Let's argue over it: Are argumentation skills better learned collaboratively or individually?

Irene-Angelica Chounta, Bruce M. McLaren, Maralee Harrell

Carnegie Mellon University, Pittsburgh, PA U.S.A.
{ichounta, bmclaren}@cs.cmu.edu, mharrell@andrew.cmu.edu

Abstract. In this paper we present a study of the use of computers for teaching argumentation through argument diagramming. We specifically focus on whether students, when provided with an argument-diagramming tool, create better diagrams, are more motivated, and learn more when working with other students or on their own. Related research has shown that the construction of visual representations, such as diagrams, may have a positive effect on understanding, deeper learning and other cognitive skills, including critical thinking and argumentation. Keeping this in mind, we explore collaborative learning, how teachers integrate computers and software tools to enhance learning, and the effect this has on students. For the purpose of our research, we use learning analytics to evaluate a variety of student activities: pre and post questionnaires to explore motivational changes; the argument diagrams created by students to evaluate richness, complexity and completion; pre and post knowledge tests to evaluate learning gains. Finally, we conclude with a discussion of how learning analytics could inform learning design and improve the learning experience when students create argument diagrams.

1 Introduction

In this paper we present a study of the use of computers for teaching students argumentation through argument diagramming. Our specific interest was to assess whether students would produce better argument diagrams, be more motivated, and learn more when working in small collaborative groups versus working individually. Related research has shown that the construction of visual representations, such as diagrams, may have a positive effect on understanding, deeper learning and other important cognitive skills, including critical thinking and argumentation [1, 2]. In addition, the benefits of collaboration have been observed and, in particular, for learning to argue and to co-construct knowledge [3, 4]. Thus, providing students with a tool that can support both argument diagramming and collaboration might result in deeper learning about argumentation and, potentially, in helping students become better arguers. Additionally, we explore how teachers integrate computers and software tools to enhance the learning process and the effect this has on students. For the purpose of our research, we study various aspects of the learning activity with the use of learning ana-

adfa, p. 1, 2011.

lytics such as: a) pre and post questionnaires to explore motivational aspects; b) the argument diagrams created by students to evaluate richness, complexity and completion; c) pre and post knowledge tests to evaluate learning gains. Additionally, we study the effects the learning design (i.e. the processes undertaken by teachers and trainers while learning) had on the activity.

To explore the advantages that a software tool with collaborative features (LASAD) can provide to college-level students, we conducted a small classroom study with Carnegie Mellon University (CMU) undergraduate students. Prior studies of collaborative argumentation have almost exclusively been lab studies of short duration. However, our aim was to study the learning impact of a full semester of use of the collaborative argumentation tool with approximately 25 philosophy students. The study compared the practice and performance of 2 – to – 4 students who collaborated with LASAD to learn argumentation for an entire semester to students who worked alone with LASAD to practice argumentation. Our overall goal was to answer the basic research question: *Does collaborative, computer-supported argument diagramming lead to better understanding of arguments and better argumentation skills than individual argument diagramming*?

2 Related work

2.1 Argument diagrams and collaboration

Students need to learn to argue and debate in a well-founded, rational way in order to succeed in a variety of academic subjects, including science, mathematics, philosophy, and writing. Argumentation skills are also vital to everyday life in our complex, democratic society. Yet, these skills are often lacking in students and, hence, need to be explicitly trained and exercised. Philosophy in particular, is an academic discipline that emphasizes argumentation skills: arguments, in fact, are the focus of philosophical inquiry and are essential to academic reading and writing. Being able to fully understand, analyze, discuss, and develop arguments are essential skills required of students in these type of courses. A key tool for learning about argumentation in both courses is argument diagramming in which students take arguments, read them carefully, and reconstruct the arguments in a graphical form [1]. In particular, the students are expected to identify distinct premises and visually indicate how the premises, when linked together, lead to a particular conclusion. Argument diagramming, supported by computer-based tools, plays a large role in many philosophy courses.

