Providing effective feedback, monitoring and evaluation to on-line collaborative learning discussions

Santi Caballé a,*, Thanasis Daradoumis a, Fatos Xhafa b, Angel Juan a

a Department of Computer Sciences, Open University of Catalonia, Barcelona, Spain
b Department of Languages and Informatics Systems, Polytechnic University of Catalonia, Barcelona, Spain

ARTICLE INFO
Article history:
Available online 14 August 2010

Keywords:
Teaching/learning strategies
Cooperative/collaborative learning
Computer-mediated communication
Interactive learning environments
Architectures for educational technology system

ABSTRACT
Learning and knowledge building have become critical competences for people in the knowledge society era. In this paper, we propose a sociolinguistic dialogue model for understanding how learning evolves and how cognitive process is constructed in on-line discussions. The knowledge extracted from this model is used to assess participation behavior, knowledge building and performance. The ultimate purpose is to provide effective feedback, evaluation and monitoring to the discussion process. Seven hundred students from the Open University of Catalonia in Spain participated in this study. Results showed that learning and knowledge building may be greatly enhanced by presenting selected knowledge to learners as for their particular skills exhibited during interaction. In addition, this valuable provision of information is used as a meta cognitive tool for tutors and moderators for monitoring and evaluating the discussion process more conveniently. This contribution presents our conceptual model for interaction management as well as key design guidelines and evaluation results. Implications of this study are remarked and further research directions are proposed.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

In the present knowledge society context (Corso, Martini, Pellegrini, Corso, & Martini, 2009; Lytras & Garcia, 2008; Lytras & Ordóñez de Pablos, 2007; Lytras, Rafaeli, Downes, Naeve, & Ordóñez de Pablos, 2007; Lytras, Sakkopoulos, & Ordóñez de Pablos, 2009; Lytras & Sicilia, 2005; Stehr, 2007), interactive learning environments (ILEs) (Wang, Woo, & Zhao, 2008) have provided a huge improvement of distance learning development, mainly in the last ten years. A vast literature production and outstanding contributions have been published reporting on studies where the learning interaction process has been exhaustively analyzed (Chou, 2004; Fahy, 2006; Michinon & Michinon, 2005; Ordóñez de Pablos, 2005; Perakyyla, 2004). Interaction analysis is certainly a key issue in the field of collaborative learning to ensure full support to the on-line learning activity and specifically to provide monitoring and evaluation capabilities based on the information captured from the participants’ actions during the collaborative process (Dillenbourg, 1999).

ILEs usually include the participation of students in-class discussions with the aim of sharing and discussing their ideas (Lytras, Damiani, & Ordóñez de Pablos, 2008; Lytras & Ordóñez de Pablos, 2009; Stahl, 2006; Wang et al., 2008). Given the added value and the extensive use of on-line discussions in the current educational institutions’ pedagogical models, great research efforts have been carried out to understand the cognitive processes underlying the collaboration. (Schellens & Valcke, 2006) investigated whether collaborative learning in asynchronous discussion groups results in enhancing academic discourse and knowledge construction. Their research work showed that students in the discussion groups were fundamentally task-oriented and that higher proportions of high phases of knowledge construction were observed. It was evidenced by (Stahl, 2006; Wang et al., 2008) and (Puntambekar, 2006) that students were able to construct their own understanding based on their interactions with others during the discussion while shared knowledge building become richer over time.

Following this increasing interest, current ILEs are incorporating advanced interactive support to on-line discussions resulting in the generation of large amounts of interaction data, which include complex issues of the collaborative work and learning process (e.g., group well-being (McGrath, 1991; Pillania, 2009) as well as self, peer and group activity evaluation (Daradoumis, Martinez, & Xhafa, 2006). As a consequence, manual monitoring and evaluation of large on-line discussion processes, typically carried out by tutors and moderators, become tedious, error-prone, and highly unreliable. Moreover, since the evaluation process is done after the completion of the learning activity, it has less impact on it (McDonald, 2003). Indeed, the lack of constantly feeding back immediate assessment from the tutor on the dynamics and performance of the collaborative activity may negatively impact on participant’s...
motivation, emotional state and problem-solving abilities, and as a result diminish the performance and acquisition of knowledge (Zumbach, Hillers, & Reimann, 2003).

