

A rating scheme for assessing collaboration quality

A. Meier, H. Spada, N. Rummel

to appear in ijCSCL (2007)

This final authors' version does not contain all corrections. The original publication is available Online First at

www.springerlink.com under DOI: 10.1007/s11412-006-9005-x

Please cite accordingly.

A Rating Scheme for Assessing the Quality of Computer-Supported Collaboration Processes

Abstract: The analysis of the process of collaboration is a central topic in current CSCL research. However, defining process characteristics relevant for collaboration quality and developing instruments capable of assessing these characteristics are no trivial tasks. In the assessment method presented in this paper, nine qualitatively defined dimensions of collaboration are rated quantitatively: sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction, and individual task orientation. The data basis for the development of these dimensions was taken from a study in which students of psychology and medicine collaborated on a complex patient case via a desktop-videoconferencing system. A qualitative content analysis was performed on a sample of transcribed collaboration dialogue. The insights from this analysis were then integrated with theoretical considerations about the roles of communication, joint information processing, coordination, interpersonal relationship, and motivation in the collaboration process. The resulting rating scheme was applied to process data from a new sample of 40 collaborating dyads. Based on positive findings on inter-rater reliability, consistency, and validity from this evaluation, we argue that the new method can be recommended for use in different areas of CSCL.

The development of appropriate methods for analyzing interactive processes is a major research topic in CSCL. Numerous papers published in the proceedings of the CSCL 2005 conference as well as in this journal address this issue (e.g., Clark & Sampson, 2005; Dönmez, Rose, Stegmann, Weinberger, & Fischer, 2005; Kapur, Voiklis & Kinzer, 2005; Lee, Chan, & van Aalst, 2006; Spada, Meier, Rummel, Hauser, 2005; Zumbach, Schönemann & Reimann, 2005). These authors and other groups of CSCL researchers, often combining different fields of expertise, strive for insights into processes relevant for computer-supported collaborative learning and work, and for the development of assessment methods that are capable of capturing these aspects.

There are several motivations for analyzing the collaboration process in CSCL. For example, specific challenges of collaborative learning and computer-supported communication have to be identified in order to find out where support is needed in the first place, and which aspects of the collaborative process are crucial for successful learning and problem-solving in CSCL (Rummel & Spada, 2005a). In the future, support measures may even be adaptive to real-time analyses of the interaction process, which can be either automated (Dönmez et al., 2005) or performed online by a human tutor (Zumbach et al., 2005). Assessment methods are further needed in order to evaluate the effects that computer support and instruction may have on learners' interactions as opposed to exclusively evaluating learning outcomes (e.g., Carell, Herrman, Kienle, & Menold, 2005; De Wever, Schellens, Valcke, & Van Keer, 2006; Dillenbourg, Baker, Blaye, & O'Malley, 1995; Weinberger & Fischer, 2006). In addition, students may be taught principles of successful collaboration and asked to evaluate their own collaboration in order to scaffold learning and foster meta-cognitive skills (Lee et al., 2006; Prins, Sluijsmans, Kirschner, & Strijbos, 2005).

Any researcher with interest in studying collaborative processes has to answer two basic questions: 1) which aspects of the collaborative process are relevant for its success and should therefore be observed? And 2) how (employing what kind of instrument, producing what kind of data) should these process aspects be assessed? The first question refers to the model of “good” collaboration the researcher employs; the second question is a methodological one. As a truly interdisciplinary field, CSCL offers a fruitful diversity of perspectives to choose from when looking for an answer to any of these two questions. At its current stage of development, however, the field of CSCL is still lacking a comprehensive theory as well as a shared methodology that would allow for comparisons across studies from all of its different sub-fields. So far, many specific aspects of collaboration have been assessed with very different tools, ranging from ethnographic studies (e.g. Koschmann et al., 2003) to automated corpus and log file analyses (e.g. Donmez et al., 2005; Nurmela, Palonen, Lehtinen, & Hakkarainen, 2003). Therefore, efforts are being made to achieve greater convergence regarding both theoretical models and methodology within CSCL. We too see our paper as a contribution towards the development of more generic assessment methods in CSCL. Its main concerns are to identify process dimensions that determine the quality of computer-supported problem solving and learning in a broad variety of collaboration settings, and to present a rating scheme that can be used to quantify the quality of these process dimensions.

The first part of this paper describes how our process dimensions were defined based on data-driven analyses of collaborative problem-solving as well as general theoretical considerations. In other words, we will first answer the question concerning our model of “good” collaboration. Also, the motivation for choosing a rating scheme rather than a coding scheme for the purpose of assessing process

quality is explained (answering the methodological question). The second part of the paper presents the results of an evaluation of this rating scheme based on process data of a sample of 40 dyads from a study on computer-supported interdisciplinary problem solving in a videoconferencing setting (Rummel, Spada, & Hauser, 2006).¹

A new instrument for assessing the quality of computer-supported collaboration: Development

Which aspects of the collaborative process are relevant for its success and should therefore be observed? In principle, there are two complementary approaches to answering this question: the researcher can either start with the data at hand or with a theoretical model in mind. The researcher who tries to bracket out all *a priori* assumptions and categories and strives to describe phenomena as they emerge from the data will gain insights that are deeply rooted in the characteristics of a given collaborative situation (e.g., by describing typical actions shown by members of particular types of groups, like the “problematizing move” in problem based learning groups [Koschmann et al., 2003]). However, these phenomena will probably be hard to transfer to other collaborative situations. On the other hand, researchers who define what they want to observe on the basis of theoretical assumptions will be able to compare a wider range of collaborative situations against the background of their theoretical model (e.g., by judging the level of perspective taking realized in online discussions against a theoretically derived standard [Järvelä & Häkkinen, 2003]). In turn, they will be in danger of overlooking what makes a given collaborative situation special.

¹ A preliminary version of the rating scheme and its evaluation are described in Spada et al. (2005)

In our research, we combined a bottom-up and a top-down approach in order to arrive at dimensions that were both grounded in the data and defined abstractly enough to be transferable to a broader range of computer-supported collaboration scenarios. In particular, *a qualitative content analysis of transcribed dialogue* from an empirical study on computer-supported collaborative problem solving (see “research context”) was combined with *theoretical considerations based on literature* from areas such as collaborative learning, computer-mediated communication, and group decision making. In the following, the empirical study of computer-supported collaborative problem solving that constitutes the empirical basis for the development of the rating scheme is briefly described. The qualitative content analysis performed on the empirical data from this study and its results are presented next. After that, five broad aspects of successful collaboration that were identified from the literature review are described from a theoretical viewpoint. For each aspect, it is set forth which dimensions resulted as a synthesis of the empirically induced categories with the theoretical considerations.

Research context

The development of our method for assessing the quality of collaborative processes was embedded in a larger research project on instructional support for computer-supported, collaborative, interdisciplinary problem solving. The primary aim of this research project was to develop instructional measures to promote students’ subsequent collaboration. Two studies have so far been conducted within this project (Rummel & Spada, 2005b, Rummel et al., 2006). Data from Study 1 were used in the development of the rating scheme’s dimensions and data from Study 2 in its evaluation (see the second part of this paper: evaluation).

In both studies, dyads consisting of a medical student and a student of psychology collaborated via a desktop videoconferencing system. They worked on hypothetical patient cases that had been carefully designed to require the combined application of both medical and psychological expertise to be solved correctly. The desktop videoconferencing system allowed participants to see and hear each other while discussing the case. It included a shared workspace they could use to prepare a written joint solution as well as two individual text editors. An instructional approach was taken in order to improve collaboration: dyads underwent a learning phase (experimental phase) before they collaborated freely during a test phase. The main goal was to evaluate two methods of instructional support that were implemented in the learning phase. In the *model conditions*, participants observed a model collaboration in which two collaborators solved the first patient case. The model presentation consisted of recorded dialogue and animated text clips that allowed participants to follow the development of a model solution in the shared text editor. In the *script conditions*, participants were provided with a script guiding them through their collaboration on the first case. Study 2 also investigated the effects of *elaboration support* provided in addition to model or script.

Data for the bottom-up analysis were taken from Study 1. In Study 1, four experimental conditions were compared (Table 1). Students in the *model condition* observed the model collaboration during the learning phase. Students in the *script condition* followed the collaboration script during the learning phase. There were two control conditions: students in the *unscripted condition* collaborated on the first case without receiving any specific instruction for their collaboration. Students in the *control condition* did not take part in the learning phase at all, but collaborated only on the second case. Dyads in all four conditions were asked to develop a diagnosis

and set up a therapy plan for the second case. The collaboration was videotaped. A post-test assessed individual knowledge about relevant aspects of collaboration in the present setting.

