that the online version still needs substantial improvement on the issues mentioned. Collaborative learning online is not easy and depends on the richness and intensity of interactions (emergence of elaborated explanations, negotiation of meanings, and mutual regulation of cognitive processes) as enabled by the collaboration structure. A too high level of structure might be contra-productive and harm students' creativity or spontaneity. The holy grail of CSCL is to find an optimal balance and establish environments that (in)directly favour the emergence of rich interactions, which is commonly referred to as 'design for converzation' (Dillenbourg & Fischer, 2007).

There were some constraints and problems with this study that have to be mentioned here. Besides technical problems and unclear instruction with the online version we already mentioned, students complained that the timing of the study during the end-of-year exams period was not very optimal. Although we did gather strong indications that playing the game increases learning outcomes, we in fact have no direct evidence for this. We cannot exclude that students would have obtained the same results without playing the game, either face-to-face or online. Due to practical reasons, it was not feasible to carry out a pre-test with a comparable problem case, or to include a control group. We have to note a potential situational dependency in the experimental set-up, since the participants were normally playing the game in a face-to-face fashion and had to adapt to the online condition. Satisfaction measures might have been more favourable after having become more used to the medium.

The study revealed that clear instruction and simple structure are especially important for online learning without direct teacher intervention. We, therefore, intend to continue our work with a comparative study differentiating high-structure (as in this version, but improved), medium-structure and low-structure in the online mastership game. In this next study, the level of structure will be determined by switching on or off 'handles', like have been defined by 'amount of phases', 'time available' (for each phase or step), 'awaiting turns', '(in)dependency on each others feedback', and other parameters that are expected to be equally valid in other collaborative scenarios. Eventually this might lead to a more systematic approach and toolset for defining and implementing structure in collaborative gaming. We also intend to look into the generalizability of findings, and apply a similar script in a different domain and scenario (e.g. the domain of dealing with dilemmas in designing serious games). Dillenbourg and Hong (2008) propose *script families* as an higher level of abstraction that discriminate classes of scripts that use the same pattern, e.g. JigSaw (distributing knowledge among group members), ArgueGraph (raising a conflict pattern), or Reciprocal Teaching (using mutual regulation). The mastership game belongs to the latter family, but it might be useful to explore others, like we already successfully implemented and studied a conflict script in a game on water management (Hummel et al., 2011). Collaboration scripts seem to offer potential to be further adapted and examined in serious gaming research. The complexity can be further reduced and reusable design patterns could become available (Westera, Nadolski, Hummel, & Wopereis, 2008). The collaboration pattern (script) we described in this study produces code that can be instantiated in different settings and domains (where mutual regulation and various perspectives play a role).

Acknowledgements

This study was carried out as an activity within the 'Learning Media' program of CELSTEC (Open University of the Netherlands), and activity within the Lectorate 'Workplace Learning and ICT' (NHL University of Applied Science, The Netherlands). Authors and developers from both institutes worked closely together during development of the online version and this study. We express our thanks to both institutes for their funding, and to NHL students for participating.

Notes on contributors

Hans Hummel works as an associate professor at the Open university. He also works as lector at the NHL University of Applied Science where he leads the Workplace learning and ICT program. His main research interests focus on electronic learning environments for professional learning.

Walter Geerts works as teacher-researcher at the NHL University of Applied Science. He teaches didactics and was the chief inventor of the mastership game. He carries out PhD research into videocases for the acquisition of professional teacher skills.

Aad Slootmaker works as a senior programmer at the Open university. He was the chief software architect of the EMERGO toolkit to develop and distribute serious games for complex learning, also being the topic of his PhD research.

Derek Kuipers works as teacher-researcher at the NHL University of Applied Science. He teaches communication and multimedia design, and carries out PhD research into serious games and factors that enable transfer to professional practice.

Wim Westera works as full professor at the Open university where he leads the Learning Media program.

