Philippe Dessus, Stefan Trausan, Sonia Mandin, Traian Rebedea, Virginie Zampa, Mihai Dascalu, Emmanuelle Villiot-Leclercq (2010):
Assessing Writing and Collaboration in Learning: Methodological Issues.
In Gunnar Stevens (Eds.),
International Reports on Socio-Informatics (IRSI),
Workshop Proceedings of 9th International Conference on the Design of Cooperative Systems - COOP 2010, Aix-en-Provence, France, May 2010
(Vol. 7, Iss. 1, pp. 86-94)

Assessing Writing and Collaboration in Learning: Methodological Issues

Philippe Dessus
LSE, Grenoble University, France
Philippe.Dessus@upmf-grenoble.fr
Stefan Trausan-Matu
"Politehnica" University of Bucharest, Romania
Stefan.Trausan@cs.pub.ro
Sonia Mandin
LSE, Grenoble University, France
Sonia.Mandin@msh-alpes.fr
Traian Rebedea
"Politehnica" University of Bucharest, Romania
trebedea@gmail.com
Virginie Zampa
LIDILEM, Grenoble University, France
Virginie.Zampa@gmail.com
Mihai Dascalu
"Politehnica" University of Bucharest, Romania
mikedascalu@yahoo.com
Emmanuelle Villiot-Leclercq
IUFM, Grenoble University, France
Emmanuelle.Villiotleclercq@gmail.com

Abstract. We describe the main tasks students usually complete when working in an e-learning platform, across five mean features that have to be taken into account in research efforts (writing-based activities, individual/collective level, knowledge/ pedagogy orientation, feedback, multiple stakeholders account). Ways to analyse and assist these tasks by (semi)-automatic assessments using NLP techniques is discussed. Two services aiming to assist writing-based tasks are presented along with their first validation.

1 Introduction

Current e-learning platforms allow rich collaborative learning activities that are now very well detailed and documented (Dillenbourg 2002; Kollar et al. 2006). However, the ways to record, study and analyse these activities yield methodological issues often debated in the literature (Strijbos & Fischer 2007) and theoretical frameworks to tackle these issues are lacking. The learning activities engaged in collaborative e-learning contexts share some specificities. First, they are based on writing. Second, their manifestations are both at individual and collective level. Third, their aim is twofold: at covering (learning) a given knowledge domain but also at leading a pedagogy-related activity. Fourth, they require to be analysed in order to provide an adequate feedback. Fifth and last, the stakeholders to be considered are not only the learners and the teachers, but also the researchers studying the activity.

Taking into account all these specificities requires devising ad hoc methodologies and overcoming research challenges. Strijbos and Fischer (2007) listed five main methodological challenges close to those pointed out in this paper, the goal of which is to present a comprehensive framework drawn from Bakhtin's work and a set of NLP-based tools that can help analyze learners' tasks according to these five points. The following sections shed light on each of them.

2 Five Specificities and Features for E-learning Tasks

The tasks every learner performs in an e-learning platform share five features:

Writing for learning. Every learner engaged in individual and collaborative learning in a virtual platform performs a set of writing-based activities (e.g., abstract writing, note taking, chatting, writing in forums), which are both evidences for, and products of, learning (Emig 1977). We can integrate the different writing-based learning activities in a comprehensive framework, based on Bakhtin's dialogism theory (Bakhtin 1981). As Koschmann (1999) put it, quoting Bakhtin: "[…] the voices of others become woven into what we say, write and think". We thus can take into account all these activities within a unique framework: everything—written, read or spoken—has a dialogic nature, which is expressed through writing and relates to learning.

Multilevel Tasks: from Individual to Collective. Tasks carried out by students are often separated in two independent ones, individual and collective. As Stahl (2006) puts it, learners engaged in a collaborative task in an e-learning platform have to cope with two *recursive and interrelated* main tasks: first, they are involved in an *individual* knowledge-building process; second, they are publicly engaged in a process of *collaborative* discussions about the notions at hand in the

first loop. Bakhtin's ideas of dialogism and inter-animation suit with these intertwined and multiple tasks (inner dialogs and debates).

Two aims: knowledge and pedagogy. The multiple tasks in which students are engaged in e-learning do not share the same goals. The complexity of any learning situation is partly due to the fact that two different and often conflicting aims interact with each other (Shulman 1986): learning a knowledge domain and in parallel being confronted with pedagogy-driven activities. On the side of knowledge, learners are given information they process in order to acquire knowledge. On the side of pedagogy, learners' behavior is directed as 'moves' within the classroom environment and pedagogical methods can be inferred from these moves.

