



Group awareness of social and cognitive performance in a CSCL environment: Effects of a peer feedback and reflection tool

Chris Phielix^{a,*}, Frans J. Prins^a, Paul A. Kirschner^{a,b}, Gijsbert Erkens^a, Jos Jaspers^a

^a Utrecht University, The Netherlands

^b Open University of the Netherlands, The Netherlands

ARTICLE INFO

Article history:

Available online xxx

Keywords:

Groupware
Peer assessment
Peer feedback
Reflection
Group performance

ABSTRACT

A peer feedback tool (Radar) and a reflection tool (Reflector) were used to enhance group performance in a computer-supported collaborative learning environment. Radar allows group members to assess themselves and their fellow group members on six traits related to social and cognitive behavior. Reflector stimulates group members to reflect on their past, present and future group functioning, stimulating them to set goals and formulate plans to improve their social and cognitive performance. The underlying assumption was that group performance would be positively influenced by making group members aware of how they, their peers and the whole group perceive their social and cognitive behavior in the group. Participants were 108 fourth-year high school students working in dyads, triads and groups of four on a collaborative writing task, with or without the tools. Results demonstrate that awareness stimulated by the peer feedback and reflection tools enhances group-process satisfaction and social performance of CSCL-groups.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Collaborative learning supported by computer networks (computer-supported collaborative learning; CSCL), while enjoying considerable interest at all levels of education (Strijbos, Kirschner, & Martens, 2004) is sometimes hampered by social problems that arise between team members (Hobman, Bordia, Irmer, & Chang, 2002; Lipponen, Rahikainen, Lallimo, & Hakkarainen, 2003). To this end, self and peer assessments can be used to provide useful feedback on group functioning, as they allow group members to better judge their own and other's behavior and contributions to the group and thus avoid the social problems often encountered (Dochy, Seegers, & Sluijsmans, 1999). Though positive effects of self and peer assessments have been reported (e.g., Cutler & Price, 1995; McDowell, 1995; Phielix, Prins, & Kirschner, 2010), they are also prone to biases. Saavedra and Kwun (1993) found, for example, that group members tend to overestimate their own performance which affects their peer assessments. To counteract this, Dochy et al. propose that a combination of peer feedback and reflection could enhance the validity of self and peer assessments, and possibly also enhance behavioral change (e.g., Prins, Sluijsmans, & Kirschner, 2006). In the study described here, a CSCL environment was augmented with a

peer feedback tool (Radar) and a reflection tool (Reflector) to help make group members better aware of individual and group behavior and stimulate them to set goals and formulate plans for improving the group's social and cognitive performance. The study has two main goals, namely to examine the effects of these tools on perceived social and cognitive group behavior and to examine social and cognitive performance (i.e., its development, the group satisfaction, the level of group conflict, member attitude towards collaborative problem solving, and the quality of the product).

1.1. Social and cognitive processes in CSCL

Collaborative learning is the 'mutual engagement of participants in a coordinated effort to solve the problem together' (Rochelle & Teasley, 1995, p. 70). A key to successful collaborative learning is social interaction (Kreijns, Kirschner, & Jochems, 2003; Liaw & Huang, 2000; Northrup, 2001), not only for the cognitive (i.e., task-related) processes in collaboration, such as discussion, reasoning, reflection, critical thinking and creating a shared understanding of the problem (e.g., Garrison, Anderson, & Archer, 2001; Johnson & Johnson, 1999; Kreijns et al., 2003), but also for the social (i.e., non-task-related) processes, such as developing positive affective relationships, feelings of group cohesiveness and trust, and achieving a sense of community (e.g., Boud, Cohen, & Sampson, 1999; Johnson, Johnson, & Smith, 2007; Kreijns & Kirschner, 2004). These social processes allow group members to get to know and understand each other so as to become a 'healthy'

* Corresponding author. Address: Research Centre Learning in Interaction, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands. Tel.: +31 (30) 253 4941; fax: +31 (30) 253 2352.

E-mail address: c.phielix@uu.nl (C. Phielix).

community of learning (Gunawardena, 1995). Both cognitive and social processes are necessary to collaboratively complete a task, solve a problem or construct knowledge (Kreijns et al., 2003). These processes, however, will not automatically occur by simply bringing learners together (Fischer, Bruhn, Gräsel, & Mandl, 2002), thus CSCL should provide an environment that supports both cognitive and social processes (Kreijns et al., 2003).

1.2. Social and cognitive benefits of groups in CSCL environments

The rapid development of information and communication technologies has led to computer applications such as e-mail, chat tools, discussion forums, video conferencing, and so forth which have proven useful to supporting learning (e.g., Janssen, Erkens, Kanselaar, & Jaspers, 2007; Phielix et al., 2010). While often considered inferior, recent research has shown both cognitive and social benefits for groups in CSCL environments as compared to contiguous (i.e., face-to-face) groups such as more equal participation between team members (Fjermestad, 2004; Janssen et al., 2007), higher satisfaction (Fjermestad), better team development, less conflict, and a more positive attitude (Phielix et al., 2010).

There are, however, also contradictory results. Students learning in CSCL environments have been shown to experience lower participation levels (Lipponen et al., 2003), more conflict (Hobman et al., 2002), less group cohesion (Straus, 1997; Straus & McGrath, 1994) and less satisfaction (Baltes, Dickson, Sherman, Bauer, & LaGanke, 2002) than students in contiguous groups. In other words, students working in CSCL environments do not always reach their full potential.

1.3. Effects of CSCL-design on social and cognitive behavior

Two important reasons for the disparity between the potential and the performance of groups learning in CSCL environments lie in the design of the CSCL environment, and the actual or perceived social and cognitive behavior of the group members. With respect to the former, most CSCL environments focus on supporting cognitive or task-related processes in collaboration and limit possibilities for social or non-task-related processes (Kreijns and Kirschner, 2004). For instance, despite technological advances, most CSCL environments are still text-based computer-mediated communication systems using e-mail, chat and/or discussion boards which cannot easily convey visual, non-verbal cues (Kreijns et al., 2003). The absence of these cues can cause specific communication and interaction problems since there are few possibilities to exchange socio-emotional and affective information, and there is little information about group members' presence, self-image, attitudes, moods, actions and reactions (Short, Williams, & Christie, 1976). According to Short et al. these non-verbal cues are related to forming, building and maintaining social relationships. Such lean systems can, thus, negatively affect impression formation and social behavior (e.g., Garton & Wellman, 1995; Walther, Anderson, & Park, 1994).

With respect to the latter (i.e., behavior), group members form interpersonal perceptions during interaction (Kenny, 1994) based on what they see and experience. They form impressions (e.g., norms, values, beliefs) about themselves, the group, other group members, and what the other group members think of them. These impressions are based on perceived cognitive behaviors (e.g., a person's or the team's productivity) and social behaviors (e.g., the dominance and/or friendliness of team members) during interaction. Based upon these perceptions, group members determine their own social and cognitive behavior, and develop relationships with each other. However, research has shown that self perceptions of performance and perceptions of group performance are generally unrealistically positive in contiguous groups (Saavedra and Kwun, 1993; Stroebe, Diehl, & Abakoumkin, 1992; Yammarino & Atwater,

1997) and in computer-mediated groups (Phielix et al., 2010; Weisband & Atwater, 1999). These unrealistic positive perceptions often result in an *illusion* of group productivity (Stroebe et al., 1992); a tendency to believe that a group is performing effectively, while it often is not. This can result in reduced effort by group members (i.e., social loafing; Williams, Harkins, & Latané, 1981), further undermining the group's social and cognitive performance. Unfortunately, group members are often not aware that they are loafing or are unwilling to admit it (Karau & Williams, 1993).

To overcome these unrealistic interpersonal perceptions of the group's social and cognitive performance, CSCL environments can be augmented with tools or widgets that influence the social interaction (Kirschner, Strijbos, Kreijns, & Beers, 2004). Such tools, also known as 'social affordance devices', can positively affect social and cognitive performance in a CSCL environment (Kirschner et al.). Two operationalizations of such tools are used in this study, namely a *peer feedback tool* and a *reflection tool*. These tools are intended to make group members aware of their individual behavior and that of the group. It is assumed that this awareness is necessary for group members to set goals and formulate plans for improving social and cognitive group performance. The next sections deal with aspects central to these tools, namely peer feedback and reflection.

1.4. Peer feedback

Feedback is information provided to increase performance (Kluger & DeNisi, 1996). This feedback can (1) relate to outcome or process feedback, (2) be given by individuals (i.e., teachers, tutors, peers) or groups, and (3) be received by individuals or groups. The study reported on here centers on peer process feedback at the individual and group level to enhance interpersonal behavior. It is expected that enhancing interpersonal behavior will positively affect a group's social performance (Geister, Konradt, & Hertel, 2006; McLeod & Liker, 1992; Phielix et al., 2010), indirectly affecting its cognitive performance (Kreijns et al., 2003).

This process feedback can be cognitive or task-related (e.g., related to task behaviors, actions and strategies), or social or non-task-related (e.g., giving information on interpersonal behavior or teamwork) (Geister et al., 2006). McLeod and Liker (1992) found that group-level process feedback on interpersonal behavior influenced dominance behavior of individual group members. Two other studies investigating individualized peer feedback on interpersonal group member behavior (e.g., communication and collaboration) found that it led to increased cooperation, communication, satisfaction, and motivation (Dominick, Reilly, & McGourty, 1997; Druskat & Wolff, 1999).

Self and peer assessments are a form of feedback often used for formative assessment and has been found to foster reflection on the student's own learning process and learning activities (Dochy et al., 1999). Boud and Falchikov (1989) define self assessment as students making judgments about their own learning, mainly about their achievements and learning outcomes. Peer assessment, defined by Falchikov (1995) as the process through which groups of individuals rate their peers is specified in this study as students monitoring and assessing the social and cognitive performance of their peers. Peer assessment becomes peer feedback when the information is shared among peers to increase performance. Somervell (1993) stresses that providing peer assessment can be seen as a part of the self assessment process, informing self assessment. However, self and peer assessments are seldom objective. During completion of self and peer assessments, students do a lot of mental comparisons (Goethals, Messick, & Allison, 1991), which are selected, interpreted, and/or biased (Saavedra & Kwun, 1993). Students tend to emphasize their strengths and positive performances, and perceive weakness and negative performances as

common in and caused by others (Campbell, 1986; Goethals, 1986; Klein, 2001; Saavedra & Kwun, 1993; Suls & Wan, 1987).

Combining peer feedback and reflection can enhance the validity of self and peer assessments (e.g., Dochy et al., 1999) and enhance behavioral change (e.g., Prins et al., 2006).

1.5. Reflection

Providing group members' with information on their cognitive and social behavior is not enough to positively alter their behavior (Prins et al., 2006). Group members also need to process this information and ask themselves whether they understand, accept, and agree with the feedback. In other words, they must reflect upon the feedback. Reflection is the intellectual and affective activities individuals engage into explore their experiences to reach new understandings and appreciations of those experiences (Boud, Keogh, & Walker, 1985). The feedback receiver needs to be challenged to reflect on his/her own performance, and determine whether the feedback provides clues for behavioral change (Prins et al., 2006). Peer feedback with reflection should, thus, be more effective than feedback alone (e.g., Schön, 1987).

