STRUCTURING COLLABORATION SCRIPTS
Optimizing online group work on classroom dilemmas in teacher education

Hans Hummel[1 2], Walter Geerts[2], Aad Slootmaker[1], Derek Kuipers[2] and Wim Westera[1]
[1]*Open University of the Netherlands, Valkenburgerweg 177, 6419AT Heerlen, The Netehrlands*
[2]*NHL University of Applied Science, Rengerslaan 10, 8917DD Leeuwarden, The Netherlands*
{hans.hummel, aad.slootmaker,wim.westera}@ou.nl, {h.hummel, w.m.geerts, d.a.kuipers}@nhl.nl

Keywords: serious games, scripted collaboration, structure, classroom dilemmas, teacher education

Abstract: Serious games can facilitate workplace learning, for instance when collaboration on solving professional problems is involved. The optimal structure in collaboration scripts for such games has appeared to be a key success factor. Free collaboration does not systematically produce effective learning, but imposing too much structure by design might easily disturb the genuine notion of spontaneous collaborative learning. In this study we compare a 'high- structured' and 'low-structured' version of a mastership game where teachers-in-training discuss solutions on classroom dilemmas. Adequate solutions are discussed and elaborated during small group play, reported in individual advisory reports, and independently assessed by both two teachers and the group peers. We collected data on the differences in learning effects and student appreciation. The most interesting result shows that reports delivered by students that played the low-structured version received significantly higher teacher grades when compared to the high-structured version. Clear differences in peer-assessments were not found. Practical implications of these findings for future research into collaborative scripting will be discussed.

1 INTRODUCTION

Serious games not only support individual learning but also foster the acquisition of soft skills like collaboration and reflection about wicked problems, that are usually not addressed by other learning platforms (Gee, 2004). Educators call these games 'serious' to denote that they are *not just* fun to play, but *also* hold potential as cognitive tools for learning and professional development (e.g., Michael & Chen, 2006).

This introduction will describe (1) how serious games may facilitate professional workplace learning when collaboration on practical problems is involved, (2) what is the role of collaboration scripts in such social games, (3) what are ways to define and optimize the structure of such collaboration scripts, and (4) introduce the collaboration game under study (i.e., teachers learning to deal with classroom dilemmas). We will hypothesize that providing structure is a key success factor for effective collaborative learning, and that imposing too much structure by design might have disruptive effects.

1.1 Serious Games for Collaboration in Workplace Learning

Workplace learning is shifting focus from individuals acquiring and updating domain knowledge towards selecting and using this knowledge for certain problem situations in daily practice. Such learning deals with competences like information skills and media literacy, problem-solving, communication and collaboration, and above all critical reflection (e.g., Garrick, 2007). Today's professionals are becoming lifelong learners that continuously face problem situations that often change dynamically and rapidly. Furthermore, organisations' tacit knowledge plays a crucial role in solving their problems but such knowledge can only be expressed and accessed in direct collaboration on professional tasks (e.g., Nonaka & Takeuchi, 1995). Professional tasks that include collaboration, argumentation and negotiation are crucial for vocational education, especially when they aim to connect school knowledge to practical work.

Games are heavily inspired by experiential learning principles which hold potential for contextualised workplace learning. Serious games appear suitable as flexible learning environments where professional tasks can be carried out with little or none direct intervention of experts or teachers (e.g., Bell *et al*., 2008). We are interested to find out to which extent this holds true in practice and what kind of learner support these learning environments should contain. Such games will take a workplace context (practical problem or authentic case lead) as starting point to stimulate learners acquire new knowledge by sharing and co-creating (Bell *et al*., 2008). 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. Collaboration support *within* a game has to be enabled by a didactic 'script' which we will name 'scripted collaboration'.

1.2 Collaboration scripts

Collaboration scripts (Kobbe *et al*., 2007) are an instructional method that structures the collaboration by guiding the interacting partners online through a sequence of interaction phases with designated activities and roles. Scripting collaboration has been examined in CSCL, where it positively influenced learning (e.g., Gunawardena, Carabajal, & Lowe, 2001). The group interaction in Computer-Supported Collaborative Learning between learners 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).

However, such collaboration scripts have hardly ever been implemented and tested in more open learning environments like serious games (Dillenbourg & Hong, 2008). No research has focused on defining or optimizing the essential elements (e.g., of structure) or has measured the learning effects of including such scripting in serious game play. 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. Dillenbourg (2002) mentions that the technological 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. Regarding the distribution of the task, for instance when a player has to read and understand while another player has to ask questions and give feedback.

