The case of designing and validating a tool to assess 11-14 year old students written argumentation

Maria Evagorou, University of Nicosia, Cyprus
Elena Papanastasiou, University of Nicosia, Cyprus
Jonathan Osborne, Stanford University, USA

Abstract

The purpose of this paper is to present the process of designing and validating a tool to assess 11-14 year old students written argumentation in science. Given the emphasis in argumentation in science education in the recent years, methodologically the assessment of argumentation has become one of the dominant issues in the field. In this paper we present two written argumentation tools that were designed for 11-14 year old students, the validation process, and the main outcomes from testing the argumentation tools with 246 students in the UK. The analysis of the data from two versions of the questionnaire (Test A and Test B) show that both tests are reliable for evaluating written argumentation when the first item is removed (Cronbach's alpha .674 for Test A and .705 for Test B). Additionally the analysis of the data implies that choosing a convincing argument is more difficult for the students than any of the other three aspects of argumentation that were evaluated in these tests (choosing a convincing counterargument, writing an argument, writing a counter-argument). Finally, the results from the questionnaire suggest that argumentation might be content specific. Implications for research include the design of a new tool combining the questions from both tests, using aspects of the questionnaire to explore whether argumentation is context specific, and exploring whether deciding on a convincing argument is a higher sub-skill compared to writing an argument.

Keywords: argumentation, evaluation, written assessment

Paper Presented at the European Science Education Research Association (ESERA), September 5-9, 2011, Lyon, France

Background, Framework and Purpose

Argumentation has been a prominent field in the science education community for the past years (Jiménez-Aleixandre & Erduran, 2008) and recent reports in the field identify the need for a science education that places an emphasis on understanding argumentation and explanation as part of the practice of science (Duschl, Schweingruber, & Shouse, 2007). Students' arguments have been analyzed using a variety of methods and data sets – researchers have analyzed interviews in which people engaged in argumentation, videos of people constructing arguments, students' artifacts created during the instruction, and written essays (e.g. Jimenez-Aleixandre, Rodriguez, & Duschl, 2000; Erduran, Simon & Osborne 2004; Sampson & Clark, 2008). However, according to published research no single valid instrument exists to evaluate young students written argumentation skills, a tool that can potentially be used as a measure of whether students improve or not their argumentation during the instruction. Hence, the main purpose of this study was to design and evaluate a questionnaire to assess students' arguments through a combination of open-ended and Likert scale questions. Therefore, the research questions explored in this paper are:

(a) Is the current assessment tool an appropriate tool for measuring students' argumentation skills*?*
(b) What can we say about students' written argumentation skills based on this test?

Rationale/Importance

As stated above, argumentation has been the emphasis of many studies in recent years, however we do not have valid tools to help us evaluate students' written argumentation and enable us as researchers to follow the progress of students, and measure the impact of research programs. Consequently, teachers do not have a consistent way to measure their students' performance in argumentation. Hence this study is important since it seeks to explore this relatively new area, and additionally help us understand whether such a tool is appropriate for the evaluation of argumentation. The theoretical framework underlying the design of theassessment test was that of Toulmin's (1958) view of the elements of an argument. According to Toulmin's framework the essential elements of an argument are claims, data, warrants and backings. Counter-arguments are also important, especially in a dialogue, since one of the reasons that students should develop their argumentation skills is to enable them to evaluate claims and data and decide, in their everyday life, whether an argument is valid or not (Millar & Osborne, 1998).

Methods

In order to test whether the current assessment tool is an appropriate tool for measuring students' argumentation skills a mixed-methods approach was used. The initial pilot testing of the questionnaire was based on open-coding, whilst the validation of the questionnaire was based on a statistical analysis of the responses provided. Several versions of the assessment tool were designed and pilot tested before finalizing the two versions presented in this proposal (Test A and Test B). Each test includes four items, each one designed to test an aspect argumentation as shown in Table 1 below. The choice of these four parts is based on the structure of argumentation, and the fact that argumentation is seen as a means of understanding knowledge and advancing new knowledge (e.g. Driver, Newton & Osborne, 2000). Hence the decision to ask students to choose the most convincing arguments and counter-arguments and then construct their own arguments and counter-arguments lies in the fact that both are part of the argumentation framework that informs the design of this study (Toulmin, 1958). Table 1 below provides an overview of the parts of the two questionnaires, and the context for each of the questions.

Table 1: Overview of structure and content of the two questionnaires

	Choosing a convincing argument	**Choosing a convincing counter-argument**	**Constructing an argument**	**Constructing a counter argument**
Test A	Q1a: Is current used up in a simple electric circuit?	Q2a: Light travels from our eyes to objects?	Q3a: Using chemicals to kill mosquitoes?	Wind energy vs. nuclear factories
Test B	Q1b: Explaining a breath vs. heart rate graph.	Q2b: Weather conditions under which it is possible to snow.	Q3b: Choosing the hardest rock from a list and explaining.	Q4b: Should we use mobile phones at school?