Bottom-up: Empirically induced categories

In a bottom-up approach, a multi-step analytical procedure built on the qualitative methodology developed by Mayring (2003) was followed in order to identify aspects of successful collaboration (Sosa y Fink, 2003). Starting points were the video recordings of collaboration from Study 1. Four dyads were selected for analysis. Two were taken from the unscripted condition, and two from the control condition. By means of this selection, we were able to observe naturally occurring collaboration that had not been influenced by the instructions and our underlying model of collaboration. In order to maximize variance, one successful and one unsuccessful dyad were selected from each of the two conditions. Their collaborative dialogue was transcribed. The qualitative content analysis performed on the transcripts involved a stepwise reduction of the material, through paraphrasing, elimination and generalization according to the rules established by Mayring (2003). Each step was documented and a final set of six categories was described and completed with anchoring examples (Sosa y Fink, 2003). Three of these categories tapped into the interpersonal relationship of the collaborators: “*goal conformity*” (e.g., agreeing upon a shared goal), “*self presentation*” (e.g., demonstrating one’s expertise by using technical terms), and “*handling of conflicts*” (e.g., uttering dissent matter-of-factly). One category assessed “*task alignment and performance orientation*” (e.g., approaching a given problem in a systematic fashion) and another one the “*construction of a shared knowledge base*” (e.g., pooling information). Coordination

of both the communication and the problem-solving process were subsumed under one category, “*coordination*” (e.g., making a plan for how to solve the case). In order to validate the categories, two coders, including the first author, applied them to the collaboration records of dyads from Study 1 that had not been used in the content analysis. This procedure has been proposed by Mayring (2003) as a way to safeguard the validity of inductively derived categories. However, inter-observer agreement proved to be hard to achieve because the categories were still too close to the content of the four specific dialogues from which they had been derived. Therefore, they were difficult to apply to new material. Some aspects (e.g., grounding the conversation on a moment-to-moment basis) that would have been relevant for assessing the new dyads were missing. Thus, a complementary top-down approach was taken in order to refine these categories and arrive at process dimensions that would be relevant in a broader range of CSCL scenarios. We reviewed literature on computer-supported collaborative learning and working in order to identify aspects of successful collaboration under the conditions of video-mediated communication and complementary expertise. The search was guided by the results of the bottom-up approach.

Top-down and synthesis: Aspects of successful collaboration

The theoretical considerations that guided the refinement of our empirically induced categories and the development of the rating scheme addressed five broad aspects of the collaboration process: communication, joint information processing, coordination, interpersonal relationships, and individual motivation. In the following, relations with the empirically induced categories are identified and the resulting rating scheme dimensions are introduced for each aspect. In total, the final rating scheme comprises

nine dimensions that cover the essence of the six empirically induced categories and all of the five aspects of collaboration considered important from a theoretical point of view (Table 2). A more detailed description of the rating scheme's dimensions can be found in the Appendix.

Communication

The success of any kind of collaborative activity depends, first of all, on effective *communication*. A “common ground” of mutually shared concepts, assumptions and expectations has to be actively established and enlarged during conversation (Clark, 1996). To do so, speaker and listener must collaborate in ensuring understanding and in “grounding” their conversation (Clark & Brennan, 1991). Speakers try to make their contributions understandable. In particular, they must tailor their utterances to their partner's presumed knowledge level, a task that seems to be particularly hard to accomplish for experts talking to lay-persons or experts from other domains; they generally find it hard to ignore their own, specialized knowledge (Nickerson, 1999; Jucks, Bromme & Runde, 2003). The listener, on the other hand, is responsible for giving positive evidence of his or her understanding (Clark & Brennan, 1991). In face-to-face conversation, this is usually achieved via eye contact or short verbal and nonverbal acknowledgments. However, in video-mediated communication, eye-contact usually is impossible and much non-verbal information is lost (Angiollilo, Blanchard, Israelski, & Mane, 1997; Rummel & Spada, 2005a). Thus, participants need to employ more explicit feedback strategies, like verbal acknowledgements or paraphrases (Clark, 1996), and to check on their understanding more often than in face-to-face conversations (Anderson et al., 1997).

As a prerequisite for a successful grounding process, participants need to ensure mutual attention (Clark, 1996). A participant wishing to start a new episode of conversation has to check his or her partner's availability first (Whittaker & O'Conaill, 1997). Further, turn-taking needs to be managed during conversation. Although turn-taking is governed by implicit rules (Sacks, Schegloff, & Jefferson, 1974) that normally ensure relatively smooth transition in face-to-face communication, even small transmission delays in video-mediated communication can severely disrupt these implicit mechanisms (O'Conaill & Whittaker, 1997). Thus, more explicit strategies have to be employed by participants, like handing over turns explicitly by asking a question or naming the next speaker (O'Conaill & Whittaker, 1997). To summarize, communicators have to coordinate both the *content* and the *process* of their conversation (Clark, 1996).

In our empirically derived categories, the coordination of both communicative process and content had been subsumed under the broad category of "coordination." For the purpose of a more detailed analysis of dyads' activities it was decided to distinguish basic communication processes from higher-level coordination. Further, the distinction between the coordination of communicative content and communicative process was adopted from Clark's (1996) communication theory. Thus, the first two dimensions of the rating scheme were defined as "*sustaining mutual understanding*" (which assesses grounding processes) and "*dialogue management*" (which assesses turn taking and other aspects of coordinating the communication process).

Joint information processing

Collaborative problem solving requires participants to pool and process their complementary knowledge in a process of group-level *information processing* (Hinsz,

Tindale, & Vollrath, 1997; Larson & Christensen, 1993). Like face-to-face groups, partners in computer-supported collaboration must avoid falling prey to the general tendency of discussing primarily such pieces of information that were known to all group members from the start (Stasser & Titus, 1985)---even more so in interdisciplinary collaboration where the relevant information is distributed between experts (Rummel & Spada, 2005a). Meta-knowledge about each others' knowledge bases and domains of expertise, i.e., a transactive memory system (Wegner, 1987), will facilitate the pooling of information (Larson & Christensen, 1993; Moreland & Myaskovsky, 2000; Stasser, Stewart & Wittenbaum, 1995). In this way, participants are able to use one another as a resource for problem solving and learning (Dillenbourg et al., 1995). Information can be pooled by eliciting information from one's partner or by externalizing one's own knowledge (Fischer & Mandl, 2003). However, explanations must be timely and given at an appropriate level of elaboration in order to be helpful (Webb, 1989).

On the basis of the pooled information, collaborators must then reach a decision concerning the solution alternatives. This decision should be preceded by a process of critically evaluating the given information, collecting arguments for and against the options at hand, and critically discussing different perspectives (Tindale, Kameda, & Hinsz, 2003). Pressure towards group conformity (e.g., Janis, 1982) as well as the tendency to avoid conflict and agree on a precipitate, illusory consensus (Fischer & Mandl, 2003) can be counteracted by group norms valuing critical thinking (Postmes, Spears, & Cihangir, 2001) and monitoring strategies emphasizing the quality of the group's solution (Tindale et al., 2003).

The aspect of joint information processing had been reflected in the empirically derived category of "construction of a shared knowledge base." The focus had

primarily been on the processes of eliciting and externalizing information, while little attention had been given to the process of decision making. For the rating scheme, two separate dimensions were defined: “*information pooling*” (eliciting information and giving appropriate explanations) and “*reaching consensus*” (discussing and critically evaluating information in order to make a joint decision).

Coordination

Particularly in complex, non-routine tasks, the *coordination* of joint efforts is a crucial factor for the success of collaboration (Malone & Crowston, 1990, 1994; Wittenbaum, Vaughan & Stasser, 1998). Coordination is necessary because of interdependencies that arise when subtasks build upon each other, when time is limited, or when group members depend on the same resources (Malone & Crowston, 1990, 1994). Discussing plans for how to approach a task and negotiating the joint efforts have been shown to be important for the quality of students’ collaborative activities and outcomes (Erkens et al., 2005; Barron, 2000). In planning their work, collaborators must take into account the nature of the task (Steiner, 1972) as well as their individual resources and fields of expertise (Hermann, Rummel, & Spada, 2001). For divisible aspects of the task, individual work phases should be scheduled so that collaborators can bring their individual domain knowledge to bear, while joint phases are necessary for working on more integrative aspects of the task and ensuring a coherent joint solution (Hermann et al., 2001). In order to manage time constraints, a time schedule should be set up (Malone & Crowston, 1994). In computer-mediated collaboration the aspect of technical coordination needs to be addressed in addition to task division and time management (Fischer & Mandl, 2003). Shared applications, for example, constitute resource interdependencies that can be managed by setting up allocation rules (Malone & Crowston, 1990).