Reflection can lead to new perspectives on experience, changes in behavior, readiness for application, and commitment to action (Boud et al., 1985). Reflection on peer feedback, thus, should make group members more aware of their own behavior, how it affects others, and whether they should alter it. This awareness allows "understanding of the activities of others, which provides a context for your own activity" (Dourish & Bellotti, 1992, p. 107).

Phielix et al. (2010) developed a shared reflection tool (Reflector) to stimulate group members to reflect on their own past and present performances as well as that of the group as a whole to enhance group performance. In that study, no significant main effects were found for Reflector on group performance. This was ascribed to the fact that the tool was not focused on future functioning and goal setting (Hattie & Timperley, 2007; Mento, Steel, & Karren, 1987; Neubert, 1998; Tubbs, 1986). For feedback to be effective, the receiver needs to answer three major questions: (1) Where am I going?/What are the goals? (feed up), (2) How am I going?/What progress is being made toward the goal? (feed back), and (3) Where to next?/What activities need to be undertaken to make better progress? (feed forward) (Hattie & Timperley, 2007). Therefore, in this follow-up study, Reflector was redesigned to make group members better aware of their individual and group behavior, and to stimulate them to set goals and formulate plans to enhance social and cognitive group performance. Group members using Reflector individually reflect and provide information on (1) their own perspective on their personal performance (feed up), (2) differences between their self perception and the perception of their peers concerning their personal performance (feed back), (3) whether they agree with those perceptions (feed back), and (4) their individual perspective on group performance (feed up). Because group performance is determined by the individual effort of all group members, Reflector also (5) stimulates group members to collaboratively reflect (i.e., co-reflect) on group performance and reach a shared conclusion on this (feed back). Based on their shared conclusion, group members (6) set goals to improve group performance (feed forward). Co-reflection is defined as "a collaborative critical thinking process involving cognitive and affective interactions between two or more individuals who explore their experiences in order to reach new intersubjective understandings and appreciations" (Yukawa, 2006, p. 206).

1.6. Research questions

This study investigated whether peer feedback tool and reflection would enhance social and cognitive group performance in a

CSCL environment. To this end, an existing CSCL environment was augmented with two independent, but complementary, tools. The first was an individualized peer feedback tool – *Radar* – to stimulate and provide group members with information about the social and cognitive behavior of themselves, their peers, and the group as a whole. This information was presented from the perspectives of the group members themselves (i.e., self perceptions), their peers (i.e., peer perceptions) and the group as a whole. The second was a shared reflection tool – *Reflector* – to stimulate group members to co-reflect on social and cognitive group performance and to set goals and formulate plans to enhance group performance. In the first experimental condition, group members used Radar at the start of the collaboration (T1), and Radar and Reflector halfway through (T2) and at the end (T3). In order to examine the effect of the tools halfway through the collaboration process, in the second experimental condition, group members used Radar and Reflector halfway through the collaboration (T2) and at the end (T3). In the control condition, the group members used the tools only at the end (T3).

The following research questions were addressed:

(1) Do group members in condition 1 (tools at T1, T2, and T3) perceive that there is less social and cognitive behavior at the second assessment (T2) compared to their perceptions of that behavior at the first assessment (T1)?

Hypothesis. Peer feedback should make group members aware of unrealistically positive self and peer perceptions, resulting in more realistic perceptions of their functioning and, thus, a decrease of self and peer assessment scores at a subsequent assessment (e.g., Phielix et al., 2010). In other words, group members in condition 1 should exhibit lower self assessment and peer assessment scores at T2 as compared to T1.

(2) Do group members in condition 1 (tools at T1, T2, and T3) perceive that there is less social and cognitive behavior halfway through (T2), than group members in condition 2 (tools at T2 and T3) who used Radar for the first time at T2?

Hypothesis. Peer feedback provided by Radar should make group members aware of their unrealistically positive self and peer perceptions (Phielix et al., 2010), resulting in more realistic perceptions of their functioning and, thus, in a decrease in self and peer assessment scores at a subsequent assessment. In other words, group members in condition 1 should exhibit lower self and peer assessment scores at T2 than group members in condition 2.

(3) Do group members in condition 1 (tools at T1, T2, and T3) and condition 2 (tools at T2 and T3) perceive that there is more social and cognitive behavior at the end of the collaboration process (T3) compared to halfway through (T2) and, for condition 1, the beginning (T1)?

Hypothesis. Information from Radar and Reflector halfway through the task (T2) stimulates group members to set goals to improve their own and the group's social and cognitive performance. In other words, group members using Radar and Reflector at T2 should exhibit higher self assessment and peer assessment scores at the end (T3) compared to scores halfway through (T2) and, for condition 1, at the beginning (T1).

(4) Do group members in condition 1 (tools at T1, T2, and T3), perceive that there is better social and cognitive behavior than group members in condition 2 (tools at T2 and T3) and condition 3 (tools at T3)?

Hypothesis. Information from Radar and Reflector stimulates group members to set goals to improve their own and the group's social and cognitive performance. In other words, group members in condition 1 should perceive the social and cognitive behavior at T3 to be better than group members in conditions 2 and 3. Also, group members in condition 2 should perceive social and cognitive behavior at T3 to be better than group members in condition 3.

(5) Do group members using Radar and Reflector show more congruence between self and peer assessment scores at a subsequent assessment?

Hypothesis. Group members need time to adjust their unrealistically positive self perceptions, and thus non-significant or small correlations should be found between self and peer assessments after the first completion of Radar, but significant and higher correlations should be found at a subsequent assessment. Also significant differences should be found between self and peer assessments after the first completion of Radar, and these differences should become non-significant or smaller at a subsequent assessment.

(6) Do group members using Reflector set goals and formulate plans to enhance social and cognitive group performance?

Hypothesis. The reflective questions in Reflector stimulate group members to co-reflect on social and cognitive performance and to set goals and formulate plans to enhance social and cognitive group performance.

(7) Do group members in condition 1 (tools at T1, T2, and T3), perceive that there is higher social performance (i.e., better team development, higher group satisfaction, less group conflict, and more positive attitudes towards collaborative problem solving) at T3, than group members in condition 2 (tools at T2 and T3) and condition 3 (tools at T3)?

Hypothesis. Radar and Reflector should positively affect social behavior in the group leading to increased social performance of the group. In other words, group members in condition 1 should perceive higher social performance at T3 than group members in conditions 2 and 3. Also, group members in condition 2 should perceive higher social performance at T3 than group members in condition 3.

(8) Do groups in condition 1 (tools at T1, T2, and T3), exhibit higher cognitive performance (i.e., produce higher quality group products) at T3, than groups in condition 2 (tools at T2 and T3) and condition 3 (tools at T3)?

Hypothesis. Radar and Reflector, by positively affecting social behavior in the group, should indirectly increase cognitive performance of the group. In other words, groups in condition 1 should exhibit higher cognitive performance at T3, than groups in conditions 2 and 3. Also, groups in condition 2 should exhibit higher cognitive performance at T3 than groups in condition 3.

2. Methods

2.1. Participants

Participants were 108 fourth-year Dutch high school students (58 male, 50 female) in four classes with an average age of 16

($M = 15.85$, $SD = .50$, $Min = 15$, $Max = 18$). Prior to the experiment, they were randomly assigned by the teacher to dyads ($n = 16$), triads ($n = 84$) and groups of four ($n = 8$), and randomly assigned by the researcher to one of three conditions (see Section 2.2). Groups were heterogeneous in ability and gender.

2.2. Design

For this study two experimental conditions and one control condition were used (Table 1). The first experimental condition ($n = 59$) received Radar and Reflector at the beginning (T1), halfway (T2) and at the end (T3) of the collaboration process. The second experimental condition ($n = 23$) received these tools halfway (through T2) and at the end (T3) of the collaboration. The control condition ($n = 26$) did not receive the tools during collaboration, but completed them at the end of the collaboration (T3) as measurement instruments.

2.3. Measures (see Table 2)

2.3.1. Social behavior

Perceived group social behavior is measured by the self and peer assessments in Radar on four variables, namely 'influence', 'friendliness', 'cooperativeness' and 'reliability'. These variables are rated on a continuous scale ranging from 0 to 4 (0 = none, 4 = very high). To simplify data-analysis, the ratings are transformed to a scale from 0 to 100 by multiplying the ratings (0–4) by 25.

2.3.2. Cognitive behavior

Perceived group cognitive behavior is measured by the self and peer assessments in Radar on the variables 'productivity' and 'quality of contribution', rated on a continuous scale ranging from 0 to 4 (0 = none, 4 = very high). The same transformation was carried out here.

2.3.3. Cognitive performance

The grade given to each group's collaborative writing task (i.e., an essay) was used as a measure of cognitive performance. The essays were graded by two researchers, both experienced in grading essays. The inter-rater reliability was substantial (Cohen's Kappa = .80).

2.3.4. Social performance

Four previously validated instruments (Strijbos, Martens, Jochems, & Broers, 2007) were translated into Dutch and transformed into five-point Likert scales (1 = totally disagree, 5 = totally agree; see Table 2). The team development scale provides information on perceived level of group cohesion. The group-process satisfaction scale provides information on perceived satisfaction with general group functioning. The intra-group conflicts scale provides information on perceived level of conflict between group members. The attitude towards collaborative problem-solving scale provides information on perceived level of group effectiveness and how group members felt about working and solving problems in a group. The 30 items in the four scales were subjected to principal component analysis. Prior to performing this analysis, the suitability of data for factor analysis was assessed. Inspection of the correlation matrix showed that all coefficients were .5 and higher. The Kaiser–Meyer–Oklin value was .73, exceeding the recommended value of .6 and Bartlett's test of sphericity reached statistical significance, supporting the factorability of the correlation matrix. The analysis revealed the presence of one main component with eigenvalues exceeding 1, explaining 76.6% of the variance respectively. Cronbach's alpha of the composed 'Social Performance (total)' scale is .90 (Table 3).

Table 1

Design overview.

Condition	T1 – beginning	T2 – halfway	T3 – end
(1) Tools at T1, T2, and T3 (n = 59)	Radar	Radar and Reflector	Radar, Reflector, and Questionnaire
(2) Tools at T2 and T3 (n = 23)		Radar and Reflector	Radar, Reflector, and Questionnaire
(3) Tools at T3 (n = 26)			Radar, Reflector, and Questionnaire

Table 2

Overview of scales, subscales, and instruments.

Scale	Subscales	Instrument
Social behavior	Influence, friendliness, cooperation, reliability	Radar
Cognitive behavior	Productivity, quality of contribution	Radar
Social performance	Team development, group-process satisfaction, intra-group conflicts, attitude towards collaborative problem solving	Questionnaire
Cognitive performance	–	Essay grade

Table 3

Examples of social performance scales.