There are a variety of benefits to argument diagramming, including, that the argument diagrams make arguments explicit and inspectable. Since the diagrams are persistent and readily shared, students can review and reflect upon them, leading to (potentially) deeper learning of the arguments than from simply reading and studying textual arguments. .

Yet, students typically work individually, not collaboratively, on argument diagramming exercises. Nonetheless, we believe that students can benefit from discussing arguments and working collaboratively as they diagram. Literature in the Learning Sciences has shown the potential benefits of collaboration, such as the benefits of explanation [5] and co-construction of knowledge [6] and, in particular, how these benefits have been observed in collaborative argumentation and learning to argue, when the argumentation is structured [4]. Thus, providing students with a tool that can support both argument diagramming and collaboration might result in deeper learning about argumentation and, potentially, in helping students become better arguers. We hypothesize that the LASAD tool will provide benefits for students, helping them to both reconstruct arguments and effectively collaborate.

2.2 Learning Analytics

The goal of *learning analytics* is to identify meaningful and beneficial practices with respect to learning outcomes and knowledge gains [7]. The type, the level and the specific objective of analysis varies dependent on the context and setting. For example, analysis of classroom activities typically focuses on the interaction between students and the teacher whereas analysis of a MOOC course typically focuses on the interaction between peers or among peers and learning resources. Learning Analytics can be applied in various settings in order to investigate either the relations between peers, or students and teachers, or learners or the learning material itself [8–10].

The use of metrics of activity and interaction is quite popular in computer supported collaborative learning, technology-enhanced learning and educational data mining. In particular, such metrics and indicators are used for mirroring, monitoring and guiding purposes [11], i.e. in order to assess the evolution of the learning process and to provide feedback to learners and to instructors for reflection and improvement [9]. These metrics usually derive from data-driven studies and represent activity volume (sum of messages, average number of words per message) or more advanced measures such as symmetry of contribution, temporal proximity, etc. [12]. In the same fashion, metrics that derive from Social Network Analysis (SNA) (such as centrality and density) have been used to evaluate the communication and knowledge flow between users during learning activities [12].

3 Method of the study

3.1 Study setup

The study took place as part of an "Introduction to Philosophy, Interpretation and Argument" course at CMU, over an entire semester (but with only three intervention sessions throughout the approximately four-month semester). The course participants were university students, between 17 and 21 years of age, from various departments (computer science, engineering, humanities and social sciences, etc.). The goal of the intervention – and a central component of the course – was to introduce students to argument diagramming and, in particular, on how to transform textual passages to argument diagrams. The creation of the diagrams was supported by a web-based argumentation system (LASAD) that allows users to argue in a structured fashion using graphical representations [3]. LASAD supports both individual and collaborative use (two – or more – users working synchronously on the same diagram). In the latter case, a chat tool facilitates communication between users. The interface of LASAD is presented in **Fig. 1**.

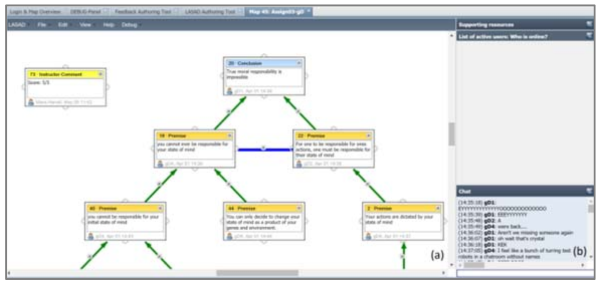

Fig. 1. The interface of the web-based argumentation tool LASAD. On the left, the tool offers a workspace for the creation of diagrammatic representations (a) and on the right, the tool provides a chat-area for the facilitation of communication between users (b)

