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
Argumentation

• Argumentation, the ‘social process, where co-operating individuals try to adjust their intentions and interpretations by verbally presenting a rationale of their actions’ (Patronis, Potari & Spiliotopoulou 1999, p.747-748), has been the emphasis of many studies in the field of education and the learning sciences during recent years (deVries, Lund & Baker 2002; Kuhn 2011, Kuhn & Udell 2003; Pontecorvo 1993).

• Argumentation is part of the practice of science for evaluating, refining and establishing new theories (Duschl, 1990) and is considered a core element of the scientific enterprise.

• Recent educational reforms have called for a science education that places an emphasis on teaching science as argument (i.e. Duschl, Schweingruber & Shouse 2007).
Research in Argumentation

a) Argument as a means of critical thinking (e.g. Kuhn, 1991, 2005; Kuhn, Goh, Iordanou & Shaenfiel 2008);

b) Argument as a vehicle for enhancing conceptual understanding (e.g. Herrenkohl & Guerra 1998; Jimenez-Aleixandre et al. 2002; Zohar & Nemet 2002);

c) Argument as a means of epistemic understanding (Bell 2004; Sandoval 2003; Sandoval & Reiser 2004);

d) Argument as a means of enhancing engagement with science, and developing citizenship, especially in socio-scientific issues (e.g. Jimenez-Aleixandre et al. 2000; Sadler & Zeidler 2005), and

e) Methodological concerns for studying arguments (Erduran, Osborne & Simon, 2004; Sampson & Clark 2008).
What we know about argumentation

- What we have gained from the aforementioned studies in argumentation is an understanding of the difficulties that the learners face when they try to construct their arguments, either scientific or socioscientific.

- 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.
The main purpose of this study was to design and evaluate a questionnaire to assess students’ written argumentation through a combination of open-ended and Likert scale questions. Therefore, the research questions explored in this study 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?
Importance of the study

- Argumentation has been the emphasis of many studies and reforms in recent years.

- 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 assessment tool

- Toulmin’s (1958) Argumentation Pattern [TAP]
- Counter-arguments are also important - students should develop their argumentation skills to enable them to evaluate claims and data and decide, in their everyday life, whether an argument is valid or not (Millar & Osborne, 1998).
- Based on TAP, the importance of counter-arguments, and the function of argumentation, an initial tool was designed with the following parts:
  - Writing a convincing argument
  - Writing a convincing counter-argument
  - Choosing a convincing argument from a list
  - Choosing a convincing counter-argument from a list
**Claim**: The conclusion whose merits are to be established.

**Warrant**: The reason that is used to justify the connections between the data and the conclusion.

**Backing**: The basic assumption that provides the justification for particular warrants.

**Qualifier**: Specifies the conditions under which the claim is true and are phrases that show what kind of degree of reliance is to be placed on the conclusions, given the arguments available to support them.

**Rebuttal**: Specifies the conditions in which the claim is not true.
The assessment tool

- Toulmin’s (1958) Argumentation Pattern [TAP]

- Counter-arguments are also important - students should develop their argumentation skills to enable them to evaluate claims and data and decide, in their everyday life, whether an argument is valid or not (Millar & Osborne, 1998).

- Based on TAP, the importance of counter-arguments, and the function of argumentation, an initial tool was designed with the following parts:
  - Writing a convincing argument
  - Writing a convincing counter-argument
  - Choosing a convincing argument from a list
  - Choosing a convincing counter-argument from a list
The assessment tool

• An initial version of the tool (8 questions, 2 for each category) was pilot tested with 23, 12 year old students.

• Two versions of the questionnaire where designed (Test A & Test B) with 4 questions in each.

• Test A questionnaire was pilot tested with 67 students, Test B questionnaire was pilot tested with 50, 11-14 year old students.
The assessment tool

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

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

• Each one of the four items in each was designed and evaluated based on the argumentation levels, 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, contradicts the claim or no response.
Question 1:
Directions: In this question you are given a claim and 5 different ways that a person could justify the claim. Your job is to rank each justification from 1 to 5: 1 being the most convincing justification and 5 being the least convincing justification.

The figure at right illustrates a simple series circuit, which consists of two light bulbs, some wire, and a battery.

Claim: The current that goes through the two identical bulbs shown in the picture (connected in series) in a simple electric circuit is not used up because...

