Abstract: This design study developed and tested a peer assessment tool and reflection tool for enhancing group functioning in a computer-supported collaborative learning environment. The underlying assumption was that group functioning can be positively influenced by making group members aware of how their behavior is perceived by themselves, their peers, and the group as a whole. This awareness, which is conditional for behavioral change, is achieved through a peer assessment tool and a reflection tool. A 2x2 factorial between-subjects design was used. Participants were 39 fourth-year high school students who worked in groups of 3 or 4 on a collaborative writing task. Results show that groups with peer assessment tool developed better teams, had lower levels of group conflicts, and had a more positive attitude towards collaborative problem solving, than groups without a peer assessment tool. Thus, peer feedback on social behavior of group members can enhance group functioning in CSCL-groups.

Introduction

Collaborative learning, often supported by computer networks (computer supported collaborative learning, CSCL) is enjoying considerable interest at all levels of education. Collaborative learning, defined as the "mutual engagement of participants in a coordinated effort to solve the problem together" (Rochelle & Teasley, 1995, p. 70) has, among other things, been found to enhance the learners’ cognitive performance (Johnson & Johnson, 1999) and to stimulate them to engage in knowledge construction (Stahl, 2004). CSCL environments, though originally simple, text-based, computer mediated communication systems, have been strongly influenced by the rapid development of information and communication technology tools and widgets (e.g., e-mail, chat, video conferencing, and discussion forums). These applications have proven to be useful for supporting education and collaborative learning (Janssen, Erkens, Kanselaar, & Jaspers, 2007; Kreijns, Kirschner, & Jochems, 2003), leading to the design and implementation of more sophisticated CSCL environments.

Though CSCL environments have been shown to be promising educational tools and though expectations as to their value and effectiveness are high, groups learning in CSCL environments do not always reach their full potential. One of the most important reasons for this disparity between their potential and their results can be found in the social interaction between the group members, which is influenced by (1) the characteristics (i.e., the design) of the CSCL environment and/or (2) the cognitive and socio-emotional characteristics (behaviour) of the group members (Kreijns, Kirschner, & Jochems, 2003).

First, the design of CSCL environments is often solely functional, focusing on the cognitive processes needed to accomplish a task and/or solve a problem and achieving optimal learning performances (Kreijns & Kirschner, 2004). These functional CSCL environments force (coerce; Kirschner, Beers, Boshuizen, & Gijselaars, 2008) group members to solely carry out these cognitive processes and thus limit the possibility for socio-emotional processes to take place. These socio-emotional processes, which are the basis for group forming and group dynamics, are essential for developing strong social relationships, strong group cohesiveness, feelings of trust, and a sense of community among group members (i.e., for creating a sound social space). Without such a sound social space, the group will not reach its full potential (Jehng, 1997). Groups in CSCL environments that lack social functionalities will ultimately perform poorly (Cutler, 1996; Kreijns & Kirschner, 2004).

Second, group members can show non-cooperative behaviour (i.e., free-rider effect and sucker-effect), which can have negative effects on group functioning and group performance. The free-rider effect or hitchhiking effect occurs when group members think that their individual effort is unnecessary, because the task can be performed by the other group members. This often occurs when the individual group members receive a grade that is based on the performance of the whole group (Kerr, & Bruun, 1983). The sucker effect occurs when productive group members believe that they invest more time and effort in the group product than their co-members. The productive group members will often reduce their individual efforts, because they refuse to support the non-contributing members (Kerr & Bruun, 1983).

To this end, CSCL environments can be augmented with computer tools or widgets that support social functions. These tools, also known as ‘social affordance devices’, can positively effect group functioning and group performance in a CSCL environment (Kreijns & Kirschner, 2004). Social affordances are defined as those properties of the CSCL environment that act as social contextual facilitators relevant for the learner’s social
Peer feedback
Peer feedback can be used to provide group members with information concerning their behaviour in a group (i.e., their interpersonal behaviour). This peer feedback can be focussed on evaluation and/or development. Topping (1998) has a more evaluative perspective on peer feedback and defines peer feedback as an “arrangement in which individuals consider the mount, level, value, worth, quality or success of the products or outcomes of learning of peers of similar status” (p. 250). In comparison, Earley, Northcraft, Lee, and Lituchy (1990) have a more developmental perspective on feedback which is focussed on performance improvement, and is described as information provided to an individual to increase performance. In this study, the developmental perspective on feedback will be used because the goal is to improve group functioning and group performance.

