Measuring the effect of collaboration in an assessment environment

Beatriz BARROS*, Ricardo CONEJO**, Eduardo GUZMAN**
** UNED, Spain,*
Juan del Rosal, 16, 28040 Madrid
*** Universidad de Málaga, Spain*
Boulevard Louis Pasteur, 35, 29071 Malaga

Abstract. This paper describes a computerized collaborative testing environment in which students can take an assessment session in groups. For each question, an individual response is firstly requested, then students are allowed to view the answers of the others and discuss the question. After that a final response is requested. This collaborative environment has been used to measure the effect of collaboration in test performance obtaining promising results that indicates that all students improve their performance no matter what their knowledge levels are.

1. Introduction

Collaborative learning scenarios should engage learners in cognitive and metacognitive activities to promote the conscious cooperative development of shared knowledge and to enrich their individual understanding of the world. Therefore, these scenarios should involve the learners in situations which require reflecting on their own knowledge as well as their colleagues' in a grounding process [2] in order to get, as result of this process, more refined and mature knowledge by themselves. There are several open challenges in this field, some of them related to this paper: first of all, the representation and quantitative measure of the process of collaboration; and second, the development of learning environments that focus the students in the learning objectives instead of dispersing the collaboration in aspects not related to the topic. The collaborative environment proposed in this paper defines semi-structured learning activities as collaborative scripts [9], which are not very rigid so they can have as rich, open and flexible a collaboration as possible.

Tests have been widely used for assessment. When they are carried out in groups, totally or partially, they are called collaborative testing [19]. Most of the research in this area has been done with paper and pencil tests. In this paper, an experiment with computerized tests in a collaborative scenario is presented. For each question in the test, students are allowed to give an individual answer that is shown to others; they can discuss and reflect on it, and subsequently, they have a second chance to submit an individual answer. For students, it is motivating [10] to know the answers of their colleagues and to chat about the answers during an assessment session. In this context, it is to be expected that the students are going to centre the discussion on the question meaning, on understanding their failures as well as their colleagues' and, in the case of conflict, to argue which of the answers is the right one. This scenario is interesting for studying the collaboration the partners discuss about the content to be learned and do not scatter the conversation. Berry[3]observed that working in small groups on assessment activities gives students an opportunity to speak about the topic and thereby sharpen their skills and understanding.

The collaborative assessment environment also incorporates a structured communication module, a chat tool that allows having argumentative discussions with labelled utterances in groups [17]. The system registers all the collaborative actions into a log file for off-line analysis. This environment is especially attractive for research because: a) it is designed as a learning activity that combines individual tasks with collaborative tasks b) it is a scenario that facilitates students' reflection and discussion on its own knowledge and the others' in a situated learning environment [11]; c) it forces students to adopt socially-acceptable ways of resolving conflicts [18]; d) students can build upon each other's knowledge with positive performance outcomes [8]; e) it introduces collaboration in scenarios that commonly are meant for individual learning; and f) it is a first step to study empirically how collaboration has influence on the performance of the learner, because (s)he is evaluated before and after the collaboration process.

In section 2, some points related with this work are going to be discussed. Next, the research scope and motivation of the paper are presented. In section 4, the collaborative environment is described and then, the experiments done with groups of students and the most significant results obtained from the evaluation process. The paper finishes with some conclusions and future proposals for working.

2. Related work

Some work has implemented and played on experiences combining computerized tests and collaboration, some of them in the classroom. Some studies [5] [13], observed that retentions and results of course content increases with the use of collaborative testing. Furthermore, Simkim [19] performed two experiments in the classroom with multiple-choice questions and obtained higher group scores, compared with individual scores. He also found that "*group exams encourage faculty to ask more challenging questions than they might otherwise and potentially increasing the amount of learning in the classroom*". [4] argue that there is no effect of collaborative testing performance in questions about theories but students should gain from collaborative testing when questions are about concepts.