Feedback delivery. In an e-learning context, *students* spend lot of time waiting for feedback from teachers or tutors about their writings, whatever are the goals and levels pursued. They encounter some problems: they stagnate themselves in the writing process; the limited feedback opportunities do not stimulate explorative approaches ("what if-trials"), but force them to hand in mainly completed versions; during writing, it is difficult to self-assess ongoing work and understanding. *Teachers* have a limited overview of the learners' processes, and assessments of students' understanding or collaboration are difficult and time-consuming. Feedback is thus necessary in e-learning contexts and can partly be automated by computer-based procedures.

Accounting for stakeholders' viewpoint. E-learning contexts are populated by numerous stakeholders (students, tutors, teachers, researchers) whose tasks may differ, overlap or be contradictory to each other. These tasks can also strongly interfere with the kind of tool used for analysing a given learning situation. Since most of the tools aiming at analysing collaborative software are devised for research purposes, they are more difficult to be used by other stakeholders.

3 NLP-Based Tools

Web-based services using NLP techniques can take into account the five features of e-learning situations presented above:

(1) detection of relations between utterances can be processed to reveal the voices engaged in writing or dialog;
(2) account for both the individual and collective level of knowledge acquisition;
(3) sensitivity to both knowledge (cognitive models) and 'moves' (dynamic situations) (Dessus et al. 2005; Wolfe et al. 1998);
(4) possibility to deliver just-in-time feedback allowing self-paced learning;
(5) deliver generic feedback to account for all the stakeholders' categories.

Let us now present two instances of web-based services designed from this viewpoint, *Pensum* and *PolyCAFe*.

Pensum supports learners at an *individual level* in the automatic assessment of their essays (summaries, syntheses). *Pensum* analyses how well learners understand course texts through their textual productions. It provides different kinds of feedback (see Figure 1) all based on LSA (Latent Semantic Analysis, Landauer & Dumais 1997) on two important features influencing writing quality: topic coverage (semantic links between sentences source texts and synthesis) and inter-sentence coherence.

PolyCAFe (Chat & Forum Analysis and Feedback System, Trausan-Matu & Rebedea 2010) functions at a *collective level* using a NLP pipe (stemming, POS tagging, chunking, etc.), advanced pattern matching, social network analysis and LSA for detecting discussion topics, threads and inter-animation in chat logs. Feedback (textual and graphical) is generated emphasizing collaboration degree, discussed topics and evaluation of the participants' contributions (see Figure 1). The graphical visualization is interactive, that means the tutors and students may choose to see different threads in the conversation, with zooming and other options.

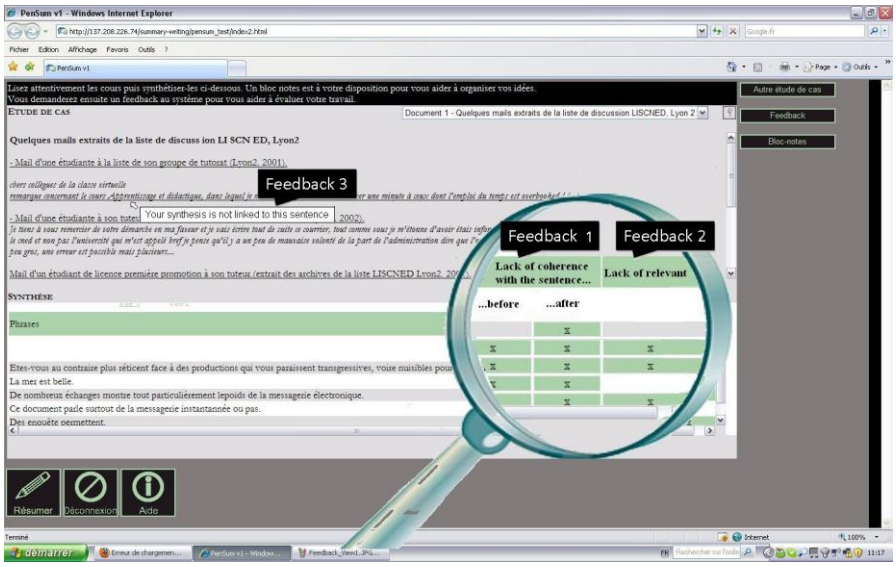

Figure 1. The different pieces of feedback delivered by *Pensum*.

