Analyzing and Presenting Interaction Data: A Teacher, Student and Researcher Perspective

Wouter van Diggelen, Jeroen Janssen & Maarten Overdijk
Research Centre Learning in Interaction, Utrecht University, Heidelberglaan 1, 3584CS Utrecht
w.vandiggelen@uu.nl, j.j.h.h.Janssen@uu.nl, m.overdijk@uu.nl

Abstract: Students’ actions within a CSCL-environment can be recorded and stored as interaction data. This data can be accessed and analyzed automatically. Teachers, students and researchers may benefit from these analyses. It gives teachers and students immediate feedback about performance indicators and it can help researchers to identify meaningful patterns in the interaction data. In the symposium we focus on these three groups of users – the teacher, the student and the researcher – and explore three issues: 1) how can we extract meaningful information from the interaction data, 2) how can this information be used in practice, and 3) how should this information be presented to the user?

Introduction

Computer Supported Collaborative Learning (CSCL) refers to those situations where the interactions between the students are mediated by computer tools. Students’ ongoing actions within these tools are recorded and stored in log files as interaction data. The availability of this data offers new opportunities for support: the data can be accessed and analyzed automatically. It may provide teachers, students and researchers with valuable insights about students’ performance. In the symposium we will address three issues with regard to this kind of support: 1) how can we extract appropriate information from the interaction data, 2) how can this information be used, and 3) how should this information be presented to the user? We make a distinction between three types of users: teachers, students, and researchers.

The first presentation focuses on the teacher as the user of interaction data. It presents a study where teachers were asked what kind of information they needed about students’ performance when they would moderate an electronic discussion (e-discussion). These requirements where used for the development of awareness tools that provide the teachers with feedback about the group processes. In the second presentation, two awareness tools are discussed that students used during their collaborative activity. This presentation addresses how awareness of group process – degree of participation and agreement – affects the performance of the students. Finally, the third presentation takes the perspective of the researcher and describes how researchers can identify and represent meaningful patterns of interactions from interaction data.

The three presentations differ with regard to the intended users but have three themes in common that will be highlighted in the symposium.

− A first theme refers to the relation between the user and the nature of feedback. In the case of teachers and students, the results of the analyses act as a form of immediate feedback. This feedback informs those user groups about students’ performance in the CSCL-environment. As such, this support may function as a form of awareness (Dourish & Belloti, 1992). It may help students to collaborate more effectively and efficiently. For teachers the support enables them to moderate an e-discussion. In the case of researchers, the support helps them to identify and represent meaningful patterns of interactions.

− A second theme that is visible in the three presentations concerns the question which data should be analyzed and presented to the users. A CSCL-environments ‘produces’ large quantities of data: some parts of this data may be relevant to some users, while other parts may be irrelevant. Which data should be analyzed and what kind of information should be presented is an educational issue rather than a technical one. To answer these questions the presenters examine what kind of problems the users face and what kind of information they need to address these problems. The second presentation by Janssen et al., for example, discusses feedback about social aspects of the collaboration because research indicates that students often have difficulties collaborating productively (Salomon & Globerson, 1989).

− A final theme concerns the issue of visualization. As stated before, large quantities of data are available for analysis. The interpretation of these analyses is a cognitively demanding task. Visualizations can make the analyses more accessible because “it is possible to have a far more complex concept structure represented externally in a visual display than can be held in visual and verbal working memories” (Ware, 2005, p. 29).

Each presenter addresses the three themes from the perspective of a specific group of users. They present the choices that they made with regard to the analysis and presentation of interaction data. The participants of the symposium will have the opportunity to discuss the consequences of these choices for research and practice.
Presentation 1: Supporting e-moderation of synchronous discussions: Awareness tools for teachers
Christa S. C. Asterhan, Rakheli Hever & Baruch B. Schwarz
School of Education, Hebrew University of Jerusalem, Israel

Collective argumentation can be a powerful means for learning and development (e.g., Schwarz & Asterhan, in press). However, high-quality argumentation is not easily elicited in learning settings. Therefore, argumentative design (Andriessen & Schwarz, in press), i.e. the design of a learning environment that may lead to productive argumentation, is an important yet complicated precondition. One way to facilitate productive argumentation is to provide students with a synchronous, graphic based CSCL-environment that enables students to co-construct an argumentative discussion map, such as Digalo (see a/o Schwarz & De Groot, 2007).