The above-mentioned experiences have been done by means of paper and pencil tests, posing a whole set of questions to each student and later, a similar set in the collaborative mode. As far as we know, there are no published data about existing computerized collaborative assessment tests that interleaves individual and collaborative responses. Two interesting experiments [12] have been carried out in a web-based environment that combine argument-based collaboration with questionnaires (about opinions that are not correct or incorrect answers) and sharing facilities to work and reflect in groups. It uses a collaborative script, called ArguedGraph, with five phases that combines individual, collective and collaborative phases, all of them supported by a computer. During group phases, students have a chat tool to discuss and write their arguments. Similarly, they use a collaborative script that combines individual an collaborative phases (with a chat tool), but with different purposes. There has been also some research on peer assessment where students evaluate the answers of others. Our research is different from all these, because: (1) questions are used for assessment; (2) our script interleaves individual and collaborative phases, (3) the answers are finally evaluated by the computer, and (4) the objective of our experiments is measure the effect of collaboration in test performance.

It has been reported that on-line interactions among students make positive contributions to students' learning [14]. Furthermore, [7] observed that the analysis of students'

contributions to online discussions provides evidence of effective collaboration in an interactive e-learning environment. A text-based online conversation tool that allows performing on-line interactions could be implemented using a chat [15]. Chat tools are being used in several collaborative learning environments, as a free conversation tool [21], or with sentences openers [20], or with typed messages, that represent conversation acts with [1] [19] or with explicit reference to a shared document [16]

3. Research scope and motivation

An experiment for a computerized collaborative test has been designed, implemented and evaluated with real users. The first step was to design a collaborative script (Figure 1). The activity is designed for small groups of (2-5) students that take a test of *n* questions. The answer to each question is organized in three phases. Students interact synchronously and the conditions to pass to the next phase are to complete the previous one. The first phase consists of answering a question on their own. During the second phase, students share the results with their colleagues and discuss the answers given in the previous phase. In the third phase, each student answers the same question again individually. It is not compulsory to get to an agreement for this second answer. Each student can decide to keep or modify his/her first answer. This script is repeated for all questions in the test.

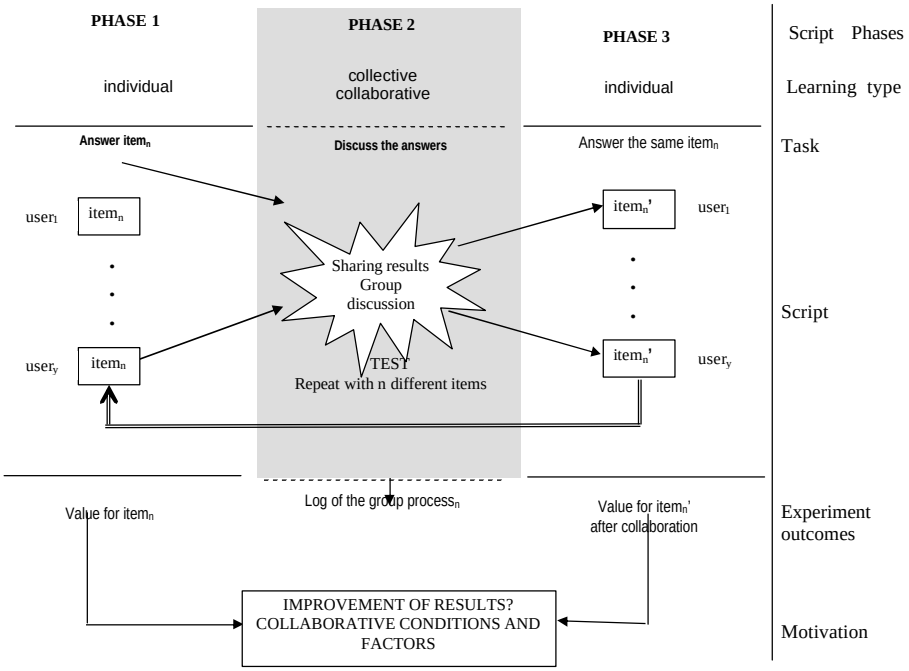

Figure 1. *Collaborative Script for the collaborative computerized tests, outcomes and motivation*

The objective of the experiment is to explore if students increase their performance when they work in groups and to study if the collaborative discussion process is related with this improvement.