Scale	k	Example	Cronbach's alpha
Team development	10	Group members contribute ideas and solutions to problems	.77
Group-process satisfaction	6	I felt that my group worked very hard together to solve this problem	.71
Intra-group conflicts	7	I found myself unhappy and in conflict with members of my group	.84
Attitude towards collaborative problem solving	7	Collaborating in a group is challenging	.74
Social performance (total)	30	(See all items of four scales stated above)	.90

2.3.5. Coding scheme output co-reflection

To improve the social and cognitive performance of the group, group members reflect together (i.e., co-reflect) to set goals and formulate plans to improve their social and cognitive activities. Categories for the coding scheme were derived from studies on social interaction and coordination processes in CSCL and were added until there were no 'rest categories'. Finally, two independent researchers coded and categorized the goals and plans in nine categories (see Table 4). Inter-rater reliability was substantial (Cohen's Kappa = .79). The first three categories are *communication*, *focusing on task* and *task coordination*, activities which are crucial for successful collaboration (Barron, 2003; Erkens, Jaspers, Prangma, & Kanselaar, 2005; Slof, Erkens, Kirschner, Jaspers, & Jansen, in press). Furthermore, students need to carry out meta-cognitive activities such as *planning* and *monitoring* to employ a proper problem-solving strategy and reflect on its suitability (Lazonder & Rouet, 2008; Narciss, Proske, & Koerndle, 2007). Students also must develop positive affective relationships with each other (Kreijns et al., 2003), thus *friendliness* is a sixth category. *Productivity* and *quality* are the seventh and eighth categories because in effective groups, group members mutually depend on the willingness, effort and participation of their peers (Janssen et al., 2007; Karau & Williams, 1993; Williams et al., 1981). The category *no suggestion* is added for students who do not have any suggestions to improve their performance.

2.4. Task and procedure

Students collaborated in dyads and groups of three or four on a collaborative writing task in sociology. Every student worked at a computer. Each group had to write an essay on a highly relevant current-events topic. Prior to this collaborative writing task, students collaborated for 1 month choosing the topic, searching for relevant sources of information, writing a short paper and giving a presentation. Thus, all information needed to write the essay was available for all groups. The collaborative writing task consisted of three 45-min sessions over a period of one week. The groups collaborated in a CSCL environment called Virtual Collaborative Research Institute (Jaspers, Broeken, & Erkens, 2004), a groupware program that supports collaborative learning on research projects and inquiry tasks (see Section 2.5). Students were instructed to use the environment to communicate with other group members and to make complete use of the tools for peer feedback and reflection when the experimental condition allowed this. Students received content information and definitions regarding the six traits on which they had to assess themselves and their peers. Students were told that they had three lessons to complete the task, that it would be graded by their teacher, and that it would affect their grade for the course. The introduction to the task stressed the importance of working together as a group and pointed out that each individual group member was responsible

Table 4

Coding scheme for output co-reflection: specific goals to improve group performance.

Label	Code	Description	Example
Communication	Com	Improve communication or teamwork	We will discuss the content of our work more often
Focusing on task	Focus	Improve concentration or focus on task	We will focus more on our work
Task coordination	Task	Improve coordination, task- or role-planning	We will divide the tasks more effectively
Planning	Plan	Improve time planning	We will decide faster, because it's taking too long
Monitoring	Mon	Improve peer monitoring	We will let our peers monitor our progression
Friendliness	Friend	Improve friendliness towards each other	We should not be so unfriendly towards each other
Productivity	Prod	Improve productivity	We will increase our productivity
Quality	Qual	Improve quality of work	We will improve the quality of our work
No suggestions	None	No suggestions for improvement	We have no suggestions for improvement

for the successful completion of the group task. To successfully complete the task, all group members had to participate.

While groups used the tools, groups without them worked on the collaborative writing task. Time-on-task (i.e., writing the essay) was equal for all conditions. At the end of the final session (T3), both tools were made available for all conditions so that all participants could assess their peers and reflect on their behaviors. Finally, all participants completed a 30-item questionnaire on the social performance of the group.

2.5. Instruments

2.5.1. 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 et al., 2004). VCRI contains more than 10 different tools, but only five were used for this experiment (see Fig. 1). Co-Writer (top left) is a shared word-processor for writing a group text. Using Co-Writer, students can simultaneously work on different parts of their texts. The chat tool (top center) is used for synchronous communication. The chat history is automatically stored and can be re-read by participants at any time. Notes (bottom right) is a note pad which allows the user to make notes and copy and paste selected information. Radar for peer feedback (bottom left) and Reflector for reflection (top right) will be described in the following sections. Windows of the available tools are automatically arranged on the screen when students log onto the VCRI.

2.5.2. Peer feedback tool (Radar)

VCRI was augmented with a peer feedback tool for eliciting information on group members' social and cognitive behavior. This

information is visualized in a Radar diagram; therefore the peer feedback tool is named Radar (see Fig. 2). Radar provides users with anonymous information on how their cognitive and social behaviors are perceived by themselves, their peers, and the group as a whole with respect to specific traits found to tacitly affect how one 'rates' others (see Brok, Brekelmans, & Wubbels, 2006). Radar provides information on six traits deemed important for assessing behavior in groups. Four are related to social or interpersonal behavior, namely (1) influence, (2) friendliness, (3) cooperation, (4) reliability, and two are related to cognitive behavior, namely (5) productivity and (6) quality of contribution. These traits are derived from studies on interpersonal perceptions, interaction, group functioning, and group effectiveness (e.g., Bales, 1988; Brok et al., 2006; Kenny, 1994; Salas, Sims, & Burke, 2005).

Influence is directly derived from Wubbels, Créton, and Hooymayers's (1985) influence dimension (i.e., dominance vs. submissiveness) in their model for interpersonal teacher behavior. This dimension is also used by Bales (1988) and represents the prominence, status, power, and personal influence that the individual is seen to have in relation to other group members. The variable is labeled 'influence', and not 'dominance' or 'submissive', because the latter two can be perceived of as negative traits.

Friendliness is one of the eight behavior categories from Wubbels et al.'s (1985) model for interpersonal teacher behavior. Bales (1988) used a similar dimension (i.e., friendliness vs. unfriendliness). Bales and Cohen (1979) defined this as the extent to which individual members are friendly with and respectful to each other.

Cooperation, which denotes the degree to which someone is willing to work with others, is derived directly from Wubbels et al.'s (1985) dimension proximity (i.e., opposition vs. cooperation) They defined proximity as the property of being close together, or in group settings as the feeling of being a group (i.e., group cohesiveness).

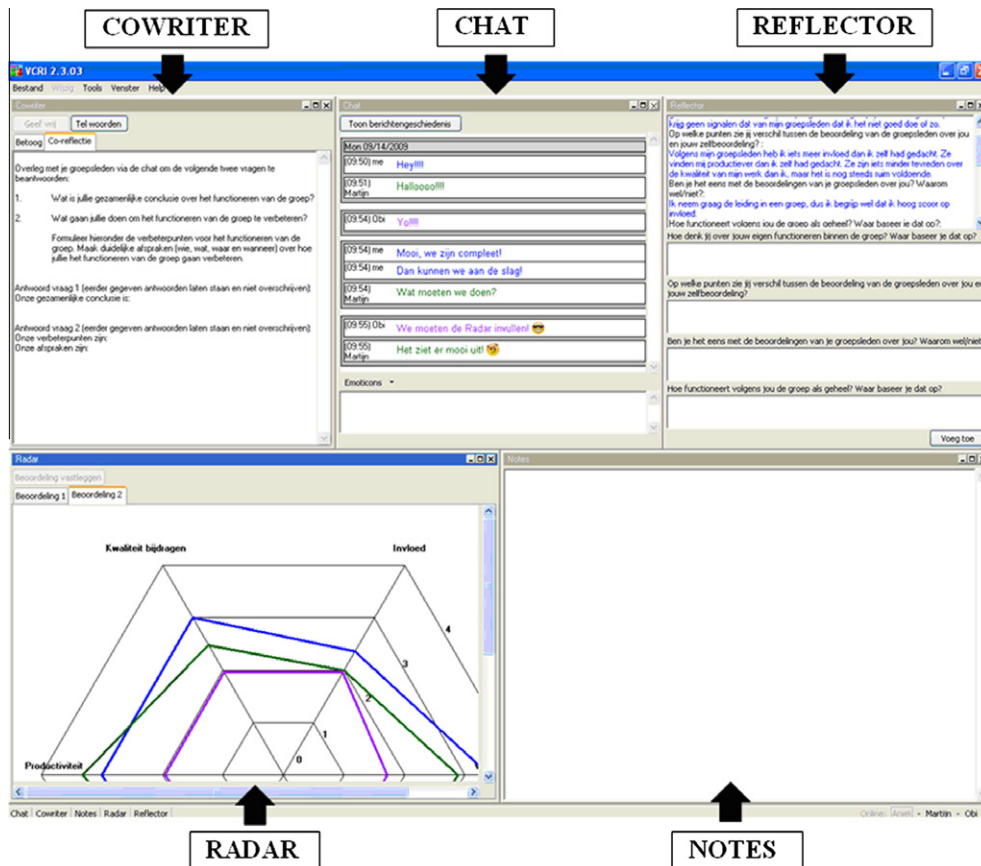


Fig. 1. Screenshot of VCRI with the five tools used in this experiment.

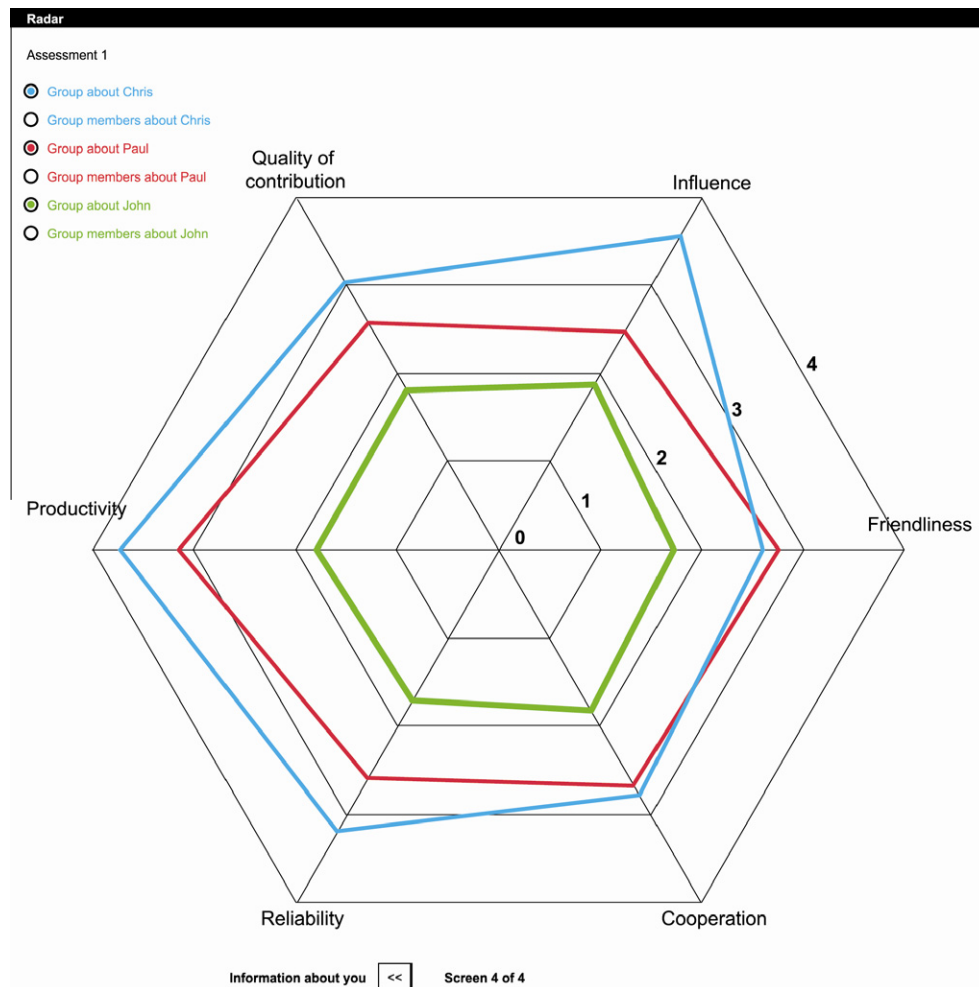


Fig. 2. Radar output: group assessment.