Overall, 22 students participated in the study but three dropped out before completion of the course. Thus, the dataset consists of 19 students (8 Females and 11 males). The students were further subdivided into two conditions: the experimental condition, where students worked in groups of 2-4 members to create a common argument diagram (Group A) and the control condition, where students worked individually to create an argument diagram (Group B). Students in both conditions had to construct argument diagrams of the following arguments/theses: a) "*Of Miracles*" by D. Hume [14], b) "*Intelligent Design is Falsifiable*" by the Discovery Institute [15] and c) "*The Impossibility of Moral Responsibility*" by G. Strawson [16]. In particular, the participants had to read the arguments, point out the premises and formulate a conclusion through a diagrammatic representation that reflects the underlying relations between them. In this study we studied 32 argument diagrams that were produced from this process.

3.2 Methodology

During the sessions where students used LASAD, we collected student actions (LASAD keeps detailed records of user actions). The data were used to analyze the argument diagrams created by the students in terms of user activity. Furthermore, we used pre and post-questionnaires to assess motivational factors and pre and post-knowledge tests to evaluate learning gains.

Pre and post Questionnaires

In order to assess the way motivational factors affected the students' performance, we asked the participants to fill in two questionnaires, one before the beginning of the study and one right after the end of it. The pre-questionnaire consisted of three parts:

- Perceived background knowledge (4 questions on a 7-point Likert scale)

- Prior experience (4 Yes/No questions)
- Motivational Aspects (13 questions, partly adopted from the MSLQ – Motivated Strategies for Learning Questionnaire – instrument [12], capturing disposition towards class-work. Also rated on a 7-point Likert scale)

The post-questionnaire was only about motivational aspects. To that end we adapted the related question group from the pre-questionnaire in order to reflect potential changes in participants' disposition during the study. The question group that referred to motivational aspects was constructed on 6 dimensions, as portrayed in **Table 1**.

Table 1. Dimensions and questions for assessing motivation using pre and post-questionnaires

Dimension	Question
1. Orientation towards challenge	- I preferred the class work we did that was challenging
2. Effort regulation	- When work is hard I either give up or study only the easy parts - Even when study materials are dull and uninteresting, I keep working until I finish - I work hard to get a good grade even when I don't like a class
3. Perceived usefulness	- I think that what I learned in this class is useful for me to know
4. Perceived importance	- It was important for me to learn what was taught in this class
5. Preference	- I liked what I learned in this class - I think that what we learned in this class is interesting
6. Satisfaction for personal performance/confidence	- I'm certain I understood the ideas taught in this course - I believe that I did very well in this class - I did an excellent job on the problems and tasks assigned for this class - I believe that I did very well in this class - I know that I learned the material we covered in this class

Learning Outcomes

To assess the benefits of collaborative argumentation in terms of leaning gains, the participants were evaluated through knowledge tests, before and after the study. In particular, the participants were evaluated for their performance in argumentation on four dimensions: a) Analysis-Conclusion (on a 3-point scale); b) Premises (on a 3-point scale); c) Connections (on a 3-point scale); d) Evaluation (on a 3-point scale). Additionally, a domain expert assessed the argument diagrams created by the participants in terms of correctness and completeness on a 5-point scale.

Metrics of activity

As part of this study, we designed and developed metrics of activity based on the events that are logged by LASAD. These metrics were grouped into three sets: action-type activity, object-type activity and content-related activity. By action-type activity, we mean the log file events that refer to the LASAD types of action. These events can be of three types – create, edit and delete – thus resulting in three action-type metrics:

- *Number of Creates* (#create), i.e. how many times an object has been created on the common workspace;
- *Number of Edits* (#edit), i.e. how many times an object has been edited on the common workspace;
- *Number of Deletes* (#delete), i.e. how many times an object has been deleted from the common workspace.

By object-related activity, we mean all the events that relate to the LASAD object types. The LASAD objects can be of two types – boxes and relations – thus resulting in two object-type metrics:

- *Number of Box-related actions* (#box_actions), i.e. how many times a box has been manipulated;
- *Number of Relation-related actions* (#relation_actions), i.e. how many times a relation has been manipulated.