4. The assessment environment

The script has been implemented in a web-based collaborative environment that manages the script sequencing, activates the collaborative interface and invokes the assessment tool SIETTE [6]. The assessment tool collects the answers of the students, and the events that

describe the activity of the users during the collaboration phase, that is, if they have seen the answers of others, the chat messages exchanged, etc.

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.

The empirical studies presented in this paper have been carried out using an extended version of SIETTE supporting student collaborative testing. We have built an envelope around SIETTE which provides all the mechanisms needed to allow and synchronize the collaboration among students. Briefly, this envelopment, that we have 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 server, implemented by means of a Java servlet. Such server implements a HTTP-based communication protocol between SIETTE and the collaborative frame of each student. Moreover, this server traces all the actions accomplished by the students as well as the messages they interchange. All tests available in SIETTE can be posed in the collaborative mode. The system with or without the collaborative support is available at http://www.lcc.uma.es/siette.

Figure 2 shows the environment while two students are taking a collaborative test. The right frame, the assessment-frame (labelled with ①), is used by SIETTE to pose questions, and to show the answers given by other students. The left frame is composed by the awareness-frame (upper, labelled with ②) and the communication-frame (below, labelled 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 the order of the question he/she is answering at this moment. Next to it, a set of status bars displays information about which stage the student is at, i.e. *individual response*, *discussion*, *group response* and *finished*. Such bars have different colours depending on which question the student is at, regarding the current student. Accordingly, 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 hierarchy of messages submitted by all the students involved in the test. As it can be seen in the lower part, students can send four types of message, i.e. comments, justifications, questions and answers. 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 two questions. That is, when the chat is enabled, the post panel does not contain the messages sent in former questions. Furthermore, a student can only see those messages sent during the question which he/she is currently contemplating. This means that 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.