Reliability is a trait reflecting 'trust' which has been identified as an important precursor for successful collaboration, in face-to-face teams (Castleton_Partners/TCO., 2007) and in CSCL (Jarvenpaa & Leidner, 1999). According to Emans, Koopman, Rutte, and Steensma (1996) trust can be seen as the cognitive and affective assurance of group members that they respect each other's interests and, therefore, can orient themselves towards each other's words, actions, and decisions with an easy conscience.

Productivity and *quality of contribution* are the extent to which individual group members contribute quantitatively and qualitatively to tasks or duties central to group performance or group efficiency. These traits, representing cognitive or task-related behavior, were selected because research has shown that group members (1) do not always participate equally (Karau & Williams, 1993) and (2) monitor the performance (i.e., quantity and quality) of other group members (Salas et al., 2005).

In Radar, group members are both assessors and assessees. As assessor, to-be-assessed peers in the group can be selected and their profile will appear as dotted lines in the center circle of the Radar diagram. Each group member is represented by a specific color. Assessors rate themselves and all of the other group members on each of the six traits using a continuous scale ranging from 0 to 4 (0 = none, 4 = very high). Each range, (e.g., from 0 to 1) contains 10 points, so every scale contained 40 points of assessment. Care was taken to ensure that all assessors use the same definition of the six traits. Prior to the experiment the researcher notified participants of text balloons with content information and definitions

that would appear when they moved the cursor across one of the traits in the tool. For example, when the assessor moves the cursor across 'influence' a balloon pops up with the text 'A high score on influence means that this person has an influence on what happens in the group, on the behavior of other group members, and on the form and content of the group product (the essay)'.

Ratings are automatically saved in a database. To simplify data-analysis, ratings are transformed to a 100-point scale by multiplying the ratings (0–4) by 25.

For groups of 3 and 4 members, the assessment is anonymous. Group members can see the assessments of the other group members, but not who entered the data. To stimulate students to complete Radar, they can only gain access to the individual and average assessments of their peers after they have completed the assessment themselves. When all group members have completed their self and peer assessments, two modified Radar diagrams become available. 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 for computing the average scores. To provide more information about the variance in the average score of their peer assessment, students can also choose to view the individual peer assessments about their own behavior (e.g., Group members about Chris). The second – information about the group (see Fig. 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 color. Participants can include or exclude group members from the diagram by clicking a name in the legend.

It is assumed that information (peer feedback) from Radar makes group members aware of the differences between their intended behavior (measured by self assessment) and how this behavior is perceived by their peers (measured by peer assessment). It is also assumed that every group member will be stimulated to improve his/her social and cognitive behavior, knowing that (1) every group member will be assessed by his/her peers and (2) these scores will be shared (anonymously) among the group. Therefore, it is expected that group members using Radar throughout will exhibit higher self and peer assessment scores on all six traits at the end compared to halfway through and the beginning.

2.5.3. Reflection tool (Reflector)

VCRI was also augmented with a reflection tool (Reflector) for stimulating group members to co-reflect on their individual behavior and overall group performance. This tool contained six reflective questions:

- (1) What is your opinion of how you functioned in the group? Give arguments to support this.
- (2) What differences do you see between the assessment that you received from your peers and your self assessment?
- (3) Why do or don't you agree with your peers concerning your assessment?
- (4) What is your opinion of how the group is functioning? Give arguments to support this.
- (5) What does the group think about its functioning in general? Discuss and formulate a conclusion shared by all the group members.
- (6) Set specific goals (who, what, when) to improve group performance.

The first four questions are completed in the Reflector, with completion 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 their peers' answers after they have added their own so as not to influence each another. The last two questions are completed in Co-Writer, in a specific section named Co-Reflection, which allows writing a shared conclusion and formulating shared goals. Responses made by the students in the Reflector are not scored or evaluated.

2.6. Data Analyses

To examine whether group members in condition 1 (tools at T1, T2, and T3) exhibit lower self assessment and peer assessment scores halfway (T2) compared to the beginning (T1), a one-way repeated measures ANOVA with dependent variables related to perceived social and cognitive behavior (i.e., influence, friendliness, cooperation, reliability, productivity, and quality of contribution) is used to compare self assessment and peer assessment scores at T1, T2, and T3.

To examine whether group members in condition 1 (tools at T1, T2, and T3) exhibit lower self assessment and peer assessment scores at T2 than groups in condition 2 (tools at T2 and T3), intra-class correlations will be calculated to examine group effects, after which multilevel analyses will be carried out to examine the effect of the tools halfway through (at T2) with respect to perceived social and cognitive behavior as measured by Radar.

To examine whether group members using Radar and Reflector perceive more social and cognitive behavior at the end of the task compared to previous assessments, a one-way repeated measures ANOVA with dependent variables influence, friendliness, cooperation, reliability, productivity, and quality of contribution is used to compare self and peer assessment scores for condition 1 at T1, T2, and T3. This is followed by a paired samples *t*-test (one-tailed) to compare self and peer assessment scores for condition 2 at T2 and T3.

To examine whether group members in condition 1 (tools at T1, T2, and T3) perceive better social and cognitive behavior at T3 than their peers in condition 2 (tools at T2 and T3) or condition 3 (control; tools at T3), intra-class correlations will be calculated to determine whether group effects exist. After this, multilevel analyses will be used to examine the effect of the tools at the end of the experiment (T3) with respect to perceived social and cognitive behavior as measured by Radar.

To examine whether group members using Radar and Reflector show more congruency between self and peer assessment scores at a subsequent assessment, a Pearson product-moment correlation coefficient is used to test congruency between self and peer assessments at T1, T2, T3 with respect to perceived social and cognitive behavior. Additionally, an independent *t*-test (one-tailed) will be used to examine the differences between self and peer assessments scores per condition at T1, T2, and T3 with respect to perceived social and cognitive behavior.

To examine whether group members using Reflector set goals and formulate plans to enhance social and cognitive group performance, goals and plans were independently coded and categorized by two researchers. Mean frequencies per group of goals and plans will be presented per condition in a table.

To examine whether group members in condition 1 (tools at T1, T2, and T3) perceive higher social performance (i.e., better team development, higher group satisfaction, less group conflict, and more positive attitudes towards collaborative problem solving) than group members in condition 2 (tools at T2, T3) and condition 3 (tools at T3), intra-class correlations will be calculated to examine group effects. Then, multilevel analysis will be used to examine the effect of condition on the dependent variables social performance, team development, group satisfaction, level of group conflicts, and attitude towards collaborative problem solving, as measured by the questionnaire at the end of the experiment.

Finally, to examine whether groups in condition 1 (tools at T1, T2, and T3) exhibit higher cognitive performance than groups in condition 2 (tools at T2 and T3) and condition 3 (tools at T3), a one-way between-groups ANOVA (one-tailed) with planned comparisons is conducted with the grade on the essay as dependent variable.

Except where noted, tests were one-tailed. The rule of thumb (Cohen, 1988, pp. 284–287) for effects sizes (η^2) was small $\geq .01$, medium $\geq .06$, and large $\geq .14$.

3. Results

Hypothesis 1. Peer feedback leads to awareness of unrealistically positive self and peer perceptions, resulting in lower self and peer assessment scores halfway (comparing T1 and T2 for condition 1).

Table 5 shows the mean scores and standard deviations of self assessment scores at T1, T2, and T3 for each condition.

A one-way repeated measures ANOVA was conducted to compare self assessment and peer assessment scores for condition 1 at the beginning (T1), halfway through (T2), and at the end (T3), with respect to perceived social and cognitive behavior. Unexpectedly, students in condition 1 (tools at T1, T2, and T3) exhibited

Table 5

Mean and standard deviations of self assessments per condition.

T	Condition	N	Influence		Friendliness		Cooperative		Reliability		Productivity		Quality of contribution	
			M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
1	1	59	64.85	13.90	72.75	15.96	68.97	17.19	70.42	18.01	64.68	13.82	65.98	13.75
2	1	59	70.20	16.76	76.46	16.54	73.78	20.97	72.86	18.05	71.93	18.28	72.75	14.78
	2	23	62.74	14.37	64.96	17.85	62.13	14.02	68.70	16.51	63.52	11.04	65.70	10.02
3	1	59	71.15	15.97	77.90	16.66	73.59	19.18	75.36	16.74	72.98	16.82	72.19	15.97
	2	23	64.26	9.89	66.39	17.12	62.91	16.15	65.57	14.87	66.00	10.47	71.48	11.61
	3	26	65.85	13.34	67.00	16.69	68.42	16.90	68.31	14.75	60.96	12.70	63.81	11.38

significantly higher self assessment scores halfway through (T2) compared to the beginning (T1). There were significant effects for influence (mean difference = 5.36, 95% CI: 1.21–9.50), Wilks' Lambda = .79, $F(2, 57) = 7.39$, $p < .001$, partial $\eta^2 = .21$; productivity (mean difference = 7.25, 95% CI: 1.84–12.66), Wilks' Lambda = .81, $F(2, 57) = 6.73$, $p < .005$, partial $\eta^2 = .19$; and for quality of contribution (mean difference = 6.76, 95% CI: 2.39–11.14), Wilks' Lambda = .80, $F(2, 57) = 7.13$, $p < .005$, partial $\eta^2 = .20$. Group members that used the tools throughout perceived themselves halfway through as having more influence, being more productive, and making higher quality contributions.

Table 6 shows the mean scores and standard deviations of average peer assessments scores at T1, T2, and T3 per condition. Unexpectedly, students in condition 1 (tools at T1, T2, and T3) exhibited significantly higher average peer assessment scores halfway through (T2) compared to the beginning (T1). There were significant effects for influence (mean difference = 6.30, 95% CI: 2.39–10.20), Wilks' Lambda = .77, $F(2, 57) = 8.38$, $p < .005$, partial $\eta^2 = .23$; cooperation (mean difference = 5.02, 95% CI: .43–9.60), Wilks' Lambda = .89, $F(2, 57) = 3.58$, $p < .05$, partial $\eta^2 = .11$; productivity (mean difference = 9.43, 95% CI: 4.75–14.11), Wilks' Lambda = .63, $F(2, 57) = 1.69$, $p < .0005$, partial $\eta^2 = .37$; and quality of contribution (mean difference = 4.94, 95% CI: 1.65–8.21), Wilks' Lambda = .78, $F(2, 57) = 7.90$, $p < .005$, partial $\eta^2 = .22$. Group members that used the tools throughout perceived their peers halfway through as having more influence, being more cooperative, productive, and making higher quality contributions.

Hypothesis 2. Peer feedback leads to awareness of unrealistically positive self and peer perceptions, resulting in lower self and peer assessment scores halfway through (comparing condition 1 with condition 2).