Each one of the four items was evaluated based on argumentation levels. These levels indicate the quality of the argument based on a modified version of the Erduran et al., (2004) framework: Level 4: (most convincing argument) data, warrant, and rebuttal; Level 3: warrant and data; Level 2: warrant only OR data only; Level 1: appeal to authority; Level 0: contradicts the claim or no response.

The questionnaires were administered to a total of 246 students (11-14 years old) in public and private schools in London and suburbs, with 114 students completing Test A and 132 students completing Test B. The decision to design two shorter versions of the tool instead of a longer one was based on observations from the initial piloting of the tool (e.g. the longer version was time consuming). The responses were coded based on the modified version of the Erduran et al. (2004) framework, and then the data were analyzed in SPSS.

Results

Cronbach's alpha, which is a measure of the internal consistency of the items, was calculated for each of the tests. Cronbach's alpha for Test A was .605, and for Test B was .616. The inter-item correlation for both tests (Test A and Test B) which explores if the four items in each test correlate is presented in Tables 2 and 3.

Table 2: Inter-Item Correlation Matrix for Test A

	Q1	Q2	Q3	Q4
Q1	1	0.095	0.187	0.129
Q2	0.095	1	0.459	0.317
Q3	0.187	0.459	1	0.474
Q4	0.129	0.317	0.474	1

Table 3: Inter-Item Correlation Matrix for Test B

	Q1	Q2	Q3	Q4
Q1	1	-0.002	0.14	0.126
Q2	-0.002	1	0.483	0.333
Q3	0.14	0.483	1	0.524
Q4	0.126	0.333	0.524	1

As evident by Cronbach's alpha for both tests, the internal consistency was relatively low in the two tests, with Test B having a slightly higher consistency than Test A. Additionally, the correlation of items in Tables 2 and 3 show than there is a low correlation between Q1 and the other items in both tests. Therefore the scores for Question 1 were removed from both tests and Cronbach's alpha was calculated again as .0674 for Test A and 0.705 for Test B. The internal consistency is relatively high for both tests when the first question is removed from the total scores.

Conclusions

The analysis of the data from two versions of the questionnaire (Test A and Test B) show that both tests are reliable for evaluating written argumentation when the first item (Question 1) is removed. Additionally, the analysis of the data suggests that argumentation is evaluated best

with open ended-questions (Q3 and Q4) since by removing either items Q1 and Q2 from both tests Cronbach's alpha is relatively higher than when removing either items Q3 and Q4 (see Table 4 below).

Table 4: Cronbach's Alpha if Item deleted for both tests

Cronbach's Alpha if Item Deleted for Test A		Cronbach's Alpha if Item Deleted for Test B
Q1	0.674	0.705
Q2	0.508	0.532
Q3	0.394	0.361
Q4	0.489	0.478

The above implies that choosing a convincing argument is more difficult for the students than any of the other three aspects of argumentation that were evaluated in these tests (choosing a convincing counterargument, writing an argument, writing a counter-argument). Finally, the results from the questionnaire imply that argumentation might be content specific since for the same structure of the two tests, Test A had a lower Cronbach's alpha, and that was higher when Question 1 (which was heavily relying on students' knowledge on electric circuits) was removed. The hypothesis that argumentation is context specific is also supported from findings in previous studies (e.g. Means & Voss, 1996).

Implications

Implications for research include the design of a new tool combining the questions from both tests to see if this provides a higher internal consistency measure, and also using aspects of the questionnaire to explore (a) whether argumentation is context specific, and (b) that deciding which is a convincing argument is a higher sub-skill. Since the findings from this study suggest that choosing a convincing argument from a list of given arguments is more difficult that constructing an argument, implications for practice include finding ways to support students when deciding upon convincing arguments, since this is a skill useful in their everyday life.

References

Duschl, R., H. A. Schweingruber, et al., Eds. (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8*. Washington, D.C.: The National Academies Press.

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

Jimenez-Aleixandre, M. P., Rodriguez, A. B. & Duschl, A.R. (2000). "Doing the Lesson" or "Doing Science": Argument in High School Genetics. *Science Education 84*(6), 757-792.

Jiménez-Aleixandre, M. & Erduran, S. (2008). Argumentation in Science Education: An overview. In S. Erduran and M. Jiménez-Aleixandre (Eds). *Argumentation in Science Education: Perspectives from Classroom-Based Research*, 3-27. Springer.

Means, M. L. & Voss, J.F. (1996). Who reasons well? Two Studies of Informal Reasoning Among Children of Different Grade, Ability, and Knowledge Levels. *Cognition and Instruction 14*(2), 139-178.

Millar, R. and Osborne, J. (1998) *Beyond 2000: Science Education for the Future.* London: King's College.

Sampson, V. & Clark, D, (2008). Assessment of the Ways Students Generate Arguments in Science Education: Current Perspectives and Recommendations for Future Directions. *Science Education 92*(3), 447-472.

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