Formative Evaluation of the SIETTE collaborative Testing Environment

Ricardo Conejo, Beatriz Barros, Eduardo Guzmán, Jaime Gálvez
Dept. of Lenguajes y Ciencias de la computación, University of Málaga, Spain
{conejo,bbarros,guzman,jgalvez}@lcc.uma.es

Abstract: This paper describes the development of a computerized collaborative testing environment upon the SIETTE system. Using this environment, students can take a test in groups. They can exchange messages, view the answers of their peers and discuss about the correct answer. We have made a formative evaluation of the system from March 2005 to January 2008. The incidences happened during the experiments and the users' feedback, have been used to improve the system. During these years, three different prototypes have been released and tested with undergraduate students from two Spanish Universities.

Keywords: Collaborative AIED systems, Evaluation of AIED systems, assessment

Introduction

Students' interaction and peer help has demonstrated to be very useful for learning. There are different types of combining collaboration and assessment. If one student evaluates his partners and he is evaluated by them, it is commonly known as *peer assessment* [5]. On the other hand, if two or more partners discuss and grade something, we call it *collaborative assessment* [2, 4]. It could be carried out with paper and pencil combined with conversations in the classroom [6] or supported by a computer [1, 5]. In this research, assessment refers to testing. It concerns to a *computer supported collaborative testing* environment that allows the realization of group activities in which partners give an individual answer that is shown to others; they can discuss and reflect about it, and subsequently, they have a second chance to submit an individual answer.

Preliminary results showed that the collaboration increased the performance in the assessment, for all students. Even those students who are the best of their group increase their performance as a consequence of the dialogue with others. Probably the reason of this is that they have to reflect on their answers in order to explain them to others [1].

This paper describes the development and evaluation process of the collaborative environment built upon SIETTE assessment tool [3], and the formative evaluation we have conducted in order to tune up the system for real class use.

1. The Collaborative Environment

SIETTE is a Web-based system for building and administering computerized tests. In fact, SIETTE is a suite of tools which implements all the stages of test construction, delivery and result analysis. We have built an environment around SIETTE, which provides all the mechanisms needed to allow and synchronize the collaboration among students while they take a test. Briefly, this environment, that we called *collaborative frame*, has been

implemented by means of a Java applet shown as a plug-in in the left side of the web browser window. This frame is only shown in the SIETTE's *virtual classroom* when the student is taking a collaborative test, and it is in charge of controlling all the aspects related to the student collaboration. Because of the use of synchronous communication, many awareness facilities have been included in the interface so the students can know where their colleagues are. The *collaborative frame* submits and retrieves information from a collaborative server, implemented by means of a Java servlet and a middleware layer. The system is available at *http://www.siette.org*.

Let us assume a student called John is going to take a collaborative test in SIETTE. Once he has logged in the system, he has to select the subject about which he is going to take the test. After this, a list of available tests on this subject is shown. Some of these tests can be taken using the collaborative environment. If he selects a collaborative test, he enters a "waiting room" where an initial collaborative framework is shown. This framework allows him to talk to other student using a chat tool. At this stage, students may agree to create a group of any number of participants to take the test. Let us assume John has decided to take a test with Jane who is also connected and waiting in the room. They have agreed the name of their group and enter into it. Once both have entered, the test starts automatically. Groups of users can be also pre-defined by the teacher. In that case, when the user enters the "waiting room" and selects the test to take, he is directly conducted to the test initial page.

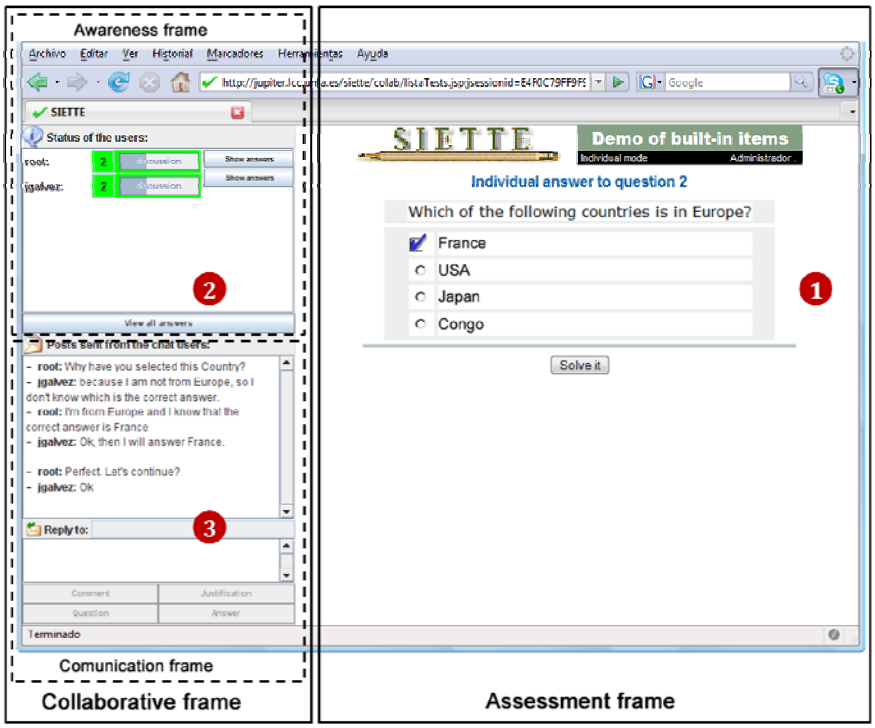

Figure 1. Interface of the collaborative assessment environment. (Third prototype).