Intra-class correlations were calculated to examine group effects, followed by multilevel analysis to examine the effect of condition halfway through (at T2) with respect to perceived social and cognitive behavior as measured by the Radar (Table 7).

As significant intra-class correlations were found for friendliness, cooperation, and reliability, indicating that there is a group effect, multilevel analyses were used to examine the differences

on these variables between conditions 1 and 2. The associated significant χ^2 -values indicate that conditions (tools) have an effect on how group members perceived their own social and cognitive behavior (i.e., influence, friendliness, etc.). Unexpectedly, the significant β -value shows that group members in condition 1 (tools at T1, T2, and T3) perceived themselves as having more influence, being more friendly, cooperative, and productive, and making higher quality contributions, than group members in condition 2 (tools at T2 and T3). No significant differences were found for reliability (Table 8).

As significant intra-class correlations were found indicating that there is a group effect for all dependent variables, multilevel analyses were carried out to examine the differences on these variables between conditions 1 and 2. The associated significant χ^2 -values indicate that conditions (tools) have an effect on group members perceived social and cognitive behavior (i.e., influence, friendliness, etc.). Unexpectedly, the significant β -value shows that group members in condition 1 (tools at T1, T2, and T3) perceived more social behavior (i.e., influence, friendliness, cooperation), and cognitive behavior (i.e., productivity and quality of contribution), than group members in condition 2 (tools at T2 and T3). No significant differences were found for reliability.

Hypothesis 3. Using Radar and Reflector leads to higher self assessment and peer assessment scores at the end compared to previous assessments.

A one-way repeated measures ANOVA was conducted to compare self assessment and peer assessment scores for condition 1 at the beginning (T1), halfway (T2), and at the end (T3), with respect to perceived social and cognitive behavior. See results of Hypothesis 1 for the comparison self assessment scores and peer assessment scores at the beginning (T1) and halfway through (T2) for condition 1. Group members in condition 1 (tools at T1, T2, and T3) exhibited significantly higher self assessment scores at the end (T3) compared to the beginning (T1), there were significant effects for reliability (mean difference = 4.93, 95% CI: .10–9.76), Wilks' Lambda = .89, $F(2, 57) = 3.64$, $p < .05$, partial $\eta^2 = .11$; productivity (mean difference = 8.30, 95% CI: 2.74–13.87), Wilks' Lambda = .81, $F(2, 57) = 6.73$, $p < .005$, partial $\eta^2 = .19$; and quality

Table 6

Mean and standard deviations of average peer assessments per condition.

T	Condition	N	Influence		Friendliness		Cooperative		Reliability		Productivity		Quality of contribution	
			M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
1	1	59	65.52	10.86	75.90	13.58	69.02	12.62	69.79	13.49	64.10	11.77	68.40	9.57
2	1	59	71.82	12.29	77.81	15.56	74.04	15.90	72.19	16.31	73.53	13.32	73.33	11.36
	2	23	63.35	8.38	67.54	11.91	61.57	12.70	69.93	8.63	64.98	7.12	65.43	9.51
3	1	59	71.36	12.48	77.29	15.26	73.09	15.77	74.96	14.13	74.31	10.67	73.76	11.31
	2	23	64.91	8.18	70.35	14.32	64.98	9.75	67.30	9.78	70.43	9.82	68.67	9.17
	3	26	66.62	12.86	72.73	8.67	68.88	15.68	68.52	12.34	65.13	12.35	68.31	12.45

Table 7
Multilevel analyses for comparing condition 1 and condition 2 on self assessment scores at T2.

Scale	Intra-class correlations	Comparing condition 1 vs. 2		Chi-square
	<i>r</i>	β	SE	χ^2
Influence	-.10	7.46*	3.92	8.15*
Friendliness	.56***	10.94*	5.61	8.94*
Cooperation	.29*	11.98*	5.53	9.77**
Reliability	.36**	3.67	5.41	5.67
Productivity	.14	8.46*	4.38	8.42*
Quality of contribution	.05	7.03*	3.50	8.25*

* $p < .05$ (one-tailed).
 ** $p < .01$ (one-tailed).
 *** $p < .001$ (one-tailed).

Table 8
Multilevel analyses for comparing conditions 1 and 2 on peer assessment scores at T2.

Scale	Intra-class correlations	Comparing condition 1 vs. 2		Chi-square
	<i>r</i>	β	SE	χ^2
Influence	.46***	8.47*	3.53	9.72**
Friendliness	.70***	10.37*	5.31	9.14*
Cooperation	.66***	12.47*	5.33	11.00**
Reliability	.57***	2.26	4.97	5.26*
Productivity	.59***	8.55*	4.06	9.10**
Quality of contribution	.53***	7.90*	3.53	9.61**

* $p < .05$ (one-tailed).
 ** $p < .01$ (one-tailed).
 *** $p < .001$ (one-tailed).

of contribution (mean difference = 6.20, 95% CI: 1.33–11.07), Wilks' Lambda = .80, $F(2, 57) = 7.13$, $p < .005$, partial $\eta^2 = .20$. No significant differences in self assessment scores were found comparing scores halfway (T2) with scores at the end (T3).

For peer assessment scores, groups in condition 1 (tools at T1, T2, and T3) exhibited significantly higher average peer assessment scores at the end (T3) compared to the beginning (T1). There were significant effects for influence (mean difference = 5.84, 95% CI: 1.72–9.95), Wilks' Lambda = .77, $F(2, 57) = 8.38$, $p < .005$, partial $\eta^2 = .23$; reliability (mean difference = 5.17, 95% CI: .57–9.77), Wilks' Lambda = .88, $F(2, 57) = 3.88$, $p < .05$, partial $\eta^2 = .12$; productivity (mean difference = 10.22, 95% CI: 5.92–14.51), Wilks' Lambda = .63, $F(2, 57) = 1.69$, $p < .0005$, partial $\eta^2 = .37$; and quality of contribution (mean difference = 5.36, 95% CI: 1.67–9.05), Wilks' Lambda = .78, $F(2, 57) = 7.90$, $p < .005$, partial $\eta^2 = .22$. No significant differences in average peer assessment scores were found comparing scores halfway (T2) with scores at the end (T3).

To compare self and peer assessment scores at T3 with T2 for condition 2, a paired samples t -test (one-tailed) showed significant

differences for students in condition 2 for self assessments halfway (T2) and at the end (T3). As expected, students in condition 2 (tools at T2 and T3) perceived a significantly higher quality of contribution at T3, (mean difference = 5.78, $SD = 10.66$, 95% CI: 1.17–10.39), $t(22) = 2.60$, $p < .01$. Significant differences were also found between peer assessment scores halfway (T2) and at the end (T3). As expected, students perceived significantly more productivity at T3, (mean difference = 5.46, $SD = 9.43$, 95% CI: 1.42–9.50), $t(22) = 2.80$, $p < .01$.

Hypothesis 4. Groups using Radar and Reflector perceive more social and cognitive behavior at the end than groups not using these tools.

In comparing self assessment scores for condition 1 (tools at T1, T2, and T3) with condition 2 (tools at T2 and T3) and condition 3 (tools at T3), significant intra-class correlations were found for friendliness, cooperation, and reliability indicating a group effect (see Table 9). Multilevel analyses were, thus, used to examine the differences between conditions 1 and 2. The associated significant χ^2 -values indicate that conditions (tools) have an effect on how group members perceived their own social and cognitive behavior (i.e., influence, friendliness, etc.). As expected, the significant β -values show that group members in condition 1 perceived themselves as having more influence, being friendlier, more cooperative, more productive, and making contributions of higher quality, than group members in condition 2. No significant differences were found for reliability. Comparing conditions 1 and 3, group members in condition 1 perceived themselves as being more friendly, more productive, and making contributions of higher quality than group members in condition 3. No significant differences were found for influence, cooperation, and reliability.

Comparing self assessment scores for condition 2 (tools at T2 and T3) with condition 3 (tools at T3), the significant β -value (see Table 9) shows that group members in condition 2 perceived themselves as making contributions of higher quality than group members in condition 3. No other significant differences in self assessment scores were found between conditions 2 and 3.

In comparing peer assessment scores for condition 1 (tools at T1, T2, and T3) with condition 2 (tools at T2 and T3) and condition 3 (tools at T3), significant intra-class correlations indicated a group effect for all dependent variables (see Table 10). Multilevel analyses were, thus, used to examine the differences between conditions 1 and 2. The associated significant χ^2 -values indicate that conditions (tools) have an effect on group members perceived social and cognitive behavior (i.e., influence, friendliness, etc.). As expected, significant β -values show that group members in condition 1 perceived their peers having more influence and being more reliable than group members in condition 2. Unexpectedly, no significant differences in peer assessment scores were found for

Table 9
Multilevel analyses for effects of condition on self assessment scores at T3.

Scale	Intra-class correlation	Comparing condition 1 vs. 2		Comparing condition 1 vs. 3		Chi-square χ^2	Comparing condition 2 vs. 3		Chi-square χ^2
		β	SE	β	SE		β	SE	
Influence	.13	6.89*	3.81	5.30	3.65	13.10***	-1.59	4.37	4.88*
Friendliness	.42**	11.51*	5.13	10.90*	4.91	17.16***	-.61	5.15	5.10**
Cooperation	.30**	11.17*	5.26	5.17	5.04	14.80***	-5.51	5.51	6.35**
Reliability	.33**	9.79*	4.71	7.05	4.51	15.03***	-2.74	4.68	5.25*
Productivity	.02	6.98*	3.59	12.02*	3.44	21.07***	5.04	3.45	6.41**
Quality of contribution	.09	.71	3.61	8.38*	3.46	14.65***	7.67*	3.96	8.29**

* $p < .05$ (one-tailed).
 ** $p < .01$ (one-tailed).
 *** $p < .001$ (one-tailed).

Table 10
Multilevel analyses for effects of condition on peer assessment scores at T3.

Scale	Intra-class correlation r_t	Comparing condition 1 vs. 2		Comparing condition 1 vs. 3		Chi-square χ^2	Comparing condition 2 vs. 3		Chi-square χ^2
		β	SE	β	SE		β	SE	
Influence	.34***	6.45*	3.53	4.74	3.38	12.89***	-1.71	3.55	4.59*
Friendliness	.74***	6.94	5.13	4.56	4.91	12.43**	-2.38	5.30	5.51**
Cooperation	.77***	8.11	5.63	4.21	5.40	12.99***	-3.90	5.84	6.01**
Reliability	.62***	7.66*	4.50	6.44	4.31	13.82***	-1.22	4.19	4.82*
Productivity	.27*	3.88	3.04	9.18**	2.91	17.37***	5.30	3.36	6.72**
Quality of contribution	.35**	5.09	3.37	5.45*	3.23	12.36**	.36	3.79	4.48*

* $p < .05$ (one-tailed).

** $p < .01$ (one-tailed).

*** $p < .001$ (one-tailed).

friendliness, cooperation, productivity, or quality of contribution. Comparing conditions 1 and 3, as expected, group members in condition 1 perceived more productivity and higher quality contributions than group members in condition 3. No significant differences were found for perceived social behavior (i.e., influence, friendliness, cooperation, and reliability).

Finally, in comparing peer assessment scores for condition 2 (tools at T2 and T3) with condition 3 (tools at T3) no significant differences in peer assessment scores at T3 were found.

Hypothesis 5. Group members need time to adjust their unrealistic positive self perceptions.