Intensive and successful research from the interaction analysis field has been achieved over the last years to facilitate the manage-
ment by computers of the large amounts of interaction data from on-
line discussions. Current efforts (Angehrn, Maxwell, Marco Luccini, & Rajola, 2009; Bardis, Drigas, Doukas, & Karadimas, 2009; De Wever, Schellens, Valcke, & Van Keer, 2006; Ordóñez de Pablos, 2004; Pena-Shaff & Nicholls, 2004; Schriere, 2006; Soller, 2001; Strijbos, Martens, Prins, & Jochems, 2006; Vargas-Vera & Lytras, 2008) aim to alleviate manual procedures while considering relevant aspects of the collaboration, such as how all participants are actually performing during the discussion and the dynamics of each participant with respect to the group. To this end, two levels of inter-
action analysis are considered, quantitative and qualitative level.

Quantitative indicators measure the participants' performance and dynamics (e.g., number of contributions written and read by each participant) as relevant information to model the group func-
tioning and task performance (Daradoumis et al., 2006). According to (De Wever et al., 2006), quantitative content analysis has been increasingly used to surpass surface level analyses in collaborative learning (e.g., counting messages) and several content analysis schemes have been employed to analyze transcripts of on-line asynchronous discussion groups in formal educational settings. (Soller, 2001) also proposed to manage large amounts of quantita-
tive information by applying a structured process where the users' interactions are tagged with certain indicators according to a co-
laborative learning conversation skill taxonomy that models the various types of interactions at different levels. Although this re-
search technique has been often used, standards are not yet estab-
lished. As a consequence, the empirical base of the validity of the
instruments is limited. Several open questions still exist, especially as concerns the unit of analysis and segmentation procedure to be followed (Pena-Shaff & Nicholls, 2004).

Qualitative information has been also considered valuable to complete the labored task of interaction analysis and evaluation of contributions. Strijbos, Martens, Prins, & Jochems, 2006 and Schriere, 2006, used a merging view of quantitative analysis within a qualitative methodology to build a model for the analysis of collaborative knowledge building in asynchronous discussion. Quantitative analysis was used to examine participation and interaction rates, at a number of levels, focusing on the discussion forum itself, the discussion threads, the messages, and the ex-
changes and moves among the messages. Qualitative procedures were used to analyze knowledge construction processes and to re-
fine a category system of indicators and descriptors. Results showed that students got engaged in a knowledge construction process by means of integrating the interactive, cognitive and dis-
course dimensions in collaborative learning. However, the mere consideration of the depth of discussion threads, the number of
messages, and he relation among messages do not guarantee by it-
self the quality of the discussion; students' postings can be simply driven by socialization reasons and not directly linked to the devel-
opment of the learning tasks.

In overall, we believe that there are more evident key discourse elements and aspects that play an important role both for promot-
ing student participation and enhancing group and individual per-
formance, such as, the impact and effectiveness of students' contributions, among others, that we explore in our work. By explicitly feeding back these elements to the participants our dis-
cussion model accomplishes high students' participation rates and contribution quality in a more natural and effective way. In-
deed, our approach goes beyond a mere interaction analysis of asynchronous discussion in the sense that it builds a multi-func-
tional model that fosters knowledge sharing and construction, develops a strong sense of community among students, provides the tutor a powerful tool for students' monitoring and evaluation, discussion regulation, while allowing for peer facilitation through effective self, peer and group feedback.

The paper is organized as follows. Section 2 presents the aims and the theoretical background to the research and the develop-
ment of our study. Section 3 describes the collection methodolo-
gies, adopted analysis procedures and the statistical models for elaboration on the resulting data. Analysis and discussion of the re-
sults, conclusions, and further research follow and end the paper.

2. Aims and theoretical background

The aim of this work is to provide a conceptual sociolinguistic dialogue model for understanding how learning evolves and how cognitive process is constructed during on-line discussions. The knowledge extracted from this model is used to assess participa-
tion behavior, knowledge building and performance for the ulti-
mate purpose of providing effective feedback, evaluation and moni-
toring to the discussion process.

In particular, this section examines how the building and distri-
bution of knowledge is manifested in student–student interaction and how it can be studied in the context of ILEs. This involves the definition of appropriate collaborative learning situations and the focus on student interaction among peers, which is the corner-
stone of our model. Participants need indeed to interact with each other to plan an activity, distribute tasks, explain, clarify, give-
information and opinions, elicit information, evaluate and contrib-
ute to the resolution of problematic issues, and so on.

To satisfy course assessment requirements, discourse contribu-
tions also need to be evaluated as effectively as possible. Evalu-
ation of hundreds of contributions in a multi-member discussion can be a tedious monitoring task for tutors and should be ade-
quately supported. Moreover, self and peer assessment should be also encouraged and facilitated by intuitive means. To this end, a dialogue model of asynchronous discourse is to be provided, which is capable of capturing, analyzing and evaluating both the process and the result of the building and distribution of knowledge. This model is described next, which is mainly defined in terms of types and structure of student–student interaction.