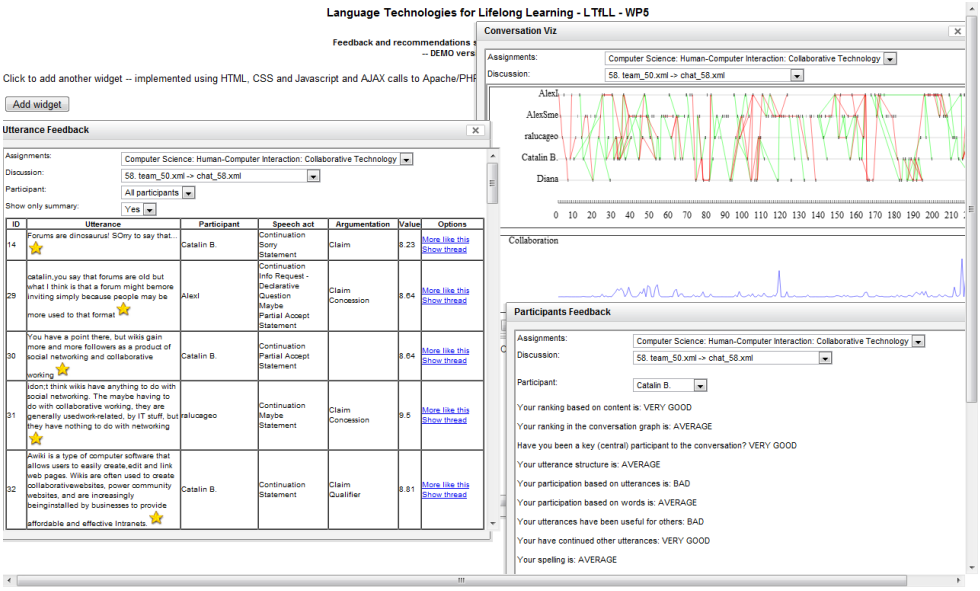

Figure 2. *PolyCAFe* main interface: utterance feedback and conversation visualization.

4 Validation Study

These two services have been subject to a first validation study. The main goal of this study was to have a closer look on the usability of the services for learners in authentic settings, since they provide complex feedback on equally complex tasks.

4.1 *Pensum* Validation

Participants. The students participating to the validation experiment were from three different university courses: Master 2nd year students in educational sciences, (with an ICT focus, $N = 6$); Master 1st year students in linguistics ($N = 3$) or language didactics and pedagogical design ($N = 2$). The average age of the participants was 34.5 ($SD = 12.1$) and 3 of which was male. They were rather proficient in computer use (5.1 h per week of use, $SD = 3.0$), mainly for Internet search and e-mail.

Task Description. Participants were given the following tasks to be performed at distance (ecological settings): to view an on-line screencast (4-minute long) describing the main functionalities of *Pensum*[1]. Then they had to use *Pensum* to write out a synthesis of a given set of two documents on ICT and Internet use in African countries. No length constraints were given and 18 days were left to perform the task. They eventually had to fill in a closed questionnaire on *Pensum*'s use (mixing questions on pedagogical soundness, usability, subjective

[1] http://www.youtube.com/watch?v=vnKLcmxq5hw

cognitive load, and overall satisfaction) and to participate to a phone interview with more qualitative questions.

Quantitative Results. The completion of the activity (i.e., writing a synthesis) took between 1 to 4 days ($M = 2$; $SD = 1.4$). They asked for feedback between 0 to 10 times ($M = 4.27$; $SD = 2.97$). The syntheses written by the participants were from 4 to 28 sentence-long ($M = 15$; $SD = 8.32$), and were modified 0 to 16 times ($M = 5.09$; $SD = 4.91$). The students made a very variable number of textual modifications: $M = 27.46$; $SD = 23.73$ (one student performed 83 modifications). Table 1 shows students' opinions on *Pensum*'s use (from closed questions). Overall, most of the answers are in the middle of the range (item 3 answer for a scale from 1 to 5), indicating a mixed opinion. For three questions, students expressed opinions were statistically different from the mid one. They think that *Pensum* gives feedback and guiding different from humans, but also that *Pensum* is rather easy to use and that errors are easy to recover. Briefly, this questionnaire showed that students had a better opinion on the usability of *Pensum* than on its effects on learning or its pedagogical capabilities. Eventually, we analysed data related to participants' subjective cognitive workload (from NASA-TLX, Hart & Staveland 1988). Their most important efforts were in trying to understand how to get a better use of *Pensum* and how to use it, and the frustration level compared to the four other factors (mental pressure, physical pressure, time pressure, achievement). These points appear to be normal considering it was the very first uses of the software.