In the bottom-up analysis, most coordinative activities had been subsumed under the broad category of “coordination.” To better differentiate between different kinds of dependencies and thus different kinds of coordinative activities, three dimensions were chosen to represent this aspect in the rating scheme. The dimension of “*task division*” was defined to assess how well participants manage task-subtask dependencies. The dimension of “*time management*” assesses how participants cope with time constraints and the dimension of “*technical coordination*” assesses how they cope with technical interdependencies.

Interpersonal relationship

Successful collaborative interactions are characterized by constructive *interpersonal relationships*. Collaborators often hold complementary knowledge that must be integrated in order to arrive at an optimal solution. They will be best able to do so in a relationship in which each of them holds the same status, and in which perspectives are negotiable in a critical discussion (Dillenbourg, 1999). Dillenbourg has termed this a “symmetrical” relationship. Further, a respectful and polite tone of the conversation will help communicators to maintain “face” (i.e., feelings of self-worth and autonomy) and thus avoid negative emotions that would distract their attention from the task (Clark, 1996). A constructive interpersonal relationship may be threatened by arising conflicts, e.g., if partners disagree on how to reach a shared goal. However, they can promote productivity if managed constructively (Deutsch, 2003). To achieve this, Deutsch advises collaborators to avoid stereotyped thinking and aggression, and instead to define conflicts as problems to be solved collaboratively.

A collaborative orientation toward the task and towards one’s partner had been reflected in the empirically induced categories of “goal conformity” and “handling of

conflicts,” while interacting in a professional tone, and thus taking on the roles of collaborating experts, had been the essence of the category of “self presentation.” In the rating scheme, however, only one dimension was defined for this aspect of collaboration, reflecting Dillenbourg’s (1999) concept of the relational symmetry underlying collaborative interactions. This dimension, termed “*reciprocal interaction*,” denotes respectful, collaboratively oriented social interactions and the partners’ equality in contributing to problem solving and decision making, both of which should result from a symmetrical interpersonal relationship.

Motivation

Last but not least, the collaboration process will reflect participants’ individual *motivation* and their commitment to their collaborative task. Motivated participants will focus their attention on the task and co-orientate their actions around it, resulting in shared task alignment (Barron, 2000). Possible motivation losses due to the group situation can be counteracted, for example, by strengthening individual accountability through mutual feedback (Johnson & Johnson, 2003). Individual collaborators may employ volitional strategies to keep up a high level of expended effort in their contribution toward the joint task, including focusing their attention on solution-relevant information, keeping their environment free of distractions, or nurturing positive expectations regarding the collaborative outcome (Heckhausen, 1989).

The motivational aspect of collaboration had been reflected in the empirically induced category of “task alignment and performance orientation,” which was assessed on the level of the dyad. However, from further observations of the dyads’ collaboration it became clear that participants sometimes differed substantially in their levels of task engagement, their willingness to spend effort on the task and to give feedback, and in

their application of volitional strategies. Thus, the decision was made to assess participants' motivation individually in our rating scheme. The resulting dimension of "*individual task orientation*" was rated separately for each participant.

Instrument development: How to quantify process quality?

A rating scheme was chosen as the most suitable method of assessing the quality of the collaborative process for two main reasons: 1) the possibility to judge quality instead of frequency, and 2) the possibility to apply the method to video recordings without the need for time-consuming transcription. First, compared to coding schemes (e.g., De Wever et al., 2006), which are employed to assess the *frequency* of specific behavioral indicators or types of utterances, a rating scheme allows a more direct assessment of process *quality*. Even though coding schemes have proven very useful in studies focusing on the relevance of specific indicators for the success of collaborative learning (for example, particular kinds of meta-cognitive statements, as studied by Kneser and Ploetzner [2001]), a general problem with these approaches is that the number of behavioural indicators often does not inform one about the success of collaboration (Rummel & Spada, 2005a). For example, if a task has to be finished within a certain time-limit, more coordinative utterances do not necessarily indicate better collaboration, because too much coordinative dialogue reduces the time available for the task itself. Too many coordinative utterances might even be an indicator of failed attempts to coordinate collaboration efficiently, and thus indicate ineffectual coordination. In contrast, a rating scheme allows judging the observed behaviors against a defined standard (Kerlinger & Lee, 2000), and thus yields a direct evaluation of the quality of the collaborative process. As a trade-off, details of the collaboration process are lost due to the aggregation processes involved in rating

process quality. However, since our goal was to provide a method that could be used to evaluate the quality of collaboration processes on a relatively global level, a rating scheme constituted the most effective type of instrument. Second, a rating scheme is economical because it does not require the transcription of dialogue, but allows one to work with video recordings of the collaboration process. After sufficient training, the ratings for each video can be obtained from a single round of viewing the tape (though some extra time needs to be allotted for breaks and the reviewing of selected passages). Thus, this method is also time efficient.

The rating scheme

Our rating scheme comprises nine process dimensions (Table 2). The assessment of process quality requires a certain amount of interpretation by the rater, and thus might result in low objectivity if raters are not carefully trained. To counteract this problem, a rating handbook was written and used in rater training in order to standardize judgment and improve objectivity. The rating handbook contained a detailed description of each of the nine dimensions, along with illustrative examples and questions intended to guide raters' attention toward specific aspects of the collaborative process. The descriptions of the collaborative dimensions built on distinct behavioral acts that could be observed from video recordings of the collaboration process. Rating instructions were given by describing the "ideal" version of the dimension at hand, regarding both desirable characteristics that ought to be present as well as undesirable characteristics that ought to be absent. The raters' task was to judge to what extent the observed behaviour matched the description in the rating handbook. In this way, the endpoints of the rating scales were defined as a "very good" match on the positive side and a "very bad" match on the negative side.

Rating scales yield data that can be treated as approximately interval-level, in particular if “only the endpoints of the scale are named and denote the extremes of a continuum” (Wirtz & Caspar, 2002, p. 124; translation by the authors). Therefore, only the endpoints of our rating scales were anchored verbally, while gradations were represented numerically. Even though for some dimensions (e.g., dialogue management) the dyads’ performance may have varied from episode to episode, the raters were required to base their judgment on the aggregated impression of how well a dyad performed in general on the dimension at hand. A shortened version of the rating handbook can be found in the Appendix.

A new instrument for assessing the quality of computer-supported collaboration: Evaluation

The rating scheme was evaluated in the complete sample ($n = 40$ dyads) of Study 2 (Rummel et al., 2006), which investigated the effects of *elaboration support* provided in addition to the instructional measures that had already been employed in Study 1. In Study 2, five experimental conditions were compared (Table 3). As in Study 1, students in the two *model conditions* observed a model collaboration, and students in the two *script conditions* followed a collaboration script during the learning phase. In the conditions with elaboration support (the “*plus*” conditions), participants received instructional explanations and prompts for individual and collective self-explanations in addition to either the model or the script. Students in the *control condition* worked on both patient cases without receiving any specific instruction regarding their collaboration.

All dyads were asked to collaboratively develop a diagnosis for the second case during the test phase. A post-test assessed individual knowledge about relevant aspects of collaboration in the present setting. The rating scheme was applied to the video recordings taken of dyads' collaboration during the test phase. For one hour of videotaped collaboration, about two hours of time were needed for viewing and rating.

Method

Data

The sample of Study 2 consisted of 40 dyads, i.e., 80 participants. Both the medical and the psychology students had a mean age of 25 years and were in an advanced phase of their studies. Collaboration in the test phase had been videotaped for all dyads. Each tape contained approximately 55 minutes of recorded collaboration. All tapes were viewed completely. Thus, the total sample consisted of about 37 hours of videotaped collaboration.

Rating procedure

The rating sheet listed ten scales, one for each of the first eight dimensions, and two scales for the dimension of "individual task orientation," which was assessed separately for each member of the dyad. The scales had five steps that went from -2 (very bad) to +2 (very good). The rating sheet left some room under each dimension, and raters were encouraged to take notes on their impression of the dyad's performance in order to aid their memory and disambiguate the ratings. The videos were watched and rated in random order.

Eight dyads were rated by a trained second rater². In the co-rated sample, each of the five experimental conditions was represented by at least one dyad. Raters were not informed about the experimental condition a dyad had participated in; however, sometimes the experimental condition could be inferred from the dyad's dialogue. In order to reduce the memory load, each video was split into three blocks that were rated separately. Later, the mean value for the three sections was calculated for each dimension and served as the overall rating for the dyad.

Measures

For the empirical evaluation of the rating scheme, measures of inter-rater reliability and consistency, as well as measures reflecting the relationship between the dimensions were used. In addition, we report results from comparisons of the experimental conditions that demonstrate the rating scheme's usefulness in assessing differential effects of instruction on students' collaboration, as well as correlations of the process ratings with two outcome measures (see Table 4 for an overview of reported measures).