1.3 Optimizing the Structure of Collaboration scripts

The distribution of the task, the mode of interaction and the timing of phases all deal with structure, which we define here as the amount of restriction imposed on the freedom that is allowed in the group collaboration process. An optimal level of structure appears to be a key success factor for effective learner support. A low degree of 'coercion' in 'induced scripts' is elegant but often not sufficient to influence the collaborative process. On one hand, free collaboration does not systematically produce learning (e.g., Dillenbourg, 2002). But too much educational engineering on the other hand might kill the spontaneity and motivation of genuine collaboration, an effect that Rothkopf (1996) denoted as 'mathematanic' (as opposed to 'mathemagenic'). 'Prompted' or 'follow me' scripts have higher levels of 'coercion' and will steer the collaborative process, at the risk of being perceived as too complex or rigid. In a previous study we found that students complained about the complexity and task instruction within for the mastership game (Hummel *et al*., 2013). Dillenbourg (2002) earlier reported on other scripts (e.g., UniverSanté) that were perceived as too complex. His advice is to keep scripts as simple as possible so that all actors are able to appropriate them. He mentions following risks of over-scripting: disturbing 'natural' interactions by adding too many interaction breakdowns; disturbing 'natural' problem solving processes by segmenting the task; increased cognitive load by the necessity to understand, memorize and execute the script itself as well; 'didactising' collaborative interactions when interactions are played like in the teacher-learner context; and 'goalless' interactions when the script prevents the group to establish a shared goal.

In order to optimize structure, we first have to further define what is structure and what are its elements. Although this might appear evident, a clear definition of structure is not well documented in CSCL literature. Building on Dillenbourg's risks of over-scripting we argue that *segmentation* and *inter-dependency* within the task constitute the main structure elements. Clearly, a holistic task is less structured than a task that has been segmented in various consecutive subtasks. Clearly, a task that can be carried out independently is less structured than a task that depends on synchronisation or approval of peers and / or teachers. We will operationalize these structure elements for this study.

Collaborative learning by some researchers (e.g., Glachan & Light, 1981) is considered to be overly optimistic or naive. Can we assume that two learners with little knowledge in a domain of study would naturally gain knowledge in miraculous interaction? The recent evolution of CSCL in the direction of using scripting to make collaboration more effective might drift away from this ‘natural’ process and get us closer to more directive teaching methods. We may question if it is at all possible to combine two pedagogical traditions (collaborative learning and traditional instruction) without losing what makes ‘natural’ collaborative learning different. In the world of semi-structured communication, optimizing structure plays a key role in this design dilemma. On the positive side, the notion of script might bridge collaborative learning and traditional instructional design.

1.4 Mastership game: working together to solve classroom 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 losing 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. Teachers without doubt will have to face unexpected situations that demand them to find solutions on the fly. Not all classroom problems present themselves as dilemmas, however as practical contexts dilemmas offer best opportunity for seeing various perspectives, solutions and discussion. Some experienced didactics teachers developed the Mastership game which helps students to find solutions to the most prevailing practical classroom management dilemmas in a playful and collaborative way, a way that will help them become better teachers. The game was originally developed as a card game to be played face-to-face in small groups (Geerts, Mitzsche, & Van Laeken, 2009), and was later transformed into on online game to be played synchronously with freedom of place (Hummel *et al*., 2013).

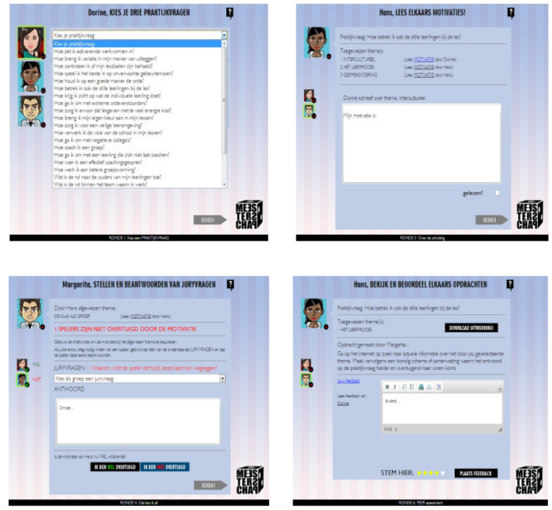

Figure 1. Screens of the online version of the Mastership game : selecting three practical dilemmas in phase 1 (upper left hand), assigning and motivating themes in phase 3 (upper right hand), motivating and discussing declined themes in phase 4 (lower left hand), and peer assessment of elaborated assignments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

The Mastership game can be played in small groups of two till six students and does not require any intervention by teachers. After selecting their avatars, students 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 each other’s problems. 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 practical classroom dilemmas, either out of a pile of twenty-four, 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”). 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 and use their co-players' input to further elaborate their solution in a short advisory report.

