

Using a Reflection Tool to Increase Reliability of Peer Assessments in a CSCL Environment

C.Phielix, F.J. Prins, J. Janssen, Utrecht University, P.O. Box 80.140, 3508 TC Utrecht, The Netherlands,
Email: C.Phielix@uu.nl, F.J.Prins@uu.nl, J.J.H.M.Janssen@uu.nl

P.A. Kirschner, Open University, P.O. Box 2960, 6401 DL Heerlen, The Netherlands, P.A.Kirschner@ou.nl

Abstract: To examine the reliability of students' peer assessments, two contiguous study groups used a peer assessment tool (Radar) with or without reflection tool (Reflector) 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 individually and collaboratively on their past, present, and future functioning. The underlying assumption was that Radar in combination with Reflector would lead to (1) more reliable peer assessment scores, and (2) more valid perceptions of the social performance of the group. Participants were 191 second year academic students working in groups of three, four or five, on a collaborative writing task. As expected, results showed that the use of a reflection tool in a CSCL environment leads to more reliable peer assessment scores and more valid perceptions of the groups' social performance.

Introduction

Computer supported collaborative learning (CSCL) environments, though originally simple, text-based, computer mediated communication systems, have been strongly influenced by the rapid development of information and communication technology, tools and widgets (e.g., chat, video conferencing, discussion forums, group awareness widgets and shared participation tools). These applications have proven to be useful for supporting education and collaborative learning (Janssen, Erkens, Kanselaar, & Jaspers, 2007; Kreijns, Kirschner, & Jochems, 2003), leading to the design and implementation of more sophisticated CSCL environments.

Though CSCL environments have been shown to be promising educational tools and though expectations as to their value and effectiveness are high, groups learning in CSCL environments do not always reach their full potential. One of the most important reasons for this disparity between their potential and their results can be found in the social interaction between the group members, which is the key for successful collaboration (Kreijns, et al., 2003).

CSCL environments can be augmented with tools or widgets that influence social interaction (Kirschner, Strijbos, Kreijns, & Beers, 2004). Such tools, also known as 'social affordance devices', can positively affect the social performance (e.g., team development) and cognitive performance (quantity and quality of work) of a group (Kirschner, et al.). An example of such tools are self and peer assessments tools, which are increasingly applied during formative assessments and to evaluate collaborative processes during group work (Dochy, Segers, & Sluijsmans, 1999; Prins, Sluijsmans, Kirschner, & Strijbos; 2005; Strijbos & Sluijsmans, 2010). Self and peer assessment tools can provide students (and teachers) with information about group members' social and cognitive performance. However, providing group members' with information on their social and cognitive performance is not enough to positively alter their behavior (Prins, Sluijsmans, & Kirschner, 2006). Information provided by self and peer assessments are namely seldomly objective. During completion of self and peer assessments, students make many 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 (Klein, 2001; Saavedra & Kwun). Therefore, group members need to reflect individually and collaboratively (co-reflect) upon the performance of their peers and the group as a whole, before they rate themselves and their peers. Thus, in this study it is hypothesized that the use of (co-)reflection could enhance the reliability of peer assessments (e.g., Dochy, et al., 1999) and enhance behavioral change (e.g., Prins, et al., 2006). To test this hypothesis, two contiguous groups used a self and peer assessment tool with or without a reflection tool in a CSCL environment.

Self and Peer Assessment

Self assessment and peer assessment have become increasingly popular in education and CSCL (e.g., Prins, et al., 2005). Boud and Falchikov (1989) define self assessment as students making judgments about their own learning, mainly about their achievements and learning outcomes. Peer assessment can be defined as an educational arrangement where students judge a fellow student's performance qualitatively and/or quantitatively, which stimulates students to share responsibility, reflect, discuss and collaborate (Topping, 1998;