Group members can use peer feedback to monitor group processes or functioning (i.e., group processing). Group processing occurs when group members discuss how well their group is functioning and how group processes may be improved (Webb & Palincsar, 1996). These discussions may help groups pinpoint, comprehend, and solve collaboration problems (e.g., free riding, lurking) and may contribute to successful collaborative behaviour (Yager, Johnson, Johnson, & Snider, 1986). This is in line with McLeod and Liker (1992), who found that peer feedback on the interpersonal behaviour in a group can change the behaviour of individual group members. For example, giving group members peer feedback on their individual behaviour in the group (e.g., their degree of communication and collaboration), led to an increase in motivation, understanding, and solution of collaboration problems (Prins, Sluijsmans, Schreurs, & Kirschner, 2006). Thus, the peer assessment- and reflection tool will be used as a basis for stimulating and supporting peer feedback dialogues which can help groups to pinpoint, comprehend, solve collaboration problems, and may contribute to successful collaborative behaviour (Prins et al., 2006; Yager, Johnson, Johnson, & Snider, 1986). Bales (1988) concurs with this, arguing that open group discussions in which explicit decisions are made to modify behaviour can encourage transformation.

Reflection
Reflection on group behaviour based on the information from the peer assessment tool can be seen as a feedback dialogue (Askew & Lodge, 2000). During a feedback dialogue, peers can discuss whether the feedback receiver understands the feedback, whether s/he accepts the feedback, whether s/he agrees with the feedback, whether the receiver is challenged to reflect on his/her own performance, and whether the feedback provides clues for behavioural change (Prins, Sluijsmans, Schreurs, & Kirschner, 2006). Thus, the peer assessment- and reflection tool will be used as a basis for stimulating and supporting peer feedback dialogues which can help groups to pinpoint, comprehend, and solve collaboration problems, and may contribute to successful collaborative behaviour (Prins et al., 2006; Yager, Johnson, Johnson, & Snider, 1986). Bales (1988) concurs with this, arguing that open group discussions in which explicit decisions are made to modify behaviour can encourage transformation.
and/or the reflection tool affect social interaction, group functioning, and group performance within a computer supported learning environment. It is assumed that peer feedback in combination with reflection on individual behaviour and group functioning, will be most effective. The specific goal of this study is to determine whether the designed peer assessment tool and reflection tool are useful. Usefulness incorporates utility - the set of functionalities that a tool incorporates - and usability - whether a tool allows for the accomplishment of a set of tasks in an efficient and effective way, that satisfies the user (Kirschner, Strijbos, Kreijns & Beers, 2004).

Method

Participants
Participants were 39 fourth-year students (19 male, 20 female), with an average age of 16 (M = 15.54, SD = .60, Min = 14, Max = 17), from an academic high school in The Netherlands. Students came from two classes and were enrolled in the second stage of the pre-university education track which encompasses the final three years of high school. The participants were randomly assigned by the researchers to groups of three or four, and to one of the four conditions (see Design). Group compositions were heterogeneous in ability and gender.

Design
A 2x2 between-subjects factorial design was used with the factors Radar unavailable (~Ra) – available (+Ra), and Reflector unavailable (~Rf) – available (+Rf). This leads to four conditions (~Ra~Rf, +Ra~Rf, ~Ra+Rf, +Ra+Rf). The condition with Radar and Reflector (+Ra+Rf) consisted of 11 students (2 groups of 4, and 1 group of 3), without Radar but with Reflector (~Ra+Rf) of 12 students (3 groups of 4), and with Radar but without Reflector (+Ra~Rf) and without both tools (~Ra~Rf) of 8 students (2 groups of 4).

Dependent variables
To measure changes in interaction between group members the communication between the group members saved in the data base (chat-history) is analysed. The dialogues between the group members is automatically coded by the Dialogue-act coding system (e.g., Erkens, Jaspers, Prangsma, & Kanselaar, 2005) which indicates the communicative function of an utterance (e.g., words, statements, expressions, et cetera) along five communicative functions (i.e., argumentative, responsive, informative, elicitative, and imperative).

