Assessing collaboration quality in synchronous CSCL problem-solving activities: Adaptation and empirical evaluation of a rating scheme

Georgios Kahrimanis[1], Anne Meier[2], Irene-Angelica Chounta[1], Eleni Voyiatzaki[1], Hans Spada[2], Nikol Rummel[2], Nikolaos Avouris[1]

[1] Human-Computer Interaction Group, University of Patras, GR-26500 Rio Patras, Greece
{kahrimanis@ece, houren@ece, evoyiatz@ece, avouris@}upatras.gr
[2] Institute of Psychology, University of Freiburg, Engelbergerstr. 41
D-79085 Freiburg, Germany
{anne.meier, nikol.rummel, hans.spada}@psychologie.uni-freiburg.de

Abstract. The work described is part of an ongoing interdisciplinary collaboration between two research teams of the University of Patras, Greece and the University of Freiburg, Germany, which aims at the exchange of analysis tools and data sets in order to broaden the scope of analysis methods and tools available for Computer-Supported Collaborative Learning (CSCL) support. This article describes the adaptation, generalization and application of a rating scheme which had been developed by the Freiburg team for assessing collaboration quality on several dimensions [1]. The scheme was successfully adapted to suit data gathered by the Patras team in a different CSCL scenario. Collaboration quality is assessed by quantitative ratings of seven qualitatively defined rating dimensions. An empirical evaluation based on a dataset of 101 collaborative sessions showed high inter-rater agreement for all dimensions.

Keywords: Computer-Supported Collaborative Learning (CSCL), rating scheme, interaction analysis, collaboration quality.

1 Introduction

Research on technology-enhanced collaborative learning has become more and more interested in studying not only the conditions and outcomes, but also the interactions and processes involved in collaborative knowledge building. As a consequence, analysis tools have been developed for studying interaction processes in a wide variety of technology-enhanced learning settings, using a variety of methodologies [2, 3, 4]. Recently, however, efforts are being made to achieve greater convergence regarding both theoretical models and analysis tools [5]. One way towards achieving convergence is to adapt existing analysis methods, which have been developed and tested

1 Please note that the LNCS Editorial assumes that all authors have used the western naming convention, with given names preceding surnames. This determines the structure of the names in the running heads and the author index.

in one learning setting, to a novel learning setting. Tool adaptation does not only save time and effort that would be necessary for developing a new analysis tool from scratch; it is also a test of whether the theoretical model underlying the tool is capable of capturing important aspects of technology-enhanced collaborative learning across different settings. In this paper, we report on the successful adaptation of an established rating scheme for assessing collaboration quality to data from a novel learning setting, demonstrating that the rating scheme's dimensions, and thus its underlying theoretical model, are capable of capturing the main aspects of collaboration quality across different technology-enhanced learning settings. The work described is part of an ongoing interdisciplinary collaboration between our two research teams at the University of Patras, Greece, and the University of Freiburg, Germany

3 Tool adaptation

Originally, the rating scheme had been developed by the Freiburg team for the purpose of analyzing collaboration quality in the context of interdisciplinary problem-solving between medical students and students of psychology who communicated over a desktop video-conferencing system [6].The scheme employed a multidimensional model of collaboration covering aspects of communication, joint information processing, coordination, relationship management, and motivation [1]. The scheme was adapted to suit data gathered by the Patras team in a very different CSCL scenario: dyads of first-year computer science students interacted through Synergo [7], a network-based synchronous collaborative drawing tool that includes a shared whiteboard and a textual communication facility. Students collaborated in the scope of 45'-75' laboratory sessions without having face-to-face contact. The learning domain was algorithm building in computer science. Each dyad was asked to solve an elementary algorithm exercise by developing a flow-chart representation of the algorithm described in Synergo's shared whiteboard.

The rating scheme was adapted to this specific task and setting by adjusting the number and definitions of the dimensions of the original scheme. Two main phases of adaptation were followed; the first resulted in an adapted definition of the rating scheme's dimensions, and the second served to fine-tune the rating instructions. In the first phase of adaptation a bottom-up approach, which involved identification of "best practice" examples in the sample data, was combined with a top-down process, during which the definitions of all original dimensions were reformulated taking into account constraints arising from the specific collaboration setting (e.g. chat communication; design task). In the second phase of adaptation, the dimensions' definitions were fine-tuned and illustrated with more detail, grounding each dimension's theoretical concepts in specific examples of collaboration practice from the data pool of the first round of adaptation.