Strijbos & Sluijsmans, 2010). Somervell (1993) stresses that providing peer assessment can be seen as a part of the self assessment process, informing self assessment. Sharing self and peer assessments with others can be seen as providing information to increase group performance, therefore, self and peer assessments can be seen as a form of peer feedback (e.g., Topping, 1998; Strijbos, Narciss, & Dunnebie, 2010). Self and peer assessments can (1) provide students and teachers with a more accurate perception of students' individual behavior and performance in collaborative group work (Cheng & Warren, 2000), (2) support students in forming judgments about what can be referred to as good group behavior and high-quality performance (Topping, 1998), and (3) foster reflection on the student's own learning process and learning activities (Dochy, et al., 1999). Thus, self and peer assessments can provide students with useful information on their social and cognitive performances at both individual and group level. For this (feedback) information to be effective, students need to be challenged to reflect individually and collaboratively on their performance. Students need to ask themselves whether they understand the feedback, accept it, and determine whether it provides clues for behavioral change (Prins, et al., 2006).

Reflection

Reflection can be defined as the intellectual and affective activities individuals engage in to explore their experiences to reach new understandings and appreciations of those experiences (Boud, Keogh, & Walker, 1985). Hattie and Timperley (2007) found that for feedback to be effective, students need to answer three major questions; (1) Where am I going? (feed up), (2) How am I going? (feed back), and (3) Where to next? (feed forward). However, reflection processes are not only useful on an individual level, but also on a group level. The process in which group members collaboratively reflect on their experiences can be referred to as co-reflection. Yukawa (2006; p. 206) defined co-reflection 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". 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). Thus, reflection can lead to new interpersonal perceptions, perspectives on experience, changes in behavior, readiness for application, and commitment to action (Boud, et al.).

Social Relations Models to Measure Reliability of Peer Assessments

In this study it is hypothesized that the use of (co-)reflection could enhance the reliability of self and peer assessments (e.g., Dochy et al., 1999) and enhance behavioral change (e.g., Prins et al., 2006). The reliability of peer assessments can be defined as the extent to which the scores for person X are consistent across all peer assessors (Bonito & Kenny, 2010). Research has shown that when students assess the performance of their peers, their assessments often dependent on one another's peer assessment (e.g., Kenny, 1994). For example, in group work, Chris's assessment of Paul is likely related to Paul's assessment of Chris. When these interdependencies are ignored, meaningful information about the interdependencies among peers is lost, and results of statistical analyses may be distorted (Bonito & Kenny; Kenny). Social Relations Models (SRM) can be used to examine these interdependencies among ratings and provide both a theoretical basis and a statistical tool (Kenny). SRM allows for variance to be partitioned into partner (assessee), actor (assessor), dyad (relationship between two assessors), and residual effects. Actor effects represent an individual's tendency to see all other group members as high or low on a particular trait, whereas partner effects reflect an individual's tendency to be seen as high or low on a particular trait by all other group members. Dyad effects represent the interaction effects of the partner and the actor at the dyadic level (i.e., does the relationship between Chris and Paul have a unique effect on the assessment even when actor and partner effects are taken into account? cf., Bonito & Kenny). In this study, SRM will be used to examine whether peer assessment scores of students with a co-reflection tool will show higher partner variances compared to students without this tool. Students with high partner variances receive more consistent (i.e., there is consensus about their cognitive or social performance) and more reliable peer assessment scores, compared to students with low partner variances.

Hypotheses

Hypothesis 1: Reflector enables to reflect upon individual and group behavior, and support students in forming judgments about what can be referred to as good group behavior and high-quality performance. Thus, students with Reflector (+Re) will perceive and receive more consistent (reliable) peer assessments scores (show higher partner variance), than students without Reflector (−Re).

Hypothesis 2: Students with Reflector (+Re) will exhibit more realistic peer assessments scores, resulting in higher correlations between the peer assessment scores and the perceived social performance, compared to groups without Reflector (−Re).

Hypothesis 3: Groups with Reflector (+Re) will score higher on social group performance compared

to groups without Reflector (\neg Re), because groups with Reflector set goals and formulate plans to enhance their social performance.