The group functioning awareness scale for both peer feedback tool (k = 4, α = .83) as reflection tool (k = 2, α = .61), provides information about whether the tools were able to make group members aware of how their behavior is perceived by themselves, their peers, and the group as a whole. The usefulness scale for both peer feedback tool (k = 10, α = .70) as reflection tool (k = 9, α = .78), provides information about whether the tools were considered as useful by the users. The feedback dialogues scale for both peer feedback (k = 2, α = .69) as reflection tool (k = 2, α = .66), provides information about whether the tools stimulated dialogues on group functioning.

To measure group functioning, previously validated instruments were translated into Dutch and transformed into 5-point Likert scales (1 = totally disagree, 5 = totally agree). The Team Development scale (k = 4, α = .83) provides information on the perceived level of group cohesion. The Group-process Satisfaction scale (k = 4, α = .83) provides information on the perceived satisfaction with general group functioning (both cf. Strijbos, Martens, Jochems, & Broers, 2007); transformed into 5-point Likert scales). The Intra-group Conflicts scale (k = 4, α = .83; cf. Strijbos et al.) provides information on the perceived level of conflict between group members. The Attitude towards Collaborative Problem Solving scale (k = 4, α = .83; cf. Strijbos et al.) is self-evident.

The grade given to the groups’ collaborative writing task (i.e., the essay) was used as a measure of Group performance. The essays were graded by two researchers, both experienced in grading essays. The inter-rater reliability was high (n = 10, Cronbach’s α = .86).

Task and procedure
The participating students collaborated in groups of three or four on a writing task in the history domain. Every student works at one computer. They had to write an essay about the film ‘Fitna’ by the Dutch parliamentarian Geert Wilders which argues that Islam encourages, among other things, acts of terrorism, anti-Semitism, sexism and violence against women, and Islamic universalism. This task was considered historically and civically highly relevant by the school. The collaborative writing task consisted two sessions of 90 minutes each, and the time between the first and the second session was one week. The groups collaborated in a CSCL environment called Virtual Collaborative Research Institute (VCRI; Jaspers, Broeken, & Erkens, 2002) which is a groupware program designed to support collaborative learning on research projects and inquiry tasks. VCRI will be further described in the Instruments section.

During collaboration, groups with a peer feedback tool, (+Ra~Rf, +Ra+Rf) made use of the tool at the beginning of the experiment (T1), halfway through the experiment which was at the end of the first session (T2),
and at the end of the second and final session (T3). Note that the ‘time-on-task’ was the same for all four conditions. The groups with a reflection tool, (−Ra+Rf, +Ra+Rf), had to fill in the tool twice, namely halfway through the experiment (T2) and at the end of the final session (T3).

At the end of the final session (T3), the peer assessment- and reflection tool became available for all conditions so that all participants could assess their peers and reflect on their behaviours. Finally, all participants completed an evaluation questionnaire.

**Instruments**

**Virtual Collaborative Research Institute (VCRI)**

The Virtual Collaborative Research Institute (VCRI) is a groupware program that supports collaborative working and learning on research projects and inquiry tasks (Jaspers, Broeken, & Erkens, 2004). The VCRI contains more than 10 different tools, but only 6 were used for this experiment (see Figure 1).

**Peer assessment tool (Radar)**

VCRI was augmented with a peer assessment tool for stimulating and facilitating group-functioning awareness. This tool provided group members with information about their own behaviour towards the other group members, and the functioning of the group as a whole. Group members assessed themselves and their peers by rating several variables in an interactive radar diagram; named Radar.

The goal of this design-study was to develop a peer feedback tool that is easy to use and to interpret. A radar diagram is appropriate because it is capable of visualising the output of multiple persons on multiple variables. The radar diagram for self- and peer assessment consisted of five variables, namely: (1) influence; (2) friendliness; (3) cooperation; (4) reliability; and (5) productivity (see Figure 2). These variables are based on
studies that focus on interpersonal perceptions, interaction, and group functioning (e.g., Bales, 1988; Brok, Brekelmans, & Wubbels, 2006; Kenny, 1994), and studies that focus on group dynamics, group processes, and group effectiveness (Forsyth, 1999; Salas, Sims, & Burke, 2005). These variables, as well as the reasons for their choice, are discussed in Phielix, Prins, and Kirschner (in preparation).