4 Dimensions of good collaboration

In the Synergo algorithm task, good collaboration can be characterized on seven rating dimensions (Table 1), covering the same five aspects of collaboration quality that had been defined in the original tool. The first two dimensions assess the aspect of students’ *communication* in the Synergo learning environment. First of all, the success students have in achieving a seamless and efficient communication is determined by observing in how far they maintain *collaboration flow*, i.e. manage dialogue and actions in a way that facilitates references to earlier utterances and actions and helps students maintain a joint focus. For example, students must make sure to react to each others' messages and actions, and must coordinate between the verbal discussion in the chat and the ongoing design of the algorithm in the shared whiteboard. Second, students need to *sustain mutual understanding*, i.e. work towards “common ground” [8]. For example, students should strive to make their actions and chat messages understandable for their partner, e.g. by telling them which object in the whiteboard they are referring to, by explaining the variables they are using, or by informing their partner about the purpose of their actions in the shared whiteboard. Students should also give each other feedback on their level of understanding, e.g. by sending short affirmative messages, or by asking clarifying questions.

Two further dimensions cover the aspect of *joint information processing*. One dimension, *knowledge exchange*, assesses how effectively students exchange information and give explanations. Information in this setting refers mainly to elementary knowledge of algorithm concepts and flowchart notation restrictions. For example, students typically develop small parts of the solution in the form of pseudocode notes individually, which they are expected to exchange and explain to each other. Further, a second information processing dimension assesses students’ *argumentation* quality, e.g. when defending a proposed solution for a part of the algorithm. Negotiating alternatives to the solution and exchanging arguments on the optimal formation of the algorithm also constitute desirable argumentation practices. Another kind of good practice that pertains to the algorithm building domain and relates to this dimension is the “simulation” of the algorithm’s behavior by applying values in the variables

Concerning the aspect of *coordination*, only one dimension was defined: *structuring the problem solving process*. It assesses the extent to which students follow a coherent and efficient plan for jointly developing the algorithm. For example, students can improve their efficiency by defining subtasks and working on different parts of the algorithm in parallel for some time. Further, they should efficiently distribute their time resources to subtasks of the problem. The aspect of *relationship management* is covered by a dimension assessing students’ *cooperative orientation*. For example, students are expected to assist each other, and to handle of conflicts and disagreements in a constructive fashion. Finally, the *motivational aspect* of collaboration is assessed by the dimension of *individual task orientation*, which is rated for each student separately. It assesses the extent to which students are actually committed to solving the task and actively in engage in its solution.

5 Empirical evaluation

The rating scheme was evaluated in a sample of 101 dyads from the Patras data set in order to test whether the instructions provided in the final version of the adapted rating scheme would allow satisfactory inter-rater agreement, and to prepare future activities.

5.1 Rating procedure

Ratings were made while reviewing a dyad's collaboration based on the data logged by Synergo. Logged activities can be reproduced in a video-like format in Synergo's playback mode, allowing raters to jump back and forth and to replay particularly rich collaboration episodes. Each dimension is rated on a 5-point scale ranging from "very low" to "very high" collaboration quality on that dimension. A rating handbook stated the scope and purpose of each dimension, gave an operational definition in a short paragraph, and provided raters with illustrative examples.

The rating procedure was conducted by two raters, one of which had already gained experience from the first round of adaptation. One third of the data was rated by both raters jointly as a training phase for the new rater. After finishing this training phase, another 34 dyads were rated by both raters separately in order to establish inter-rater reliability. The rest of the dataset was split into two so that each rater assessed half of the remaining activities.

5.1 Results

In the co-rated sample, absolute agreement of the ratings ranged between 65% (argumentation) to 85% (knowledge exchange); differences of more than one point on the five-point rating scale were very rare. Accordingly, measures of inter-rater reliability (intra-class correlations for absolute values) were high for all dimensions (Table 1). Thus, this part of the empirical evaluation was considered successful.

An analysis with the complete sample 101 dyads further showed that all dimensions (except the individually rated dimension of "individual task orientation") inter-correlated quite highly ($r > .60$). On the one hand, this shows that the rating scheme is a useful means of obtaining consistent measures of overall collaboration quality; on the other hand, lower inter-correlations would be desirable for obtaining differential assessments of specific aspects of collaboration. This is, however, only possible for dyads that show a medium level of collaboration quality overall, while the sample in which the inter-correlation results were obtained also contained many dyads who collaborated either extremely well or extremely bad and thus obtained extreme ratings on nearly all dimensions (as one would expect, inter-correlations are much lower empirically if only dyads of medium collaboration quality are considered).