By content-related activity, we mean the metrics that relate to the content of the LASAD objects. To that end, we defined two metrics:

- *Number of content modifications* (#content_mod), i.e. how many times there has been some content modification
- *Average number of words per object* (#avg_words), i.e. how many words on average a LASAD object contains per argument diagram.

4 Results

Metrics of Activity

From the analysis of the argument diagrams, it was evident that the groups had more intense activity in terms of actions than the individual participants. On average, the groups performed 66% more action-type related events (they create more objects and they edit objects more often) than the individual participants and they performed 60% more object-type events than the individuals. This means that the groups created bigger diagrams (more objects that in argument diagrams stand for more premises) and more elaborated diagrams (more relations between the elements of the diagrams).

Although the groups also modified the content of the argument diagrams more than the individuals did, they didn't use more words to elaborate on the premises. The average number of words per object and per argument diagram was similar for both conditions. **Fig. 2** presents the metrics of activity used in this study for both conditions, i.e. groups (Group A) and individual participants (Group B).

Fig. 2. Action-type, Object-type and Content-related activity metrics for groups (Groups A) and individual participants as computed from the log files of LASAD

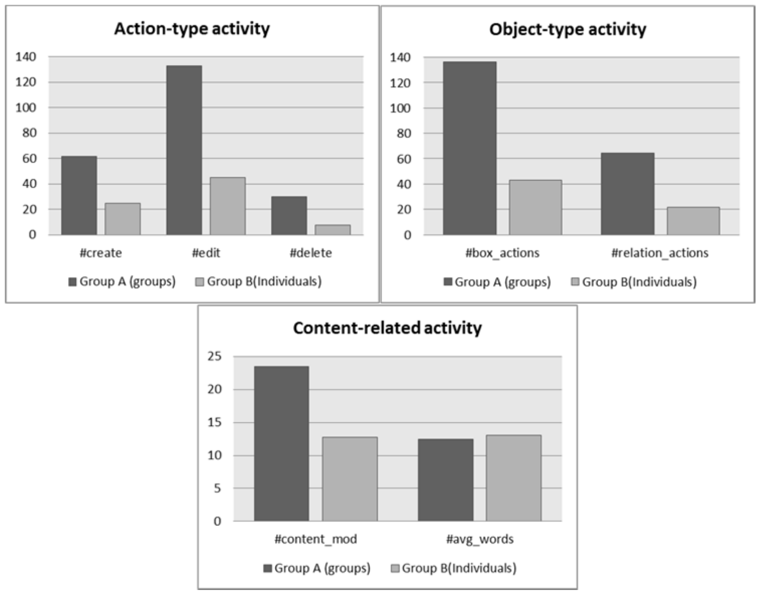

Table 2 presents the results for the log file metrics of activity for both conditions. In most cases, these differences were statistically significant (presented in bold letters). One could argue that it is expected that the more people work on a common argument diagram the more content they would create, either in terms of more words (which was not the case here) or in terms of more objects (which was indeed confirmed). However this does not necessarily indicate argument diagrams of better quality. To assess the quality of the diagrams, the third author – who is the teacher of the philosophy course and a domain-expert – rated the argument diagrams for correctness and completeness on a [0, 5] range. Comparing the grades of groups to the grades of individual participants, it was shown that the diagrams created by groups were on average rated higher than the diagrams created by individuals (avg_Grade[groups] = 4.45 > avg_Grade[individuals] = 3.875). However this difference was not statistically significant.

Next, we carried out a correlation analysis between the metrics of activity and the grade of the argument diagrams. From the analysis of the results, it was shown that the grade correlated positively and significantly to user activity related to the creation and manipulation of objects and content (*#create*, *#edit*, *#content_mod*) but did not correlate with the number of deletions (*#delete*). Furthermore, grades did not correlate with the volume of textual information (*#avg_words*). The results of the correlation analysis are presented in **Table 3**.