In the Radar, all group members are both assessor and assessee. In the role of assessor, the to-be-assessed peer in the group can be selected and her/his profile will appear as dotted lines in the centre circle of the radar diagram. Each group member is represented by a specific colour in the Radar. The assessor rates her/himself and all of the other group members on the five variables mentioned earlier, using a continuous scale ranging from 0 to 4 (0 = none, 1 = less, 2 = average, 3 = high, 4 = very high). The ratings are automatically saved in a database. To simplify data-analysis, the ratings are transformed to a scale from 0 to 100 by multiplying the ratings by 25. The assessment is anonymous; group members can see the output of the assessments of the other group members, but cannot see who entered the data.

After all group members have completed their self- and peer assessments, two modified radar diagrams become available on the screen. The first - Information about yourself - shows the output of the self assessment (e.g., Chris about Chris) along with the average scores of the peer assessments of her/him (e.g., Group about Chris). The self-assessment is not taken into account when the average scores are computed. To provide the students with more information about the variance in the average score their assessment by their peers, they can see the individual assessments of the other group members about their own behaviour (e.g., Group members about Chris). The second - Information about the group (see Figure 2) - represents the average scores of the group members, so that group members can get a general impression about the functioning of the group.

All group members are represented as a solid line in the diagram, each with a different colour. It is possible for the student to exclude data in the diagram. The student can decide which group member to include in or exclude from the diagram by clicking a name in the legend. It is also possible for group members to compare their self-assessments with the average scores of the assessments of their peers.

Figure 2. Output group assessment

Reflection tool (Reflector)
VCRI was also augmented with a reflection tool (Reflector) containing five reflective questions designed to stimulate reflection on different aspects of the group processes taking place. The questions were:
- What is your opinion on how the group functioned? Give arguments to support this.
- What do you contribute to the functioning of the group? Give examples.
- What do the other members of your group think about your functioning in the group? Why do you think
What is your opinion on how you functioned in the group? Give arguments to support this.

What does the group think about its functioning in general? Discuss and formulate a conclusion that is shared by all the group members.

The first four questions are completed in the Reflector, and completion is indicated by clicking an ‘Add’-button. This allows the student to share her/his answers with the rest of the group and allows her/him to see the answers of the others. Students can only gain access to the answers of their peers after they have added their own answers so as not to be influenced by one another. The fifth question is completed in Co-Writer which allows writing a ‘shared’ conclusion.

Results

Group Radar

Table 1 shows the means and standard deviations of the scales usefulness, group functioning awareness and feedback dialogues, concerning the Group Radar. For all the scales a 5-point Likert scale was used. The scale usefulness consisted of 10 items \((\alpha = .70)\). Overall students are positive about the usefulness of the tool \((M = 3.53, SD = .51, N = 35)\). According to the majority of the students the data entry is easy (69%), as well as the interpretation of the output was easy to understand (89%). A small majority (51%) of the students would like to have the Radar during on-line collaboration, and 74 % think that the Radar is useful.

Table 1: Means and Standard Deviations of the Scales Usefulness, Group Functioning Awareness and Feedback dialogues.

<table>
<thead>
<tr>
<th>Scales</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness Radar</td>
<td>35</td>
<td>3.53</td>
<td>.51</td>
</tr>
<tr>
<td>Group functioning awareness Radar</td>
<td>35</td>
<td>3.48</td>
<td>.73</td>
</tr>
<tr>
<td>Feedback dialogues</td>
<td>35</td>
<td>3.21</td>
<td>.86</td>
</tr>
</tbody>
</table>

Students were positive about the achieved group functioning awareness \((k = 4, \alpha = .83)\), by the Radar \((M = 3.48, SD = .73, N = 35)\). Students stated that Radar provided them with information about their own functioning in the group, and that it stimulated them to reflect on their own functioning in the group. Students also stated that Radar provided them new information about the functioning of their group members and the group as a whole. The perceptions of whether Radar affected interaction was measured with a 5-point Likert scale \((k = 2, \alpha = .69)\). Students using Radar stated that it stimulated dialogues on group functioning \((M = 3.21, SD = .86, N = 35)\). However, an independent samples t-test showed no significant differences in percentage frequencies of the five types of utterances in the chat history, between groups with and without Radar.

To examine whether the Radar had any effect on group functioning during the collaboration process, the average ratings between the first, second and third peer assessments were compared. Table 2 shows the results of a paired samples t-test between the first, second, and third assessments, with influence, friendliness, cooperation, reliability and productivity as dependent variables. Self-assessments are excluded.