This previous subsection mentioned (a) segmentation and (b) inter-dependency as constituting elements of structure. Following structure elements were identified within the mastership game: (a1) the scenes within a phase (two or three) are consecutive (true) or might be carried out in parallel (false); (a2) the obligatory number of cards to draw (possible values 1-3); (b1) players have to take turns (true) or may work in parallel (false) within the various scenes and phases; (b2) players are being assessed by group members before they may proceed (true), or may decide themselves (false); and (b3) at moments players are dependent on which cards others draw for them (true), or at any time may draw their own cards (false). Most elements only pertain to some of the scenes and phases, with the exception of (b1) which occurs in most phases (see Table 1).

Table 1. Possible values of structure elements for each scene / phase

Scene	**a1 (has to finish scene)**	**a2 (number of cards to draw)**	**b1 (wait for others)**	**b2 (feedback required)**	**b3 (others draw cards)**
1_1	true	3/2/1	true/false	false	false
1_2	true	1	true/false	false	false
2_1	false	-	true/false	false	false
2_2	true/false	-	true/false	false	false
2_3	true/false	-	true/false	true/false	false
3_1	true	3/2/1	true/false	false	true/false
3_2	true/false	-	true/false	false	false
4_1	true/false	-	true/false	false	false
4_2	true	-	true	true	false
4_3	true	-	true	true	false
5_1	true	-	true/false	true/false	false
5_2	-	-	-	-	-
6_1	-	-	-	-	-
6_2	-	-	-	-	-

Based on this operationalization of structure we define high-structure as all elements having the value 'true' (and the value of a2 being 3), and low-structure as all elements having the value 'false' (and the value of a2 being 1). We also defined and developed a medium-structured version with a1 being false, a2 being 1, b1 being false, b2 being true, and b3 being true.

The main hypotheses (research questions) to be answered in the next sections are twofold: (1) Will less structure lead to more 'natural' and effective collaborative learning? (We hypothesize that the individual reports of those that played the low-structured game will be objectively graded higher by their teachers (1a) and peers (1b)); and (2) Will less structure in the collaboration be appreciated more by students? (We hypothesize that students will subjectively appreciate the low-structured game higher on a number of aspects.)

2 METHOD

After describing the participants and the learning materials we used, we explain the procedure and assessment instruments we used to measure the effectiveness of learning and the satisfaction with the scripted collaboration in both conditions.

2.1 Participants

Twenty-nine teachers-in-training, third year students of the NHL University of Applied Science in the Netherlands, participated in this case study as part of their regular curriculum. This case study was awarded a study load of about 10 hours (half an EC point) as part of workplace learning. 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 (M = 39.5, SD = 8.10), ranging between 26 and 54 years. Twenty-two were female and seven were male students, equally divided over the conditions. Gender and age showed neither to differ over conditions nor to be related to learning outcomes. The effects of gender and age on learning found were $F(2,27) = 0.755$, $p = 0.704$ and $F(2,27) = 1.286$, $p = 0.267$ respectively.

2.2 Learning materials

The 'Mastership' game contains a total of 66 'cards' or instructions: 24 practical dilemmas; 13 exploratory assignments; 10 themes, with 10 jury cards containing questions per theme; and 8 final assignments. Instructions and rules for playing the game are provided at the start of each phase. The game will be played by a small group of two till six players, and will take about two hours to play. The five phases with various types of cards involved were explained in the previous section.

We developed the online version of the 'Mastership' game using the ZK toolkit (http://www.zkoss.org). The EMERGO toolkit for serious game development (http://www.emergo.cc) was used for game run management and for storing and analysing 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 easily be configured by a game author on several aspects like structure.

