Designing Automated Assessment FOR Collaborative Argumentation in Science Classroom: A Pilot Study

Wenli Chen, National Institute of Education, Singapore, wenli.chen@nie.edu.sg
Chee Kit Looi, National Institute of Education, Singapore, cheekit.looi@nie.edu.sg
Wenting Xie, National Institute of Education, Singapore, wenting.xie@nie.edu.sg
Yun Wen, National Institute of Education, Singapore, happywenyun@hotmail.com

Abstract: In the realm of CSCL research, collaborative argumentation is regarded as a key type of knowledge construction process that should be mastered by students to enable knowledge advancement. We designed an automated assessment system to support students' collaborative argumentation in Science learning. This paper describes the system design and explores how it was used by teachers and students in Singapore classrooms from a pilot study.

Keywords: collaborative argumentation; assessment for learning; automated assessment

Introduction

There is an ever-growing need to engage students in collaborative argumentation as coherent arguments to justify solutions and actions are important 21st century skills. Argumentation per se can be regarded as a significant knowledge construction activity that leads to knowledge advancement (Weinberger & Fisher, 2006). In the science classroom, the necessity to engage students in argumentation experience is more stressed as the process of arguing is an essential aspect of the epistemic practice in doing science, the discourse processes in talking science, and scientific thinking (Jimenez-Aleixandre et al., 2000) that are need to be appropriated.

Being aware of the significance of both "learning to argue" and "arguing to learn" (Scheuer et al., 2010) to the development of 21st century competencies and scientific content knowledge in students, the absence of argumentation in the current pedagogy of science learning, and the difficulties teachers and students face in enacting argumentation activities (Clark et al., 2007), support for collaboration and argumentation is needed to bring the learning processes into fruition. Such support can be envisioned and engendered with in-situ assessment of group learning. Being cognizant of the inadequacy of current assessment practices of collaborative learning which are mostly summative, individual-based, cognition-oriented, and conducted only by the teacher (Strijbos, 2011), we aim to develop an assessment-oriented system for supporting collaborative learning in secondary school classrooms. This is an on-going project which is still at the first cycle of the design research. This paper describes the design of the system and explores how the system was used in Science classroom from a pilot study.

Design considerations

Teachers do need technological and pedagogical support for designing, implementing and assessing collaborative argumentation activities in classrooms. CSCL researchers have developed a good number of online learning environments to support students' collaborative argumentation (Scheuer et al., 2010), and established a variety of analytical frameworks to assess the quality of this type of learning activity (Clark et al., 2007). However many research in technology-enhanced collaborative argumentation and assessment projects are not conducted in real school classroom context. This project is motivated by the need from teachers in designing and implementing collaborative argumentation in classroom. The research team did several focus group discussions with Singapore teachers to find out what are needed in supporting classroom collaborative argumentations. The focus group discussion results become the design considerations of the system. Due to the page limit the detailed focus group discussion results are not presented in this paper. These considerations are summarized as follows:

1. Make argumentation intuitive for students by providing graphical scaffolding. A diagram-based argumentation which includes an organized set of argument elements represented by nodes and/or directed links will help students focus on important tasks in argumentation.
2. Provide multi-dimensional assessments to have holistic assessment on collaborative argumentation
    a. both the individual level and the group level (to achieve inter-dependence and avoid free-rider);
    b. both the learning outcomes (usually reflected in group co-constructed artifacts) and the collaborative learning processes (to enable timely regulation);

    c. both the cognitive and social processes enacted in collaborative learning;

3. Provide assessment FOR learning instead of assessment OF learning so that students can make meaning out of the assessment immediately and act upon it.
    a. providing real-time assessment (students as autonomous and responsible learners; students themselves should make the decision for their future learning; self-awareness and meta-cognition);
    b. provide dynamic visualizations of assessments to help teachers and students make interpretations and decisions (managing cognitive load).