Table 1. Rating dimensions and inter-rater agreement for the adapted rating scheme

Rating dimensions	*Inter-rater agreement (ICCs) in the co-rated sample (n=34 dyads)*
Collaboration Flow	.88
Sustaining Mutual Understanding	.92
Knowledge Exchange	.96
Argumentation	.91
Structuring the Problem Solving Process	.96
Cooperative Orientation	.96
Individual Task Orientation	.92

6 Conclusions and future plans

We have described how a rating scheme that had been developed to assess the quality of students' collaboration in one collaborative learning setting (involving students with complementary knowledge backgrounds engaged in medical decision-making and collaborating over a desktop-videoconferencing system) was successfully adapted to assess the quality of students' collaboration in a novel collaborative learning setting (involving students with similar knowledge backgrounds engaged in algorithm building and collaborating using the chat and shared whiteboard facilities in the Synergo learning environment). The adaptation and application of the rating scheme was successful in terms of establishing high inter-rater reliability. Although some significant modifications to the rating scheme were made, the resultant version was close enough to the original so as not to violate it's core rationale. Thus, the multidimensional model of collaboration underlying the tool has been shown to be applicable across very dissimilar CSCL settings, and to be a useful basis for assessment.

The development of the adapted rating scheme has also paved the way for further research paths with both practical and methodological implications. Two studies using the rating scheme and the model of collaboration quality underlying it are currently under way. One project studies whether feedback provided based on the ratings in specific dimensions can lead to improvement of students' subsequent collaboration. A pilot study has already been conducted, in which a human tutor assessed collaboration quality using the rating scheme, and then gave feedback assembled according to a corresponding feedback scheme. Feedback that was based on a profile of high and low ratings achieved by a dyad, and thus tailored to students' specific strengths and weaknesses, was effective in improving students' collaboration [9].

As long-term goals we further aim to develop technical support for facilitating the rating process so that the rating may be used more efficiently in real classroom settings. An interesting new research thread we follow in this context is to use a larger set of activities evaluated with the rating scheme as a rigorous and mature way for predicting collaboration quality in unrated activities based on automatic interaction analysis metrics provided by the Synergo tool. Such metrics can refer to different aspects of collaboration that are reflected in summations of events, like the total number of alterations in students' chat messages or the balance of workspace contributions between two participants. Automatic measures can then be used in statistical methods like regression for the prediction of the scores of collaboration quality according to the rating scheme. In more sophisticated means, machine learning algorithms can be trained for the same purpose. Possible success in this study would provide an efficient way to automatically evaluate activities based not only in log-file based measures, but in in-depth interpretations of researchers as well. This would then provide new opportunities for assessing collaboration quality that demands less work from teachers or researchers. Furthermore, integration of prediction algorithms in Synergo or a tool for supervising multiple CSCL activities could give the opportunity for metacognitive elaboration on the part of students or a scaffold for sophisticated and efficient feedback on behalf of the teachers.

References

1. Meier, A., Spada, H., Rummel, N.: A rating scheme for assessing the quality of computer-supported collaboration processes. International Journal of Computer-Supported Collaborative Learning 2 (1), 63--86 (2007)
2. De Wever, B., Schellens, T., Valcke, M., & Van Keer, H.: Content analysis schemes to analyze transcripts of online asynchronous discussion groups: a review. Computers & Education, 46, 6--28 (2006)
3. Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C.: 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 (1995)
4. Weinberger, A., & Fischer, F.: A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. Computers & Education, 46, 71--95 (2006)
5. Suthers, D., Law, N., Rose, C. & Dwyer, N.: A common framework for CSCL interaction analysis. Pre-conference workshop at the 2008 ICLS conference, Utrecht, The Netherlands (2008)
6. Rummel, N., Spada, H.: Learning to collaborate: An instructional approach to promoting problem-solving in computer-mediated settings. The Journal of the Learning Sciences 14(2), 201--241 (2005)
7. Avouris, N., Margaritis, M., Komis, V.: (2004). Modelling interaction during small-group synchronous problem solving activities: the Synergo approach. In Proceedings of 2nd International Workshop on Designing Computational Models of Collaborative Learning Interaction, ITS2004, 7th Conference on Intelligent Tutoring Systems, Maceio, Brasil, pp. 13--18 (2004).
8. Clark, H. H., Brennan, S. E.: Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (eds.), Perspectives on socially shared cognition, 127--148, Washington, DC: American Psychological Association (1991)
9. Meier, A., Voyatzaki, E., Kahrimanis, G., Rummel, N., Spada, H., Avouris, N.: Teaching students how to improve their collaboration: Assessing collaboration quality and providing adaptive feedback in a CSCL setting. Submitted as part of the symposium "New Challenges in CSCL: Towards adaptive script support" (Nikol Rummel and Armin Weinberger), ICLS 2008, Utrecht, June 2008 (2008)