Research from classroom discussions, on the other hand, identifies the teacher as an important mediator for productive student argumentation (e.g., Yackel, 2002). However, teacher moderation of synchronous e-discussions is highly complex: The dynamic flow of simultaneously posted (and sometimes overlapping) contributions may quickly fill up the computer screen. It means that the teacher has tremendous difficulties reading the contributions, understanding their interweaving, and evaluating how to intervene in relation to such a flow. This may explain why teachers’ attitudes towards on-line moderation are often negative and they may leave discussants without substantive interventions (Gil, Schwartz & Asterhan, 2007). These difficulties are expected to increase when teachers are required to monitor and moderate several small group discussions at once.

We will describe a system that aims to support teachers and tutors in their on-line moderation practices, by, among others, providing them with tailored awareness tools. In addition to describing the nature of these awareness indicators we will also present research on teachers’ preferences concerning the presentation of interaction data and discussions features that may facilitate their on-line moderation practices.

A system for moderating e-discussions

The aim of the ARGUNAUT project (www.argunaut.org) is to develop a computerized system that supports moderators in their endeavor to increase the quality of argumentative e-discussions. The most salient features of ARGUNAUT’s Moderator’s Interface are: (a) awareness tools that provide immediate representations of aspects of e-discussions; (b) pre-defined alerting rules; (c) a remote control intervention panel from which the moderator can send textual comments and images to targeted students or the whole group in a variety of ways; and (d) tools for off-line reflection (see Figure 1). These aids are envisioned to help the teacher monitor, evaluate and direct the discussion without disrupting the flow of the on-going collective argumentation.

Figure 1. The ARGUNAUT Moderator's Interface at its current state of development.

The design of the tools is based on a user-centered design approach, involving teachers, tutors and (high school and university) student discussants. Currently, the system supports moderation of e-discussions within two different platforms, but the system could also be relevant for other synchronous discussion tools as well, particularly graphical tools. One of the currently supported platforms is Digalo (see figure 5), who provides its users with a shared workspace based on a concept-mapping interface. The tool enables synchronous, textual talk through mediation of geometrical shapes that represent different dialogical moves.
Awareness tools

The term awareness is defined as "an understanding of the activities of the others which provides context for your own activity" (Dourish & Belloti, 1992). A great variety of tools for e-discussion are available today, many of which offer awareness features for participants or moderators. Awareness feedback is based on various properties of the discussion, such as social interaction patterns, participation, temporal stages and text analysis. Yet, a systematic integration and combination of structural, process-oriented, and textual aspects has only recently been discussed in initiatives such as the Interaction Analysis project in the European Kaleidoscope network. The awareness tools that are developed for the ARGUNAUT system are derived from structural, process-oriented and textual elements of Digalo discussion maps. The structural elements are the direct or computable attributes of each shape or arrow object (such as type, creator, and number of characters) and any combination using these attributes as building blocks. The process-oriented elements are comprised of user actions on the discussion objects, and sequences thereof (stressing the dimension of time and the process of discussion rather than the end product). The textual elements are the text contributions typed within each shape.

We address several levels and methods of processing these elements. The lower levels are those of simple statistics (e.g., average text length, distribution of contribution types, and the like) or the use of statistical relationships between elements as relatively simple composite indicators (e.g., using the ratio between the number of connectors and the number of shapes as a measure of connectivity). The higher levels may involve more complex units, combining several types of elements (e.g., a cluster of shapes with certain characteristics, a sequence of actions indicative of a certain phenomenon) and/or the use of artificial intelligence methods and intensive input from pedagogical researchers (e.g., machine-learning classifications, query by example, search for pre-defined patterns). We refer to previous publications on these issues (e.g., Hever, et al., 2007).

Table 1. Interaction data indicators for on-line moderation, as indicated by experienced teachers (N=14).

<table>
<thead>
<tr>
<th>Source category</th>
<th>Details and possible indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence</td>
<td>Who is online, who has been active for the last [time interval], who is lurking.</td>
</tr>
<tr>
<td>Participation</td>
<td>Number of contributions per discussant, by the different discussion ontology categories.</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>The extent of interpersonal interaction and responsiveness to each other. For example: social network diagrams, numerical indicators of connectivity on discussion map-level, frequency of questions shapes within a map, or indicators of dialogicity based on machine learning classifications.</td>
</tr>
<tr>
<td>Quality of reasoning: shape-level</td>
<td>Whether discussants use reasoned arguments as opposed to simple claims (based on machine learning classifications).</td>
</tr>
<tr>
<td>Quality of reasoning: discussion map-level</td>
<td>Whether the group considered multiple perspectives. For example: ratio of supporting and opposing links, map-level classifications based on content analyses.</td>
</tr>
<tr>
<td>Impact of contributions</td>
<td>Identification of contributions that were neglected (i.e., without links) and contributions that catalyzed or opened up the discussion (i.e., with many links).</td>
</tr>
<tr>
<td>Textual length</td>
<td>Identification of particularly long contributions.</td>
</tr>
</tbody>
</table>

What type of interaction data do teachers need?