2.3 Procedure

All students 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 were randomly allocated to one of three conditions (high-structured, low-structured, control). Participants in the control group had to solve the practical classroom dilemma individually without playing the collaboration game. Each gaming condition contained two groups (of four or five students each). The players received an e-mail before the meeting, containing the URL and their personal account. All playing participants received a questionnaire about their appreciation of the game by e-mail a day after playing the game. At the time of the meeting, playing participants went to a computer room to work together online. A teacher was present in this computer room to control for direct (non)verbal communication beyond the program. During the time of the meeting, students in the control group individually worked on their practical task, without playing the game.

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. It was estimated that the elaboration of the reports would take about half a day. 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). Playing students were allowed to deliver and grade reports either online or by mail. Students were able to pass through phases without technical problems, with the exception of a group playing the medium-structured version of the game (for this reason the medium-structured condition had to be left out of the research design and primary analyses, although we will mention what was found for this version of the game). All data could be collected two weeks after the meeting took place.

2.4 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 by the teacher / topic expert (being one of the authors of this article). The elaborated reports can be assessed on 'growth in professional productivity', and 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 2 contains indications for the possible scores on these criteria, with total scores ranging from 0 to 10. Sufficient inter-rater reliability of the instrument was determined in a previous study (Hummel et al. , 2013).

2.5 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

Table 2. Sub-scales and scoring categories of the learning effect correction model

Subscales	Insufficient (0 points)	Sufficient (1 point)	Good (2 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

the structure (S, 5 questions), user-friendliness and clarity (U, 5 questions), the timing of the phases (T, 2 questions), the quality of the dilemmas and assignments (Q, 5 questions), and the interaction during collaboration (I, 2 questions). The focus on structure and clarity of instruction was inspired by previous studies (e.g., 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 3 were not listed in the original questionnaire.) All these items used a Likert-scale with five 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

Table 3. 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.

and all values above as (slightly) positive / negative appreciations. Item 20 was an open question, allowing room for comments and suggestions. Table 3 contains the list of items.

3 RESULTS

This results section provides answers to the twofold research question we posed at the end of the introduction: (1) Will less structure lead to more 'natural' and effective collaborative learning? We hypothesized that the individual reports of those that played the low-structured game will be objectively graded higher by their teachers (1a) and peers (1b); and (2) Will less structure in the collaboration be appreciated more by students? We hypothesized that students will subjectively appreciate the low-structured game higher on a number of aspects. We present the objective learning effect measures (answering the first question) and the subjective questionnaire measures (answering the last question).

3.1 Learning effect measures

We found that most individual reports (76%) could be graded as sufficient. Grades below 6.0 were considered not sufficient, and only seven students received either an 4.5, 5.0 or 5.5 (three times in the high-structured and control conditions, and one time in the low-structured condition). The average grade for all participants was M = 6.62, SD = 1.29. We added a control group to establish if playing the game does contribute *at all* to learning. As you see in Table 4 the average teacher grades for the control group were indeed lowest, so there appears to be an effect of playing the game. This effect appears significant when we compare the non-playing group to the low-structured (t (18) = 2.97, $p < 0.01$) and the medium-structured condition (which we left out of the analyses). However, we could not observe a significant difference between non-players and those playing the high-structured version (t (17) = 0.67, p = 0.51).

Table 4. Average report grades for all conditions, both from teachers and peers

	High structure (n = 9)		Low structure (n = 10)		Control (n = 10)		All (N = 29)	
Assessment	M	SD	M	SD	M	SD	M	SD
Teacher grade	6.44	1.59	7.35	1.03	6.05	0.93	6.62	1.29
Peer rating	7.93	.66	7.52	1.04	7.68	0.89	7.70	0.87

When looking for an overall effect of condition (N = 29) on learning effect we see a clear trend: low-structure scores best, than high-structure, and finally the control group. This effect is 'marginally' significant (F (2, 26) = 3.072, MSE = 4.428, p = 0.063, η_p^2 = 0.18), with values of the partial-eta-squared above .13 showing large effect size according to Cohen (1988). On top of this and even more importantly for the central research question, a significant difference (t (17) = 4,86, p = 0.042) is found in favour of low-structure when comparing with high-structure (N= 19). When looking at the peer ratings, we do not find any significant differences between conditions. We therefore *accept our first hypothesis* with a positive answer to the first sub question (yes, playing the low-structured game leads to higher teacher grades). We could not find evidence to accept the second sub question (no, peer grading does not seem to differentiate that well between conditions). The average and normalized ratings awarded by peers (co-players) do appear to be correlated to the grades awarded by teachers (r_s = 0.36, p = 0.052, two-tailed), although this relation is only 'marginally' significant. Ratings by peers (M = 7.70, SD = 0.87) overall are higher than teacher grades (M = 6.62, SD = 1.29).