Based on these design considerations, a set of multi-dimensional indicators for assessing collaborative argumentation have been derived. The assessments of both the argumentation process and outcome were pursued. The outcome assessment is concerned with the evaluation of the argumentation artifacts that a student or a group of students created when asked to articulate and justify claims or explanations; the process assessment is about judging the process of constructing the argumentation artifacts (Sampson & Clark, 2008). The assessment addresses both individual performance and group performance. The current assessments are mainly used for evaluating secondary school students doing argumentation in science. The particularity of the learning context dictates the advancement of scientific knowledge and the engagement in the epistemic practices in doing science (e.g., coordinating between theory and evidence; taking alternative perspectives into consideration), the application of reasoning and epistemic strategies, and the participation in the collaborative knowledge construction processes as the prior pedagogical goals. This prescribes the interpretation and operationalization of the selected indicators for each dimension and the specification of the measurements. The establishment of the indicators and their measurement in essence are the adaptation and translation of existing valid and reliable analytic frameworks of collaborative argumentation while accommodating the specific demands and characteristics of the user and the environment involved in our study (e.g., Erduran et al., 2004; Weinberger & Fisher, 2006; Zhang et al, 2011). The operationalization of the indicators and the specifications of the measurements are introduced in Table 1.

Table 1: The indicator and operationalization of assessment

	Indicator	Operationalization	Level
Social Process	Engagement Level (Unit: Group / Individual)	The frequency of contribution to the group work either in the form of constructing the shared argument graph or participating in group chat. The higher the frequency, the higher the level of engagement is.	The setting for the critical value for each level depends on the normal distribution of all acts of student.
	Centralization (Unit: Group)	The degree to which the group members equally participate in group interaction. It is measured by the inequality of interactions by different members within the group. The higher the inequality, the lower the centralization is.	The setting for the critical value for each level depends on the normal distribution of all centralization value.
Learning outcome (argument as the outcome)	Structural Completeness (Unit: Group)	The presence of the essential structural components in an argument generated. The better an argument is, the more components are included. In the present context, an argument contains a claim, one (or more than on) evidence for, and more than one evidence against is regarded as a complete argument. The structural completeness of the argument diagram is measured by the sum of scores of all the arguments it contains.	Level 1: Claim Level 2: Claim+ Evidence for; Claim+ 1 Evidence against Level 3: Claim + Evidence for (>=1) + 1 Evidence against Level 4: Claim + Evidence for (No. >=1)+ Evidence against (No.>1)
	Relevance (Unit: Individual argument element produced by a group/ individual)	Whether the evidence provided is related to the topic under argumentation and whether it can support the claim or the evidence that it is directed to.	Level 1: Irrelevant information/facts. Level 2: Some relevance but no logic coherence. Level 3: Relevant and logic but not reflect the key points. Level 4: Relevant and logic, and reflect the key points.
	Scientific Sophistication (Unit: Individual argument	The extent to which students have moved from an intuitive toward a scientific framework. Scientific sophistication represents the level of success a student has achieved in processing an idea at a certain complexity level. The higher the sophistication, the	Level 1: Misconception; naive conceptual framework. Level 2: Misconceptions that have incorporated scientific information but show mixed misconception/scientific

	Indicator	Operationalization	Level
	element produced by a group/ individual)	more scientific the idea that produced is.	frameworks. Level 3: Basically scientific ideas based on scientific framework, but not precisely scientific. Level 4: Scientific explanations those are consistent with scientific knowledge.
	Epistemic Complexity (Unit: Individual argument element produced by a group/ individual)	The extent to which students make effort to produce theoretical explanations and articulations of hidden mechanisms central to the nature of science (i.e., providing and elaborating explanations or justifications) besides providing descriptions of the material world (i.e., providing unelaborated facts). It represents the level of complexity at which a student\group chooses to approach an issue. Epistemic complexity of an argument is measured by the cognitive effort taken to processing it as reflected in the content. The greater the cognitive effort, the higher the complexity is.	Level 1: Unelaborated facts: Description of terms, phenomena, or experiences without elaboration. Level 2: Elaborated facts: Elaboration of terms, phenomena, or experiences. Level 3: Unelaborated explanations: Reasons, relationships, or mechanisms mentioned without elaboration. Level 4: Elaborated explanations: Reasons, relationships, or mechanism elaborated.

System design

According to the derived design considerations and conceptualized assessment indicators, we have developed the system. The graphic representations (i.e. the shared argument diagram) are supported (see Figure 1). On the diagram-based argumentation space, an argument refers to an organized set of argument elements represented by nodes and/or directed links. The specific types of argument elements adopted are in accordance with Toulmin's Argumentation Pattern (TAP) (1958). For pragmatic considerations (e.g., understandability of secondary school students) (Scheuer et al., 2010), the original TAP model is simplified. Three argument elements, namely claim, evidence for (support), and evidence against (rebuttal) are identified as the essential components of an ideal argument. These elements are represented by: 1) the type of Node: Claim vs Evidence and/or; 2) the type of directed Link: For vs. Against.