2.1. A dialogue model for managing discussion interaction

The model proposed in this paper is based on the integration of several models and methods: the Negotiation Linguistic Exchange Model (Martin, 1992); a model of Discourse Contributions (Clark & Schaefer, 1989); and the types of learning actions underlying a partici-
 pant turn (Self, 1994). The structure of a long interaction is con-
structed cooperatively by using the exchange as the basic unit for communicating knowledge. Following (Martin, 1992), three gen-
eral exchange structure categories are considered: give-information exchange, elicit-information exchange and raise-an-issue exchange, which consist of different types of moves (Schwartz, 1999) and de-
scribe a generic discourse goal.

More specifically, the goal of the actor who initiates the give-
information exchange is to inform his/her partners about a certain situation with the aim to change the partners' mental states. Informing includes moves that explain, give an opinion, describe or remind a situation in different ways. The actor goal of the sec-
ond exchange is to elicit the partners' state of mind (knowledge, beliefs, attitude, desire or abilities) of a situation, in which the ac-
tor is not aware or certain about. The actor goal of the third ex-
change is to raise-an-issue (a problem or question) to be resolved by the participants, which causes to explore their state of mind (knowledge, beliefs, etc.).
According to Martin (1992), there is a move that constitutes the "obligatory move" of the exchange, since it either carries or indicates completion of the discourse goal for which the exchange is initiated. According to Clark and Schaefer (1989), each move is seen as a contribution to discourse. This means that in a cooperative conversation, contributions are regarded as collective acts performed by the participants working together, resulting in units of conversation – typically turns (moves) – that aim to make a success of the discourse they compose. Yet, not all moves contribute in the same way toward the successful completion of the exchange. According to Self (1994), some moves have a pure contributing function toward the realization of the obligatory move of the exchange. In fact, without the presence of those moves, the obligatory move cannot be realized; thus, those moves really contribute toward the realization of the obligatory move. Consequently, it is stated that successful realization of the obligatory move conveys evidence of (initial) success of the exchange. In contrast, others moves have a rather supporting function (provide evidence of support) toward the definite completion of the obligatory move and consequently of the exchange. This is the case of the follow-up moves of the three exchanges. Supporting moves are optional, so they may not be realized. In such a case, they convey an implicit support toward the obligatory move, that is, toward the definitive completion of the exchange.

Based on the work of Self (1994), Pilkington (1999), and Soller (2001), partners are involved in a process of realizing a number of learning actions which lead to the completion of the exchange goal. Each move type captures and controls the evolution of the learning action performed by a participant by setting the expectations of the type of learning actions which has to be realized next by the other participants so that the goal set by the initial move be accomplished.

Both the quantity and the quality of the several move types performed are measured by the collaborative effort of the members involved to achieve the discourse goal of an exchange. The term collaborative effort means both the number of contributing and supporting moves issued by a participant, which indicates an active participation (distinguishing between proactive and reactive one) or passive one. It is also considered the type and effectiveness of these moves, which indicate the way a participant contributes toward the achievement of the shared discourse goal, as regards knowledge possession and transfer, reasoning capability and positive attitude. The tutor measures move effectiveness by assessing the quality of their content. In addition, peer assessment can be effected to complete the evaluation of each contribution made. The roles that these moves play in the exchange as well as the degree of success of that role determine the successful completion of the exchange goal.

Completion of an exchange expresses the mutual beliefs of all participants about the accomplishment of its discourse goal. Moreover, it implies the achievement of a certain degree of knowledge building and distribution among the different participants. This degree can be deduced and measured by exploring the principal interaction indicators proposed by this model. For each participant, the model measures: the total number of moves created, his/her participation behavior (proactive, reactive, supportive, or passive), their effectiveness and impact that each move has in the discourse, and in the achievement of the current discourse goal, as well as the evaluation of the move content and significance by his/her peers and the tutor.

In general, the three general types of exchanges presented represent standard discourse structures for handling information and suggest a certain type of knowledge building, as a result of giving and eliciting information or working out a solution on an issue set up. These discursive structures enable the participants to take turns, share information, exchange views, monitor the work done and plan ahead. Most importantly, they provide a means to represent and operationalize the cognitive product at individual level, that is, the way the reasoning process is distributed over the participants as it is shared in a collaborative discourse.

Consequently, interaction analysis takes into account both the way the interaction is structured and the types of contributions, which are explicitly defined and expressed. The analysis results yield very useful conclusions on aspects such as individual and group working, dynamics, performance and success, which allows the tutor to obtain a global account of the progress of the individual and group work and thus to identify possible conflicts and monitor the whole learning process much better.