As a measure of inter-rater reliability, the intra-class correlation (ICC, adjusted, single measure) for each dimension was calculated in the sample of co-rated dyads ($n = 8$).

While the ICC cannot be applied to dichotomous or nominal-level coding data, its use is recommended for approximate interval-level rating data (Wirtz & Caspar, 2002).

According to Wirtz and Caspar (2002, p. 234), ICCs above .7 allow for meaningful

² For rater training, the co-rater read the rating handbook and clarified questions with the trainer (first author). In addition, video sequences were selected in order to illustrate each of the dimensions described in the rating handbook (only videos were selected that were not part of the sample to be rated by the co-rater). The tape of one dyad whose members collaborated especially well was viewed completely. All video examples were accompanied by oral explanations from the trainer. The co-rater rated two additional videos for training purposes (these videos were not part of the sample in which inter-rater reliability was determined), and differences between her and the trainer's ratings were discussed.

group-level analysis. Before the ratings for the three separate blocks were collapsed for each dimension, their internal consistency (Cronbach's α) was analyzed. This was done for the whole sample ($n = 40$). Collaboration quality may of course change in the course of a dyad's collaboration. Therefore, low consistency may indicate a rating problem, but also a real change over time. For descriptive purposes, the correlations between the dimensions were also calculated (product-moment correlation r) for the complete sample ($n = 40$).

The rating scheme was used in data analysis for Study 2 in order to test for effects of the instruction on dyads' collaboration process (Rummel et al., 2006). Dyads from the five experimental conditions in Study 2 were compared by means of a MANOVA with subsequent, post-hoc ANOVAs with the experimental condition as an independent variable. In this way, the inflation of the type-I-error that would result from a series of independent ANOVAs was prevented. The results are repeated here because they point toward the rating scheme's sensitivity for detecting differential effects of instruction on collaborative processes, and thus are interesting for the evaluation of the rating scheme itself. It would have been desirable to test the rating scheme's sensitivity for measuring collaboration quality by comparing it with other measures of process quality. However, since process analyses are very time-consuming, especially for a body of over 30 hours of recorded collaboration, we were not able to conduct such additional analyses.

In order to evaluate the rating scheme's predictive validity, the process ratings were correlated with an expert rating of the quality of the dyad's joint solution. The expert read the written diagnosis and assigned grades as a school teacher would have done, taking into account the argumentation structure and coherence of the explanations the

students gave for the patient's symptoms. The post-test that participants had to work on individually after collaboration was used as an additional outcome measure. In this test, participants were asked to describe the elements and work phases that should be present in a fictive, "ideal" collaboration on the same type of task that they had just completed themselves. We included this test in order to assess participants' knowledge about central aspects of good collaboration in the given scenario, i.e., what they had learned from the instruction provided in the learning phase and from their own collaboration during the test phase. For each dyad, the mean value of the two individual test scores was calculated. Correlations with the process ratings were then determined using this mean value, except for the dimension of "individual task orientation": here, correlations were calculated separately for the medical and the psychology students using individual test scores.

Results

Inter-rater reliability and consistency

Inter-rater reliability (Table 5) was satisfactory for the majority of the dimensions. The ICC was found to exceed .7 for the three coordinative dimensions ("task division," "time management," and "technical coordination") and to be close to .7 for "sustaining mutual understanding" and "reaching consensus." Nevertheless, all dimensions were included in further analyses. However, results from dimensions with low inter-rater reliability must be interpreted carefully. The rating instructions for these dimensions were once more revised, but remain to be tested in a new sample and with improved rater training.

Cronbach's α for the three consecutive ratings for each dyad was satisfactory (Table 5). As we were interested in the dyads' overall performance, the three ratings were

collapsed by calculating the mean value, which served as the basis for all further analyses.

Interrelation between dimensions

The process ratings correlated moderately to highly, with the highest correlations between those dimensions designed to assess related concepts (Table 6). For example, some of the highest correlations were found between the three dimensions assessing coordination: “task division,” “time management” and “technical coordination.” All correlations were positive, indicating that good dyads collaborated well and bad dyads collaborated badly on most of the dimensions.

Instructional effects

The rating scheme was successfully applied to detect the effects of the instruction given in the learning phase on the subsequent, free collaboration that took place during the test phase: The MANOVA revealed a significant difference between the experimental conditions (Wilk’s Lambda: $F = 1.77$; $df = 100.44$; $p = .01$, partial $\eta^2 = .39$), indicating an overall effect of the instructional measures on the quality of collaboration (Rummel et al., 2006). In comparing the groups’ mean values for each of the dimensions with post-hoc ANOVAs (Table 7), two distinct patterns were identified. On several dimensions, the control group obtained the lowest ratings, the script groups substantially better ones, and the model groups received the best ratings. In the model groups, the dyads who had received additional elaboration support (Model plus) obtained even higher ratings than those who had not. This first pattern is shown by the three coordinative dimensions---“task division,” “time management” and “technical coordination”---as well as by “information pooling,” even though significance was only reached in the case of “task division” and “time management”

(Table 7). This first pattern is illustrated by the dimension of “task division” in Figure 1. A second pattern became visible for the dimension of “individual task orientation” (Figure 2). Here, the two script conditions obtained the lowest ratings, followed by the control condition. The model conditions still obtained the best ratings. Differences reached significance for both the students of psychology and the medical students (Table 7). A similar trend was visible in the dimension of “sustaining mutual understanding,” but didn’t reach significance.

The ratings revealed that the instructional methods employed in the learning phase had differential effects on the quality of the collaboration during the test phase (see Rummel et al., 2006, for a more detailed discussion). While no systematic differences were found concerning the presence of additional elaboration support, the collaboration of the dyads differed according to the kind of instructional support they had received. The two model conditions profited most. They did not only show the best coordination of their collaboration regarding task division, time management, and the management of technological constraints, but also the highest individual task orientation. Dyads in the script conditions seem to have profited regarding the coordination of their collaboration as well, though not as much as the dyads in the model groups. However, having to follow a collaboration script during the learning phase seems to have lowered the participants’ interest and engagement in the task, leading to a relatively low individual task orientation.

These results are in accordance with results from Study 1, where the model condition outperformed the scripted condition and the two uninstructed conditions on several variables (Rummel & Spada, 2005b). Thus, they point towards the rating scheme’s sensitivity for detecting effects of instruction on subsequent collaboration, even though no second measure of process quality was available to confirm these effects.

Process-outcome validity

The expert ratings of solution quality (Figure 3) showed a pattern similar to the one found for the dimensions of “individual task orientation” (Pattern 2; compare Figure 2). The differences between the experimental conditions did not reach significance ($F(4;35) = 1.89$; $p = .13$; $\eta^2 = .18$).

No substantial correlations between process ratings and solution quality were found. Of course, these process-outcome correlations are not only contingent on the reliability of our process ratings but also on the reliability with which the joint outcome was assessed. Since the participants of our study had to solve complex tasks, assessing the quality of the solution was not trivial. Process and outcome measures might show a stronger relation when applied to problems whose solution quality is easier to evaluate.

Higher correlations were obtained between the quality of participants' collaboration and their score on the individual post-test. As can be seen from Table 8, participants who collaborated well with their partners, particularly regarding the coordination of their work, or exhibited a high individual task orientation, were also able to state principles of good collaboration in the post-test. Thus, the process ratings corresponded with the mental representation of good collaboration held by the participants.

Discussion

In this paper, a rating scheme for assessing the quality of collaborative processes in computer-supported problem-solving and learning settings was presented, and results from its evaluation on a sample of dyads from a collaboration experiment were presented. The rating scheme assesses collaboration quality on nine process

dimensions that integrate results from a qualitative analysis of transcribed collaboration dialogue with theoretical considerations based on the relevant literature. The nine dimensions of the rating scheme are: sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction, and individual task orientation. Together they cover important aspects of cooperation: communication, joint information processing, coordination, interpersonal relationship, and motivation. Thus, the rating scheme enables the user to assess the quality of the collaborative process on a broad scale. The rating instructions are both grounded in observations of typical behaviors and rooted in theoretical concepts. They allow a thorough and systematic review of collaborative process data. The rating scheme affords a direct assessment of the quality of collaborative processes that could not be achieved with coding schemes that merely count frequency. It is time efficient, as it does not require the transcription of dialogue, but can be applied to video recordings of collaboration.