Figure 1 shows a screenshot of the environment while two students are taking a collaborative test of Geometry. On the right, the assessment-frame (labeled with ①), is used by SIETTE to pose questions, and to show the answers given by students (in the figure, John is having a look at the answer of Jane). On the left, the collaborative frame is composed by the awareness-frame (upper, labeled with ②) and the communication-frame (below, labeled with ③). The awareness-frame depicts the evolution of the students involved in the test (including himself). The information of each student is shown in a different row. The first row always corresponds to the current student. Each row begins with the student nickname, followed by a color bar that shows the number of question that the student is currently

answering, and its phase. There are three possible phases within a question: *individual response*, *discussion*, and *group response*. The bar has different colors depending on the students' question number. The current student bar is always shown in green. All those students answering the same question are shown in green, those answering a former question are in red, whereas those in posterior questions in blue. Finally, each row has a button which allows the student to query the answers selected by the others. This button is only enabled during the collaborative stages of discussion and group response.

The communication frame has a chat window. Its upper part is the post panel. It is formed by a list of messages submitted. The lower part, is the writing panel, where students can type and send four types of messages: *Comments*, *Questions* and *Answers* and *Justifications*. The chat is only enabled during the *discussion* and *group response* stages of each question. In addition, every time the chat is disabled, the post panel contents are cleared and its messages are not kept between questions. That is, the post panel does not contain the messages sent in former questions. Furthermore, if two students (A and B) are trying different questions (Q_A and Q_B respectively), the one who is in the former question (for instance A), will only be able to read the messages sent by the other (student B) when student A arrives at question Q_B.

2. Usability and Formative Evaluation

The development of SIETTE collaborative environment started in 2004. Since then, we have released three different prototypes. Figure 1 includes on the left the collaborative frame corresponding to the current prototype. Figure 2 shows the two previous prototypes. The assessment framework was the same in all prototypes

2.1 First Prototype Evaluation

The first prototype of the collaborative frame was released around March 2005. It had notable differences regarding the second and third prototypes. However most of the current functionalities were already present. The chat tool was structured as a tree, but instead of displaying the message content directly, the tree simply displayed a message header. When the user clicked on the header, the message body was displayed in the panel below. In order to send a message, the user should select the previous message he wanted to answer (from the message tree), categorize his own message as "*Comment*", "*Question*", "*Answer*" or "*Justification*" (by using a radio button), write down the message body, and submit it. The student could also query other users' responses, by clicking on the nickname which appears on the bottom right panel (labeled with the text "*Users*").

To evaluate this prototype, we conducted an experiment with a small group of 12 Master and PhD students from the Informatics Schools at Malaga and Madrid Universities. The experiment consisted in a collaborative test about English grammar. They were organized in groups of two people: one located at Malaga and the other at Madrid. The experiment completely failed, but was very useful to us, since it addressed some non-qualitative information that was carefully taken into account in the following prototypes.

First of all, people found difficult to synchronize with their partners. They used the chat room for that, but they were continuously asking to the other about their current question number. The delay in the message exchange contributed to a total mess, and after the first three or four questions were posed, the partners disengaged and finished the test by

their own. This fact pointed out the important of awareness and established a desired goal: *Each user should know where he is and where the partners are*.

The second blocking problem found in the first prototype was related to the system performance. SIETTE has been previously tested under different user conditions, and proved to be stable with more than 100 students taking tests simultaneously. However the performance of the collaborative frame was not tested before in real conditions. The system had a problem of scalability, and by the time the third or fourth group tried to begin a new session, the whole system crashed. We analyzed the problem and realized that the more time consuming feature was the log of the users' actions. Logs were written in a separate XML file in the server side, and included a large set of information for further studies. To solve this problem we adopted realistic criteria: *Do not register unnecessary data, and replace XML files by a database as a persistence mechanism.*

The observation of the users' behavior during the test addressed a third problem. The system assumed that the student had to provide an individual answer first. After that, using the chat room, they should discuss the solution about the current test question, and then provide a final answer. This cycle is repeated in the next questions. Some tests in SIETTE are configured to show the correct answer after the student response. A pair of students cheated the system by avoiding the synchronization. That is, one student was answering question n, while his partner was answering question *n+1*, and the second sent a message to the first through the chat room, giving the right answer. This problem generated a new system constraint specification: *Forbid users' communication when they are not at the same question*.

2.2 Second and Third Prototype Evaluation

Due to the many problems that arose during the experiment with the first prototype, we were not able to carry out a formal evaluation. We had to reconstruct and recode several part of the software, to solve them. The second prototype was released in October 2006. It solved the main technical problems found in previous prototypes, and included the awareness features. It was the first fully operative, but it still had some scalability problem that we will discuss later. Third prototype was released in March 2007 and included some minor changes compared to the second one, but it was significantly more stable and useful.