Method

Participants

Participants were 191 second-year Dutch academic Educational Science students (37 male, 154 female) with an average age of 23.64 years ($SD = 7.16$, $Min = 19$, $Max = 55$). Prior to the experiment, they were randomly assigned by the teacher to groups of three ($n = 21$), four ($n = 160$), and five ($n = 10$), and randomly assigned by the researchers to one of two conditions (see Design). Groups were heterogeneous in ability.

Table 1: Design of the study.

Condition	T1 – week 1	T2 – week 3	T3 – week 6	T4 – week 8
1. With Reflector (+Re)	Radar	Radar Reflector	Radar Reflector	Radar, Reflector, Questionnaire
2. Without Reflector (\neg Re)	Radar	Radar	Radar	Radar, Questionnaire

Design

For this study, an experimental design was used with one experimental and one control condition (see Table 1). The experimental condition ($n = 105$) received a self and peer assessment tool (Radar) and a co-reflection tool (Reflector). The control condition received ($n = 86$) only Radar. During a period of 8 weeks, participants in both conditions completed the Radar four times. Additionally, from the second measurement occasion (T2: week 3), participants in the experimental condition also had to complete the Reflector.

Measures

Social behavior. Perceived group social behavior is measured by the self and peer assessments in Radar on four variables (influence, friendliness, cooperativeness, reliability). These variables are rated on a continuous scale ranging from 0 to 4 (0 = none, 4 = very high).

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).

Social performance. The perceived social performance was measured by the questionnaire at the end of the collaboration process (week 8). Four previously validated instruments (Strijbos, Martens, Jochems, & Broers, 2007) were translated into Dutch and transformed into 5-point Likert scales (1 = totally disagree, 5 = totally agree; see Table 3). 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-Olkin 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 Eigen values exceeding 1, explaining 76.6% of the variance respectively. Cronbach’s alpha of the composed ‘Social Performance (total)’ scale was .90.

Task and Procedure

Students collaborated in groups of three, four, and five, on a collaborative writing task in educational psychology. To successfully complete the collaborative writing task, each group had to write a paper about a pilot-study which they conducted over a period of eight weeks. During this period, they had to complete a self and peer assessment tool (Radar) four times, with or without a supplement co-reflection tool (Reflector). The groups used a CSCL environment called Virtual Collaborative Research Institute (VCRI; Jaspers, Broeken, & Erkens, 2004), a groupware program that supports collaborative learning on research projects and inquiry tasks. Students were instructed to make complete use of the available tools (e.g., self and peer assessment tool and reflection tool). During use of the tools, students were instructed to use the chat tool to communicate with other

group members. 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 eight weeks to complete the task, that it would be graded by their teacher, and that it would affect their final 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 for the successful completion of the group task (i.e., interdependence). At the end of the final session all participants completed a 30-item questionnaire on the social performance of the group.

Instruments

Self and Peer Assessment Tool (Radar)

Radar is a self and peer assessment tool for eliciting information on group members' social and cognitive behavior visualized in a radar diagram. Radar provides students 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 (Den Brok, Brekelmans, & Wubbels, 2006). Radar provides information on six traits 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., Den Brok, Brekelmans, & Wubbels; Kenny, 1994; Salas, Sims, & Burke, 2005). These variables, as well as the reasons for their choice, are discussed in Phielix, Prins, and Kirschner (2010) and Phielix, Prins, Kirschner, Erkens, and Jaspers (in press).

Students rate themselves and their peers 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 a rating of 0 to 1) is divided into tenths so that every scale contained 40 points of assessment. To simplify data analysis, ratings are transformed to a 100-point scale by multiplying the ratings (0-4) by 25. Students can only access 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 in which students anonymously can (1) compare their self and received (average) peer assessments, (2) compare average peer assessment of all group members, and (3) see all personal (self and peer) assessments of their group members.

Co-reflection Tool (Reflector)

Reflector assists group members in becoming aware of their individual and group behaviour, and stimulates 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).

The 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 (i.e., who, what, when) to improve group performance.

The first four questions are completed individually, with completion indicated by clicking an 'Add'-button. This allows students to share their answers with the rest of the group and allows them 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 a specific frame (Co-Reflection), which allows writing a shared conclusion and formulating shared goals. Responses made by the students in Reflector are not scored or evaluated.