Qualitative Results. The analysis of the open questions (interviews) showed that the students found that *Pensum* was useful for revising courses and helped them focus on the gist of the course text they read. However, their opinion on feedback quality was mixed: some of them complained that *Pensum*'s feedback was confusing because too many sentences were underlined, without sufficient explanations.

Overall, whereas the opinion of *Pensum*'s first users was mixed, this first validation study provides some indications to improve its usability further; first, in enhancing the quality of the feedback (particularly with regard to the inter-sentence coherence), second, in enabling teachers to put comments on the synthesis and to enrich the kinds of feedback given by the system (teachers may set the severity degree of the feedback themselves), third, in giving students control over the system (self-assessment and synthesis annotations).

Table 1. Usability data from the validation study.

Software service	Pedagogic effectiveness	Cognitive load*	Usability	Satisfaction
Pensum	2.4/5	3.9/5	n/a	2.1/5
PolyCAFe	3.9/5	3.6/5	4.1/5	3.9/5

*Scores measuring the users' cognitive load are not comparable across the two systems, since the questions from which they were processed are not similar (an aggregated score from NASA-TLX was used for *Pensum* whereas *PolyCAFe*'s score used answers to closed questions).

4.2 *PolyCAFe* Validation

Participants. A group of 9 students and 4 tutors participated to the validation of the service. The students were enrolled at the Human Computer Interaction course (4th, senior undergraduate year) at the Department of Computer Science and Engineering at the "Politehnica" University of Bucharest. Their average age was 23 years without having an important deviation. All of them have very good computer skills as they will be graduating with an engineering degree in computer science.

Task Description. The learning scenario consisted of the following tasks:

- Tutors grouped students in teams of 4-5 participants. Each student has been assigned a topic to study individually and afterwards to support it during an online chat debate with the colleagues from the same team. The subject of the chat discussion at the HCI course was "collaborative tools available on the web". The four topics assigned to the four members of the teams were: 'discussion forum', 'chat', 'wiki' and 'blog'. The team that had 5 members received an extra topic: 'google wave'.
- Students scheduled by themselves a date for the chat. As they were instructed, they stayed online and discussed for about two hours, structuring the conversation in two steps. In the first part, each of them had to support his/her assigned topic by presenting its features and advantages and criticize the others' topics by invoking their flaws and drawbacks. In the final part of the chat, they had to discuss on how they could integrate all these tools in a software environment.
- Students met with the validation team, watched a screen-cast describing *PolyCAFe* and then used the software in order to get feedback about their participation in the chat conversation. During their use of the software, they were encouraged to think-aloud about the usability of the tools.
- Students filled in a questionnaire on *PolyCAFe*'s use and participated in a focus group conducted by the validation team.

Quantitative Results. Due to the nature of the instant messaging technology, each student participated to the conversation only once, for 90–120 minutes. Then, they consulted the results provided by *PolyCAFe* once for each chat, for a period of about 60–90 minutes. It should be taken into account that this was also the first time when they used the software. The questionnaire that each student had to answer consisted of 32 items: 13 general questions related to the use of *PolyCAFe* as a whole and 19 questions related to specific functionality of the system's components. Table 1 offers an overview of the results to the generic questions grouped by category. There was a further category of questions not displayed in the table that considered the efficiency of the implemented solution which has an average score of 4.2/5. As can be seen from this data, the students considered the system to be effective, efficient and easy to use, with an average score of slightly above 4/5. They were also satisfied by the results provided by the system and the cognitive load was not very high taken into consideration the fact that it was the

first time they have seen the system. The highest scores were obtained for the following items:

- The students considered that *PolyCAFe* provides adequate support for their learning activities ($M = 4.33$; $SD = 0.47$).
- They considered that it takes less time to complete learning tasks using *PolyCAFe* than without the system ($M = 4.22$; $SD = 0.79$).
- The system was easy to learn to use ($M = 4.56$; $SD = 0.50$).