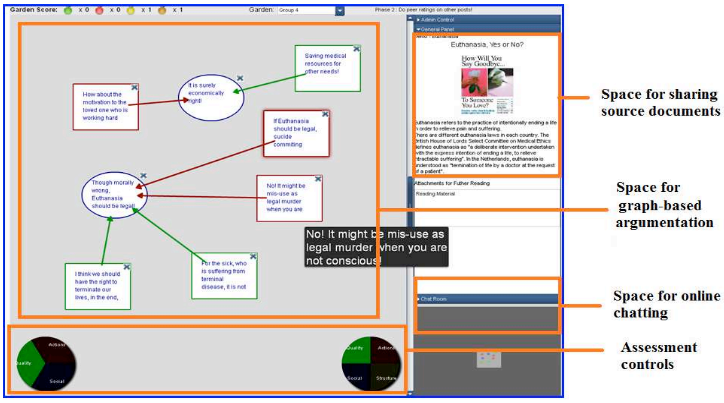


Figure 1. Interface of the System

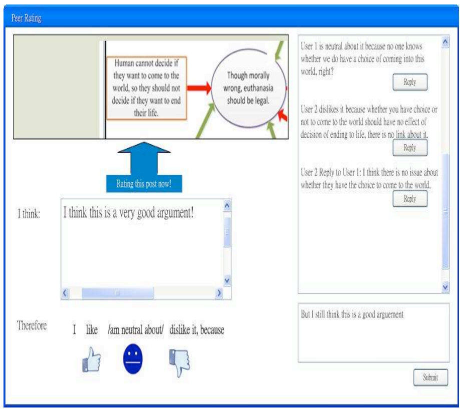
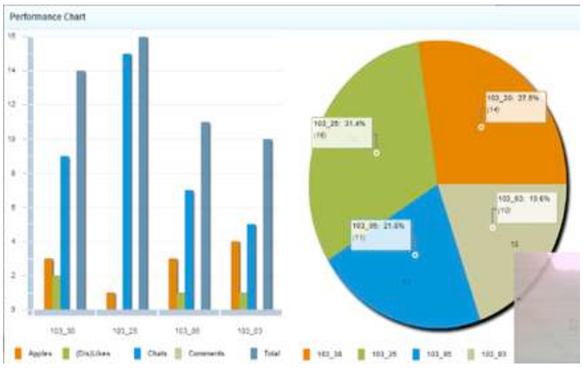


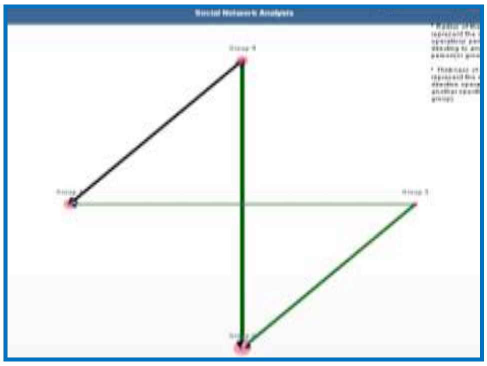
Figure 2. Peer-rating on the system

The automation of scoring complex data remains a motivating goal in educational assessment (Griffin, Care, Bui, & Zoanetti, 2013), and it is also the key to realize real-time assessment for learning. The measurement of the action frequency for the social participation is easy to achieve. Automated assessment of the quality of argumentation is a challenge. The system engages peer-rating for assessing the quality of argumentation as outcome (Figure 2). The peers who are working on the same task are prompted to provide quality ratings for the elements of the argument diagram once finished. Peer-rating is regarded as a good way to quantify the qualitative data as the feedback of peers can lead to reliable and accurate assessments (Cho & Schunn, 2007). Prompting students to rate peers' contributions may have a learning effect as they reflect on and assess their own and others' contributions (Scheuer et al., 2010).

Visualization is important for student and teachers to do real –time assessment which inform then what to do next. Figure 3 shows the visualization of various assessment indicators.