A Pearson product-moment correlation coefficient was used to test congruency between self and peer assessments at T1, T2, T3 with respect to perceived social and cognitive behavior. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity and/or homoscedasticity. The rule of thumb (Cohen, 1988) for the strength of the correlation (r) was small = .10–.29, medium = .30–.49, and large = .50–1.0. Table 11 shows the Pearson correlations between average peer assessments and self assessments at T1, T2, and T3.

As expected, results in Table 11 show non-significant or relatively small correlations for self and peer assessment scores at group members' first assessment, and significant medium to large correlations at subsequent assessments. For students in condition 1 (tools at T1, T2, and T3), nearly all correlations between self and peer assessment scores are significantly positive, except for productivity at T1. Compared to T1, all correlations increased significantly at T2. Compared to T2, correlations for friendliness, productivity and quality of contribution show small decreases at T3, but remain significantly positive. Compared to T2, correlations for influence, cooperativeness and reliability increased at T3. Compared to the correlations at the beginning (T1) this indicates a higher convergence of self and peer assessments at the end of

the collaboration process. For students in condition 2 (tools at T2 and T3), no significant correlations were found between self and peer assessment scores at T2. Compared to T2, significant correlations were found for influence, friendliness, productivity and quality of contribution at T3. For students in condition 3 (tools at T3), no significant correlations were found between self and peer assessment scores at T3.

An independent t -test (one-tailed) was used to examine differences between self and peer assessments at T1, T2, and T3 with respect to perceived social and cognitive behavior. Tables 5 and 6 show the mean scores and standard deviations of self and peer assessments per condition. No significant differences in self and peer assessments for condition 1 were found at T1, nor for conditions 1 and 2 at T2. At T3; students in condition 2 perceived their productivity significantly lower (mean difference = -4.43 , $SD = 9.02$), $t(22) = -2.36$, $p < .01$, than their peers. No significant differences in self and peer assessments were found at T3 for conditions 1 and 3.

Hypothesis 6. Groups using Reflector set goals and formulate plans to enhance social and cognitive group performance.

It was expected that the reflective questions in Reflector would stimulate group members to (co)reflect on social and cognitive performance, and set goals and formulate plans to enhance the social and cognitive group performance. To examine whether groups using Reflector set goals and formulated plans to enhance social and cognitive group performance, their responses were independently coded and categorized by two researchers. Mean frequencies, percentages and cumulative percentages of plans and goals per group and per condition, are presented in Table 12.

Groups in conditions 1 and 2 formulated twice as much plans and goals as condition 3, since groups in conditions 1 and 2 completed Reflector twice, compared to once for condition 3. Groups in condition 1 (tools at T1, T2, and T3) formulated goals focused on improving communication, task coordination and on better focusing on the task. On average, groups in condition 1 formulated

Table 11
Pearson correlations between average peer assessments and self assessments at T1, T2, and T3.

Condition	T	N	Influence		Friendliness		Cooperative		Reliability		Productivity		Quality of contribution	
			r	p	r	p	r	p	r	p	r	p	r	p
1	1	59	.28	.03*	.27	.04*	.52	.00**	.28	.03*	.23	.08	.32	.01*
	2	59	.62	.00**	.72	.00**	.54	.00**	.53	.00**	.62	.00**	.55	.00**
	3	59	.69	.00**	.66	.00**	.59	.00**	.57	.00**	.34	.01*	.46	.00**
2	2	23	-.18	.42	.07	.74	.13	.57	.02	.94	.24	.28	.06	.78
	3	23	.65	.00**	.47	.02*	.38	.07	.02	.92	.61	.00**	.52	.01*
3	3	26	.36	.07	-.2	.91	.30	.14	.14	.51	.01	.97	-.4	.86

* $p < .05$ (two-tailed).

** $p < .01$ (two-tailed).

Table 12
Mean frequencies, percentages and cumulative percentages of future goals per group per condition.

Label	Condition 1 tools at T1, T2, and T3 (n = 20)			Condition 2 tools at T2 and T3 (n = 8)			Condition 3 tools at T3 (n = 9)		
	Mean f	%	Cum. %	Mean f	%	Cum. %	Mean f	%	Cum. %
Communication	0.70	35.00	35.00	0.25	13.33	13.33	0.22	20.00	20.00
Task coordination	0.40	20.00	55.00	0.75	40.00	53.33	0.22	20.00	40.00
Focussing	0.20	10.00	65.00	0.13	6.67	60.00	0.22	20.00	60.00
Productivity	0.15	7.50	72.50	0.13	6.67	66.67	0.33	30.00	90.00
Quality	0.10	5.00	77.50	0.13	6.67	73.33	0.00	0.00	90.00
Planning	0.05	2.50	80.00	0.38	20.00	93.33	0.00	0.00	90.00
Friendliness	0.05	2.50	82.50	0.00	0.00	93.33	0.00	0.00	90.00
Monitoring	0.05	2.50	85.00	0.13	6.67	100.00	0.00	0.00	90.00
No suggestions	0.30	15.00	100.00	0.00	0.00	100.00	0.11	10.00	100.00
Goals per group	2.00	100.00		1.88	100.00		1.11	100.00	

nearly three times as many goals focused at improving communication than groups in conditions 2 or 3. Groups in condition 2 (tools at T2 and T3) formulated goals that were focused on improving task coordination, planning and communication. Groups in condition 3 (tools at T3) formulated goals focused on improving productivity, communication, task coordination and on focusing on the task. On the average, groups in condition 3 formulated three times as many goals focused on improving productivity than groups in conditions 1 or 2. The five most often mentioned goals relate to focusing on improving task coordination, communication, productivity, and on focusing on the task. This indicates that groups formulated goals to achieve better teamwork, which can enhance their social and cognitive performance.

Hypothesis 7. Using Radar and Reflector leads to increased social performance.

It was expected that groups in condition 1 would perceive higher social performance at T3 than groups in conditions 2 and 3 and that groups in condition 2 perceive higher social performance at T3 than groups in condition 3. To examine whether groups in condition 1 (tools at T1, T2, and T3) perceive higher social performance (i.e., better team development, higher group satisfaction, less group conflict, and more positive attitudes towards collaborative problem solving) than groups in condition 2 (tools at T2 and T3) and condition 3 (tools at T3), first, intra-class correlations were calculated to examine any group effects on the social performance scales, and second, multilevel analysis was used to examine the effect of condition on dependent variables social performance (total), team development, group satisfaction, level of group conflicts, and attitude towards collaborative problem solving, as measured by the questionnaire at the end of the experiment, after collaboration in the CSCL environment (VCRI). Table 13 shows intra-class corre-

lations and multilevel analyses for effects of condition on social performance scales.

In comparing condition 1 (tools at T1, T2, and T3) with condition 2 (tools at T2 and T3) and condition 3 (tools at T3) on perceived social performance, significant intra-class correlations were found for all social performance scales except attitude towards collaborative problem solving, indicating that the group has an effect on the perceived social performance of individual group members. Multilevel analyses are, thus, needed to examine the effect of condition (tools) on perceived social performance. As expected, the significant β -value shows that groups in condition 1 (tools at T1, T2, and T3) perceived their team as being better developed than groups in condition 2 (tools at T2 and T3). However, no significant differences were found for team development between condition 1 (tools at T1, T2, and T3) and condition 3 (tools at T3), as indicated by a non-significant β -value. As expected, the significant β -value shows that groups in condition 1 (tools at T1, T2, and T3) experienced higher levels of group satisfaction, than groups in condition 2 (tools at T2 and T3) and condition 3 (tools at T3). These effects should be interpreted with caution as the associated χ^2 -value was only marginally significant ($p = .07$). As expected, groups in condition 1 (tools at T1, T2, and T3) experienced lower levels of conflicts, than groups in condition 2 (tools at T2 and T3), but no significant differences were found between condition 1 and condition 3 (tools at T3), as indicated by a non-significant β -value. Also this effect should be interpreted with caution, as the associated χ^2 -value was not significant ($p = .11$). As expected, groups in condition 1 (tools at T1, T2, and T3) had significantly more positive attitude towards collaborative problem solving than groups in condition 2 (tools at T2 and T3), but no significant differences were found for attitude between condition 1 (tools at T1, T2, and T3) and condition 3 (tools at T3). As expected, groups in condition 1 experienced significantly higher social performance than groups in conditions 2 and 3.

Table 13
Multilevel analyses for effects of condition on social performance scales.

Scale	Condition 1 tools at T1, T2, and T3 (n = 54)		Condition 2 tools at T2 and T3 (n = 22)		Condition 3 tools at T3 (n = 25)		Intra-class correlation r_i	Comparing condition 1 vs. 2		Comparing condition 1 vs. 3		Chi-square χ^2	Comparing condition 2 vs. 3		Chi-square χ^2
	M	SD	M	SD	M	SD		β	SE	β	SE		β	SE	
Team development	4.09	0.56	3.48	0.62	3.80	0.37	.60***	.61***	.18	.27	.18	7.84**	-.32	.18	2.01
Group-process satisfaction	4.04	0.52	3.59	0.61	3.71	0.59	.41**	.45**	.17	.31*	.17	3.90	-.13	.20	-1.00
Intra-group conflicts	1.95	0.58	2.43	0.67	2.23	0.50	.62***	-.48*	.21	-.27	.20	3.07	.20	.22	-0.19
Attitude	3.85	0.50	3.59	0.60	3.65	0.55	.03	.26*	.14	.19	.13	-.07	-.06	.17	-1.57
Social performance (total)	4.01	0.46	3.56	0.56	3.73	0.44	.49***	.45**	.15	.26*	.14	4.88*	-.17	.17	-0.49

* $p < .05$ (one-tailed).

** $p < .01$ (one-tailed).

*** $p < .001$ (one-tailed).

In comparing condition 2 (tools at T2 and T3) with condition 3 (tools at T3) on perceived social performance (see Table 13), no significant differences were found between conditions 2 and 3. The non-significant χ^2 -values indicate that the condition (tools) had no effect on the perceived social performance for condition 2.

Hypothesis 8. Using Radar and Reflector leads to higher cognitive performance.

It was expected that groups in condition 1 would exhibit higher cognitive performance at T3 than groups in conditions 2 and 3 and that groups in condition 2 would exhibit higher cognitive performance at T3 than groups in control condition 3. A one-way between-groups ANOVA (one-tailed) with planned comparisons was conducted to explore the effect of Radar and Reflector on group cognitive performance, as measured by the grade given to their (group) essays. Table 14 shows means and standard deviations for group performance per condition. No significant effects were found.

4. Discussion and conclusion

This study examined the effects of a peer feedback tool (Radar) and a reflection tool (Reflector) on (1) perceived social and cognitive behavior, and (2) social and cognitive performance of the group. Most of the expectations were met. Results showed that groups using tools throughout perceive better social and cognitive behavior halfway through, show more convergence between self and peer assessments, and report higher social group performance, than groups not using tools. Results did not show a decrease of self and peer assessment scores halfway through, indicating that there is no support that Radar and Reflector can help to reduce group members unrealistic positive perceptions of self and peer performance. There was also no support that using Radar and Reflector indirectly leads to higher cognitive performance. Below, the eight research questions of this study, accompanied with results and explanations will be addressed.