A further innovation of this model is that it allows participants to end up an exchange, which took several moves to conclude by "replaying" the main contributing move of the exchange. For instance, in a set-up-an-issue exchange, a solution move may not be sufficiently complete and thus has to be further elaborated, corrected or extended. To that end, another participant has the option to provide an amplify-solution move, which completes the initial solution. In general, a "replay" move can be used to resume all the changes produced from the initial appearance of an exchange goal to be achieved to its final conclusion and acceptance by all participants. This can be useful both to reinforce the fact that the goal of the exchange has been completed successfully and to explicitly indicate the progress achieved in the participants' process of knowledge building (especially as regards the participant who provided the main contributing move of the exchange).

Finally, the participant is required to commit certain action to indicate s/he has read a certain contribution, such as send a reply and assent the contribution. The aim is both to provide reliable indicators on the number of contributions read and to promote the discussion's dynamics by increasing the users' interaction with the system.

In overall, our model annotates and examines a variety of elements that contribute to the understanding of the nature of the collaborative interactions, such as the students' passivity, proactivity, reactivity as well as the effectiveness and impact of their contributions to the overall goal of the discussion. The aim is to provide both a deeper understanding of the actual discussion process and a more objective assessment of individual and group activity.

3. Research methodology

This section presents a methodological approach to validate the previous conceptual model for interaction management. To this end, first, a multi-experiment carried out at the Open University of Catalonia is described. Then, a new interactive discussion tool that was used to collect the experimental data is presented along with the description of a set of indicators we incorporated to measure and ultimately analyze participation behavior, knowledge building, and performance during the discussion. Finally, the statistical models used for the elaboration on the data collected are described.

3.1. Experiences using real learning context

The real context of this study is the virtual learning environment of the Open University of Catalonia (UOC), which offers higher education over the Internet. Given the added value of...
asynchronous discussion groups, the UOC have incorporated on-line discussions as one of the pillars of its pedagogical model. To this end, great efforts are being made to develop adequate on-line tools to support the essential aspects of the discussion process, which include students' monitoring and evaluation.

Six experiences in all took place at the UOC over the last two academic terms. A total of 730 graduate and undergraduate students from three courses in Computer Science were involved directly or indirectly forming the experiment sample. For each experience, students were equally distributed into two classrooms and participated in the experience with the same rules, at the same time and during the same time (about a fortnight). Students from one classroom were required to use the well-known asynchronous threaded discussion forum offered by the UOC virtual campus while the other group of students used a new discussion tool, which incorporated our model of interaction management. This discussion tool is presented next.

3.2. Data collection through an effective structured discussion forum

All data from these experiences were collected by means of a prototype of an ad hoc web-based structured collaborative learning system, called Discussion Forum (DF). This tool incorporates our conceptual model for interaction management, which gives new opportunities to learn by discussion (Caballé, 2008; Caballé & Xhafa, 2009; Caballi & Xhafa, 2010). For the sake of understanding how the collected data was generated, certain key design aspects of this tool are described here.

3.2.1. Collection of post tagging, assent, and rating

The design of the DF includes certain thematic annotation cards based on the general exchange types identified in Section 2, namely give-information, elicit-information and raise-an-issue. Six exchange moves and quite a few low-level categories (see Table 1 for a complete list) have been identified to qualify each exchange move in the discussion processes occurring at our university though they are not conclusive since more experimentation process has to be undertaken.

Consequently, DF's users are urged to qualify their contributions before sending a new or reply post (see Fig. 1 and refer to Section 3.2.3 for an improvement of this categorization process).

Contribute may be assented and also evaluated by both the tutor and other participants in terms of content quality and the utility in their progress in the discussion (see Fig. 2).

3.2.2. Provision of feedback

Based on the previous assumptions, all contributions are recorded in the DF as exchange moves, which are later on analyzed and presented as knowledge to participants either in real time (to guide directly students during the learning activity) or after the task is over (in order to understand the collaborative process). Finally, relevant feedback is provided to the discussants and tutors based on the data collected and the following methodology that identifies and measures relevant dimensions of the discussion process (see Table 2 and Fig. 3):

1. Participation behavior indicators are distinguished into proactive, reactive and supportive (or assentive). Participants are proactive when they take the initiative to open a new exchange of the type give-information, or raise-an-issue. Participants are reactive when they reply to moves such as elicit-information, set-up-an-issue/problem, or provide-solution. Participants are supportive if they give their assent to previous contributions. In that case, a supporting value is defined which is assigned a default numerical value 1, which means that the move fully supports and recognizes the value, contribution and effectiveness of a previous move it refers to. If several supporting moves refer to a particular move M, it implies a broader consensus about the impact of M, which increases M's impact value to 1.