Table 2. Metrics of activity on average for groups (Group A) and individual participants (Group B) per type of activity and per argument diagram. Asterisks (*) denote the cases where the difference between two conditions are statistically significant

Type-related activity			Object-related activity			Content-related activity		
metrics	Group A	Group B	metrics	Group A	Group B	metrics	Group A	Group B
#create*	**61.64**	**24.90**	#box actions*	**136.45**	**42.90**	#content mod*	**23.45**	**12.71**
#edit*	**109.27**	**32.10**	#relation actions*	**64.64**	**21.90**	#avg words	12.46	13.10
#delete	30.18	7.81						
								**p-value < 0.05*

Table 3. Results of the correlation analysis between metrics of activity and the grade per argument diagram. Statistically significant correlations are marked as (**) for statistical significance on the 0.01 level and as (*) for statistical significance on the 0.05 level

	Spearman's ρ
#create	0.451*
#edit	0.361*
#delete	0.261
#content_mod	0.493**
#avg_word	0.163
	***p-value<0.01*
	**p-value<0.05*

Pre and post Questionnaires

The pre and post questionnaires were used to gain insight on the way motivational factors affect the activity. With respect to pre-questionnaires, participants in both Group A (groups) and Group B (individuals) rated similarly their perceived background knowledge (*BK*) in a [1, 7] range (BK_{GroupA}=4.386, sd_{GroupA}=0.875 for group and BK_{GroupB}=4.344, sd_{GroupB}=0.514 for individuals). In total, most participants took a course in *argument writing* (12 participants out of 19) and *persuasive learning* (14 participants out of 19) prior to this course but had no experience in *logic* (14 participants out of 19) and *argument diagramming* (14 participants out of 19). Participants, on average, were positively motivated and disposed towards class work (*Mot_pre*), including learning material and course's importance (Mot_pre =5.130, sd=0.724 in a [1, 7] range). Individuals scored, on average, higher (Mot_pre_{GroupB} = 5.317, sd_{GroupB}=0.29) comparing to group participants (Mot_pre_{GroupA} = 4.993, sd_{GroupA} = 0.89) but this can be due to higher variance groups' distribution.

With respect to the post-questionnaires, participants' motivation decreased in the post-test while the variance increased (Mot_post =5.020, sd = 0.924 on a [1, 7] scale). The picture was similar for both groups (Mot_post_{GroupA}=4.909, sd_{GroupA} = 0.98) and individuals (Mot_post_{GroupB}=5.173, sd_{GroupB} = 0.815). The difference in motivation between groups and individuals was maintained but the variation in the distribution was increased for the individual participants (5 out of 8 participants rated motivation lower in post than pre-questionnaire).

Table 4. Pre and Post Questionnaires comparison with respect to motivational aspects for both conditions (groups and individuals)

	Group A (groups)		**Group B (individuals)**	
	pre	**post**	**pre**	**post**
orientation towards challenge	4.636	4.545	5.125	4.375
effort regulation	4.758	4.576	5.000	5.083
perceived usefulness	5.545	5.273	5.375	4.625
perceived importance	5.909	4.818	5.625	5.000
preference	5.227	5.091	5.563	5.500
Satisfaction for personal performance/confidence	4.974	5.078	5.411	5.339

Both group and individual participants rated lower the items referring to curriculum (e.g. "*I liked what I learned in the class*", "*It was important to learn what was taught in the class*") in the post questionnaire, pinpointing that their expectations might have not been met. Participants who worked in groups maintained the same attitude with respect to giving up when work got harder than individuals and they gave higher ratings to items referring to perceived personal performance (e.g. "*I believe I did very well in this class*"). This indicates that working in groups made participants feel more confident about their performance and achievement in terms of grading and it also encouraged them to insist on their tasks instead of giving up.

