A general framework for tracking and analysing learning processes in computer-supported collaborative learning environments

Francesca Pozzi, Stefania Manca, Donatella Persico & Luigi Sarti

To cite this article: Francesca Pozzi, Stefania Manca, Donatella Persico & Luigi Sarti (2007) A general framework for tracking and analysing learning processes in computer-supported collaborative learning environments, Innovations in Education and Teaching International, 44:2, 169-179, DOI: 10.1080/14703290701240929

To link to this article: http://dx.doi.org/10.1080/14703290701240929

Published online: 20 Mar 2007.

Submit your article to this journal

Article views: 210

View related articles

Citing articles: 29 View citing articles
A general framework for tracking and analysing learning processes in computer-supported collaborative learning environments

Francesca Pozzi*, Stefania Manca, Donatella Persico and Luigi Sarti
Istituto per le Tecnologie Didattiche, Consiglio Nazionale delle Ricerche, Italy

This paper describes a method for analysing the learning processes that take place in a computer-supported collaborative learning (CSCL) environment. The approach is based on tracking the interactions between learners and tutors. Keeping track of meaningful events serves three main purposes: evaluation of the quality of the process, monitoring students’ performance in real time and the assessment of individual learning performances. The results can be of benefit to course designers, tutors/instructors and researchers involved in collaborative learning experiences. The method is based on a five-dimensional model that includes a participative, an interactive, a social, a cognitive and a teaching dimension. Each dimension can be analysed through a set of indicators obtained by tracking course participants’ behaviour within the learning environment. Examples of significant indicators, drawn from the literature and the authors’ field experience, are discussed for each dimension. The paper also proposes a taxonomy of the data needed to obtain the described indicators. Finally, possible directions for further research are outlined.

Introduction

In this paper the problem of how to track and analyse the learning processes occurring in an online course is addressed. The theoretical background stems from the field of computer-supported collaborative learning (CSCL) and the social constructivist theories of collaborative learning (The Cognition and Technology Group at Vanderbilt, 1991; Scardamalia & Bereiter, 1994; Dillenbourg, 1999; Kanuka & Anderson, 1999). The proposed approach takes a pragmatic point of view in that it has been adopted and validated in a number of online and blended courses designed and delivered by the authors’ institution in recent years, mostly in the context of pre-service and in-service teacher training (Bocconi & Pozzi, 2002; Midoro, 2004; Persico & Delfino, 2004). The pedagogical model underlying these courses is based on the

*Corresponding author. Istituto Tecnologie Didattiche, Consiglio Nazionale delle Ricerche, via de Marini 6, 16149 Genova, Italy. Email: pozzi@itd.cnr.it
assumption that knowledge can be constructed through social negotiation and that discussion with others—peers or tutors—is a primary way to learn because it encourages critical thinking and hence understanding (Garrison et al., 1999). Another important element of this approach is that learners should engage in a task with a very tangible outcome, be it the solution of a problem or the production of an artefact such as a document, concept map, schema or hypertext. In this perspective, the need to achieve a common aim acts as a catalyst for concentration and effort. As a consequence, courses are usually structured in various phases, and in each phase participants are required to engage in collaborative activities aimed at producing an artefact (or part of it). To do so, they will need to have discussions, negotiate meanings and finally reach an agreement in order to reify a shared product that is the expression of the groups’ achievements. The quality of the product is not so important, because it is the process leading to its realization that really generates learning.

In order to pursue and sustain a high-quality educational experience, a deep understanding of the learning process is required. Designers and tutors are required to go beyond the surface level of interactions in order to understand the social and the cognitive processes involved. As computer conferencing systems usually record and maintain a history of the events occurring during the learning process, a wealth of information that can be analysed is available.