Table 2. Paired Samples t-test between Peer Assessments on T1, T2 and T3 \((n=54)\).

<table>
<thead>
<tr>
<th>Paired differences between assessment T1-T2</th>
<th>assessment T2-T3</th>
<th>assessment T1-T3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Influence</td>
<td>3.15</td>
<td>14.45</td>
</tr>
<tr>
<td>Friendliness</td>
<td>6.02**</td>
<td>13.49</td>
</tr>
<tr>
<td>Cooperativeness</td>
<td>3.50*</td>
<td>12.10</td>
</tr>
<tr>
<td>Reliability</td>
<td>5.01*</td>
<td>17.85</td>
</tr>
<tr>
<td>Productivity</td>
<td>-2.23</td>
<td>12.03</td>
</tr>
</tbody>
</table>

* \(p < .05\) (2 tailed)  
** \(p < .01\) (2 tailed)

On average, all ratings on the second assessment decreased compared to the first assessment, except for productivity. Students perceived significantly less friendliness \((t = 3.28; df = 53; p < .01)\), less cooperativeness \((t = 2.13; df = 53; p < .05)\), and less reliability \((t = 2.07; df = 53; p < .05)\) at the second assessment. On average, all ratings on the third assessment increased compared to the second assessment. Compared to the second
assessments, students perceived significantly more friendliness ($t = -3.04; df = 53; p < .01$) and more reliability ($t = -3.01; df = 53; p < .01$) at the third assessment. The ratings of the third assessment increased towards the values of the first assessment. No significant differences in means were found between the first and third assessment.

The effect of Radar on group functioning was also measured using four 5-point Likert scales in the questionnaire that addressed team development, group satisfaction, level of intra group conflicts, and attitude towards problem solving. A two way between-groups ANOVA was conducted to explore the effect of Radar and Reflector on team development, group satisfaction, group conflicts and attitude towards collaborative problem solving. Participants were divided into four groups according to their condition. There were no significant interaction effects between Radar and Reflector, and no significant main effects for Reflector. There was a main effect for Radar on team development, $F(1, 30) = 4.19, p = .05$, with a medium effect size (partial eta squared = .12), level of group conflict, $F(1, 31) = 4.49, p = .04$, with a medium effect size (partial eta squared = .13), and attitude towards collaborative problem solving, $F(2, 31) = 1.44, p = .04$ (one-tailed), with a medium effect size (partial eta squared = .13).

An independent t-test was conducted to examine the main effects of Radar on team development, group conflict and attitude towards problem based collaboration. Conditions +Ra~Rf and +Ra+Rf were combined into a new group named ‘with Radar’, and conditions ~Ra+Rf and ~Ra~Rf were combined into group ‘without Radar’ (see Table 3).

<table>
<thead>
<tr>
<th>Scale</th>
<th>Treatment</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Mean difference</th>
<th>p</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team development</td>
<td>with radar</td>
<td>16</td>
<td>4.08</td>
<td>.35</td>
<td>.26*</td>
<td>.04</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>without radar</td>
<td>18</td>
<td>3.82</td>
<td>.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group satisfaction</td>
<td>with radar</td>
<td>17</td>
<td>3.95</td>
<td>.55</td>
<td>.00</td>
<td>.49</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>without radar</td>
<td>18</td>
<td>3.95</td>
<td>.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level of group conflict</td>
<td>with radar</td>
<td>17</td>
<td>1.79</td>
<td>.37</td>
<td>- .38*</td>
<td>.03</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>without radar</td>
<td>18</td>
<td>2.17</td>
<td>.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude towards collaborative problem solving</td>
<td>with radar</td>
<td>17</td>
<td>3.89</td>
<td>.39</td>
<td>.32*</td>
<td>.04</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>without radar</td>
<td>18</td>
<td>3.57</td>
<td>.62</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < .05 (1-tailed)

The results in Table 3 show that groups with Radar (+Ra~Rf and +Ra+Rf) scored significantly higher on team development, $t (32) = 1.79, p = .04$ (one tailed), experienced a significantly lower level of group conflicts $t (36) = -2.03, p = .03$ (one tailed), and had a significantly more positive attitude towards collaborative problem solving, $t (29) = 1.84, p = .04$ (one tailed), than groups without Radar (~Ra+Rf and ~Ra~Rf).