2. Passive participants are considered those who just read others' contributions, as well as the ones who also evaluate the usefulness of these contributions. Passivity becomes an essential indicator for the discussion process' dynamics as it identifies certain important profiles of the participant, such as arrogance (participant who just contributes but does not read the contributions of others) and also promotes reactive attitudes and social grounding skills (Daradoumis et al., 2006) by engaging the participant in the collaborative process.

3. An impact value is assigned an initial (default) numerical value between 0 and 1, which is modified (increased or decreased) according to the impact (number of reactions received) that the move M has on the dialogue and on the achievement of the current discourse goal and task. If the reaction is positive (the move M is being assented), then M receives a positive one (+1) point. If the reaction is negative (M is not assented) then it receives a negative 0.5 points. The points received by a reaction move depends on the type of learning action underlying the move and take on the default value of the move's impact value. The final value is obtained by the mean value of all moves involved in move M.

4. The effectiveness value of a move is calculated by the mean value of the number of assent moves received. An assent move M is identified and recorded after a participant receives M and consents it. Note that only give-information and raise-an-issue exchange acts can be assented. A negative assent requires a reply move on M to provide further information to reason why M has not been assented, which generates another move in the current discourse.

5. Finally, tutor and peer assessment indicators are to evaluate both the quality of the contribution's content by the lecturer monitoring the discussion process and the usefulness of the contribution by the student participating in the discussion. Both indicators are on the scale 0–10 so as to be accurate in providing mean values of them.

All these quantitative and qualitative indicators were weighted adequately according to the specific goals and procedures of each

<table>
<thead>
<tr>
<th>Exchange moves</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support</td>
<td>Greeting</td>
</tr>
<tr>
<td></td>
<td>Encouragement</td>
</tr>
<tr>
<td></td>
<td>Motivation</td>
</tr>
<tr>
<td>Elicit-information</td>
<td>REQUEST-information</td>
</tr>
<tr>
<td></td>
<td>REQUEST-elaboration</td>
</tr>
<tr>
<td></td>
<td>REQUEST-clarification</td>
</tr>
<tr>
<td></td>
<td>REQUEST-justification</td>
</tr>
<tr>
<td></td>
<td>REQUEST-opinion</td>
</tr>
<tr>
<td></td>
<td>REQUEST-illustration</td>
</tr>
<tr>
<td>Give-information</td>
<td>INFORM-extend</td>
</tr>
<tr>
<td></td>
<td>INFORM-lead</td>
</tr>
<tr>
<td></td>
<td>INFORM-suggest</td>
</tr>
<tr>
<td></td>
<td>INFORM-elaboration</td>
</tr>
<tr>
<td></td>
<td>INFORM-explain/clarification</td>
</tr>
<tr>
<td></td>
<td>INFORM-justify</td>
</tr>
<tr>
<td></td>
<td>INFORM-state</td>
</tr>
<tr>
<td></td>
<td>INFORM-agree</td>
</tr>
<tr>
<td></td>
<td>INFORM-disagree</td>
</tr>
<tr>
<td>Set-up-an-issue</td>
<td>PROBLEM-statement</td>
</tr>
<tr>
<td>Provide-solution</td>
<td>PROBLEM-solution</td>
</tr>
<tr>
<td>Consent-solution</td>
<td>PROBLEM-extend solution</td>
</tr>
<tr>
<td></td>
<td>PROBLEM-assert solution</td>
</tr>
</tbody>
</table>
discussion. For monitoring purposes, the system proposed to the tutor an updated final mark of the progress of each student based on all the indicators presented (see last column of Fig. 3).

3.2.3. Collecting reliable data

For the sake of increasing the reliability of the data collected from this tool, we incorporated further innovations, which are briefly described here.

Typically, non-functional requirements, such as user scalability, service availability, performance, and interoperability, are not fully considered in the current ILEs. From the experience at our university, this may have considerable repercussions on the collaborative learning dynamics and performance, and outcomes, and as a result it may interfere with the normal learning flow a great deal (Caballé, 2008; Caballé, Paniagua, Xhafa, & Daradoumis, 2005). For instance, when a server hosting an ILE is down for technical reasons, all participants have to temporarily stop participating, which causes great deal of frustration, especially at our university, where students have very limited time to dedicate to their studies.2 Therefore, the lack of ILEs in the marketplace meeting these requirements encouraged us to entirely build our innovative DF system. Distributed infrastructure was added to provide permanent high quality of service even in the presence of system failures (Caballé, Xhafa, & Daradoumis, 2007).