Generally speaking, the tracking and analysing of learning processes can have three main aims:

- **To evaluate quality, be it the overall quality or aspects of it.** Designers and researchers use the results of the analysis to identify strengths and weaknesses of the instructional design and the learning environment, in order to improve the course. This activity can be carried out after the course, to re-design further editions, but also during the process, with the purpose of tuning and adapting it to unexpected needs and situations.
- **To monitor students’ performance** so as to provide them with appropriate support. Tutors, who act as learning facilitators, need to collect and analyse data to keep the situation under control and be able to take decisions on how to proceed (i.e. adjusting or re-scheduling activities, stimulating discussion, urging answers, writing public or individual postings, deadline reminders, comments on essays, etc.).
- **To assess individual learning processes and carry out formative and summative evaluation of students’ performances.**

The framework presented below identifies five relevant dimensions of a collaborative learning experience, together with descriptions of possible indicators and data needed for investigating such dimensions.

**Relevant dimensions and indicators**

Most popular methods for the analysis of CSCL processes are based on both quantitative and qualitative data. Several studies have proposed integrated approaches, in which aggregations of various natures—coming from automatic tracking and more subjective sources—are juxtaposed and compared (Henri, 1992; Hara et al., 2000; Lally, 2002; Lipponen et al., 2003; Martinez et al., 2003; Daradoumis et al., 2004). Although the analysis of quantitative aspects of students’ exchanges is easy to automate and achieve, by itself it does not allow in-depth understanding of
Tracking and analysing learning processes

On the contrary, qualitative analysis, which is much more complex and time-consuming, often allows a thorough investigation of specific cases, even though its results are harder to generalize upon. As a consequence, the use of both methods, through aggregations of heterogeneous data, provides a balanced and comprehensive way to understand the collaborative processes and a basis for the evaluation of the overall experience.

The framework here proposed builds mainly on Henri’s (1992) and Garrison and Anderson’s (2003) models, achieving a five-dimensional approach which includes the participative, interactive, social, cognitive and meta-cognitive, and teaching dimensions. Henri’s (1992) pioneering work in the field of content analysis puts forward a framework for analysing online discussions. Her study focuses on both what is said and how it is said, thus fulfilling the need to integrate quantitative and qualitative analyses. According to her model, the five dimensions for the analysis of online discussion are the participative, interactive, social, cognitive and meta-cognitive. However, in our view, a drawback of this model is the lack of attention that is given to the contribution provided by tutors and instructors—the members of the learning community that guide the cognitive and social processes towards the realization of the learning outcomes.

In the framework of the ‘Communities of Inquiry’, Garrison and Anderson (2003) investigated the three dimensions related to the social, cognitive and teaching factors. This model aims to provide a conceptual order for investigating learning processes that occur in asynchronous-based learning environments. According to these authors, text analysis can track indicators of social, cognitive and teaching presence within students’ and tutors’ messages and therefore support the monitoring process by identifying the various components of online presence in computer-mediated communication (CMC). According to this approach, social, cognitive and teaching presence are closely interconnected dimensions, their boundaries are blurred and the related indicators often overlap. Although this gives a certain flexibility to the model, it makes it difficult to apply using all three dimensions at once. Hence, the majority of studies based on it focus on only one dimension—see, for example, McKlin et al. (2001), Meyer (2003), Pawan et al. (2003), Shea et al. (2003), Lobry de Bruyn (2004) and Oriogun et al. (2005). In order to investigate all three dimensions within the same learning experience, it is necessary to contextualize and sometimes re-define indicators, so as to obtain a clear-cut classification. This causes a certain re-elaboration of the original categories and indicators, as explained in the sections below.

The five-dimensional model proposed in this paper does not mandatorily imply the use of all five dimensions: depending on the specific aims of the analysis, one may decide which dimensions are most relevant. For example, when a tutor is monitoring the learners’ level of participation, he or she is most likely to be interested in the participative and the interactive dimensions. On the contrary, all the dimensions are worth investigating when the purpose is to validate the overall learning experience.