(a)



(b)

(c)

(d)

Figure 3. Assessment visualizations

Participation (a); interaction (b); content quality (c); and structural completeness (d)

The pilot study

The system was deployed in a secondary school in Singapore. As this pilot study is exploratory in nature, we did not adopt experimental design to examine the effectiveness of the system in comparison with a control group. Data collected and analyzed in this pilot study was to investigate whether: students could improve their scientific content knowledge using the system; teacher could improve their instruction using the system; and the automated assessment was reliable. Altogether, 4 Secondary 1 classes (40 students per class) taught by two science teachers (two classes by each teacher) participated in collaborative argumentation activities using the system in 1 physics lesson (1 hour per lesson) with the topic on Colors of Light. The lesson was co-designed by the teachers and researchers. Before the lessons, two sessions of technical training and 2 teacher professional sessions were conducted. Due to the page limit, we summarize the general findings from multi-facet data. In general there is positive result of using the system on learning and teaching from three aspects.

Learning outcome. To examine the effect of the system on students' development of scientific content knowledge, pre-test and post-test was conducted. A same test paper designed by teachers and another group of science teachers help validate the test items in terms of alignment of lesson deign and difficulty level etc. Two teachers scored students' test papers independently with good inter-rater reliability (Pearson r =.91). There is significant improvement in the scores from students' pre-test score to post-test for all the 4 classes (Table 2).

Table 2. Paired-sample t tests for 4 classes

Class	Type	Mean	SD	t	Class	Type	Mean	SD	t
Class A (n=36)	Pre test Post test	3.56 7.17	1.93 1.55	8.90**	Class C (n=31)	Pre test Post test	3.48 8.24	1.92 0.99	12.33**
Class B (n=33)	Pre test Post test	3.12 5.69	1.59 1.90	6.62**	Class D (n=34)	Pre test Post test	3.45 4.66	1.79 2.36	2.69**

Note **p<.01

Students' perception. Each student wrote reflections on their experiences of using the system for collaborative argumentation. The results revealed that students held positive attitudes towards both the lessons and the system. They liked to use the system for collaborative argumentation it is a new and refreshing way to learn science and is more interesting than a normal science lesson. Their engagement and participation in the learning activity was also enhanced as every student could and would like to express their ideas as the system supports anonymous contributions from the students. This was especially encouraging for those shy and quiet ones. Apart from improving participation, engagement and interest in learning, students also indicated that the system could help them learn the scientific knowledge and skills (through argumentation and from peers) and develop collaboration and critical thinking skills. The assessment components further advanced their learning.

Teacher's feedback. The system is envisaged as a tool to support teachers' orchestration in the classroom based on the provision of immediate assessments of student work. To explore teachers' use of the system in class, teacher interviews and post-lesson reflections were administered. Data analysis showed the positive role of system played in augmenting teaching. According to teachers, the instruction which utilizes ICTs, argumentation, real-time individual/group-based assessment, and group work, could well motivate and engage the students. Another striking feature of the system that particularly benefited their teaching was the provision of opportunities to access the thoughts and ideas of every single student. In the system, all the thoughts/ideas generated by students are accumulated and documented, and the teacher could keep track of students' misconceptions with ease and address them immediately.

Validation of assessment. For the indicators from the social aspects (i.e., Participation & Interaction), the measurement is based on frequency data generated by the system, which was affirmed accurate with human checking. The measurement of the Structural Completeness is based on the kind of data that can be accurately captured using the system (this was also confirmed by human checking). Nevertheless, the big challenge was to establish the validity of the ratings by the peers for the content quality. To achieve this, correlation between the peer-rating and the researcher rating has been conducted. In doing this, firstly, a sample (N = 77) of the posts generated by students in the second lesson was formed via random selection. It was found that the correlation between the peer-ratings and researcher ratings were not very strong (Pearson r = .51). This may be caused by the lack of scientific knowledge and rating competencies of the students. There is a necessity of equipping the raters with sufficient domain knowledge and rating competencies to achieve fair and reliable ratings.

Discussion and future work

The system is a knowledge representation tool where the structure of argumentation is explicitly represented to support students' collborative argumentation. The argumentative diagrams can clarify relations (Suthers, 2003), represent structure (Schwarz, Neuman, Gil, & Ilya, 2000), provide overviews (Larkin& Simon, 1987), maintain focus (Veerman, 2000), and enhance reflection on alternative perspectives (Kolodner & Guzdial, 1996). The assessments proposed in the system are not only about establishing what the students might have learnt. They are about providing less final and judgmental (Boud, 1995) but more interactive and forward-looking (Carless, 2002), timely and with a potential to be acted upon assessments (Gibbs & Simpson, 2004) to inform the subsequent teaching and learning activities. Analyses of data collected from the pilot study are indicative of the positive role the system play for both learning and teaching. The teacher and students not only "*assess to learn" by* finding out how student perform in the process of collaborative argumentation in order to improve their learning, but also "*learn to assess"* collaborative argumentation holistically from multi-facet perspectives.