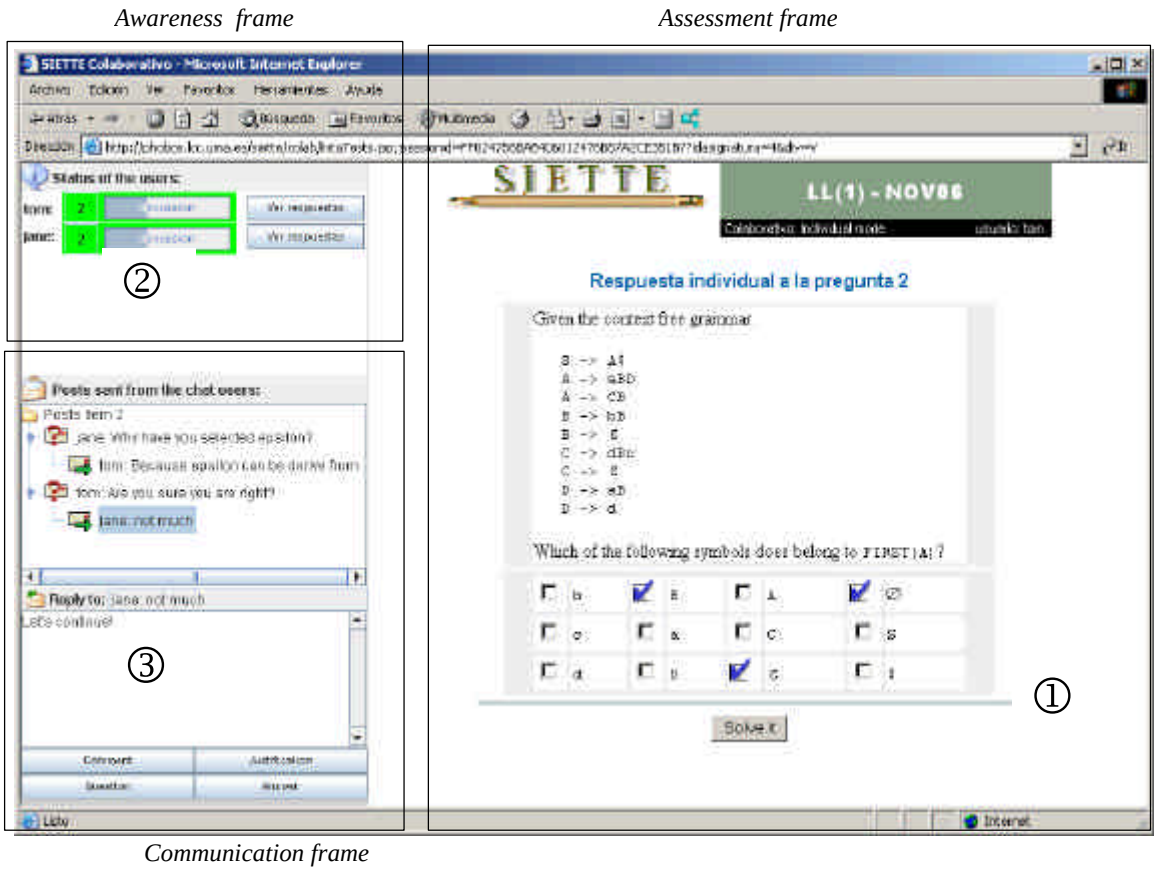

Figure 2: *Interface of the collaborative assessment environment*

5. Experiments

The evaluation was carried out by using a test that contains 10 questions about Compiler Construction. In particular, we focus on the LL(1) techniques and asked questions about finding the elements of the set obtained by FIRST() and FOLLOW() functions of a certain context free grammar. However, the system is completely domain independent. All questions were multiple choice so that the students were asked to select the appropriate set of symbols among 16 choices. That means that the probability of finding the right answer just by guessing is $1/2^{16}$, that is, almost 0.

Among these questions, the first one was easier than the others, and was included just for training the student to use the collaborative environment. This question has been removed in the results analysis, so the test can be considerd having just 9 questions. The evaluation involved a total of 24 students, in two different sessions of 7 and 17 students each. Students were randomly divided in 9 groups of 2 people and two groups of 3 people, one for each session.

First of all, we analyze the average of correct responses that has been given before the discussion phase (individual mode), and after (collaborative mode). Table 3a shows the results by each question. As it is shown a better performance is obtained for all questions.

In the first case the average success rate was 45,8% and in the second 63,7%, The difference is statistically significant ($p < 0,05$), even more so if we consider that the answers were given one after the other and that the probability of guessing is almost 0.

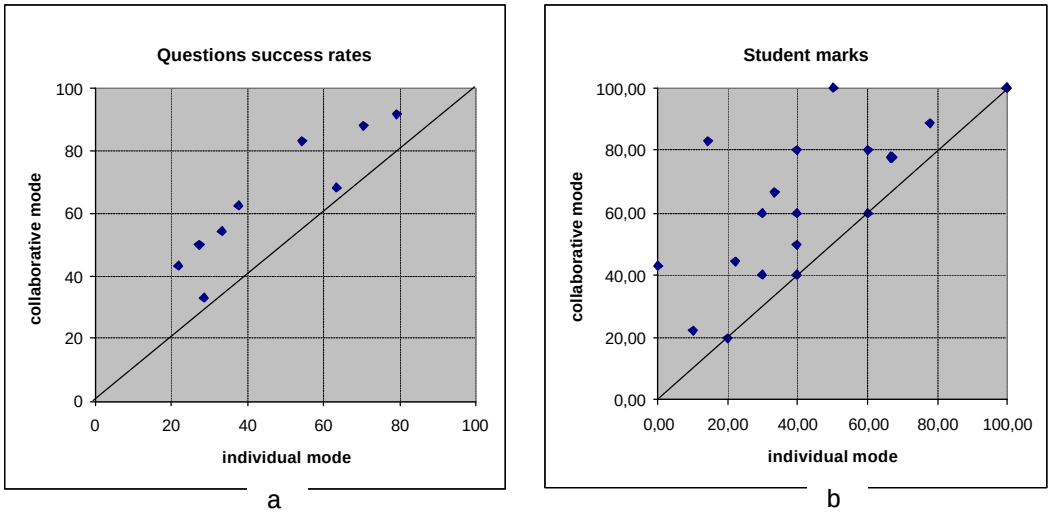

Figure 3: *Questions and students performance in individual and collaborative modes*

A reciprocal study has been done by considering the students' marks, obtaining similar results. As it was expected, figure 3b, all of the students improved their marks after the collaboration phase.