In order to bridge the gap between the theoretical framework and its practical applications, it is essential to identify suitable indicators, that is quantitative or qualitative elements that allow the evaluation of each dimension according to specific objectives. These indicators express the effective manifestations of the five dimensions in a learning community. Contrary to Henri (1992), who identifies a stable set of indicators, we argue that indicators may vary considerably, depending on two main variables: the aim of the analysis (which may use not only a subset of the dimensions but also a subset of the indicators) and the context of the learning experience (including the educational setting, the technology used, etc.).
In our paper the five dimensions, together with some examples of indicators, are described. An exhaustive description of possible indicators goes beyond the aims of this paper and is practically impossible, because different research aims and different contexts may require the definition of new indicators. However, the indicators provided here are the result of extensive field experience and cover several of the most common aims of the actors involved—evaluation, monitoring and assessment.

The participative dimension

Taking part in an online collaborative experience means ‘doing’ something (entering a platform, accessing a forum, etc.) that demonstrates that one is there, independently from any other action. The level of students’ participation in a learning experience gives a rough but important indication of their involvement in the process. For this reason, investigation of the participative dimension may help the tutor to tune the activities, inform the assessment procedures and be used as a first element of the overall evaluation of the learning experience.

Going beyond Henri’s suggestions (1992), who identifies the total number of messages and the number of statements as the only indicators for participation, we considered not only the students’ visible actions (such as sending a message or downloading a document), but also their silence and (dis)continuity. These latter may be especially useful for pointing out critical events or elements of the process.

In our experience, the following three categories are the minimum indicators needed to investigate the participative dimension:

- Indicators of active participation, which include the number of messages sent by individual participants, the number of documents uploaded, the number of chat sessions attended, etc.;
- Indicators of passive participation, which include the number of messages read, the number of documents downloaded, etc.;
- Indicators of continuity, that is the distribution of participation along time.

The interactive dimension

In CSCL contexts, interaction refers to the relationships that participants build during the learning processes. The degree of interactivity between pairs is defined by the reciprocal influences in the cognitive processes (Dillenbourg, 1999, p. 12), rather than by the frequency of interactions (this latter pertains more to the participatory dimension). From this point of view, an investigation of the interactive dimension should address aspects such as cross-reference to others’ messages and documents or co-production of artefacts. In our experience, the most significant indicators of the interactive dimension are drawn from quantitative and qualitative content analysis of messages and documents exchanged by students during the learning process. Examples are:

- Passive participation before posting, that is the number of relevant messages read by a student before posting his/her own, the number of documents downloaded before posting, etc.
- References to other students’ messages, that is the number of answers to other students’ messages, the number of implicit or explicit citations of other students’ messages, etc.
- Consideration of other students’ contributions in products, that is qualitative analysis of students’ messages and documents with the aim of finding references to others’ messages or documents.

### The social dimension

The definition of social presence given by Garrison *et al.* (1999) goes beyond previous attempts at identifying it as the expression of whatever ‘is not related to formal content or subject matter’ (Henri, 1992, p. 126). In their definition, social presence is ‘the ability of participants in a community of inquiry to project themselves socially and emotionally, as “real” people (i.e., their full personality), through the medium of communication being used’ (Garrison *et al*., 1999, p. 94).

In order to investigate social presence, it is necessary to identify cues that testify to affection and cohesiveness within communication acts. They are expressed through *thematic units*, that is units of meaning that convey a single item of information extracted from a segment of content (see Rourke *et al*., 2001, p. 17).

In particular, we consider the following indicators of the social dimension:

- Thematic units characterized by *Affection*, that include expression of emotions, expression of intimacy, presentation of personal anecdotes.
- Thematic units characterized by *Cohesiveness*, that include vocatives, references to the group using inclusive pronouns, phatics, salutations.

While Garrison *et al.* (1999) include *Interaction* as an indicator of social presence, we agree with Henri (1992) and consider it worth addressing separately.