The first question addressed in this study was to examine whether group members in condition 1 (tools at T1, T2, and T3) perceive less social and cognitive behavior halfway (T2) compared to the beginning (T1). Based on Stroebe et al. (1992) and the results of a previous study (Phielix et al., 2010), it was hypothesized that group members generally form unrealistically positive perceptions of self performance and peer performance. Therefore, we expected that the peer feedback provided by Radar at the first assessment (T1) would make group members better aware of these perceptions, resulting in a decrease of self and peer assessment scores at the second assessment (T2). Unexpectedly, most self and peer assessment scores increased significantly at the second assessment (T2). Group members using the tools throughout perceived their peers halfway through as having more influence, being more cooperative, productive, and making higher quality contributions. Therefore, the hypothesis that Radar and Reflector can reduce unrealistic positive perceptions of self and peer performance was not supported by the data. A possible explanation could be that,

in comparison to the previous study, these group members were familiar in collaborating with each other. Prior to this experiment, group members collaborated with each other for 1 month. It is likely that this former collaboration period caused students to have more realistic perceptions, and these perceptions increased over time by using Radar and Reflector.

The second question addressed was to examine whether group members in condition 1 (tools at T1, T2, and T3) perceive less social and cognitive behavior than group members in condition 2 (tools at T2 and T3), who used Radar for the first time at T2. As stated above, it was hypothesized that information provided by Radar should make group members aware of their unrealistic positive self perceptions and peer perceptions (Hattie & Timperley, 2007; Phielix et al., 2010). Therefore it was expected that group members in condition 1 (using Radar for the second time) exhibit lower self assessment and peer assessment scores at T2, than group members in condition 2 (using Radar for the first time). Unexpectedly, group members in condition 1 perceived more social behavior (i.e., influence, friendliness, cooperation), and cognitive behavior (i.e., productivity and quality of contribution) halfway through, than group members in condition 2. As stated above, an explanation could be that group members in condition 1, who completed Radar for the second time, managed to improve their social and cognitive behavior over time, resulting in significantly higher self and peer assessment scores at T2 compared to T1.

The third question addressed was to examine whether groups in condition 1 (tools at T1, T2, and T3) perceive more social and cognitive behavior at the end (T3) compared to halfway (T2) and the beginning (T1). Based on Hattie and Timperley (2007) and results of a previous study (Phielix et al., 2010) it was hypothesized that information ('feed back' and 'feed forward') from Radar and Reflector halfway (T2) would stimulate group members to set goals to improve the social and cognitive performance of themselves and the group. Therefore, it was expected that self assessment and peer assessment scores would be higher at the end (T3), compared to halfway (T2) and the beginning (T1). As expected, group members using the tools throughout perceived their peers at the end as having more influence, being more reliable, more productive, and making higher quality contributions, compared to the beginning. Unexpectedly, no significant differences were found comparing scores halfway (T2) with scores at the end (T3). An explanation could be that time was long enough for the tools to cause an effect between the beginning (T1) and the end (T3), but was too short to cause an effect between halfway (T2) and the end (T3).

The fourth question addressed was to examine whether group members in condition 1 (tools at T1, T2, and T3), perceive better social and cognitive behavior at the end (T3), than group members in condition 2 (tools at T2 and T3) or condition 3 (tools at T3). It was hypothesized that information from Radar and Reflector will stimulate group members to set goals to improve the social and cognitive performance of themselves and of the group. Significant χ^2 -values indicated that conditions (tools) have an effect on how group members perceived the social and cognitive behavior of themselves and their peers. As expected, group members using tools throughout (condition 1) perceived themselves as having more influence, being friendlier, more cooperative, more produc-

Table 14
Means and standard deviations for cognitive performance per condition.

Condition	n_{groups}	Cognitive performance (grade essay)			
		<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
(1) Tools available at T1, T2, and T3	20	6.81	0.71	4.0	8.5
(2) Tools available at T2 and T3	8	6.54	1.04	4.5	8.5
(3) Tools available at T3	9	6.36	1.61	4.0	8.5

tive, and making contributions of higher quality, than group members using tools since halfway (condition 2). Group members using tools throughout also perceived themselves as being more friendly, more productive, and making contributions of higher quality than group members not using the tools (condition 3). Group members using the tools since halfway (condition 2) perceived themselves as making contributions of higher quality than group members not using tools (condition 3). Significant χ^2 -values also indicated that conditions (tools) have an effect on how group members perceived the social and cognitive behavior of their peers. Group members using the tools throughout perceived their peers as having more influence and being more reliable than group members using tools since halfway. Unexpectedly, no significant differences were found for perceived social behavior (i.e., influence, friendliness, cooperation, and reliability), comparing group members using tools throughout with group members not using tools. Nevertheless, group members using tools throughout perceived more productivity and higher quality contributions than group members not using tools. No significant differences in peer assessment scores were found for conditions 2 and 3.

These inconclusive results indicate that no clear patterns can be found in the self and peer assessment scores of the Radar concerning perceived social and cognitive behavior. For example, it is hard to explain why significant differences in friendliness are found between conditions 1 and 2 at T2 (halfway through), but no significant differences are found within condition 1.

The fifth question addressed was to determine whether group members using Radar and Reflector show more congruency between self and peer assessment scores at a subsequent assessment. It was hypothesized that group members need time to adjust their unrealistic positive self perceptions, and thus it was expected that non-significant or small correlations would be found between self assessment and peer assessment scores after group members' first completion of the Radar, but significant and higher correlations would be found at a subsequent assessment. It was also expected that significant differences would be found between self and peer assessment scores after group members' first completion of the Radar, and that these differences would become non-significant or smaller at a subsequent assessment. One significant difference was found between self and peer assessments, students in condition 2 perceived their own productivity significantly lower than their peers. Furthermore, we did not find any significant differences between self and peer assessment scores at T1, T2, or T3, but did find, as expected, non-significant or relatively small correlations for self assessment and peer assessment scores at group members' first assessment, and significant medium to large correlations at subsequent assessments. At the end of the collaboration process, group members using tools (conditions 1 and 2) showed higher convergence of self assessment and peer assessment scores over time, than group members not using tools (condition 3). These results indicate that group members using tools establish a shared perception on the social and cognitive behavior of individual group members. Results also indicate that group members use their personal references (e.g., norms, values, beliefs) about themselves, their peers and the group during completion of the first self and peer assessments in the Radar (Kenney, 1994). As stated, it appears that after completion of the Radar for the first time, group members need time to (1) process, reflect and act upon the received feedback, (2) monitor and assess their own social and cognitive performance and that of their peers during collaboration, and (3) establish shared norms, values and beliefs.

The sixth question addressed was to examine whether group members using Reflector set goals and formulate plans to enhance social and cognitive group performance. Results show that Reflector stimulates group members to set goals and formulate plans to improve their social and cognitive group performance. Results also

show that goals and plans were mainly focused on improvement of activities, such as task coordination, communication, productivity, and focussing on the task. An explanation why goals and plans were mainly focused on these activities could be that these activities are derived from dimensions of Radar. For example, improvement of 'task coordination' and 'communication' could indicate that students want to improve their cooperation, and it is likely that these activities are derived from Radar output on the dimension 'cooperation'. The need to improve 'productivity' and 'focussing on task' could indicate that students want to improve the productivity and quality of their work, which are both dimensions of Radar. The social and cognitive activities stated above are crucial for successful collaboration (Barron, 2003; Erkens et al., 2005), and indicate that groups formulated goals to enhance their social and cognitive performance (teamwork).

The seventh question addressed was to examine whether group members in condition 1 (tools at T1, T2, and T3) perceive higher social performance, in terms of better team development, higher group satisfaction, less group conflict, and more positive attitudes towards collaborative problem solving, than group members in condition 2 (tools at T2 and T3) and condition 3 (tools at T3). First, as expected, significant intra-class correlations were found for all social performance scales, except for 'attitude towards collaborative problem solving'. An explanation for not finding significant intra-class correlation for attitude towards collaborative problem solving could be that, in comparison with the other scales (e.g., team development), this scale is not determined by a single collaboration session (e.g., a collaborative writing task), but is a summary evaluations of several collaboration sessions (Petty, Wegener, & Fabrigar, 1997). Second, unexpectedly, no significant differences were found between conditions 2 and 3. Using tools since halfway had no effect on the perceived social performance for condition 2. The period of time may be too short to find effects of the tools on social performance. Group members in condition 2, who received and used the tools halfway, had half the amount of time (67.5 min – from halfway to the end) to change their social and cognitive behavior, compared to group members in condition 1, who received and used tools from the beginning (135 min). Third, as expected, comparing conditions 1 and 2, group members in condition 1 (tools at T1, T2, and T3) perceived their team as being better developed, experienced higher levels of group satisfaction, lower levels of conflicts, and had a more positive attitude towards collaborative problem solving than group members in condition 2 (tools at T2 and T3). Comparing condition 1 and 3, group members in condition 1 (tools at T1, T2, and T3) experienced higher levels of group satisfaction and higher social performance than group members in condition 3 (tools at T3). Nevertheless, the associated χ^2 -values were low or not significant, indicating that using tools throughout can have a positive effect on social performance, in terms of better team development, higher group satisfaction, less group conflict, and more positive attitudes towards collaborative problem solving, but also indicate that there are other (not examined) factors (e.g., group member's discourse or communication) that have a major influence on these social performance scales. These results support our hypothesis that frequent use of Radar combined with Reflector focused on future group functioning and goal setting, enhances social group performance. In future research it would be interesting to examine the effect of the tools on group member's intended and actual behavior by analyzing their communication using discourse analysis (e.g., Erkens & Janssen, 2008).

The final question addressed was to examine whether using Radar and Reflector would lead to higher cognitive performance, measured by the grade given to the essays. As found in the previous study (Phielix et al., 2010), no significant effects of Radar and Reflector were found for grade given to the essays. Again, the

period of time may be too short to find effects of the tools on cognitive performance.

A few limitations of this study should be kept in mind. First, as mentioned, a limitation of this study might lie in the short period of time in which group members had to use the tools, fulfill the task, and establish shared norms, values and beliefs. Significant effects of the tools on social and cognitive behavior were found, but could have been stronger when more time would have been available. In future research Radar and Reflector will be used for a longer period (i.e., 3 months), during which the tools will be available from the beginning for all conditions, and will have to be used several times. Second, in this study self and peer assessments (interpersonal perceptions) were used in order to change the social and cognitive behavior of individual group members and the group as a whole. Students were both feedback provider (assessor) as feedback receiver (assessee), thus student's interpersonal perceptions and behavior were influenced by their provided and received feedback. In future research it would be interesting to use a larger sample size and Social Relations Models (SRM) to examine how much variance in self assessment and peer assessment scores can be explained by the actor (assessor), partner (assessee), diad (specific combination of two students), and group (specific combination of three or more students). Third, in this quantitative study the output of the Reflector, concerning group member's intentions to enhance their social and cognitive performance by setting goals and formulating plans, was only analyzed in a quantitative way. It would be interesting to examine in a qualitative study whether these intentions lead to actual changes in social and cognitive behavior and activities (e.g., using discourse analysis to find out whether the intention to become more friendly actually led to more friendly and helpful behavior in the chat).

In conclusion, the effects of Radar and Reflector on group functioning are very promising. Results show that, by adding these easy to complete and easy to interpret peer feedback and reflection tools in a CSCL environment, students (1) become aware of interpersonal perceptions and behavior, (2) exhibit higher social and cognitive behavior, (3) establish shared perceptions on interpersonal behavior, and (4) can enhance the social performance of the group.