References

Bell, B. S., Kanar, A. M., & Kozlowski, S. W. J. (2008). Current issues and future directions in simulation-based training in North America. *The International Journal of Human Resource Management*, *19*(8), 1416–1434.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Copeland, W. D., & D'Emidio-Caston, M. (1998). Indicators of development of practical theory in pre-service teacher education students. *Teaching and Teacher Education*, *14*(5), 513–534.

Dillenbourg, P. (2002). *Over-scripting CSCL: The risks of blending collaborative learning with instructional design*. Lausanne, Switzerland: CRAFT, Swiss Federal Institute of Technology.
Dillenbourg, P., & Fischer, F. (2007). Basics of computer-supported collaborative learning. *Zeitschrift für Berufs- und Wirtschaftspaedagogik*, *21*, 111–130.
Dillenbourg, P., & Hong, F. (2008). The mechanics of CSCL macro scripts. *International Journal of Computer-Supported Collaborative Learning*, *3*(1), 5–23.
Dillenbourg, P., & Jermann, P. (2006). Designing integrative scripts. In F. Fischer, I. Kollar, H. Mandl, & J. Haake (Eds.), *Scripting computer-supported communication of knowledge. Cognitive, computational, and educational perspectives* (pp. 259–288). New York: Springer.
Gee, J. P. (2003). *What video games have to teach us about learning and literacy?* New York: Palgrave MacMillan.
Geerts, W., Mitzsche, M., & Van Laeken, M. (eds.). (2009). *Wat zou je doen? Dilemma's in de onderwijspraktijk*. Bussum: Coutinho.
Gunawardena, C. N., Carabajal, K., & Lowe, C. A. (2001). *Critical analysis of models and methods used to evaluate online learning networks*. Seattle: AERA.
Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research*, *17*(4), 395–429.
Hernandez Leo, D. (2007). *A pattern-based design process for the creation of CSCL macro-scripts compatationally represented with IMS-LD* (Doctoral dissertation). University of Valladolid, Spain.
Heuvelmans, A. P. J. M., & Sanders, P. F. (1993). Inter-rater reliability. In T. Eggen & P. Sanders (Eds.), *Psychonometry in practice* (pp. 443–471). Arnhem, The Netherlands: CITO Institute for Test Development.
Hummel, H. G. K., Van Houcke, J., Nadolski, R. J., Van der Hiele, T., Kurvers, H., & Löhr, A. (2011). Scripted collaboration in serious gaming for complex learning: Effects of multiple perspectives when acquiring water management skills. *British Journal of Educational Technology*, *42*(6), 1029–1041.
Jeong, H., & Chi, M. T. H. (2000). *Does collaborative learning lead to the construction of common knowledge?* Retrieved July 9, 2004, from http://www.ircs.upenn.edu.edu/cogsci2000/PRCDNGS/SPRCDNGS/posters/jeo_chi.pdf
Kiili, K. (2007). Foundations for problem-based gaming. *British Journal of Educational Technology*, *38*, 94–40.
Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fisher, F. (2007). Specifying computer-supported collaboration scripts. *Computer-Supported Collaborative Learning*, *2*, 211–224.
Michael, D., & Chen, S. (2006). *Serious games: Games that educate, train and inform*. Boston, MA: Thomson Course Technology.
Nadolski, R. J., Hummel, H. G. K., Van den Brink, H. J., Hoefakker, R., Slootmaker, A., Kurvers, H., & Storm, J. (2008). EMERGO: Methodology and toolkit for efficient development of serious games in higher education. *Simulations & Gaming*, *39*(3), 338–352.
Prensky, M. (2006). *Don't bother me Mom, I'm learning: How computer and video games are preparing your kids for 21st century success and how you can help*. Saint Paul, MN: Paragon House.
Shaffer, D. W. (2006). *How computer games help children learn*. New York: Palgrave MacMillan.
Van Bruggen, J., Kirschner, P. A., & Jochems, W. (2002). Representation of argumentation in CSCL and management of cognitive load. *Learning & Instruction*, *12*(1), 121–138.
Westera, W., Nadolski, R. J., Hummel, H. G. K., & Wopereis, I. (2008). Serious games for higher education: A framework for reducing design complexity. *Journal of Computer Assisted Learning*, *24*(50), 420–432.