As a first step, we gathered data on teachers’ needs with regards to the different sources of information that could aid them in the moderation process. A total of fourteen teachers who had experience with conducting and/or moderating Digalo discussions, participated in one or more of the following three data collection activities: (1) Questionnaires in which they were asked to rate and rank predefined sources of information according to the extent according to which they considered each one of them as helpful in supporting on-line moderation of Digalo discussions; (2) Evaluation of screen shots of possible awareness indicators for on-line moderation; and (3) Identification of interaction and discussion features that would require on-line teacher intervention, based on two actual discussion maps.

The combined results of these three different data collection activities led us to identify seven different categories of information sources that were recurrently mentioned by the teachers to be potentially helpful in supporting on-line moderation. They are presented in Table 1, together with a short description of how these features can be operationalized into awareness indicators. Most of these features are currently implemented in the Moderator's Interface of the ARGUNAUT system as automatically detected and up-dated awareness visualizations for moderators. Currently the system has been implemented in different locations in Israel, Colombia, Great Britain and Germany. We will present preliminary findings from research with the first integrated versions of the system in both actual and simulated learning settings. This research focuses mainly on how teachers evaluate and appropriate the different ARGUNAUT features during e-moderation practices.
The research on the effectiveness of computer-supported collaborative learning (CSCL) and the processes that take place during CSCL, demonstrate that the collaboration in CSCL environments is not always free of problems. These include for example, lack of awareness of group members’ activities, communication and coordination problems, and lack of quality discussions. The central issue in this contribution is how and why visualizations of the collaborative process affect group members’ collaboration in CSCL environments. In order to tackle this issue, two different visualizations were developed. The effects of these visualizations were examined in two studies by giving one group of students access to the visualization, while the other group of students was not given access to the visualization. The two visualizations were implemented in the VCRI environment (Virtual Collaborative Research Institute, Broeken, Jaspers, & Erkens, 2006). Both studies employed 16 to 18-year old participants in the fifth year of pre-university secondary education. Both studies were carried out in the domain of history.

**Study 1: Visualizing Participation**

Study 1 describes the effects of the Participation tool (see Figure 2), which visualizes group members’ relative contributions to the online discussion through the Chat tool of the VCRI environment. When working in CSCL environments, it is often difficult to determine what other group members are doing and whether everyone is contributing equally to the group task (Kreijns, 2004). The Participation tool gives an impression of the number of messages sent by each group member and the relative length of these messages. The tool thus gives feedback about group members’ participation during the collaboration (e.g., is there equal participation in our group?) and allows them to compare themselves to other group members. This may raise their awareness about the manner in which they are collaborating and may stimulate group discussions about the collaborative process.

To examine the effects of the Participation tool, 17 groups of students ($N = 52$) were given access to the Participation tool while working in the VCRI environment, while 5 groups ($N = 17$) were not given access to the tool. We hypothesized that giving group members access to the Participation tool would increase their awareness of the group processes and activities taking place during their collaboration. This assumption was partly confirmed: group members with access to the Participation tool reported they knew better when another group member was taking a free ride. The Participation tool therefore enabled students to determine free riding behavior more easily. Additionally, the Participation tool was found to have an effect on participation: students who had the option to inspect the Participation tool had higher participation rates. Furthermore, the intensity with which the Participation tool was used by students was found to have an effect on equality of participation: more equal participation was found in groups that used the tool more often. Content analyses of the collaboration protocols furthermore showed that students who could use the Participation tool engaged more in coordination and regulation of social activities: they typed more statements that addressed the planning of their collaboration. This indicates that visualizing participation stimulates group members to devote effort to coordinating their collaboration. Finally, we expected that through stimulating participation and equality of participation, raising awareness, and facilitating coordination group performance would increase. To test this assumption, we examined the quality of the products made by the participating groups. No effect of the Participation tool on the quality of group products was found, however.

**Study 2: Visualizing Agreement and Discussion among Group Members**

A visualization called Shared Space was the focus of Study 2. The Shared Space visualizes the amount of agreement or discussion during group members’ chat conversations (see Figure 3). To do so, the Shared Space analyzes all messages that students type in the Chat tool of the VCRI environment. The Shared Space
uses discourse markers (Erkens, 1997; Schiffrin, 1987) to determine the communicative function of each message. Based on this analysis, the Shared Space moves the current chat topic to the left if it contains indications of disagreement (e.g., denials, negative evaluations) or to the right if it contains indications of agreement (e.g., confirmations, positive evaluations). Thus, the Shared Space gives feedback about the types of discussions group members are conducting (e.g., critical discussions or consensual discussions) and may raise their awareness about conversational strategies. Furthermore, by raising group members’ awareness, groups may be stimulated to discuss how well their group is functioning and how the group process may be improved.