2 Students at the UOC are 30 years old on average and 90% attending professional jobs.
Another important improvement was made in the DF system to collect more reliable data by avoiding unnecessary choice in post tagging (see Fig. 1). In each specific phase of the discussion process the DF shows a short list with just those categories that are relevant (e.g., in replying a request, just those cards involving the provision of information are provided). This makes the choice of the appropriate tag much easier and less error-prone. In addition, the tutor is to examine and assess all contributions posted based partially on the tags used by students to categorize them, and as a result students are aware of the potential repercussions of tagging posts incorrectly to optimize their own assessment instead of reflecting the true meaning of their posts.
3.3. Adopted analyses procedures and statistical models for data elaboration

Three different approaches were combined to analyze the data collected from all the experiments performed by using both the new DF discussion tool and the well-known standard discussion tool (ST) offered by the UOC virtual campus.

3.3.1. Quantitative procedure

In all experiments, students were free to open zero, one or several discussion threads where they proposed specific objectives, activities, and processes needed to appropriately address the discussion topic. Hence, there was no requirement to open a discussion thread and all students could participate in the discussion threads at convenience. At the end of the discussions, those students who had opened a discussion thread were asked to close it by sending a contribution that summarized and concluded the main points arisen in the thread.

For each of the six experiments and for both environments (DF and ST), the following variables were considered: (a) Number of participating students (ST Students), (b) Number of generated threads (ST Threads), and (c) Number of submitted posts (ST Posts). We used these data to compare average results from the DF and ST as explained in Section 4.

3.3.2. Qualitative procedure

The qualitative evaluation of the discussions was addressed by both examining those discussion threads that contained enough discussion (i.e. more than seven posts) and checking whether the student who was in charge of each thread had posted both a start and close contribution on the same issue. The content quality of both contributions was considered and compared for the purpose of evaluating the evolution of the student’s acquisition of knowledge.

Average distance from the first and last qualitative mark of the considered threads was used to elaborate the results. In addition, a structured and qualitative report was conducted at the end of each experience addressed to the DF users who were also asked to compare it to the ST tool they had already used in previous courses at the UOC. This complemented the qualitative analysis.

3.3.3. Procedure of the evaluation process

In order to evaluate the reliability of the semi-automatic evaluation, the tutor supervising the discussions was required to (i) submit to the DF system a precise assessment on content quality of every contribution posted, which was automatically presented to students as feedback information (see Fig. 3) and (ii) evaluate students’ performance manually by filling in a spreadsheet that helped score each student’s participation according to both their contributions’ content quality and the purpose and context where the contribution took place (e.g., whether it was a new argumentation or a reply, brought interesting opportunities for further discussion, it was just a greeting-type post, etc.). This entire manual evaluation task could be complemented with extra information on the discussion’s individual behavior observed by the tutor according to his/her knowledge and experience of this type of in-class assignments. The ultimate aim of this double evaluation task was to compare the manual and semi-automatic evaluation process. To this end, both evaluation processes resulted in proposing a final mark for each student as well as a list where all students were ranked according to his/her final mark (see first and last columns depicted in Fig. 3).

For the elaboration on data analysis results, we considered five relevant indicators of the discussion, namely, activity, passivity, impact and effectiveness, becoming 50% of the semi-automatic evaluation. Both tutors and peers addressed the last indicator – assessment – manually. Tutors evaluated specifically the content quality of the contributions (40%) while peers evaluated how useful they found others’ contributions (10%) (note that tutors can adjust these percentages according to the type of the discussion). On the other hand, the manual evaluation process was carried out entirely by the tutor and followed the same assessment procedure as that traditionally performed when using the standard discussion tool of the UOC.

4. Results and analytical data discussion and interpretation

Table 3 shows the data collected summarized for the three variables of interest considered, namely # Students, # Threads, and # Posts. Both the DF and the ST tools in all experiments collected the data.

Table 4 considers for both tools the derived variables: (a) Number of threads per student, (b) Number of posts per student, and (c) Number of posts per thread. Means and standard deviations are calculated in each case. Finally, whenever possible (i.e., for those populations where the normality assumption can be assessed), a hypothesis test for two independent populations has been performed on the difference between the average value associated to the ST and the average value associated to the DF, i.e.,

\[
\left\{ \begin{array}{l}
H_0 : \mu_{ST} = \mu_{DF} \\
H_1 : \mu_{ST} \neq \mu_{DF}
\end{array} \right.
\]

Note that, together with the T-values and P-values associated to each hypothesis test, also a 95% confidence interval has been obtained for the corresponding difference.