### The cognitive and meta-cognitive dimension

Cognitive presence can be defined as ‘the extent to which learners are able to construct and confirm meaning through sustained reflection and discourse in a critical community of inquiry’ (Garrison *et al*., 2001). In our framework, cognitive presence is revealed by thematic units referring to the following phases:

- *Revelation*, that is recognizing a problem, showing a sense of puzzlement, explaining or presenting a point of view.
- *Exploration*, that is expressing agreement/disagreement, sharing ideas and information, brainstorming, negotiating, exploring.
- *Integration*, that is connecting ideas, making synthesis, creating solutions.
- *Resolution*, that is real-life applications, testing solutions.

Even if the original model of ‘Communities of Inquiry’ does not specifically address the meta-cognitive dimension, Garrison (2003) considers it an important element of the cognitive process that should be the object of monitoring, evaluation and assessment activities. In online learning, in fact, meta-cognitive skills are often among the explicit or implicit objectives of the learning process. This is due to the very nature of the online collaborative approach, mostly based on asynchronous textual interaction, that makes it particularly well suited to reflection not only on content but also on the learning process itself. On the other hand, meta-cognitive objectives are
often included in the online learning initiatives because many learners are new to this approach and therefore lack some of the meta-cognitive skills that are needed to take full advantage of the training method.

Research in this area has so far drawn little on interaction analysis. The first attempts to investigate the meta-cognitive dimension through content analysis (i.e. Henri, 1992) were picked up by other researchers (Gunawardena et al., 1997; Hara et al., 2000), who stressed the difficulties arising from the application of such an approach. Some underline that inferring a person’s strategic knowledge from written messages is extremely difficult and subjective and that only some types of meta-cognitive skills exhibited within the dialogue (such as planning, regulation, evaluation and self-awareness) can be coded (Hara et al., 2000). Others question the real difference between a cognitive and meta-cognitive dimension and the possibility of distinguishing between declarative and procedural knowledge (Newman et al., 1995). Other proposals aim to redefine the distinction between cognitive and meta-cognitive dimensions in terms of strategies in the co-creation of knowledge and negotiation of meaning (Gunawardena et al., 1997).

The above-mentioned research results invite caution in proposing indicators to detect and measure the meta-cognitive dimension implied in Web-based learning processes. In addition to this, one should consider that the meta-cognitive process is rarely manifested spontaneously by the learners, hence difficulties in detecting it through the analysis of written interactions does not necessarily mean it did not occur. For these reasons the framework here proposed does not include indicators for the meta-cognitive dimension. Further research is needed to define what aspects may be captured through a content analysis approach and what necessarily remains elusive and latent. So far, we still consider interviews, where learners are requested to describe ex-post the strategies and the methods they used during the learning process, and questionnaires, aimed at eliciting learners’ opinions about the whole process, as the main sources of information concerning meta-cognition.

The teaching dimension

Teaching presence is defined as ‘the design, facilitation, and direction of cognitive and social processes for the purpose of realizing personally meaningful and educationally worthwhile learning outcomes’ (Anderson et al., 2001, p. 5). In other words, teaching presence is the binding element in creating a learning community: messages carry teaching presence when they address objectives like building understanding, challenging and stimulating participants, or providing direct instruction. Indicators for teaching presence may include:

- Thematic units containing direct instruction, that is presenting contents, proposing activities, diagnosing misconceptions, confirming understanding through assessment and explanatory feedback.
- Thematic units aimed at facilitating discourse, that is identifying areas of agreement/disagreement in order to achieve consensus, encouraging, acknowledging or reinforcing participant contribution, setting the climate for learning.
- Thematic units addressing organizational matters, that is introducing topics, planning the course, explaining methods, reminding students of deadlines.
Data for tracking learning processes

The above mentioned dimensions, together with their indicators, can be investigated by the collection and manipulation of different kinds of data. In CMC systems the interactions between the participants are usually recorded automatically and are a very important source of data. They comprise messages and documents exchanged, their ‘history’ (sending and reading timestamps, name of sender, name of readers, etc.) and log files. Other kinds of data can be gathered through questionnaires, interviews, via textual content analysis or may be drawn from design documentation. Some of these data have an intrinsically quantitative nature and can be easily processed by means of statistical methods and tools (Wang, 2004). Others may be determined by quantitatively elaborating the results of qualitative analysis of textual data (Chi, 1997).