We examined the effects of the Shared Space by giving 20 groups of students (N = 59) access to chat-tool with Shared Space visualization, while 20 other groups (N = 58) had access to the regular chat-tool of the VCRI environment. Content analyses of the collaboration protocols show the Shared Space facilitated coordination of the collaborative process. Students with access to the Shared Space were less busy exchanging information and typed fewer messages aimed at reaching and maintaining mutual understanding. This indicates that for these students it may have been easier to understand their group members’ contributions. In other words, the process of negotiation of meaning was facilitated. Furthermore, the Shared Space was found to have an effect on exploratory group norm perception. This indicates that groups that had access to the Shared Space perceived their discussions as more critical and constructive than did groups without access to the Shared Space. Additionally, students with access to the Shared Space reported more positive group behavior (e.g., helping behavior, positive group atmosphere) and more effective group task strategies than students without access to the Shared Space. In other words, they perceived their collaboration as more positive and effective. Lastly, we hypothesized that by stimulating students to adopt exploratory group norms and by decreasing communication problems, the Shared Space would also increase group performance. This hypothesis was partly confirmed as groups with access to the Shared Space only outperformed groups without access on one part of the group task (out of a total of three parts).

Conclusion

The results of the two studies show that analyzing the collaborative process in CSCL environments and visualizing the subsequent information for students can be beneficial. Students use the information these visualizations provide to raise their awareness about group members’ activities in the environment. Furthermore, students use the visualizations to coordinate the process of online collaboration: they use them to start discussions about the collaborative process itself and to facilitate the process of negotiation of meaning.

Presentation 3: Analysis and presentation of interaction data: A researcher perspective

Kristine Lund, Wouter van Diggelen, Gregory Dyke, Maarten Overdijk, Jean-Jacques Girardot & Annie Corbel

ICAR laboratory, CNRS, University of Lyon, Lyon, France; Research Centre Learning in Interaction, Utrecht University, The Netherlands; G2I / RIM, École Nationale Supérieure des Mines de St. Etienne, France

CSCL environments record and store their actions as interaction log files (e.g. Börner, Hazlewood & Lin, 2002). Researchers have access to an extensive data set that focuses on the activities students carry out. Researchers have to make several choices how to access the data set. For example, a choice must be made whether to record private unshared tool use, in addition to recording actions that are carried out within the collaborative tool. From the researcher’s point of view, this choice is guided by the research agenda. A research question concerning the role of private notes in cooperative text writing goes unanswered if the private space is
not recorded. On the other hand, many research questions focus specifically on the collaborative processes (Lund, Rossetti, Metz, 2007; Overdijk & van Diggelen, 2008). The interaction log files contain events (private or shared) that occurred within the environment of the tools; they contain the textual and graphical contributions of the students as well as the separate chronological actions that the students performed. These extensive interaction log files — on which analytical choices have already been performed — form the semi-raw material on which researchers do their analyses. The first problem is extracting relevant interpretable data from these interaction log files; this implies translating the data into a readable and understandable format. Secondly, it does not follow that such an initial transformation automatically produces results. One of the goals of analyses of CSCL situations is to identify patterns of specific actions and interactions, perhaps associated with particular tool uses, and to relate them to learning outcomes or to the quality of student’s productions (e.g. Lund, Molinar, Séjouné & Baker, 2007). If we can successfully relate a particular pattern to a desired outcome, then we can attempt to bring about the pattern in question (Ronteltap, Goodyear & Bartoluzzi, 2004). However, identifying such patterns in interaction log files that are so extensive and diverse is not self-evident.

In what follows we present an analytical tool called TATIANA (Trace Analysis Tool for Interaction ANAlysts) that supports researchers in the analysis of interaction log files obtained from various CSCL situations. We then present the general principle and show for two specific cases how TATIANA can analyze and visualize patterns of individual student actions and relate them to patterns of collective student interactions. Hereby we open up the way to correlating patterns with outcomes. Finally we conclude and discuss further work.