Results

Hypothesis 1: Reflector enables to reflect upon individual and group behavior, and support students in forming judgments about what can be referred to as good group behavior and high-quality performance. Thus, students with Reflector (+Re) will perceive and receive more consistent (reliable) peer assessments scores (show higher

partner variance), than students without Reflector (\neg Re).

Table 2: Proportion partner & actor variance for peer assessment scores per dependent variable per condition.

		Influence		Friendliness		Cooperation		Reliability		Productivity		Quality	
		<i>Partner</i>	<i>Actor</i>	<i>Partner</i>	<i>Actor</i>	<i>Partner</i>	<i>Actor</i>	<i>Partner</i>	<i>Actor</i>	<i>Partner</i>	<i>Actor</i>	<i>Partner</i>	<i>Actor</i>
+Re	T1	.20*	.35*	.09*	.68*	.11*	.64*	.06	.44*	.11	.41*	.08	.42*
	T2	.39*	.20*	.12*	.63*	.21*	.41*	.30*	.38*	.26*	.18*	.19*	.40*
	T3	.43*	.19*	.13	.49*	.29*	.34*	.43*	.18*	.40*	.22*	.38*	.21*
	T4	.28*	.35*	.04	.42*	.17*	.53*	.23*	.37*	.22*	.30*	.22	.27*
\neg Re	T1	.16*	.37*	.09	.64*	.00	.60*	.05	.51*	.16*	.39*	.16*	.60*
	T2	.09	.25*	.00	.74*	.05	.70*	.06	.55*	.19*	.44*	.21*	.36*
	T3	.14*	.48*	.01	.67*	.04	.61*	.07*	.64*	.12*	.63*	.05	.65*
	T4	.20*	.46*	.05	.54*	.00	.66*	.17*	.58*	.23*	.55*	.19	.37*

* p < .05

SRM analyses were used to examine the differences in partner (assessee) and actor (assessor) variance between students with and without Reflector. Table 2 shows the partner and actor variance per dependent variable (i.e., influence, friendliness, etc.) over time per condition. As expected, for students with Reflector, all partner variances related to social behavior (influence, friendliness, cooperativeness, and reliability) are higher compared to students without Reflector. These results indicate that students with Reflector perceived (and received) more consistent peer assessments. Concerning partner variances related to cognitive behavior (e.g., productivity and quality of contribution), students with Reflector showed higher partner variances for productivity, at T2 and T3, and quality of contribution at T3. Students without Reflector showed higher partner variances for productivity at T1 and T4, as well as for quality of contribution at T1 and T2. Except for friendliness and cooperativeness at T1, and productivity at T1 and T2, all actor variances for students with Reflector are lower compared to students without Reflector. This indicates that compared to students without Reflector, peer assessments of students with Reflector are less determined by the tendency of an actor (assessor) to see all other group members as high or low on a particular trait. In contrast to students with Reflector, partner variances for students without Reflector never exceeded actor variances. For students with Reflector, at T3, partner variances were higher than actor variances for influence, reliability, productivity, and quality of contribution. Unexpectedly, for students with Reflector, partner variance for influence, cooperativeness, reliability and productivity decreases between T3 and T4. For students without Reflector, partner variance for influence, reliability, and productivity increases between T3 and T4.

Hypothesis 2: Students with Reflector (+Re) will exhibit more objective and realistic peer assessments scores, resulting in higher correlations between the peer assessment scores and the perceived social performance, compared to groups without Reflector (\neg Re). A Pearson product-moment correlation coefficient was used to test correlations between group members' average peer assessment scores and their perceived social performance at T4. Results are shown in Table 3 for students with and without Reflector.

Table 3: Correlations for average peer assessments (with and without reflector) and social performance at T4.