**The five components of the Radar**

Factor analysis was carried out to determine whether the Group Radar measured five independent components of group functioning. One component was found for groups without Radar, two components were found for groups with Radar. The first component consisted of the variables influence, friendliness and cooperativeness. The second component consisted of reliability and productivity. Additionally, correlations were calculated between the five variables that were envisioned as affected by the peer-assessment tool (see Table 4), and compared between groups with and without Radar. All variables at the final peer assessment (T3) for the groups without Radar correlated relatively strongly with each other. However, correlations were considerably lower in the groups with Radar (see Table 4). In comparison with the groups without Radar, ‘Reliability’ no longer significantly correlated with ‘Influence’, ‘Friendliness’ and ‘Cooperativeness’. ‘Reliability’ only correlated significantly with ‘Productivity’ ($r = .55; p < .01; n = 54$).

**Reflector**

Table 5 shows the means and standard deviations of the scales usefulness, group functioning awareness and feedback dialogues for the Reflector. For all the scales a 5-point Likert scale was used. The usefulness of the Reflector was measured by a 5-point Likert scale ($k = 9, \alpha = .78$) in the questionnaire. A majority of the students (63%; $n = 20$) found that the questions in the Reflector were clear, but overall students were not very positive about the functionalities of the tool ($M = 2.86, SD = .54, N = 32$). The majority (91%) of the students stated that
they do not need a tool as the Reflector during on-line collaboration.

Table 4. Intercorrelations between the Variables on Peer Assessment T3 for Groups With Radar (n = 54).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Influence</th>
<th>Friendliness</th>
<th>Cooperativeness</th>
<th>Reliability</th>
<th>Productivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Influence</td>
<td>-</td>
<td>.58**</td>
<td>.47**</td>
<td>.25</td>
<td>.39**</td>
</tr>
<tr>
<td>Friendliness</td>
<td>-</td>
<td></td>
<td>.52**</td>
<td>.05</td>
<td>.29*</td>
</tr>
<tr>
<td>Cooperativeness</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>.41**</td>
</tr>
<tr>
<td>Reliability</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>.55**</td>
</tr>
<tr>
<td>Productivity</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** p < .01 (2 tailed)
* p < .05 (2 tailed)

Table 5. Means and Standard Deviations of the Scales Utility, Usability and Awareness

<table>
<thead>
<tr>
<th>Scales</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness Reflector</td>
<td>32</td>
<td>2.86</td>
<td>.54</td>
</tr>
<tr>
<td>Group functioning awareness Reflector</td>
<td>32</td>
<td>2.91</td>
<td>.83</td>
</tr>
<tr>
<td>Feedback dialogues</td>
<td>32</td>
<td>2.72</td>
<td>.84</td>
</tr>
</tbody>
</table>

Students were not very positive about the achieved group functioning awareness (k = 2, α = .61), by the Reflector (M = 2.91, SD = .83, N = 32). According to the students, the Reflector did not really stimulate them to reflect on their functioning in the group, or provided them new information about group functioning.

To examine the perceptions of whether the Reflector affected interaction between group members (i.e., amount of peer feedback dialogues concerning group functioning) a 5-point Likert scale was used (k = 2, α = .66). Students were asked whether the Reflector stimulated dialogues on group functioning, but the majority responded negatively (M = 2.72, SD = .907, N = 35). Analysis of the chat history showed no significant differences in percentage frequencies of the five types of utterances, between groups with and without Reflector.

A two way between-groups ANOVA was conducted to explore the effect of Radar and Reflector on group cognitive performance, as measured by the grade given to their essays. There were no significant interaction effects between Radar and Reflector, and no significant main effects for Radar or Reflector.

At the time this paper was written the qualitative analyses of the chat history and output of the Reflector were still in progress.

Discussion & Conclusion

In this design study, the usefulness and effects of a peer assessment tool and a reflection tool were examined. A CSCL-environment was augmented with a peer-assessment tool named Radar, and a reflection tool, named Reflector. It was assumed that Radar, in combination with Reflector, would positively affect social interaction, group functioning, and group performance. Note that not all data is analysed, so no definitive conclusions can be derived on how and to what extent the Radar and/or Reflector affect social interaction, group functioning and group performance. Nevertheless, several conclusions can be derived from the preliminary results concerning the design and effects of the tools, and the used method.