References

- Bales, R. F. (1988). A new overview of the SYMLOG system: Measuring and changing behavior in groups. In R. B. Polly, A. P. Hare, & P. J. Stone (Eds.), *The SYMLOG practitioner: Applications of small group research* (pp. 319–344). New York: Praeger.
- Bales, R. F., & Cohen, S. P. (1979). *SYMLOG: A system for the multiple level observation of groups*. New York: Macmillan Publishing.
- Baltes, B. B., Dickson, M. W., Sherman, M. P., Bauer, C. C., & LaGanke, J. (2002). Computer-mediated communication and group decision making: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 87(1), 156–179.
- Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, 12, 307–359.
- Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment and Evaluation in Higher Education*, 24, 413–426.
- Boud, D., & Falchikov, N. (1989). Quantitative studies of self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18, 529–549.
- Boud, D., Keogh, R., & Walker, D. (1985). Promoting reflection in learning: A model. In D. Boud, R. Keogh, & D. Walker (Eds.), *Reflection: Turning experience into learning* (pp. 18–40). London: Routledge Falmer.
- Brok, P. den, Brekelmans, M., & Wubbels, Th. (2006). Multilevel issues in studies using students' perceptions of learning environments: The case of the Questionnaire on Teacher Interaction. *Learning Environments Research*, 9, 199–213.
- Campbell, J. D. (1986). Similarity and uniqueness: The effects of attribute type, relevance, and individual differences in self-esteem and depression. *Journal of Personality and Social Psychology*, 50, 281–294.
- Castleton_Partners/TCO. (2007). *Building trust in diverse teams*. London/Cambridge, UK: Castleton Partners/TSO International Diversity Management.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (rev. ed.). Hillsdale, NJ: Erlbaum.
- Cutler, H., & Price, J. (1995). The development of skills through peer assessment. In A. Edwards & P. Knight (Eds.), *Assessing competence in higher education* (pp. 150–159). Staff and Educational Development Association.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331–350.
- Dominick, P. G., Reilly, R. R., & McGourty, J. W. (1997). The effects of peer feedback on team member behavior. *Group & Organization Management*, 22, 508–525.
- Dourish, P., & Bellotti, V. (1992). Awareness and coordination in a shared workspace. In M. Mantel & R. Baecker (Eds.), *Proceedings of the ACM conference on computer-supported cooperative work* (pp. 107–114). New York: ACM Press.
- Druskat, V. U., & Wolff, S. B. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology*, 84, 58–74.
- Emans, B., Koopman, P., Rutte, C., & Steensma, H. (1996). Teams in organisations: Interne en externe determinanten van resultaatgerichtheid [Teams in organizations: Internal and external determinants of outcome orientation]. *Gedrag en Organisatie*, 9, 309–327.
- Erkens, G., & Janssen, J. (2008). Automatic coding of dialogue acts in collaboration protocols. *Computer-Supported Collaborative Learning*, 3, 447–470.
- Erkens, G., Jaspers, J., Prangma, M., & Kanselaar, G. (2005). Coordination processes in computer supported collaborative writing. *Computers in Human Behavior*, 21, 463–486.
- Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International*, 32, 175–187.
- Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction*, 12, 213–232.
- Fjermestad, J. (2004). An analysis of communication mode in group support systems research. *Decision Support Systems*, 37, 239–263.
- Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking and computer conferencing: A model and tool to access cognitive presence. *American Journal of Distance Education*, 15(1), 7–23.
- Garton, L., & Wellman, B. (1995). Social impacts of electronic mail in organizations: A review of the research literature. In B. R. Burleson (Ed.), *Communication yearbook* (Vol. 18, pp. 438–453). Thousand Oaks, CA: Sage.
- Geister, S., Konradt, U., & Hertel, G. (2006). Effects of process feedback on motivation, satisfaction, and performance in virtual teams. *Small Group Research*, 37, 459–489.
- Goethals, G. R. (1986). Fabricating and ignoring social reality: Self-serving estimates of consensus. In J. M. Olson, C. P. Hermann, & M. P. Zanna (Eds.), *Relative deprivation and social comparison: The Ontario symposium* (Vol. 4, pp. 135–157). Hillsdale, NJ: Lawrence Erlbaum.
- Goethals, G. R., Messick, D. M., & Allison, S. T. (1991). The uniqueness bias: Studies of constructive social comparison. In J. Suls & T. A. Wills (Eds.), *Social comparison research: Contemporary theory and research* (pp. 149–176). Hillsdale, NJ: Lawrence Erlbaum.
- Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International Journal of Educational Telecommunications*, 1(2–3), 147–166.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Hobman, E. V., Bordia, P., Irmer, B., & Chang, A. (2002). The expression of conflict in computer-mediated and face-to-face groups. *Small Group Research*, 33, 439–465.
- Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning. *Computers & Education*, 49, 1037–1065.
- Jarvenpaa, S., & Leidner, D. (1999). Communication and trust in global virtual teams. *Organization Science*, 10, 791–815.
- Jaspers, J., Broeken, M., & Erkens, G. (2004). *Virtual collaborative research institute (VCRI) (version 2.0)*. Utrecht: Onderwijskunde Utrecht, ICO/ISOR.
- Johnson, D. W., & Johnson, R. T. (1999). *Learning together and alone: Cooperative, competitive, and individualistic learning* (5th ed.). Boston, MA: Allyn and Bacon.
- Johnson, D. W., Johnson, R. T., & Smith, K. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review*, 19, 15–29.
- Karau, S., & Williams, K. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65, 681–706.
- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford.
- Kirschner, P. A., Strijbos, J., Kreijns, K., & Beers, P. J. (2004). Designing electronic collaborative learning environments. *Educational Technology Research and Development*, 52(3), 47–66.
- Klein, W. M. (2001). Post hoc construction of self-performance and other performance in self-serving social comparison. *Society for Personality and Social Psychology*, 27(6), 744–754.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254–284.
- Kreijns, K., & Kirschner, P. A. (2004). Determining sociability, social space and social presence in (a)synchronous collaborating teams. *Cyberpsychology and Behavior*, 7, 155–172.
- Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, 19, 335–353.

- Lazonder, A. W., & Rouet, J. F. (2008). Information problem solving instruction: Some cognitive and metacognitive issues. *Computers in Human Behavior*, 24, 753–765.
- Liaw, S., & Huang, H. (2000). Enhancing interactivity in web-based instruction: A review of the literature. *Educational Technology*, 40(3), 41–45.
- Lipponen, L., Rahikainen, M., Lallimo, J., & Hakkarainen, K. (2003). Patterns of participation and discourse in elementary students' computer-supported collaborative learning. *Learning and Instruction*, 13, 487–509.
- McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovations in Education and Training International*, 32, 302–313.
- McLeod, P. L., & Liker, J. K. (1992). Process feedback in task groups: An application of goal setting. *Journal of Applied Behavioral Science*, 28, 15–52.
- Mento, A. J., Steel, R. P., & Karren, R. J. (1987). A meta-analytic study of the effects of goal setting on task performance: 1966–1984. *Organizational Behavior and Human Decision Processes*, 39, 52–83.
- Narciss, S., Proske, A., & Koerndle, H. (2007). Promoting self-regulated learning in web-based learning environments. *Computers in Human Behavior*, 23, 1126–1144.
- Neubert, M. J. (1998). The value of feedback and goal setting over goal setting alone and potential moderators of this effect: A meta-analysis. *Human Performance*, 11, 321–335.
- Northrup, P. T. (2001). A framework for designing interactivity into web-based instruction. *Educational Technology*, 41(2), 31–39.
- Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology*, 48, 609–647.
- Phielix, C., Prins, F. J., & Kirschner, P. A. (2010). Awareness of group performance in a CACL environment: Effects of peer feedback and reflection. *Computers in Human Behavior*, 26, 151–161.
- Prins, F. J., Sluijsmans, D. M. A., & Kirschner, P. A. (2006). Feedback for general practitioners in training: Quality, styles, and preferences. *Advances in Health Sciences Education*, 11, 289–303.
- Rochelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69–97). New York: Springer Verlag.
- Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology*, 78, 450–462.
- Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in teamwork? *Small Group Research*, 36, 555–599.
- Schön, D. A. (1987). *Educating the reflective practitioner*. San Francisco, CA: Jossey-Bass.
- Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London: John Wiley & Sons.
- Slof, B., Erkens, G., Kirschner, P. A., Jaspers, J. G. M., & Janssen, J. (in press). Guiding students' online complex learning-task behavior through representational scripting. *Computers in Human Behavior*. doi:10.1016/j.chb.2010.02.007.
- Somervell, H. (1993). Issues in assessment, enterprise and higher education: The case for self-, peer and collaborative assessment. *Assessment and Evaluation in Higher Education*, 18, 221–233.
- Straus, S. G. (1997). Technology, group process, and group outcomes: Testing the connections in computer-mediated and face-to-face groups. *Human-Computer Interaction*, 12, 227–266.
- Straus, S. G., & McGrath, J. E. (1994). Does the medium matter? The interaction of task type and technology on group performance and member reactions. *Journal of Applied Psychology*, 79(1), 87–97.
- Strijbos, J. W., Kirschner, P. A., & Martens, R. L. (Eds.). (2004). *What we know about CACL: And implementing it in higher education*. Boston, MA: Kluwer Academic/Springer Verlag.
- Strijbos, J. W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2007). The effect of functional roles on perceived group efficiency during computer-supported collaborative learning: A matter of triangulation. *Computers in Human Behavior*, 23, 353–380.
- Stroebe, W., Diehl, M., & Abakoumkin, G. (1992). The illusion of group effectivity. *Personality and Social Psychology Bulletin*, 18, 643–650.
- Suls, J., & Wan, C. K. (1987). In search of the false-uniqueness phenomenon: Fear and estimates of social consensus. *Journal of Personality and Social Psychology*, 59, 229–241.
- Tubbs, M. E. (1986). Goal setting: A meta-analytic examination of the empirical evidence. *Journal of Applied Psychology*, 71, 474–483.
- Walther, J. B., Anderson, J. F., & Park, D. (1994). Interpersonal effects in computer-mediated interaction: A meta-analysis of social and anti-social communication. *Communication Research*, 19, 460–487.
- Weisband, S., & Atwater, L. (1999). Evaluating self and others in electronic and face-to-face groups. *Journal of Applied Psychology*, 84(4), 632–639.
- Williams, K. D., Harkins, S. G., & Latané, B. (1981). Identifiability as a deterrent to social loafing: Two cheering experiments. *Journal of Personality and Social Psychology*, 40, 303–311.
- Wubbels, T., Créton, H. A., & Hooyman, H. P. (1985). *Discipline problems of beginning teachers, interactional teacher behavior mapped out*. In Paper presented at the annual meeting of the American Educational Research Association, Chicago, March–April.
- Yammarino, F. J., & Atwater, L. E. (1997). Do managers see themselves as others see them? Implications of self-other rating agreement for human resources management. *Organizational Dynamics*, 25(4), 35–44.
- Yukawa, J. (2006). Co-reflection in online learning: Collaborative critical thinking as narrative. *Journal of Computer-Supported Collaborative Learning*, 1, 203–228.