	<i>n</i>	Team development	Group process satisfaction	Intra group conflicts	Attitude towards CL problem solving	Social performance (total)
		<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>
With Reflector (+Re)						
Influence	80	.15	.31**	-.09	.13	.26*
Friendliness	80	.38**	.44**	-.28*	.19	.44**
Cooperativeness	80	.36**	.35**	-.22*	.10	.41**
Reliability	80	.24*	.19	-.16	.09	.25*
Productivity	80	.14	.17	-.12	.23*	.20
Quality of contribution	80	.14	.21	-.08	.12	.22
Without Reflector (\negRe)						
Influence	81	.15	.12	-.15	.17	.13
Friendliness	81	.04	.05	-.04	.02	.04
Cooperativeness	81	.31**	.16	-.25*	-.05	.17

Reliability	81	.25*	.21	-.23*	.05	.17
Productivity	81	.26*	.16	-.21	.04	.17
Quality of contribution	81	.28*	.20	-.21	-.02	.20

* $p < .05$ (2-tailed)

** $p < .01$ (2-tailed)

As expected, compared to students without Reflector, students with Reflector show significantly higher correlations between their ratings on influence, friendliness, cooperativeness, and their perceived social performance (in total). No significant correlations with perceived social performance (in total) were found for students without Reflector.

Hypothesis 3: Groups with Reflector (+Re) will score higher on social group performance compared to groups without Reflector (-Re), because groups with Reflector set goals and formulate plans to enhance their social performance.

Table 4: Multilevel analyses for effects of condition on social performance scales.

Scale	With Reflector (+Re, $n = 89$)		Without Reflector (-Re, $n = 86$)		Comparing +Re vs. -Re		Chi- square χ^2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	β	<i>SE</i> β	
Team development	3.60	.73	3.87	.48	-.27*	.15	3.31*
Group-process satisfaction	3.07	.41	3.14	.34	-.07	.07	.91
Intra-group conflicts	2.64	.59	2.43	.50	.21*	.12	2.85*
Attitude	3.06	.21	3.08	.18	-.02	.03	.00
Social Performance (total)	3.13	.20	3.18	.15	-.05	.04	1.60

* $p < .05$ (1-tailed)

Multilevel analysis was used to examine whether groups with Reflector (+Re) 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 without Reflector (-Re). Table 4 shows multilevel analyses for effects of condition on social performance scales. Unexpectedly, the significant β -value shows that groups with Reflector perceived their team as being less developed and having more intra-group conflicts, than groups without Reflector. However, no significant differences were found for total social performance, group process satisfaction, attitude towards collaborative problem solving.

Discussion & Conclusion

In this study it was hypothesized that the use of a co-reflection tool in a CSCL environment could enhance the reliability of peer assessments and enhance behavioral change. Social Relations Models (e.g., Kenny, 1994) were used to analyse the self and peer assessment data, which has never been done before in an educational or CSCL setting. Findings in this study support the assumption that supplementing a self and peer assessment tool (Radar) with a co-reflection tool (Reflector) can lead to more reliable peer assessments scores (higher partner variances). As expected, compared to students without Reflector, students with Reflector exhibited higher partner variances for their peer assessment scores on social behavior (e.g, influence, friendliness, cooperativeness and reliability), which indicate that these students perceived (and received) more consistent and reliable peer assessments. The highest partner variances for peer assessment scores on cognitive behaviour (e.g., productivity and quality of contribution) differed over time per condition. Except for productivity at T2, all actor variances for students with Reflector are lower compared to students without Reflector. This indicates that compared to students without Reflector, peer assessments of students with Reflector are less determined by the tendency of an actor (assessor) to see all other group members as high or low on a particular trait. Thus, the hypothesis that the use of a co-reflection tool in a CSCL environment could enhance the reliability of peer assessments, is accepted. Unexpectedly, for students with Reflector, partner variance decreased between T3 and T4 towards to the level of students without Reflector. A possible explanation could be that in final stage of the collaboration process students are less focused on the process (group members' behavior) and more on getting the product (paper) finished before the deadline (e.g., Aubert & Kelsey, 2003).