First, the design and effects of the Radar. Results show that students perceived the tool as useful, easy to use, and easy to interpret. According to the students, the Radar increased group functioning awareness and stimulated dialogues on group functioning. As expected, main effects were found for Radar on team development, group conflict, and attitude towards collaborative problem solving. However, no effects were found for group satisfaction and group performance (grade given for the essay).

Factor analysis showed that the Radar only measured two components of group functioning, instead of five. The first component consisted the variables ‘influence’, ‘friendliness’ and ‘cooperativeness’, the second component consisted ‘reliability’ and ‘productivity’. Additionally, correlations were calculated between these five variables, and compared between groups with and without Radar. All variables at the final peer assessment (T3) for the groups without Radar correlated relatively strongly with each other. However, correlations were considerably lower in the groups with Radar. In comparison with the groups without Radar, ‘Reliability’ no longer significantly correlated with ‘Influence’, ‘Friendliness’ and ‘Cooperativeness’. ‘Reliability’ only
correlated significantly with ‘Productivity’. These results indicate that three out of five variables (e.g., friendliness, cooperativeness and productivity) need to be replaced. Nevertheless, it appears that halfway the collaboration process the Radar has an effect on the individual behaviour of the group members. The results of the groups with Radar show a significant decrease in mean, on three out of five variables, halfway the collaboration process. Students perceived their group as significantly less friendly, less cooperative, and less reliable, in comparison of the first assessment (T1). The data, however, does not allow for the analysis of whether these differences were caused by the presence of Radar or Reflector. Due to the design of this study, the control groups did not have an assessment at the halfway point or at the beginning of the experiment. An explanation for the decrease in means halfway the collaboration process, could be that Radar provides the group members a more realistic, and less positive view on group functioning. This would be in line with findings of Homma, Tajima and Hayashi (1995), and Stroebe, Diehl and Abakoumkin (1992), who found that group members intuitively estimate the quantity and quality of their group product and their personal contributions, and that these estimates are generally unrealistically positive, resulting in an illusion of group productivity.

Second, the design and effects of the Reflector. Based on the results of the questionnaire, students stated that the Reflector itself has no added value during on-line collaboration. However, qualitative analyses of the chat history and output of the Reflector are still in progress. Although the effects of this study are mainly ascribed to the Radar, it is still assumed that Radar in combination with Reflector will be most effective. An explanation for finding no significant main effects for the Reflector on social group performance could be that in this study the Reflector was focussed on past and present (and not on future) group functioning, which might have caused superficial reflections, lacking reflections on future group behavior. Therefore, in further studies the design of the Reflector will be changed and will also be focussed on future group functioning, that is, on stimulating group members to formulate plans and set goals for improving social and cognitive group performance.

Research has shown that outcome feedback can increase individual and group performance, especially when it is combined with goal setting (Mento, Steel, & Karren, 1987; Neubert, 1998; Tubbs, 1986), and there is no reason to believe why this should be different for process feedback.

Third, the method design of this study. Several limitations of this study should be kept in mind. The statistical power of this study is rather low because of the relatively small sample size ($N = 39$). However, even with this small sample, significant main effects were found for Radar on team development, level of group conflict and attitude towards collaborative problem solving. In this study Radar is both intervention as measurement tool for the dependent variables (e.g., Influence and Friendliness). Therefore, with the current design it was not possible to determine whether the decrease of self-assessment and peer assessment scores halfway collaboration at T2, was caused by the Radar or Reflector, or whether this also occurred in the control group. Therefore, in future studies, an extra control group will be added in which the Radar will become available at T2.

It is possible that the effects of the Radar and Reflector can only be measured over a longer period of time. In this study, the time between the first and last assessment was one week. Therefore, a second design study will be carried out amongst 25 third-year university student, who will collaborate, over a period of five months, on their bachelorthesis.

In sum, results with Radar are promising. They show that social group functioning in CSCL environments, such as team development, level of group conflicts and attitude towards collaborative problem solving, can be enhanced by adding an easy to complete and easy to interpret peer feedback tool, such as Radar. For Reflector, it was argued that the focus of the Reflector’s questions should be directed towards future group performance and goal setting.

References