To deeply analyze what is going on inside the groups, we have divided the students in two classes according to their own results in the individual mode: the B-class which is formed by those students below the average, and the A-class, formed by those above the average. The data obtained indicates that all groups improve their mark in the collaborative mode. (table 1). The second column in this table represents the average of the differences in the marks (percentage of correct answers) obtained before and after the collaboration. It is not surprising that the B-class improve more than the A-class, simply because they can improve more. A new measure named "*relative improvement*" x is proposed to compare the results of the two classes, which is given by the formula:

$$x = \frac{M_C - M_I}{100 - M_I}$$

Where M_I represents the mark (percentage of correct answers) obtained before the collaboration, and M_C represents the mark (percentage of correct answers) after the collaboration phase. This result provides evidence that the relative improvement is similar in both classes.

Table 1: *Performance of students according to their estimated knowledge level*

	Number of cases n	Average of absolute improvement	Average of relative improvement $\bar{x}$	Standard deviation of the relative improvement s	Confidence interval at 95% $\bar{x} \pm t_{0.025} \frac{s}{\sqrt{n}}$
A-class	10	+11.7%	0,30	0,32	0,30±0,23
B-class	14	+23.1%	0,31	0,25	0,31±0,15

We can also analyze the results according to the group companion, and divide the sample in two classes: Those students that have a companion with higher individual marks (noted as L-class) and those that do not have a companion with higher individual marks (H-class). That is, in each group the student with the higher individual score is considering to be of H-class and the rest are L-class. Because there were 2 groups of 3 people the L-class and H-class have different sizes. 2 cases were discarded because they obtained the same mark as their companion.

Table 2: *Performance of students according to their relative knowledge level in their groups*

	Number of cases n	Average of absolute improvement	Average of relative improvement $\bar{x}$	Standard deviation of the relative improvement s	Confidence interval at 95% $\bar{x} \pm t_{0.025} \frac{s}{\sqrt{n}}$
H-class	10	+6,33%	0,16	0,19	0,16±0,14
L-class	12	+30,34%	0,47	0,27	0,47±0,17

The results (table 2) indicate that both classes improve their results ($p<0.05$), and that there is a significant difference ($p<0.05$) between the L-class and H-class. That is, it can be statistically concluded that those students that have at least one companion with a higher knowledge level, improve more than those that have the highest knowledge level in their groups.

The last experiment concerns the collaboration within groups and its relation to student improvements, measured as the average of the "*relative improvement*" of the student in the group as it was defined before. We have defined four indicators of the collaboration inside a group: a) The total number of messages in the chat room, b) The total length of all messages, c) The average length of messages, and d) the number of interactions, that is the number of response messages to previous messages posed by others. The last indicator tries to distinguish between those messages posted just by the same student, (a stand alone behaviour), from those that are posted as part of a conversation. We have obtained the correlation coefficient for each indicator, and found that there is a positive correlation between the average relative improvement and the total number of messages in the chat room ($r=0.74$) and with the interaction indicator ($r=0.71$). In fact both indicators were highly related in this case. These results are statistically significant with $p<0.05$. Another good indicator is that in 70% of the cases, the students have changed their previous responses after the collaboration phase, and in 66% they have agreed on a common answer.

6. Discussion and Conclusions

From our experiments it can be statistically concluded that collaboration, as has been defined in our web-based assessment environment, improves assessment performance for all students. Even those students that are the best in their group improve their performance as a consequence of the dialogue with other students, probably because they have to reflect on their answers to explain them to others.

In the experiment presented in this paper, the results before and after collaboration have been observed as well as the collaboration process in terms of three single indicators, number of messages, size of the messages and interaction level.