4.1. Quantitative results

Table 3 shows a significant impact on participative behavior by using the DF tool, which collected twice as many posts in all as those collected by the ST. In addition, despite the ST tool generated more discussion threads than the DF, most of these threads (about 85%) were actually empty since they just contained the start-thread post only. In overall, these results evidence the effects of the inherent structure and richness provided by the DF to the discussion process whereas the ST promotes large monolithic one-sided points of view.

The results in Table 4 lead us to formulate the following statements regarding the variables of interest:

- As regards the number of posts per student, the hypothesis test has shown the existence of statistically-significant differences (p-value < 0.001) between the two tools. In fact, according to the calculated 95% confidence interval, each average student participating in the DF will be likely to post between 2 and 3 posts more than the average student participating in the ST.
- Similarly, as regards the number of posts per thread, the hypothesis test shows the existence of statistically-significant differences (p-value < 0.001) between the two environments (i.e., according to the calculated 95% confidence interval, an average thread in the DF will contain between 5 and 9 posts more than an average thread in the ST).

Table 3 Variables of interest and data collected from the DF and ST in six experiments.

<table>
<thead>
<tr>
<th>Variable of interest</th>
<th>Tool</th>
<th>Experiment</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># Students</td>
<td>ST</td>
<td>40</td>
<td>40</td>
<td>66</td>
<td>141</td>
<td>35</td>
<td>44</td>
<td>366</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DF</td>
<td>40</td>
<td>40</td>
<td>66</td>
<td>139</td>
<td>34</td>
<td>45</td>
<td>364</td>
<td></td>
</tr>
<tr>
<td># Threads</td>
<td>ST</td>
<td>57</td>
<td>43</td>
<td>52</td>
<td>60</td>
<td>54</td>
<td>6</td>
<td>272</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DF</td>
<td>65</td>
<td>21</td>
<td>31</td>
<td>10</td>
<td>12</td>
<td>8</td>
<td>147</td>
<td></td>
</tr>
<tr>
<td># Posts</td>
<td>ST</td>
<td>171</td>
<td>71</td>
<td>229</td>
<td>140</td>
<td>109</td>
<td>189</td>
<td>909</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DF</td>
<td>549</td>
<td>199</td>
<td>417</td>
<td>241</td>
<td>214</td>
<td>197</td>
<td>1817</td>
<td></td>
</tr>
</tbody>
</table>
To sum up, the use of the DF tool has a significant and measurable positive impact both on the number of posts sent by students and also on the number of posts contained by each average thread, which is important in order to avoid the “empty-threads” problem (this problem can cause dispersion of contents and, therefore, a difficult an adequate development of discussion processes generated in most course forums).

4.2. Qualitative results

The results on the DF showed that on average 32% of students had improved their qualitative mark by going through the discussion in their threads, 68% kept the same mark, and no mark had dropped. On the other hand, no significant data were collected from the discussions using the ST as they were poorly contributed.

Table 5 shows an extract of the results of the questionnaire addressed to the DF users.

### Table 4

<table>
<thead>
<tr>
<th>Variable of interest</th>
<th>Tool</th>
<th>Number</th>
<th>Mean</th>
<th>St. dev.</th>
<th>Normality assumption</th>
<th>T</th>
<th>P</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td># Threads per student</td>
<td>ST</td>
<td>366</td>
<td>0.74</td>
<td>0.97</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>DF</td>
<td>364</td>
<td>0.40</td>
<td>0.90</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td># Posts per student</td>
<td>ST</td>
<td>366</td>
<td>2.48</td>
<td>4.30</td>
<td>Yes</td>
<td>–9.19</td>
<td>0.000</td>
<td>(–3.338, –2.162)</td>
</tr>
<tr>
<td></td>
<td>DF</td>
<td>364</td>
<td>9.99</td>
<td>3.12</td>
<td>Yes</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td># Posts per thread</td>
<td>ST</td>
<td>272</td>
<td>3.34</td>
<td>2.90</td>
<td>Yes</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>DF</td>
<td>147</td>
<td>12.36</td>
<td>7.72</td>
<td>Yes</td>
<td>–11.32</td>
<td>0.000</td>
<td>(–9.156, –6.438)</td>
</tr>
</tbody>
</table>

4.3. Results of the evaluation process

The results of the semi-automatic evaluation were very promising since the tutors in charge of the DF agreed with the final marks proposed by the system in more than 75% on average. In the same proportion, the DF’s rank matched the same position as in the rank appeared in the tutors’ spreadsheet. In addition, tutors reported benefits from using the DF in the monitoring process on the discussion since this new tool alleviates them from the tedious work of tracking and evaluating the discussion’s dynamics and outcomes manually. On the other hand, a clear inconsistency was identified since all final marks proposed by the system scored 1.1 points lower on average than those proposed by the tutors, thus showing the need to weight the indicators in the DF more objectively. In overall, these results are not conclusive but they encourage us to undertake more experimentation and especially validation processes on the semi-automatic evaluation approach.