We further analyzed the disposition of groups versus individuals with respect to the dimensions of the motivational questionnaires (**Table 1**). On one hand, the analysis showed that the satisfaction/confidence score increased for groups after the study but it decreased for the individuals. On the other hand, effort regulation increased for individuals but orientation towards challenge decreased. On average, individuals were eager to complete challenging tasks but they did not prefer such tasks over simpler ones. Perceived usefulness and perceived importance of what was taught in the class decreased for both conditions. The results of this analysis are presented in **Table 4**.

Learning Outcome

The teacher of the course evaluated the participants using knowledge tests before and after the completion of the study. As aforementioned, the participants were evaluated on four dimensions. To assess the overall learning gain, we compute the average grade of the four dimensions and compare the performance of the groups versus individuals. On average, the participants who worked collaboratively (Group A) scored higher in the post-knowledge tests ($Post_{GroupA}$ = 2.075 in a [0, 3] range) than the pre-knowledge tests (Pre_{GroupA} = 1.875 in a [0, 3] range), with a knowledge gain of 1.968. On the contrary, the participants who worked individually scored similarly in the pre (Pre_{GroupA} = 2.143 in a [0, 3] range) and post-knowledge tests ($Post_{GroupB}$ = 2.07 in a [0, 3] range). However, we should mention that Group A scored lower in the pre-test than Group B and thus one might argue that there was room for improvement. Even though, the difference in pre-knowledge tests between Group A and Group B was not statistically significant, the effect of collaborative argumentation on learning gain should be further studied.

5 Discussion

In this paper we presented a study of the use of computers for teaching argumentation through argument diagramming. We specifically focused on whether students, when provided with an argument-diagramming tool, create better diagrams, are more motivated, and learn more when working with other students or on their own. Our basic research question was: *Does collaborative, computer-supported argument diagramming lead to better understanding of arguments and better argumentation skills than individual argument diagramming*?

To that end, we carried out a preliminary study where 19 undergraduate students from various disciplines used a collaborative tool to construct argument diagrams based on given (written) arguments. The students were divided into two groups: those who worked collaboratively in teams and those who worked individually. For the purposes of the study, we used learning analytics such as: a) pre and post questionnaires to explore motivational aspects; b) the argument diagrams created by students to evaluate richness, complexity and completion; c) pre and post knowledge tests to evaluate learning gains.

The analysis revealed that participants, on average, were positively motivated towards the class (including class-work, learning material and course's importance) before the study but their motivation score dropped after its completion. Both groups and individual participants rated motivation towards items that referred to curriculum (e.g. "*I liked what I learned in the class*", "*It was important to learn what was taught in the class*") lower in post questionnaires than in pre-questionnaires. Participants who collaborated in groups rated higher motivation related to perceived personal performance (e.g. "*I believe I did very well in this class*") in contrast to individuals. Furthermore, they were determined to carry out activities perceived as dull or uninteresting. This may be an indication that collaborative work made participants feel confident about their performance and motivated them towards completing their tasks.

From the analysis of the diagrams, it was evident that groups created bigger and more detailed diagrams than individuals but used similar number of words to develop their arguments. However, size and complexity do not necessarily signify good quality. To that end, a domain expert assessed the argument diagrams with respect to their quality (i.e. correctness and completeness). The diagrams that were created collaboratively got higher grades than the ones created by individuals. Additionally, the grade correlated positively and significantly to user activity related to the creation and manipulation of objects and content (*#create*, *#edit*, *#content_mod*) but did not correlate with the number of deletions (#*delete*). The expert who assessed the dia-diagrams focused on the correctness and completeness of the resulting, final diagram. On one hand, it is expected that "complete" diagrams would have more object than the "incomplete" ones since this would signify the use of more premises and relations to elaborate and support argumentation. On the other hand, the expert assessed only the final diagram without taking into consideration the whole process. Actions such as "*create*" and "*edit*" leave visible traces on the diagrams. On the contrary, actions such as "*delete*" – that leave no traces – might improve the outcome but this would be observable only if someone monitored the whole process and not the final result alone.