These data have been contrasted in relationship to the performance of the students. Results show that the bigger the interaction level within the group the better the student's performance. Furthermore, an interesting property of the collaborative assessment environment presented in this paper is the possibility of a quantitative measurement of the effects of collaboration that allows a deeper analysis of what is going on inside the groups.

References

[1] Barros, B.. & M.F.Verdejo (2000)."Analysing student´s interaction process for improving collaboration: the DEGREE approach", *IJAIED* 11, pp. 221-241

[2] Baker, M.J., Hansen, T., Joiner, R. & Traum, D. (1999). The role of grounding in collaborative learning tasks.In P. Dillenbourg (Ed.), *Collaborative Learning : Cognitive and Computational Approaches*, pp. 31-63. Amsterdam : Pergamon / Elsevier Science.

[3] Berry, J. and Nyman, M.A. (2002), "Small-Group Assessment Methods in Mathematics" *Int. Journal of Mathematical Education in Science and Technology* vol. 33, no. 5, pp. 641-64 9

[4] Breedlove, W., Burkett, T. & Winfield, I. (2004) Collaborative Testing and Test Performance, *Academic Exchange Quarterly*, Fall 2004, Vol. 8, Is.3.

[5] Cortright,R.N., Collins,H.L., Rodenbaugh, D.W. & DiCarlo, S.E. (2002) "Student Retention of Course Content is Improved by Collaborative-Group Testing", *Advances in Physiololgy Education*, 26, pp. 37-41.

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

[7] Curtis, D.D. & Lawson, M.J. (2001) "Exploring Collaborative Online Learning", *JALN*, Vol. 5, Iss. 1, February 2001, pp. 21-33.

[8] Damon W. & Phelps, E. (1989) Critical distinctions among three methods of peer education; *International Journal of Education Research*, 13, PP. 9-19.

[9] Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner *Three worlds of CSCL*, pp. 61-91.

[10] Graham, R.A. & Graham, B.L. (1997) "Cooperative Learning: The benefits of Participatory Examinations in Principles of Marketing Classes", *Journal of Education Business*, Vol. 72, no. 3, January/February, pp1. 149-152.

[11] Henning, P. H. (2004) Everyday Cognition and Situated Learning, In D. H. Jonassen (Ed.), *Handbook for Research in Educational Communications and Technology*, 2nd Edition. (pp. 143-168). Mahwah, NJ: Lawrence Erlbaum Associates.

[12] Jermann, P.,& Dillenbourg, P. (2002) "Elaborating new arguments through a CSCL Script" , Andriessen, J., Baker, M., Suthers, D. (eds.), Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning environments, pp. 1- 6.

[13] Hite, P. A. (1996). "An experimental study of the effectiveness of group exams in an individual income tax class". *Issues in Accounting Education*, 11(1), 61-75.

[14] Laurillard, D. M. (1999). A conversational framework for individual learning applied to the 'learning organization' and the 'learning society.' *Systems Research and Behavioural Science, 16*(2), 113–122.

[15] Mazur, J. M. (2004). Conversation analysis for educational technologists: Theoretical and methodological issues for researching the structures, processes and meaning of on-line talk. In D. H. Jonassen (Ed.), *Handbook for Research in Educational Communications and Technology*, 2nd Edition. (pp. 1073-1098). Mahwah, NJ: Lawrence Erlbaum Associates.

[16] Mühlpfordt, M. & Wessner, M. "Explicit referencing in chat supports collaborative learning**"** ***CSCL'2005***

[17] Pask, G. (1976). *Conversation theory: Applications in education and epistemology.* Amsterdam: Elsevier.

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

[19] Sheremetov, L., Guzman-Arenas, A. (2002) "EVA: an interactive Web-based collaborative learning environment", *Computers & Education*, 39, 161–182

[20] Soller, A. (2001) Supporting Social Interaction in an Intelligent Collaborative Learning System. *IJAED*, 12, 40-62.

[21] Veerman, A.L., E.B. Andriessen & Kanselaar, G. (1999) Collaborative Learning through Computer-Mediated Argumentation, *CSCL'99*