In the ratings scheme's application to a sample of 40 dyads from a study on computer-supported collaborative problem solving, the inter-rater reliability of most, though not all, dimensions proved satisfactory. Rating instructions for dimensions with low inter-rater reliability were revised, but have not yet been tested with a new sample and improved rater training. The consecutive ratings for the three thirds of the collaboration time achieved high consistency for all dimensions, indicating that high reliability may be achieved. Results from comparisons across experimental conditions indicate that the instrument can be successfully used to detect effects of instruction on dyads' subsequent collaboration. However, no other measures of process quality were available to confirm these effects, and thus no conclusive proof of the rating scheme's sensitivity for differences in the quality of collaborative processes can be given at the

moment. Correlations with the quality of the joint solution were disappointing; however, ratings corresponded well with participants' answers on a post-test assessing meta-knowledge about collaboration.

We see our rating scheme as a contribution towards more generic assessment methods in CSCL. This does not mean that we would like to propose ratings of collaboration quality as the standard methodology for analyzing CSCL processes. For a number of research questions, more fine-grained analyses of transcribed material will be the better alternative. However, we propose that the rating scheme's dimensions efficiently capture the essentials of the collaboration process in many areas of CSCL research that involve collaborative problem solving and learning on the basis of complementary expertise. The ratings may be used to evaluate effects of an intervention on collaboration relatively quickly, and point towards aspects of the collaboration that need further improvement. They may also be helpful in identifying aspects of collaboration that should be analyzed in more detail. Further, the proposed dimensions could inform more fine-grained analyses of collaboration quality and guide data aggregation. From an educational perspective, the rating scheme and its handbook offer general principles of effective collaboration that may be used for instructional purposes. For example, students' collaborative meta-knowledge and the quality of their collaboration may improve if they are taught how to apply the rating scheme to recordings of their own or others' collaboration.

The rating scheme was developed and tested in the context of two studies in which students of psychology and medicine collaborated on complex patient cases via a desktop-videoconferencing system. We believe, however, that it could be applicable to video-recordings of most instances of computer-supported, synchronous collaboration in dyads or small groups solving problems that require the integration of

interdependent knowledge from different domains. The specific rating instructions as well as the scale's anchors will have to be adjusted to the characteristics of the collaboration scenario one wishes to study. Raters should be trained in advance in order to be sensitive to relevant characteristics of collaborative processes. Training should include collaboration examples from the population one wishes to study. Under these conditions, rating collaborative processes on all or some of the proposed dimensions is an economic method for assessing collaboration quality. Further research is needed, however, in order to test the rating scheme's usefulness in other collaboration settings.

APPENDIX

Rating handbook for the nine process dimensions

In the following, a shortened version of the rating handbook is given.

Dimension 1: Sustaining mutual understanding

Speakers make their contributions understandable for their collaboration partner, e.g., by avoiding or explaining technical terms from their domain of expertise or by paraphrasing longer passages of text from their materials, rather than reading them aloud to their partner. They make sure they have been understood by eliciting feedback from their partner. Listeners focus their attention on what the speaker is saying and give verbal feedback on their understanding (backchannels, paraphrases), demonstrate their understanding in an appropriate response, or ask for clarification. An episode in which a dyad sustains mutual understanding successfully by eliciting and giving evidence of understanding might look like this (the following is a translated example from a dyad in Study 1):

Psychology student (P): "...Did you understand what I just said?" Medical student (M): "U-hum. That is, you mean, whether now there is a psychotic component in addition to the depression and the multiple sclerosis?" P: "Exactly!"

Dimension 2: Dialogue management

A smooth "flow" of communication is maintained in which little time is lost due to overlaps in speech or confusion about whose turn it is to talk. Turn-taking is often facilitated by means of questions ("What do you think?") or explicit handovers ("You go!"). Speakers avoid redundant phrases and fillers ("um... um", "or...or") at the end of a turn, thus signalling they are done and the partner may now speak. To the observer it is always clear who is talking and who is listening. Before partners start a new conversation after a phase of parallel individual work, they make sure they have their partners attention, e.g., by calling his or her name or by using a meta-statement (e.g., "Can I ask you something?"). The following is a translated example from a dyad in Study 1 who had trouble ensuring a smooth flow of communication; utterances set in brackets within the same row are overlapping in time:

P: [It's certainly a depression] M [...has always been]

M: [Mhm, that's] sure P: [I don't understand you]

M: [But schizophrenia ... schizophrenia] I can't read that. P: [It's certainly a depression ... Sorry?]

Dimension 3: Information pooling

Partners try to gather as many solution-relevant pieces of information as possible. New information is introduced in an elaborated way, for example by relating it to facts that have already been established, or by pointing out its relevance for the solution. In this way, the provider of the information ensures that it actually enters the problem solving process. Participants elicit domain-specific knowledge from their partner, using his or her expertise as a resource. At the same time they make sure that

the aspects that are important from the perspective of their own domain are taken into account and they take on the task of clarifying any information needs that relate to their domain of expertise. For example, the psychology student should ask for information about possible medical causes for the patient's symptoms, and at the same time provide the medical student with the diagnostic criteria for possible clinical psychological diagnoses.

Dimension 4: Reaching consensus

Decisions for alternatives on the way to a final solution (i.e., parts of the diagnosis) stand at the end of a critical discussion in which partners have collected and evaluated arguments for and against the available options. If partners initially prefer different options, they exchange arguments until a consensus is reached that can be grounded in facts (e.g., information from the case materials). Even if partners agree from the start, they still evaluate their shared preference critically and search for facts that support it as well as facts that challenge it. The point at which a final decision is reached is clearly identifiable, i.e., a once agreed upon option is not called into doubt except in the light of new information. For example, dyads should take time to critically assess whether the patient actually shows all the symptoms listed in the diagnostic criteria, and to reflect on which symptoms might have medical causes and thus cannot be counted towards a psychiatric diagnosis.

Dimension 5: Task division

The task is divided into subtasks. Partners proceed with their task systematically, taking on one step toward the solution after the other with a clear goal or question guiding each work phase. Individual as well as joint phases of work are established, either in a plan that is set up at the beginning, or in short-term arrangements that

partners agree upon as they go. Partners define and take on individual subtasks that match their expertise and their resources. The work is divided equally so none of the collaborators has to waste time waiting for his or her partner to finish a subtask. An exemplary task division (within the context of our study) could, for example, involve the following phases: in the beginning of their cooperation, after clarifying initial questions, partners take some time for individual work in which they try to summarize what is important from the perspective of their own discipline and develop first hypotheses concerning the diagnosis. These thoughts are then exchanged and discussed in a joint work phase, perhaps with some phases of parallel individual work in which missing information is looked up. Partners agree upon a diagnosis in a final discussion phase, and each takes some time to write down arguments that support the diagnosis from the perspective of their own discipline. In a concluding joint work phase, individual texts are integrated into a joint written solution, and final corrections are made.

Dimension 6: Time management

Partners monitor the remaining time throughout their cooperation and make sure to finish the current subtask or topic with enough time to complete the remaining subtasks. They check, for example, whether the current topic of discussion is important enough to spend more time on, and remind one another of the time remaining for the current subtask or the overall collaboration. They might also set up (and monitor adherence to) a time schedule in which each subtask is allotted a realistic amount of time. Good time management allows partners to finish the task in time without having to rush.

Dimension 7: Technical coordination

Partners master the basic technical skills that allow them to use the technical tools to their advantage (for example, they know how to switch between applications, or how to “copy and paste”). Collaborators further arrange who may write into the shared editor at which time. At least one partner makes use of his or her individual text editor, thus allowing for phases of parallel writing.

Dimension 8: Reciprocal interaction

Partners treat each other with respect and encourage one another to contribute their opinions and perspectives. Critical remarks are constructive and factual, never personal; i.e., they are formulated as contributions toward the problem solution (e.g., “I’m not quite sure about that diagnosis---why don’t we try to list all symptoms first?”). Partners interact as equals, and decisions (e.g., regarding task division or the general procedure towards the problem solution) are made cooperatively. Since both partners hold equal amounts of unique task-relevant knowledge, they should contribute equally toward the problem solution; i.e., the solution process is not dominated by one of the partners (as, for example, in a “tutor-tutee” relationship).

Dimension 9: Individual task orientation

Each participant actively engages in finding a good solution to the problem, thus bringing his or her knowledge and skills to bear. He or she focuses attention on the task and on task relevant information, avoids distractions, and strives to mobilize his or her own as well as the partner’s skills and resources. The participant shows interest in the task or enjoyment of the work involved, for example by expressing a wish to help the patient, a general interest in the subject domains, or pride in work considered well done.

Acknowledgements

The present research was supported by the German Science Foundation (DFG) with project grants Sp 251/16-2 and 16-3. We would like to thank Sabine Hauser and Dejana Diziol for their tremendous help with data collection. Furthermore, we would like to thank the CSCL community for their valuable questions and remarks on a preliminary version of the rating scheme that was presented at the CSCL 2005 conference, as well as three anonymous reviewers for their helpful suggestions.