The complex and multifaceted range of data informing monitoring and evaluation of online learning has been categorized on the basis of their nature and origin, according to a model developed within the DPULS project (see Acknowledgements):

- Raw data;
- Additional data;
- Subjective data;
- Aggregated data.

Raw data

Online courses usually make use of CMC platforms with a client-server architecture: participants use a client application (a web browser or some proprietary application) to connect to a centralized server program, which relies on a database to provide access to messages, user directories, discussion areas, etc. (e.g. Centrinity FirstClass, http://www.centrinity.com/; Moodle, http://www.moodle.org/). Server programs record significant events in log files, which are the most widespread method of collecting raw data.

Tracking of raw data and subsequent analysis of log files—this is usually carried out by means of specialized software (e.g. Nihuo Web Log Analyzer, http://www.loganalyzer.net/; Web Log Expert, http://www.weblogexpert.com/; Visitors, http://www.hping.org/visitors/; Web Log Explorer, http://www.exacttrend.com/WebLogExplorer/)—is often used by tutors for gaining on-the-fly information about the participative dimension (e.g. summarizing participation to identify inactive students or lurkers, providing personalized feedback, supporting meta-cognitive reflection, and so on).

Examples of raw data are: the number of sessions for each participant (with starting and ending time), the number of messages sent (sorted according to participant, area/conference, given periods, etc.), the size of messages sent by each participant, the number of messages read by each participant, the number of threads activated by students and by tutors, the number of produced documents, the number of chats, etc.

Additional data

Additional data describe the learning context, including course features, organizational aspects, target population characteristics, etc. They are not usually recorded by the system but
may be drawn from the course design documentation, elicited from the course developers through interviews or collected from participants via questionnaires. When the target population is large, questionnaires are usually preferable because of their easy scalability, besides which they can be delivered online through the same platform used for the course. The nature of these data is generally factual and largely independent of personal beliefs and interpretations.

Examples are: data concerning the course structure (course schedule, including timetable and deadlines of learning activities, layout of the learning system, including number and types of discussion areas, etc.); data concerning the course participants, such as number of students, personal data (age, gender, etc.), prerequisites, etc.

**Subjective data**

Subjective data usually consist of quantitative information elaborated on the basis of qualitative aspects of the course. In this category statistics about participants’ beliefs, attitudes, opinions on the course content, methodology, organization, etc., are included. These data can be collected through interviews, questionnaires or textual analysis of the participants’ messages.

Textual analysis is usually performed by one or more raters who read the messages exchanged by course participants and classify the communication units (messages or parts of them) according to the presence of various indicators. This work may be done with the support of tools for the qualitative analysis of text (e.g. Atlas.ti, http://www.atlasti.de/).

**Aggregated data**

Aggregated data are elaborations of raw, additional and/or subjective data. Such elaborations introduce a first level of interpretation by presenting data in a concise form (e.g. statistical and/or graphical), to facilitate the tutors, the designers or the researchers. Aggregations are potentially unlimited and are usually oriented towards a specific aim.

For instance, the distribution of session durations per participant (i.e. how long each participant has been logged into the system) through time, together with the standard deviation and the variation range of the average session duration, may be useful to investigate the learners’ continuity of participation (participative dimension). Another example addresses the percentage of messages containing quotations from (or references to) other students’ messages (interactive dimension).

Aggregations may also inform the social, the cognitive and the teaching dimensions, since tables or graphical representations may support an overall view or comparisons of such dimensions.