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 = .16, δ = .09). On the level of the sub-scales of the instrument inter-rater reliability was fair to moderate for four out of five scales. Only on sub-scale B (Reflection) there was poor agreement (K = .19). Excluding this sub-scale would immediately increase the overall Kappa to moderate (K = .48) and acceptable. Closer inspection of the differences in scores between both raters revealed that rater one consistently awarded slightly lower grades (M = 6.48, p = 1.38) in comparison with rater two (M = 6.76, p = 1.41). Since the standard Kappa does not take into account degrees of disagreement between observers (all disagreement is considered as total disagreement) when having ordered categories, we decided it would be preferable to calculate Weighted Kappa's. Since the difference between the first and second category had the same importance as the difference between the second and third category (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 and acceptable (K_w = .47, δ = .08), and when excluding sub-scale B appeared even good (K_w = .68, δ = .09). On a more minute level of interpretation we would like to mention that, when using linear weighting, a good interpretation of the observed Kappa's depends on the maximal Kappa's, which on its turn depend on the marginal distribution of cells for each sub-scale. A smaller observed Kappa (e.g. $K_w = .36$) then might still be acceptable if the maximal Kappa is just a bit higher (e.g., $K_w = .45$), yielding a ratio of 0.80. On the other hand a higher observed Kappa (e.g. $K_w = .43$) might not be that acceptable if the maximal Kappa is much higher (e.g., $K_w = .95$), yielding a ratio of 0.45. We found ratios of observed and maximal Kappa's for all subscales to range between .51 and .89. We decided it was very acceptable to use the original instrument, even without correction for less agreement on sub-scale B. Kappa's between 0-20 are considered 'poor' or 'light', between 20-40 as 'fair', between 40-60 as 'moderate', between 60-80 as 'substantial' or 'good', and between 80-100 as 'almost perfect' or 'very good' (Heuvelmans & Sanders, 1993, p. 450).

3.2 Satisfaction measures

Table 5 presents the average scores on all items of the questionnaire for both conditions. The last colon presents the significances of the difference ($p\Delta$) between both group means on each item after running an ANOVA.

Table 5. Average score on the satisfaction questionnaire items, and significance of difference between versions

	High structure (n = 9)		Low structure (n = 10)		All (N = 19)		
Item	*M*	*SD*	*M*	*SD*	*M*	*SD*	$p\Delta$
1	3.00	1.41	2.70	1.25	2.85	1.30	.64
2	4.00	0.76	4.00	0.82	4.00	0.76	1.00
3	3.50	1.31	4.20	0.92	3.89	1.13	.20
4	2.88	1.25	3.50	1.08	3.22	1.17	.27
5	2.38	1.40	2.67	1.50	2.53	1.42	.67
6	2.25	0.89	1.40	0.52	1.78	0.81	.02
7	3.13	1.25	3.40	1.07	3.28	1.13	.62
8	3.25	1.49	4.20	1.03	3.78	1.31	.13
9	2.75	0.87	2.30	0.95	2.50	0.92	.32
10	2.13	0.83	2.20	1.40	2.17	1.15	.89
11	4.00	0.53	2.90	1.37	3.39	1.19	.04
12	3.38	1.51	2.80	1.23	3.06	1.35	.38
13	2.88	1.36	3.00	0.90	2.91	1.14	.88
14	3.13	0.99	3.33	0.82	3.21	0.89	.68
15	2.88	0.99	2.33	1.16	2.73	1.01	.46
16	2.50	1.31	2.56	1.24	2.53	1.23	.93
17	4.25	1.04	3.56	1.24	3.88	1.16	.23
18	3.00	1.07	3.67	1.55	3.18	1.08	.39
19	3.00	1.19	3.22	1.56	3.12	1.36	.75

Considering the formulation of items (listed in Table 3), we can observe that for both conditions aspects have been valued as slightly positive (above neutral), like items 2, 3, 6, 9, and 10, while others have been valued as slightly negative (below neutral), like items 5, 8, 11, 16 and 17. Although students do not clearly criticize the structure for each phase, they do indicate that collaboration rules were not always clear (item 16) and that mutual dependency was too high (item 17), with the last items clearly being more negative for the high-structured group. For most items we did not find significant differences between both versions of the game, with just two exceptions. The low-structured group showed to be more satisfied with the amount of time to play (item 16). The high-structured group indicated that the overall structure was too high (item 11), a finding in line with what was reported on learning effects. Based on these findings we cannot accept our *second hypothesis*. It did not become clear that low-structure was appreciated more by students on various aspects