REFERENCES

- Anderson, A. H., O'Malley, C., Doherty-Sneddon, G., Langton, S., Newlands, A., Mullin, J., et al. (1997). The impact of VMC on collaborative problem solving: an analysis of task performance, communicative process, and user satisfaction. In K. E. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), *Video-mediated communication* (pp. 133-156). Mahwah, NJ: Lawrence Erlbaum Associates.
- Angiolillo, J. S., Blanchard, H. E., Israelski, E. W., & Mané, A. (1997). Technology constraints of video-mediated communication. In K. E. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), *Video-mediated communication* (pp. 51-74). Mahwah, NJ: Lawrence Erlbaum Associates.
- Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *The Journal of the Learning Sciences*, 9, 403-436.
- Carell, A., Herrman, T., Kienle, A., & Menold, N. (2005). Improving the coordination of collaborative learning with process models. In T. Koschmann, D. Suthers, & Chan, T.W. (Eds.), *Proceedings of the CSCL 2005* (pp. 18-27). Mahwah, NJ: Lawrence Erlbaum Associates.

- Clark, D., & Sampson, V.D. (2005). Analyzing the quality of argumentation supported by personally-seeded discussions. In T. Koschmann, D. Suthers, & Chan, T.W. (Eds.), *Proceedings of the CSCL 2005* (pp. 76-85). Mahwah, NJ: Lawrence Erlbaum Associates.
- Clark, H. H. (1996). *Using language*. Cambridge MA: Cambridge University Press.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-148). Washington, DC: American Psychological Association.
- Deutsch, M. (2003). Cooperation and conflict. A personal perspective on the history of the social psychological study of conflict resolution. In M. A. West, D. Tjosvold, & K. G. Smith (Eds.), *International handbook of organizational teamwork and cooperative working* (pp. 9-43). Chichester, UK: Wiley & Sons.
- De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: a review. *Computers & Education*, 46, 6-28.
- Dillenbourg, P. (1999). Introduction: What do you mean by "collaborative learning"? In P. Dillenbourg (Ed.), *Collaborative learning. Cognitive and computational approaches* (pp. 1-19). Amsterdam: Pergamon.
- Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1995). The evolution of research on collaborative learning. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189-211). Oxford: Elsevier/Pergamon.

- Dönmez, P., Rose, C., Stegmann, K., Weinberger, A., & Fischer, F. (2005). Supporting CSCL with automated corpus analysis technology. In T. Koschmann, D. Suthers, & Chan, T.W. (Eds.), *Proceedings of the CSCL 2005* (pp. 125-134). Mahwah, NJ: Lawrence Erlbaum Associates.
- Erkens, G., Jaspers, J., Prangma, M., & Kanselaar, G. (2005). Coordination processes in computer supported collaborative writing. *Computers in Human Behavior*, *21*, 463-486.
- Fischer, F., & Mandl, H. (2003). Being there or being where? Videoconferencing and cooperative learning. In H. van Oostendorp (Ed.), *Cognition in a digital world*. (pp. 205-223): Lawrence Erlbaum Associates.
- Heckhausen, H. (1989). *Motivation und Handeln. [Motivation and behavior]*. Berlin: Springer.
- Hermann, F., Rummel, N., & Spada, H. (2001). Solving the case together: The challenge of net-based interdisciplinary collaboration. In P. Dillenbourg, A. Eurelings, & K. Hakkarainen (Eds.), *Proceedings of the first European conference on computer-supported collaborative learning* (pp. 293-300). Maastricht: McLuhan Institute.
- Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin*, *121*(1), 43-64.
- Janis, I. L. (1982). *Groupthink*. Boston: Houghton Mifflin.
- Järvelä, S., & Häkkinen, P. (2003). The levels of web-based discussions: Using perspective-taking theory as an analytical tool. In H. van Oostendorp (Ed.),

Cognition in a digital world. (pp. 77-95). Mahwah, NJ: Lawrence Erlbaum Associates.

Johnson, D. W., & Johnson, R. T. (2003). Training for cooperative group work. In M. A. West, D. Tjosvold, & K. G. Smith (Eds.), *International handbook of organizational teamwork and cooperative working* (pp. 167-183). Chichester, UK: Wiley & Sons.

Jucks, R., Bromme, R., & Runde, A. (2003). Audience Design von Experten in der netzgestützten Kommunikation: Die Rolle von Heuristiken über das geteilte Vorwissen. [Audience design of experts in net-based communication: the role of heuristics about shared knowledge]. *Zeitschrift für Psychologie*, 211(2), 60-74.

Kapur, M., Voiklis, J., & Kinzer, C.K. (2005). Problem solving as a complex, evolutionary activity: a methodological framework for analyzing problem solving processes in a computer-supported collaborative environment. In T. Koschmann, D. Suthers, & Chan, T.W. (Eds.), *Proceedings of the CSCL 2005* (pp. 252-261). Mahwah, NJ: Lawrence Erlbaum Associates.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research*. Fort Worth: Harcourt College Publishers.

Kneser, C., & Ploetzner, R. (2001). Collaboration on the basis of complementary domain knowledge: Observed dialogue structures and their relation to learning success. *Learning and Instruction*, 11(1), 53-83.

Koschmann, T., Zemel, A., Conlee-Stevens, M., Young, N., Robbs, J., & Barnhart, A. (2003). Problematizing the problem. In B. Wasson, S. Ludvigsen, & U. Hoppe (Eds.), *Designing for change* (pp. 37-46). Dordrecht: Kluwer.

- Larson, J. R., & Christensen, C. (1993). Groups as problem-solving units: toward a new meaning of social cognition. *British Journal of Social Psychology*, 32, 5-30.
- Lee, E. Y. C., Chan, C. K. K., & van Aalst, J. (2006). Students assessing their own collaborative knowledge building. *International Journal of Computer-Supported Collaborative Learning*, 1, 277-307.
- Malone, T. W., & Crowston, K. (1990). What is coordination theory and how can it help design cooperative work systems? *Proceedings of the Conference on Computer-Supported Cooperative Work* (pp. 357-370). Los Angeles, CA.
- Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys*, 26(1), 87-119.
- Mayring, P. (2003). *Qualitative Inhaltsanalyse. Grundlagen und Techniken [Qualitative content analysis. Foundations and techniques]*. Weinheim: Beltz.
- Moreland, R. L., & Myaskovsky, L. (2000). Exploring the performance benefits of group training: transactive memory or improved communication? *Organizational Behavior and Human Decision processes*, 82 (1), 117-133.
- Nickerson, R. S. (1999). How we know - and sometimes misjudge - what others know: imputing one's own knowledge to others. *Psychological Bulletin*, 125(6), 737-759.
- Nurmela, K., Palonen, T., Lehtinen, E., & Hakkarainen, K. (2003). Developing tools for analyzing CSCL process. In B. Wasson, S. Ludvigsen, & U. Hoppe (Eds.), *Designing for change* (pp. 333-342). Dordrecht: Kluwer.
- O'Conaill, B., & Whittaker, S. (1997). Characterizing, predicting, and measuring video-mediated communication: A conversational approach. In K. E. Finn, A. J.

Sellen, & S. B. Wilbur (Eds.), *Video-mediated communication* (pp. 107-132).

Mahwah, NJ: Lawrence Erlbaum Associates.

Postmes, T., Spears, R., & Cihangir, S. (2001). Quality of decision making and group norms. *Journal of Personality & Social Psychology*, 80(6), 918-930.

Prins, F. J., Sluijsmans, D. M. A., Kirschner, P. A., & Strijbos, J.-W. (2005).

Formative peer assessment in a CSCL environment: a case study. *Assessment & Evaluation in Higher Education*, 30 (4), 417-444.

Rummel, N., & Spada, H. (2005a). Instructional support for collaboration in desktop videoconferencing settings. How it can be achieved and assessed. In R.

Bromme, F. W. Hesse, & H. Spada (Eds.), *Barriers and biases in computer-mediated knowledge communication - and how they may be overcome* (pp. 59-88). New York: Springer.

Rummel, N., & Spada, H. (2005b). Learning to collaborate: An instructional approach to promoting problem-solving in computer-mediated settings. *The Journal of the Learning Sciences*, 14(2), 201-241.

Rummel, N., Spada, H., & Hauser, S. (2006). Learning to collaborate in a computer-mediated setting: observing a model beats learning from being scripted. In: S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.), *Proceedings of the International Conference of the Learning Sciences 2006* (p. 634-640). Mahwah, NJ: Lawrence Erlbaum Associates.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematic for the organization of turn-taking in conversation. *Language*, 50, 696-753.