**Conclusions**

Designers, researchers and tutors are usually involved in the analysis of data gathered at the beginning, during or at the end of learning processes with the aim of gaining a better understanding of the process itself. Designers and researchers are usually more interested in the evaluation of the overall quality of the whole process, hence the analysis is carried out at the end of the course and usually based on general statistics. Tutors, on the other hand, are more interested
in capturing each single student's performances during the course, usually for monitoring and assessment purposes. This analysis must be performed in real time to allow prompt intervention.

In this paper a general framework for analysing collaborative learning processes through tracking is proposed. This approach is applicable to all those situations where a community is engaged in collaborative tasks that require interaction among participants. Up until now researchers have often focused on well-defined problems or on specific aspects. The framework herein proposed is broad enough to capture the most significant dimensions of CSCL and to adapt to different aims and contexts. It therefore provides a common research reference and allows comparability of results. The range of examples provided in this paper, drawn from the literature and the authors' experience, testifies to the broad applicability of the approach.

However, further work is needed both to identify other indicators and to tune the theoretical model. For example, recent studies about Social Network Analysis (SNA) suggest that this technique may provide further useful indicators for the interactive dimension (Lally, 2002; Reffay & Chanier, 2002; Aviv et al., 2003; Martinez et al., 2003; Daradoumis et al., 2004), even if some weaknesses in the approach have still to be overcome. SNA yields graphs that clearly describe the interactions within a group of individuals, but it does not lend itself very well to situations where the receiver of a message is not an individual, but the whole community, as often happens in many forum-based environments.

In spite of the wide range of applicability of this framework, there are some drawbacks. One is that, being based on tracking, the approach only provides insight into aspects that are revealed through learners' interactions. There may well be feelings, competences and meta-cognitive abilities that were never expressed in messages and yet were felt or developed by participants. Interaction analysis can give us information about what was said, how it was said, how much was said, but not about everything that had actually happened. In addition, caution should be used when trying to apply the indicators proposed to contexts that are culturally very distant from those where they were conceived and successfully used. As a matter of fact, some expressions of feelings that are likely to be found in conversations among Italians are rarely found when examining messages exchanged by British people, let alone Chinese or Indian. Furthermore, some clues that work in one language simply do not work in others.

To conclude, even if the framework proposed is applicable to different contexts, new indicators should be worked out not only to enrich our proposal, but also to pursue different analysis objectives and suit different cultural contexts.

Acknowledgements

This study was carried out within the DPULS project (Design Patterns for recording and analysing Usage of Learning Systems). DPULS is a JEIRP (Joint Executed Integrated Research Project) of the Kaleidoscope Network of Excellence, supported by the European Community as part of the Information Society Technologies priority of the 6th Framework Programme.

Notes on contributors

Francesca Pozzi is a researcher at the Institute for Educational Technology of the Italian National Research Council (CNR), a public research body devoted to the study of educational
innovation through the use of ICT. Presently, her main research interests focus on: models and methods for supporting the diffusion of innovation in the knowledge society; and designing, managing and evaluating online collaborative learning communities.

Stefania Manca is a researcher at the Institute for Educational Technology of the Italian National Research Council (CNR). Her main interests are the analysis of social and cognitive processes in Computer Supported Collaborative Learning environments and the analysis of specific linguistic features used to express and construct social dimension in asynchronous-based learning environments.

Donatella Persico is a researcher at the Institute for Educational Technology of the Italian National Research Council (CNR) and lecturer in Educational Technology at the University of Genoa. She has been active in the field of educational technology, theory and applications since 1981. Her main interests include instructional design, e-learning and teacher training.

Luigi Sarti is a researcher at the Institute for Educational Technology of the Italian National Research Council (CNR). He is mainly interested in studying and developing methodologies for the application of ICT to learning processes in a social constructivist framework. He also teaches Web Programming Languages at the Faculty of Education of the University of Genova.

References