4 CONCLUSIONS

Collaboration can be successfully facilitated by scripting serious games when we take into account the importance of good instruction and optimal structure. Also in this study players reported problems with clear task instruction and collaboration rules without teacher intervention. This study found that over-scripting may indeed have disruptive learning effects. Players of the low-structured version of the mastership game produced reports that were graded significantly higher than the ones of those playing the high-structured version (and of those not playing the game). The average grade of a small group playing the medium-structured version (which for methodological reasons we had to leave out of our design) was close to the low-structure group (M = 7.25, SD = 0.61, n = 5). The structure elements that we left out when going from 'high' to 'medium', having to work in a strict order (a1) and having to take turns (b1), appear to have potential disruptive (or mathematanic) effects on the emergence of rich interactions. This result has practical implications for 'designing for conversation' which according to Dillenbourg & Fischer (2007) is the holy grail of CSCL.

For the generalizability of these findings it will be useful to carry out studies that research the effectiveness of other types of collaboration scripts and implementations in other domains. For this end we need a dedicated authoring environment that enables us to manipulate elements of scripts. Dillenbourg and Hong (2008) proposed 'script families' at a macro level of abstraction. They differentiate three classes of scripts that more or less use the same collaboration patterns: JigSaw (distributing knowledge amongst group members, e.g. by allocating various perspectives or roles), Reciprocal Teaching (using mutual regulation, e.g. by taking turns in contributing to tasks of each other), and Conflict Raising (competing or taking opposing stands, e.g. by negotiation or argumentation). In our research we are currently constructing a dedicated authoring environment to instantiate collaboration scripts for various situations, which is built on the EMERGO platform. Studies into case instances of the JigSaw scripting class (Hummel *et al.*. 2011) and Reciprocal Teaching class (Hummel *et al*., 2013) have already been realised and reported. For the third family (Conflict Raising) we are currently preparing a study instantiating an argumentation game in EMERGO, based on an existing wiki where players take and defend opposing stands (Van Rosmalen & Westera, 2012). When user-friendly authoring environments and reusable design patterns become available, the complexity and effectiveness of collaborative learning through serious game play can be further improved.

ACKNOWLEDGEMENTS

This study was carried out as research activity within the 'Learning Media' program of CELSTEC (Open University of the Netherlands), and 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 game and this study. We express our thanks to both institutes for their funding, and to NHL students for participating.

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. Three worlds of CSCL.

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. Springer.

Garrick, J. (2007). Informal learning in the workplace: Unmasking human resource development. London: Routledge.

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. (Red.). (2009). *Wat zou je doen? Dilemma's in de onderwijspraktijk*. Bussum: Coutinho.

Glachan, M.D., & Light, P.H. (1981). Peer interaction and teaching: can two wrongs make one right? In G. Butterworth & P.H. Light (Eds.) Social cognition: studies of the development of understanding. Brighton: Harvester Press.

Gunawardena, C.N., Carabajal, K., & Lowe, C.A. (2001). Critical *Analysis of Models and Methods used to evaluate Online Learning Networks*. Seattle: AERA.

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.

Hummel, H.G.K., Geerts, W.M., Slootmaker, A., Kuipers, D., & Westera, W. (2013, online). Collaboration scripts for mastership skills: online game about classroom dilemmas in teacher education. *Interactive Learning Environments*, 2013.

Jeong, H., & Chi, M.T.H (2000). Does Collaborative Learning Lead to the Construction of Common Knowledge? [Retrieved online on July 9, 2004 from: http://www.ircs.upenn.edu/cogsci2000/PRCDNGS/SPRCDNGS/posters/jeo_chi.pdf]

Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, 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*. Thomson Course Technology, Boston, MA.

Nonaka, I. & Takeuchi, H. (1995). The Knowledge-creating company: How Japanese companies create the dynamics of innovation. New York: Oxford University press.

Rothkopf. E. Z. (1996. Control of mathemagenic activities. In D.H. Jonassen (Ed.). Handbook of research for educational communications and technology (pp. 879-896). New York: MacMillan Library Reference

Van Rosmalen, P. & Westera, W. (2012). Introducing serious games with Wikis: empowering the teacher with simple technologies. Interactive Learning Environments, 2012, 1-14.