To better inform the unfolding teaching and learning practices, the assessments designed should reflect the critical aspects of the learning activity. The assessments established are prescribed by the assessment considerations derived from literature review and through the focus group discussion with teachers. The selection of assessment indicators reflects the commonly acknowledged view on collaborative argumentation as social processes that can enable knowledge construction and creation. Teacher opinions were elicited for validating the assessment indicators included. Teaches reported that the provision of information and evaluation of students' engagement level, participation level, and understanding of the content knowledge at both the individual and group levels would definitely benefit their teaching practices, and all of these could be obtained using our system. According to them, the assessments currently available on the system are useful and comprehensive. As each assessment serves their own purpose, the teachers would like to have them all for reference.

The system is designed by reflecting the critical aspects of the collaborative argumentation and addressing the need from teachers and students in classroom learning activity. Except for the attainments, inadequacy of the current design has also been identified. The main issue identified is the validity of the peer-rating. This calls for the engagement of natural language processing and machine learning techniques to explore automated or semi-automated assessment of content quality. This project is still ongoing. We seek to establish automated assessment for more indicators to render a more complete capture of the collaborative learning processes. In the following cycle when Natural Language Processing is embedded, indicators that reflect the state of students' problem solving, argumentative knowledge construction and epistemic moves can be identified and assessed. Students' application of argumentation and thinking strategies can be examined as well. Another task ahead is to polish and enrich the visualizations of system assessments.

References

Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.

Carless, D. (2002). The 'mini-viva' as a tool to enhance assessment for learning. *Assessment and Evaluation in Higher Education, 27*(4), 353-363.

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education, 48*(3), 409–426.

Clark, D. B., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic frameworks for assessing dialogic argumentation in online learning envrionments. *Educational Psychology Review, 19*(3), 343-374.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin' s argument pattern for studying science discourse. *Science Education, 88*(6), 915–933.

Gibbs, G. & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1*, 3-31.

Griffin, P., Care, E., Bui, M., & Zoanetti, N. (2013). Development of the assessment design and delivery of collaborative problem solving in the assessment and teaching of 21st century skills project. In E. McKay (Ed.), *ePedagogy in Online Learning: New Developments in Web Mediated Human-Computer Interaction* (pp. 1–24). Melbourne: Information Science Reference.

Jiménez-Aleixandre, M. Pilar; Bugallo Rodríguez, A., & Duschl, R. A. (2000). " Doing the lesson " or " Doing science ": Argument in high school genetics. *Science Education, 84*(6), 757–792.

Kolodner, J., & Guzdial, M. (1996). Effects with and of CSCL: Tracking learning in a new paradigm. In T.

Koschman (Ed.), *CSCL: Theory and practice* (pp. 307–320). Mahwah, NJ: Erlbaum.

Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education, 92*(3), 447-472.

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning, 5*(1), 43-102.

Schwarz, B., Neuman, Y., Gil, J., & Ilya, M. (2000). Construction of collective and individual knowledge in argumentative activity: An empirical study. *Journal of the Learning Sciences, 12*, 219–256.

Strijbos, J. W. (2011). Assessment of (computer-supported) collaborative learning. *IEEE Transactions on Learning Technologies, 4*(1), 59-73.

Suthers, D. D. (2003). Representational guidance for collaborative inquiry. In J. Andriessen, M. Baker, & D. D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer- supported collaborative learning environments* (pp. 27–46). Springer.

Toulmin, S. (1958). *The uses of argument*. New York: Cambridge University Press.

Veerman, A. L. (2000). Computer-supported collaborative learning through argumentation. Doctoral dissertation. Utrecht: Utrecht University.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71-95.

Zhang, J., Hong, H. Y., Scardamalia, M., Teo, C. L., & Morley, E. A. (2011). Sustaining knowledge building as a principle-based innovation at an elementary school. *Journal of the Learning Sciences, 20*, 262-307.