We believe that this study provides insight on how to support learning through learning design and learning analytics. In our case, we applied the use of a collaborative tool to support teaching of argumentation skills. It was shown that collaboration empowered participants with respect to confidence and to achieving their goals. However, this work also had limitations that need to be further addressed, such as the small number of participants and the limited use of the collaborative tool. Furthermore we focused on the activity that took place on the common workspace but did not take into consideration the chat logs that would provide information about the communication between group members. Additionally, we focused on the activity of students – since this was only preliminary research that aimed to study the effect of the tool's use – and did not study the role of the teacher or the activity itself from the teacher's perspective.

In future work we plan to carry out extensive studies with more participants in order to validate our findings. Additionally, we plan to engage teachers from various fields in order to study how the use of tools and analytics affect their practice and the classroom dynamics.

6 References

1. Harrell, M.: Argument diagramming and critical thinking in introductory philosophy. High. Educ. Res. Dev. 30, 371–385 (2011).
2. Harrell, M., Wetzel, D.: Improving first-year writing using argument diagramming. In: Proc. of the 35th Annual Conf. of the Cognitive Science Society. pp. 2488–2493 (2013).
3. Scheuer, O., McLaren, B.M., Harrell, M., Weinberger, A.: Will structuring the collaboration of students improve their argumentation? In: Artificial Intelligence in Education. pp. 544–546. Springer (2011).
4. Weinberger, A., Stegmann, K., Fischer, F.: Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). Comput. Hum. Behav. 26, 506–515 (2010).
5. Fonseca, B., Chi, M.T.H.: The self-explanation effect: A constructive learning activity. Handb. Res. Learn. Instr. 270–321 (2011).
6. Webb, N.M.: Information processing approaches to collaborative learning. (2013).

7. Ferguson, R.: Learning analytics: drivers, developments and challenges. Int. J. Technol. Enhanc. Learn. 4, 304–317 (2012).
8. Chounta, I.-A., Hecking, T., Hoppe, H.U.: Every Answer Has a Question: Exploring Communication and Knowledge Exchange in MOOCs Through Learning Analytics. In: CYTED-RITOS International Workshop on Groupware. pp. 140–147. Springer (2015).
9. Voyiatzaki, E., Avouris, N.: Support for the teacher in technology-enhanced collaborative classroom. Educ. Inf. Technol. 19, 129–154 (2014).
10. Manske, S., Chounta, I.-A., Rodriguez-Triana, M.J., Gillet, D., Hoppe, H.U.: Exploring Deviation in Inquiry Learning: Degrees of Freedom or Source of Problems? In: Proceedings of the 23rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education (2015).
11. Jermann, P., Soller, A., Muehlenbrock, M.: From mirroring to guiding: A review of the state of art technology for supporting collaborative learning. In: European Conference on Computer-Supported Collaborative Learning EuroCSCL-2001. pp. 324–331 (2001).
12. Kahrimanis, G., Chounta, I.-A., Avouris, N.: Study of correlations between log file-based metrics of interaction and the quality of synchronous collaboration. Guest Ed. 24, (2010).
13. Chounta, I.-A., Hecking, T., Hoppe, H.U., Avouris, N.: Two make a network: using graphs to assess the quality of collaboration of dyads. In: CYTED-RITOS International Workshop on Groupware. pp. 53–66. Springer (2014).
14. Hume, D.: Of miracles. (2001).
15. Discovery Institute: Intelligent Design is Falsifiable,ok http://www.discovery.org/f/494.
16. Strawson, G.: The impossibility of moral responsibility. Philos. Stud. 75, 5–24 (1994).
17. Pintrich, P.R., De Groot, E.V.: Motivational and self-regulated learning components of classroom academic performance. J. Educ. Psychol. 82, 33 (1990).