However, there were questions that received lower scores; the lowest one being when they were asked to compare the support provided by *PolyCAFe* compared to the current support provided by humans, which received an average score just above 3 ($M = 3.11$; $SD = 1.10$). However, it should be noted that the system is not designed to replace human feedback, but just to enhance it and provide a quicker alternative. On the other hand, there were 3 questions related to specific functionality of the system that had an average score below 3. This points out that although the system has been validated as a whole, specific modules should be improved in the next version of *PolyCAFe*.

Qualitative Results. The focus group results show that the students considered the system to be very useful for understanding their role in the chat conversation and the degree of collaboration, as well as the coverage of the concepts related to the topic of the discussion. As the feedback provided by the tutors for each chat is usually late and quite poor, the alternative of receiving preliminary results from *PolyCAFe* was received with enthusiasm. However, they pointed out that the usability of the system can be improved in order to provide a better guide to using the software and understanding how to use the results, indicators and textual feedback returned by the system. Moreover, not all the components were considered to be equally effective: the conversation visualization and the utterance feedback widgets were considered the most effective, while the conversation and participant feedback widgets were considered the least effective and more error prone. Several improvements to these components were suggested by students in order to be more relevant for their learning activities.

5 Conclusion and Future Work

We presented a theoretical framework arguing that learning tasks in collaborative platforms share five main features. We also designed and tested two web services supporting these tasks and accounting for these features. Our services (1) are focused on writing activities; (2) are both on individual and collective levels (3) can embed pedagogical facets through the use of web widgets; (3) propose high-level and automated feedback; (5) can be used by various stakeholders. A first validation study of these two services has been undertaken and shows promising results. Further work is planned to cross the results of these tools to uncover patterns of efficient individual or collaborative forms of writing.

6 Acknowledgments

This work was partly supported by the 7th Framework Programme (Theme FP7-ICT-2007-1-4.1) of the European Community LTFLL (STREP Project, Grant # 212578). We wish to thank the participants to the validation studies and the MSH-Alpes for providing material support.

7 References

Bakhtin, M.M. (1981): *The dialogic imagination: Four essays*, The University of Texas Press, Austin.

Dessus, P., Allègre, E., and Maurice, J.-J. (2005): 'L'enseignement en tant que supervision d'un environnement dynamique', *Ann. Rech. Sci. Educ.*, pp. 149-162.

Dillenbourg, P. (2002): 'Over-scripting CSCL: The risks of blending collaborative learning with instructional design', in P.A. Kirschner (Ed.): *Three worlds of CSCL, can we support CSCL?*, Open Universiteit Nederland, Heerlen, pp. 61-91.

Emig, J. (1977): 'Writing as a mode of learning', *College Composition & Comm.*, vol. 28, n° 2, pp. 122-128.

Hart, S.G., and Staveland, L.E. (1988): 'Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research', in P.A. Hancock, and N. Meshkati (Eds.): *Human Mental Workload*, North-Holland, Amsterdam, pp. 139–183.

Kollar, I., Fischer, F., and Hesse, F.W. (2006): 'Collaboration scripts - A conceptual analysis', *Educ. Psychol. Rev.*, vol. 18, n° 2, pp. 159–185.

Koschmann, T. (1999): 'Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration', in C. Hoadley, and J. Roschelle (Eds.): *Computer Support for Collaborative Learning (CSCL'99)*, Mahwah, Palo Alto.

Landauer, T.K., and Dumais, S.T. (1997): 'A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge', *Psychol. Rev.*, vol. 104, n° 2, pp. 211-240.

Shulman, L.S. (1986): 'Those who understand: Knowledge growth in teaching', *Educ. Researcher*, vol. 15, n° 2, pp. 4-14.

Stahl, G. (2006): *Group cognition. Computer support for building collaborative knowledge*, MIT Press, Cambridge.

Strijbos, J.-W., and Fischer, F. (2007): 'Methodological challenges for collaborative learning research', *Learn. Instr.*, vol. 17, n° 4, pp. 389–393.

Trausan-Matu, S., and Rebedea, T. (2010): 'A polyphonic model and system for inter-animation analysis in chat conversations with multiple participants', in A. Gelbukh (Ed.): *Computational Linguistics and Intelligent Text Processing (CICLing 2010)*, Springer, New York, pp. 354–363.

Wolfe, M.B.W., Schreiner, M.E., Rehder, B., Laham, D., Foltz, P., Kintsch, W., and Landauer, T.K. (1998): 'Learning from text: Matching readers and texts by Latent Semantic Analysis', *Discourse Processes*, vol. 25, n° 2-3, pp. 309-336.