Findings also supported the second assumption that students with Reflector would exhibit more consistent (reliable) peer assessments scores, resulting in higher correlations between the peer assessment scores and the perceived social performance, compared to groups without Reflector. As expected, compared to students without Reflector, students with Reflector show significantly higher correlations between their peer assessment scores measured by Radar and their perceived social performance as measured by the questionnaire. No significant correlations were found for students without Reflector, indicating that the use of a co-reflection tool in a CSCL environment could enhance the validity of students' peer assessments. Thus, also the second

hypothesis is accepted. An explanation for these findings could be that students without Reflector apply norm-referenced standards rather than criterion-referenced standards for assessing themselves and their peers, for instance based on prior experiences or personal beliefs. Apparently, students do not automatically reflect on a high cognitive level on their perceived and received peer assessments and need a reflection tool (i.e., Reflector) to do so (Kollar & Fischer, 2010).

Findings did not support the third assumption that the use of a co-reflection tool could enhance the social performance of the group. Unexpectedly, groups with Reflector perceived their team as being less developed and having more intra-group conflicts, than groups without Reflector. However, differences between the two conditions (with or without Reflector) are small and no significant differences were found for social performance in total. Furthermore, the perceived social performance of the students without Reflector can be argued, since these perceptions as measured by the questionnaire do not correlate with their perceived behavior (peer assessment scores) as measured by Radar.

Overall, first, SRM analyses proved to be a very useful tool to measure the variance and reliability in self and peer assessments. Second, results showed that the use of a reflection tool in a CSCL environment can lead to reliable peer assessment scores and more valid perceptions of the social performance of the group. Third, for future research on self and peer assessments in a CSCL environment, these results indicate that for self and peer assessments to be reliable, self and peer assessment tools need to be supplemented with a reflection tool.

References

- Aubert, B., & Kelsey, B. (2003). Further understanding of trust and performance in virtual teams. *Small Group Research, 34*, 575–618.
- Bonito, J. A. & Kenny, D. A. (2010) The measurement of reliability of social relations components from round-robin designs. *Personal Relationships, 17*, 235–251.
- 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.
- Cheng, W. & Warren, M. (2000). Making a Difference: using peers to assess individual students' contributions to a group project, *Teaching in Higher Education 5*(2), 243-255.
- 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.
- 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.
- 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.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112.
- 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.
- Jaspers, J., Broeken, M., & Erkens, G. (2004). *Virtual Collaborative Research Institute (VCRI) (Version 2.0)*. Utrecht: Onderwijskunde Utrecht, ICO/ISOR.
- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford.
- Kirschner, P., 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.
- Kollar, I., & Fischer, F. (2010). Commentary e peer assessment as collaborative learning: a cognitive perspective. *Learning and Instruction, 20*(4), 344-348.
- 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.
- Phielix, C., Prins, F. J., & Kirschner, P. A. (2010). Awareness of group performance in a CSCL-environment: Effects of peer feedback and reflection. *Computers in Human Behavior, 26*, 151-161.
- Phielix, C., Prins, F. J., Kirschner, P. A., Erkens, G., & Jaspers, J. (in press). Group awareness of social and cognitive performance in a CSCL environment: Effects of a peer feedback and reflection tool.

- Computers in Human Behavior*, doi:10.1016/j.chb.2010.06.024
- 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.
- Prins, F. J., Sluijsmans, D. M. A., Kirschner, P. A., & Strijbos, J. W. (2005). Formative peer assessment in a CSSL environment: A case study. *Assessment and Evaluation in Higher Education, 30*(4), 417-444.
- 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.
- 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.
- 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.
- Strijbos, J. W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*(4), 291-303.
- Strijbos, J. W., & Sluijsmans, D. M. A. (2010). Guest editorial - Unravelling peer assessment: methodological, functional, and conceptual developments. *Learning and Instruction, 20*(4), 265-269.
- Topping, K. (1998). Peer assessment between students of colleges and universities. *Review of Educational Research, 68*(3), 249-276.
- Yukawa, J. (2006). Co-reflection in online learning: Collaborative critical thinking as narrative. *International Journal of Computer-Supported Collaborative Learning, 1*(2), 203-228.