In order to make a formative evaluation of both prototypes, we defined a questionnaire to be posed to students after taking a collaborative test. The questionnaire was divided in four sections: (1) About the user and their expertise using a computer testing and chat tools; (2) About the activity; (3) About the system and its usability and (4) General comments and proposals, to identify the advantages and disadvantages of collaborative testing with SIETTE. All questions, but the general comments, were rated in a Likert scale from 1 to 5. The questionnaire was voluntary and anonymous. We posed the same questionnaire to different student populations after taking a collaborative test.

The first questionnaire was filled by group J1 in December 2006 (second prototype). The second was filled by group J2 in November 2007 (third prototype). In both cases the sample came from undergraduate students of the last course of the School of Telecommunication Engineering at Malaga University. The test was about Java Programming. A third questionnaire was scheduled in January 2008, for first course students at the Forestry School (Madrid Polytechnical University). The test topic was Botany. In all cases, the students had previous experience with SIETTE, but it was the first time they used the collaborative frame. Table 1 shows the means (± the standard error at 95% confidence) of the most relevant results.

Regarding the open questions, in the group J1, most students (70% of the sample) pointed out that the main disadvantage of the system was its slowness. However, in group J2, nobody complained about this issue. Concerning the strengths, only 9% of group J1 considered the system as a useful tool for improving their learning. This percent was increased in sample J2 (14%). The main strength remarked by the students of both samples is the collaboration. Students highly valued this feature (39% in J1 and 47% in J2) since, according to their words, it contributes to solve their doubts, to focus through discussions on new issues not learned before, to learn new concepts, etc.

Table 1. Comparison among the experiment results.

	J1	**J2**	**B**
Evaluated Prototype	2nd	3rd	3rd
Date	Dec.06	Nov. 07	Jan. 08
Number of users	33	36	18
(1.1) My expertise with computers	4.12±0,23	3.89±0,26	3.17±0.55
(1.2) I like computer based testing	3.64±0.36	3,83±0.23	3.67±0.64
(1.3) I use the chat frequently	3.65±0.43	3,66±0.40	3.22±0.69
(2.1) I have enjoyed taking the collaborative test	4.33±0.34	4,.47±0.21	3.41±0.58
(2.2) I think I have learnt from my partner	4.00±0,35	4.06±0.32	3.12±0.63
(2.3) The partner answer helped me	3.65±0.33	3.75±0.32	3.18±0.58
(2.4) The partner discussion helped me	4.15±0.31	4.31±0.27	3.65±0.60
(3.1) The system works fine	3.62±0.33	4.42±0.31	3.17±0.57
(3.2) I always knew where I was	3.99±0.36	4.03±0.36	4.44±0.31
(3.3) I easily got the answers of my partner	4.04±0.38	4.53±0.29	4.00±0.48
(3.4) I can easily communicate with my partner	3.59±0.38	4.51±0.26	4.33±0.48
(3.5) The chat annotation was useful	2.66±0.31	3.17±0.47	3.44±0.57
(3.6) The structured chat was useful	2.70±0.41	--	--
(3.7) It is better using SIETTE with the collaborative frame	4.38±0.37	4.58±0.23	3.78±0.47
(4.1) My overall rating of the system is	4.06±0.35	4.44±0.19	3.83±0.35

3. Conclusions

The development of the system has been guided by a formative evaluation that has indicated some missing features, like the awareness facilities, and has transformed others, like the chat room. It has also pointed out the potential advantages and disadvantages. On the one hand the interactive learning, on the other, the time required to take a test, that is longer than in the individual mode. Accordingly, the system is currently used just as a learning tool for self-assessment, and has not been used for formal grading yet.

4. References

[1] Barros, B., Conejo, R. & Guzmán, E. (2007). Measuring the effect of collaboration in an assessment environment. *AIED'2007*, Los Angeles.

[2] Brna, P., Self, J., Bull, S. & Pain, H. (1999). Negotiated Collaborative Assessment through Collaborative Student Modelling. In: Workshop on Open, Interactive and Other Overt Approaches to Learner Modelling, *AIED'1999*, 35-42.

[3] Conejo, R., Guzmán, E., Millán, E., Trella, M., Pérez-de-la-Cruz, J.L. & Ríos, A. (2004) SIETTE: A Web-Based Tool for Adaptive Testing. *IJAIED*, 14, 29 61.

[4] Kwok, R. & Ma, J. (1999). Use of a group support system for collaborative assessment. *Computers & Education*, 32, 109-125.

[5] Lin, S., Liu, E. & Yuan, S. (2001). Web-based peer assessment: feedback for students with various thinking styles. *Journal of Computer Assisted Learning*, Vol. 17, 420-432.

[6] Simkin. M.G. (2005) An Experimental Study of the Effectiveness of Collaborative Testing in an Entry-Level Computer Programming Class, *Journal of Information Systems Education*, Vol. 16, 273-280.