TATIANA (Trace Analysis Tool for Interaction ANAlysts)

TATIANA (Dyke et al., 2007) attempts to help researchers represent their data in ways that are meaningful to them. We call these representations replayables. For video this simply entails opening it with a media player. For interaction log files, TATIANA supports the creation of different forms of replayables: Excel-like tables in which each row corresponds to an event and visualizations in which each event is shown as a graphical object. Researchers describe how each event will be represented (fields of a row, appearance of a graphical object) through the use of filters, allowing them to select only the relevant facets of the data and represent them in an understandable form. TATIANA also treats categorization and annotation as being the creation of a new replayable. TATIANA allows replayables to be synchronized: navigating to a certain timestamp in the video will highlight the corresponding event (row for tabular replayables, graphical object for visualizations); selecting an event in visualization will highlight the corresponding row in the table and navigate the video to the timestamp of that event. Finally, TATIANA allows replayables to be combined to create new replayables, for example importing a categorization into an existing visualization or combining aspects of two visualizations to create a new visualization which is better suited to the researcher’s current needs.

Visualizing the different facets of student actions and interactions: the general case

A microanalysis of the collaborative learning process begins with the identification of individual actions within a given interaction between students. However, the interpretation of the form and content of these individual actions is undertaken by the analyst in the context of the interaction between students. Indeed, verbal interactions in a face-to-face discussion are often described by local principles of turn taking and adjacency (Sacks, Schegloff & Jefferson, 1978). The coherence of a discussion is based on the assumption that each act is an appropriate response to a previous act (Littlejohn, 2002) and meaningful patterns are searched for on that basis. However, the timely order of actions within a CSCL environment is often hidden - for different reasons - of which the following two are examples:

1. The actions of one student in a particular collaborative tool may overlap in time with the actions of another student in another tool (e.g. one student types in a chat, while another enters text into a shared text editor);
2. Students can both act separately within the shared environment of a single collaborative tool (e.g. write simultaneously in different parts of a shared text editor, or construct different argument sequences simultaneously in different areas of an argument grapher).

Visualizing various graphical representations of situations like the ones mentioned above enables researchers to display the relationship between actions for which adjacency is not immediately apparent. Researchers thus create a logical representation that reflects meaningful student actions and interactions. TATIANA proposes a general visualization of such graphical representations called an interaction score of actions (based on the notion of a musical score) where different facets and sub-facets of a participant’s action (participant’s name, tool, particular use of tool, type of action, etc.) can be displayed in relation to time and in part, automatically generated and then merged, based on research goals. In the next section, we illustrate representations that fall into the two categories presented above; others are possible.
Two specific cases

The example shown in Figure 4 illustrates overlapping student and teacher actions. The x-axis shows time in the form of intervention numbers and the y-axis shows the participant and the tool he or she used. The legend lists the uses observed for each tool. We see that student A chats from intervention 0 to 10 during which she introduces herself in the text editor. She takes notes in the chat around intervention 40 and participates in dialogue periodically throughout most of the interaction. During this time, student B participates with student A in the chat and the text editor, but also carries out other actions (e.g. writes a synthesis in the text editor). It’s possible to pinpoint areas of interaction during simultaneous use of a particular tool, but also across different shared tools (notes in chat for Student A and synthesis in text editor for Student B). The researcher may now ask how content is treated during those times.

![Figure 4](image1)

**Figure 4.** Overlapping actions of two students and a teacher.

Figure 5 shows how students can act separately within the shared environment of a single tool. The x-axis of the right-hand figure shows time; the y-axis of the same figure illustrates the quadrants (Q1, Q2, Q3) of an argument graph tool in which students constructed elements. The quadrants and students’ constructions can be seen to the left. Taking student A, we see he begins constructing in Q1 with box 8, then moves to Q3 to construct box 12, goes back to Q1 to construct box 15 and so forth. The researcher may now consider the relations between boxes (possible interactions between students) and examine argumentative content.

![Figure 5](image2)

**Figure 5.** Argumentative sequences constructed by three students within an argumentative tool.

Conclusions and further work

We have shown how the researcher can be assisted in analyzing complex interaction log files, issued from CSCL situations, with the help of TATIANA (Trace Analysis Tool for Interaction ANALysts). We introduced the notion of interaction scores of actions — based on the concept of a musical score — that illustrate how different facets of students’ actions move though time during a given interaction. We illustrated with two examples how the researcher is thus assisted in identifying patterns of individual action and group interaction. Further work will augment the types of corpora analyzed with TATIANA in order to explore the possible visualizations of different facets of student actions and interactions.

Endnotes

(1) The material in the paper by Asterhan et al. is based on work supported by the EC -funded ARGUNAUT project (IST-2005-027728).

(2) The material in the paper by Lund et al. is based on work supported by the EC -funded LEAD project (IST-2005-028027).

References