5. Conclusions and further development

This paper describes a conceptual framework for modeling interaction from on-line discussions that contributes to the improvement of the collaborative learning process in terms of participation behavior, knowledge building, and performance. We believe that the consideration of evident key discourse aspects, which foster student participation and enhance individual performance, such as the impact and effectiveness of contributions, provides a significant step forward in the ILE development.

The results of the experiences reported here are not conclusive due to its exploratory nature. However, from the representative sample chosen, the fair reliability of the data collected and the relevant statistical models used, it has been proved to promise significant benefits for students in the context of learn by discussion and collaborative learning in general.

Further research directions are going through three perspectives:

- **Conceptual perspective.** Incorporate Natural Language Processing (NLP) and machine-learning techniques (Vargas-Vera & Moreale, 2005), for assessing discussion contributions automatically based on a small set of intrinsic text features, such as syntactic, lexical and quantitative. An initial approach is found in Caballé, Lapedriza, Masip, Xhafa, and Abraham (2009).
- **Technological perspective.** Explore the interesting possibilities offered by adding decentralized distributed infrastructure to the prototype of our discussion tool. The gain in performance (Caballé et al., 2007) might help us, for instance, collect more complex information of the collaboration and presented it in real time, such as modeling the participants’ behavior during the discussion by combining individual and group session and navigation information.
- **Application perspective.** Next step is to validate our approach at large scale by leveraging the investigations reported hereto support the heterogeneous discussion dynamics found in the different studies and programs of the UOC.

### Acknowledgements

This work has been partially supported by the European Commission under the Collaborative Project ALICE “Adaptive Learning via Intuitive/Interactive, Collaborative and Emotional System”, VII Framework Programme, Theme ICT-2009.4.2 (Technology-Enhanced Learning), Grant Agreement n. 257639, and the IN3-UOC HAROSA Knowledge Community Program (http://dpcs.uoc.edu). Fatos Xhafa’s work is partially done at Birkbeck, University of...
London, on leave from Technical University of Catalonia (Barcelona, Spain). His research is supported by a grant from the General Secretariat of Universities of the Ministry of Education, Spain.

References


SANTI CABALLE received his PhD, Masters and Bachelors in Computer Science from the Open University of Catalonia (Barcelona, Spain). Since 2003 he has been an Assistant Professor at the Open University of Catalonia, where he became Associate Professor in 2006 teaching various on-line courses in Computer Science in the area of Software Engineering, Collaborative Learning and Information Systems. Dr. Caballé has been involved in the organization of several international conferences, conference tracks and workshops, and has published over 60 research contributions as books, book chapters, and refereed international journal and conference papers. He has also acted as editor for books and special issues of international journals. His research focuses on e-Learning and Collaborative and Mobile Learning, Distributed, Grid and Peer-to-Peer Technologies, and Software Engineering.

THANASSIS DARADOUMIS has a PhD in Computer Sciences from the Polytechnic University of Catalonia-Spain, a Masters in Computer Sciences from the University of Illinois, and a Bachelor in Mathematics from the University of Thessaloniki-Greece. Since 1984, he has been an Assistant Professor at several universities in the USA, Greece and Spain, teaching a variety of courses in Mathematics and Computer Science. Since 1998, he has been working as an Associate Professor in the department of Computer Science, Multimedia and Telecommunication at the Open University of Catalonia where he coordinates several online courses as well as the development of teaching materials appropriate for virtual learning. He is a member of the Distributed, Parallel and Collaborative Systems Research Group at the Open University of Catalonia. His research focuses on e-Learning and Collaborative and Mobile Learning, Distributed, Grid and Peer-to-Peer Technologies, and Software Engineering.
ANGEL JUAN is an associate professor of simulation and data analysis in the computer sciences department at the Open University of Catalonia (Spain). He is also a lecturer at the Technical University of Catalonia. He holds a PhD in industrial engineering (UNED), an MS in information technology (Open University of Catalonia), and an MS in applied computational mathematics (University of Valencia). His research interests include computer simulation, applied data analysis, and mathematical e-learning. He has published several papers in international journals, books, and proceedings regarding these fields. As a researcher, he has been involved in several international research projects.