Sosa y Fink, S. (2003). Merkmale gelungener Kooperation. Eine qualitative Analyse netzgestützter Zusammenarbeit. [Characteristics of successful cooperation. A

qualitative analysis of net-based collaboration.] Unpublished diploma thesis, Albert-Ludwigs-Universität, Freiburg.

Spada, H., Meier, A., Rummel, N., & Hauser, S. (2005). A new method to assess the quality of collaborative process in CSCL. In T. Koschmann, D. Suthers, & Chan, T.W. (Eds.), *Proceedings of the CSCL 2005* (pp. 622-631). Mahwah, NJ: Lawrence Erlbaum Associates.

Stasser, G., Stewart, D., & Wittenbaum, G. (1995). Expert roles and information exchange during discussion: the importance of knowing who knows what. *Journal of experimental social psychology, 31*, 244-265.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during group discussion. *Journal of Personality and Social Psychology, 48*, 1467-1478.

Steiner, I. D. (1972). *Group process and productivity*. New York: Academic Press.

Tindale, R. S., Kameda, T., & Hinsz, V. B. (2003). Group decision making. In M. A. Hogg & J. Cooper (Eds.), *Sage handbook of social psychology* (pp. 381-403). London: Sage.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Education Research, 13*, 21-39.

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185-208). New York: Springer.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*, 71-95.

- Whittaker, S., & O'Conaill, B. (1997). The role of vision in face-to-face and mediated communication. In K. E. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), *Video-mediated communication* (pp. 23-50). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. [Inter-rater agreement and inter-rater reliability]*. Göttingen: Verlag für Psychologie.
- Wittenbaum, G. M., Vaughan, S. I., & Stasser, G. (1998). Coordination in task performing groups. In R. S. Tindale et al. (Eds.), *Theory and research on small groups* (pp. 177-204). New York: Plenum
- Zumbach, J., Schönemann, J., & Reimann, P. (2005). Analyzing and supporting cooperative computer-mediated communication. In T. Koschmann, D. Suthers, & Chan, T.W. (Eds.), *Proceedings of the CSCL 2005* (pp. 758-767). Mahwah, NJ: Lawrence Erlbaum Associates.

FIGURE LEGENDS

Figure 1: Mean values and standard errors for “task division” (Pattern 1) for dyads from the five experimental conditions.

Figure 2: Mean values and standard errors for “individual task orientation, medical student” (Pattern 2) for dyads from the five experimental conditions.

Figure 3: Mean values and standard errors for expert ratings of solution quality for dyads from the five experimental conditions. High values correspond to high quality.

TABLE CAPTIONS

Table 1: Experimental conditions in Study 1 (Rummel & Spada, 2005b)

Table 2: Five aspects of the collaborative process and the resulting nine dimensions of the rating scheme

Table 3: Experimental conditions in Study 2 (Rummel et al., 2006)

Table 4: Measures used in the empirical evaluation of the rating scheme

Table 5: Intraclass correlations between the values of the two raters and internal consistency of the three consecutive ratings for all dimensions

Table 6: Correlations between the nine process dimensions

Table 7: Mean values and standard deviations of the dimensions for the five experimental conditions of Study 2

Table 8: Correlations of the process ratings with knowledge about good collaboration as assessed in the post-test

Figure 1:

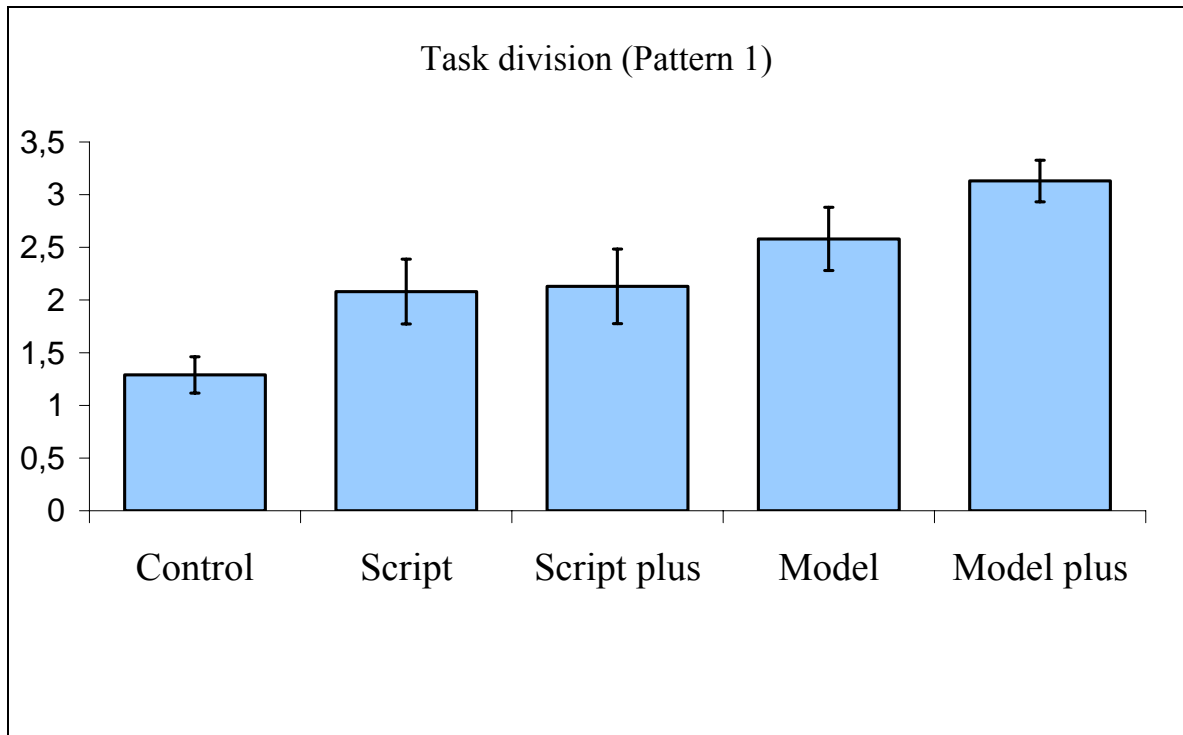


Figure 2:

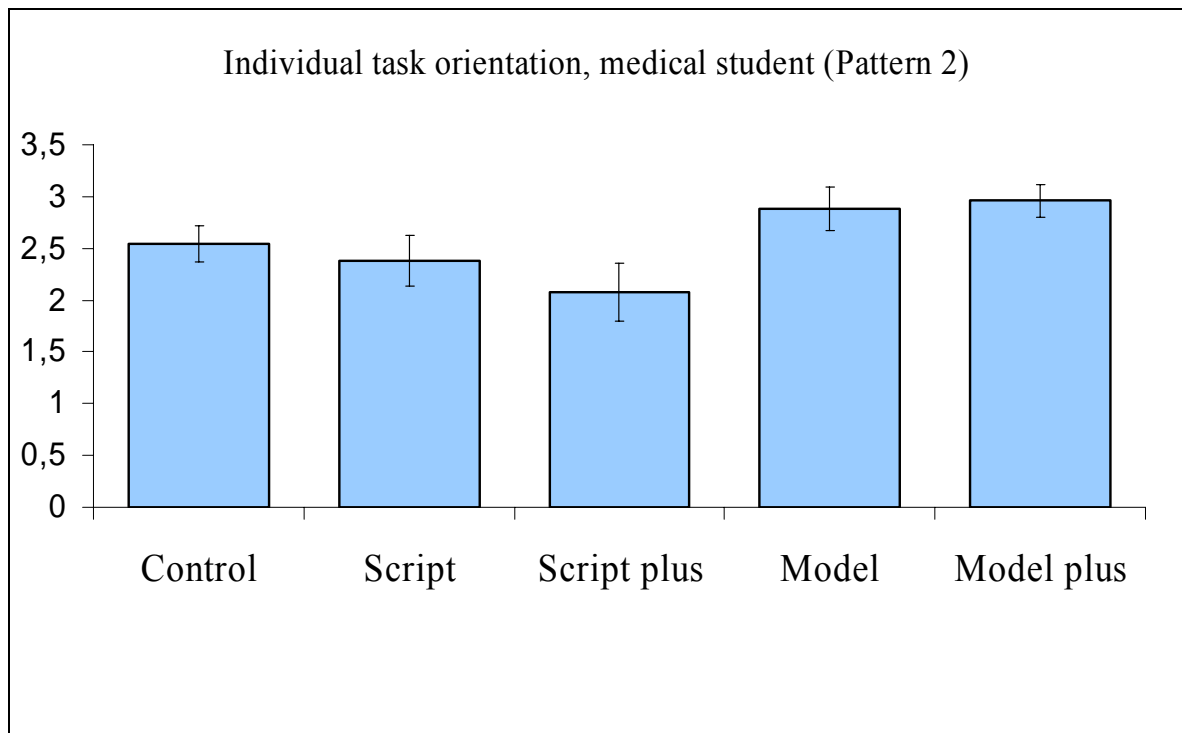


Figure 3:

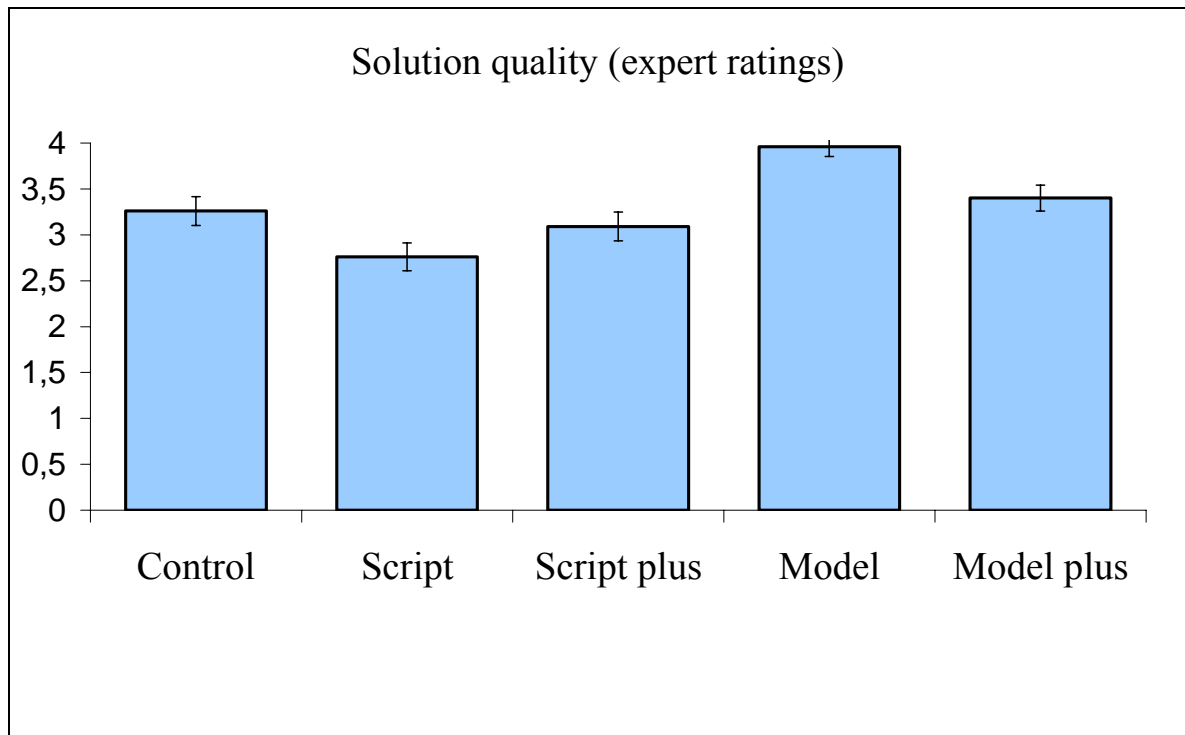


Table 1:

Condition	Learning phase	Test phase
	Diagnosis and therapy plan for Case 1 (120 min)	Diagnosis and therapy plan for Case 2 (120 min)
Model (9 dyads)	observational learning	no further instruction
Script (9 dyads)	scripted collaboration	
Unscripted (9 dyads)	uninstructed collaboration	
Control (9 dyads)	no learning phase	

Table 2:

<p>Communication</p> <ul style="list-style-type: none">1) Sustaining mutual understanding2) Dialogue management
<p>Joint information processing</p> <ul style="list-style-type: none">3) Information pooling4) Reaching consensus
<p>Coordination</p> <ul style="list-style-type: none">5) Task division6) Time management7) Technical coordination
<p>Interpersonal relationship</p> <ul style="list-style-type: none">8) Reciprocal interaction
<p>Motivation</p> <ul style="list-style-type: none">9) Individual task orientation

Table 3:

Condition	Learning phase Diagnosis for Case 1 (55 min)	Test phase Diagnosis for Case 2 (55 min)
Model (8 dyads)	observational learning	no further instruction
Model <i>Plus</i> (8 dyads)	observational learning <i>plus elaboration support</i>	
Script (8 dyads)	scripted collaboration	
Script <i>Plus</i> (8 dyads)	scripted collaboration <i>plus elaboration support</i>	
Control (8 dyads)	uninstructed collaboration	

Table 4:

	Statistical values given	n
Inter-rater reliability	- intraclass-correlation (ICC)	8
Consistency	- Cronbach's α	40
Interrelations between dimensions	- product-moment correlation	40
Instructional effects	- MANOVA - ANOVAs	40
Process-outcome correlations	- product-moment correlation with quality of diagnosis (expert rating) - product-moment correlation with post-test score	40 40

Table 5:

Dimension	ICC	Cronbach's α
Sustaining mutual understanding	.67	.71
Dialogue management	.52	.77
Information pooling	.42	.62
Reaching consensus	.66	.76
Task division	.83	.82
Time management	.86	.83
Technical coordination	.82	.61
Reciprocal interaction	.48	.73
Individual task orientation (P) ^{a)}	.19	.66
Individual task orientation (M) ^{a)}	.38	.77

^{a)} P = psychology student; M = medical student

Table 6:

Dimension	Dialogue Management	Information Pooling	Reaching Consensus	Task Division	Time Management	Technical Coordination	Reciprocal Interaction	Individual Task Orientation (P) ^{a)}	Individual Task Orientation (M) ^{a)}
Sustaining Mutual Understanding	.57**	.28	.23	.53**	.45**	.46**	.41**	.26	.38*
Dialogue Management	--	.31	.25	.43**	.36**	.35*	.49**	.19	.34*
Information Pooling		--	.57**	.58**	.59**	.45**	.34*	.46**	.66**
Reaching Consensus			--	.47**	.43**	.21	.43**	.35*	.48**
Task Division				--	.82**	.74**	.39*	.49**	.59**
Time Management					--	.56**	.28	.45**	.49**
Technical Coordination						--	.27	.34*	.37*
Reciprocal Interaction							--	.09	.53**
Individual Task Orientation (P) ^{a)}								--	.63**

^{a)} P = psychology student; M = medical student

* significant on the 0.05- level ** significant on the 0.01 –level

Table 7

Mean (SD)	Control (n=8)	Script (n=8)	Script plus (n=8)	Model (n=8)	Model plus (n=8)	F (4;35)	p	η^2
Sustaining mutual understanding	1.79 (.71)	1.79 (.47)	1.67 (.73)	2.13 (.69)	2.29 (.97)	1.03	.40	.11
Dialogue management	1.60 (.68)	2.25 (.61)	1.88 (.53)	2.33 (.73)	2.04 (.52)	1.79	.15	.17
Information pooling	1.88 (.73)	2.52 (.72)	2.25 (.61)	2.69 (.74)	2.75 (.71)	2.09	.10	.19
Reaching consensus	1.43 (.89)	1.88 (1.23)	1.67 (.89)	2.31 (.71)	1.65 (.52)	1.14	.36	.12
Task division	1.29 (.49)	2.08 (.87)	2.13 (1.00)	2.58 (.85)	3.13 (.56)	6.04	<.01	.41
Time management	.83 (.56)	1.71 (.68)	2.00 (.84)	2.25 (.98)	3.04 (.86)	8.10	<.01	.48
Technical coordination	2.42 (.58)	2.83 (.59)	2.83 (.67)	2.83 (.69)	3.33 (.31)	2.47	.06	.22
Reciprocal interaction	2.46 (.53)	2.63 (.70)	2.25 (1.07)	2.58 (.49)	2.33 (.79)	0.37	.83	.04
Individual task orientation (P) ^{a)}	2.50 (.25)	2.38 (.60)	2.38 (.55)	2.92 (.53)	3.08 (.24)	4.08	.01	.32
Individual task orientation (M) ^{a)}	2.54 (.50)	2.38 (.68)	2.08 (.79)	2.88 (.59)	2.96 (.45)	2.75	.04	.24

^{a)} P = psychology student; M = medical student

Table 8:

Dimension	Correlation with post-test score
Sustaining mutual understanding	.35*
Dialogue management	.06
Information pooling	.29
Reaching consensus	.26
Task division	.55**
Time management	.59**
Technical coordination	.41**
Reciprocal interaction	.05
Individual task orientation (psychology student)	.35*
Individual task orientation (medical student)	.48**

* significant on the 0.05- level

** significant on the 0.01 –level