...electricity comes from a stream of electrons moving through a wire. Light bulbs restrict the flow of electrons, which produces light and heat, but they do not use up these electrons. Level 2

...light bulbs only restrict the flow of electrons. We placed an ammeter before and after each of the bulbs in the circuit and the indication of the ammeter was the same after each bulb. Level 3

...as our teacher told us in the previous class, the current is never used up in simple electric circuits. The book also says that the current is not used up. Level 1

....when we placed an ammeter after each bulb in circuit the reading was the same. That means that the bulbs do not use up the current because the reading was the same in each ammeter. If the ammeter's reading for one of the two bulbs was different then the current was used up. Level 4

John
Nick
Justin
Lynn
**Question 4:**

The local council has decided to build a nuclear energy factory because a very small amount of nuclear fuel will provide a very large amount of electricity. The people in the area disagree. They prefer to use wind energy instead. You are representing the people in the area. Check the facts’ cards below and try to write a counter-argument to convince the local council to use wind energy instead.

<table>
<thead>
<tr>
<th>WIND ENERGY: FACTS</th>
<th>NUCLEAR ENERGY: FACTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Relatively easy to set up.</td>
<td>1. Very expensive to set up.</td>
</tr>
<tr>
<td>2. Does not work when there is no wind or when the wind is too strong.</td>
<td>2. Can be very harmful if it goes wrong.</td>
</tr>
<tr>
<td>3. Each wind turbine does not generate very much electricity.</td>
<td>3. A very small amount of nuclear fuel (uranium) provides a very large amount of electricity.</td>
</tr>
<tr>
<td>4. Does not produce any gases that pollute the atmosphere.</td>
<td>4. Does not produce any gases that pollute the atmosphere.</td>
</tr>
<tr>
<td>5. Low maintenance costs - does not need any fuel.</td>
<td>5. Very high maintenance costs</td>
</tr>
<tr>
<td>6. Some people think wind turbines look ugly.</td>
<td>6. Very reliable and can provide energy for a long time</td>
</tr>
</tbody>
</table>

*(example adapted by the IDEAS pack, Osborne et al., 2005)*

I believe that we should not use nuclear energy because ..........................................................................................................................
......................................................................................................................................................
......................................................................................................................................................
......................................................................................................................................................
......................................................................................................................................................

Instead, we should use wind energy because ..........................................................................................................................
......................................................................................................................................................
......................................................................................................................................................
......................................................................................................................................................
......................................................................................................................................................
The questionnaires were administered to a total of 246 students (11-14 years old) in public and private schools in London and suburbs.

114 students completed Test A, and 132 students completed Test B.

The responses were coded based on the modified version of the Erduran et al. (2004) framework, and then the data were analyzed in SPSS.

Cronbach’s alpha, which is a measure of the internal consistency of the items, was calculated for each of the tests.
Results

- 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:

\[
\begin{array}{cccc}
\text{Q1} & \text{Q2} & \text{Q3} & \text{Q4} \\
\hline
\text{Q1} & 1 & 0.095 & 0.187 & 0.129 \\
\text{Q2} & 0.095 & 1 & 0.459 & 0.317 \\
\text{Q3} & 0.187 & 0.459 & 1 & 0.474 \\
\text{Q4} & 0.129 & 0.317 & 0.474 & 1 \\
\end{array}
\]

\[
\begin{array}{cccc}
\text{Q1} & \text{Q2} & \text{Q3} & \text{Q4} \\
\hline
\text{Q1} & 1 & -0.002 & 0.14 & 0.126 \\
\text{Q2} & -0.002 & 1 & 0.483 & 0.333 \\
\text{Q3} & 0.14 & 0.483 & 1 & 0.524 \\
\text{Q4} & 0.126 & 0.333 & 0.524 & 1 \\
\end{array}
\]

Thursday, September 1, 2011
Results

• Therefore the scores for Q1 were removed from both tests and Cronbach’s alpha was calculated again as .0674 for Test A and 0.705 for Test B.

<table>
<thead>
<tr>
<th></th>
<th>Cronbach's Alpha if Item Deleted for Test A</th>
<th>Cronbach's Alpha if Item Deleted for Test B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>0.674</td>
<td>0.705</td>
</tr>
<tr>
<td>Q2</td>
<td>0.508</td>
<td>0.532</td>
</tr>
<tr>
<td>Q3</td>
<td>0.394</td>
<td>0.361</td>
</tr>
<tr>
<td>Q4</td>
<td>0.489</td>
<td>0.478</td>
</tr>
</tbody>
</table>

• The internal consistency is lower when the scores any of the open-ended questions (Q3 and Q4) were removed from the test.
Conclusions

• Test A and Test B are reliable for evaluating written argumentation when the first item (Question 1) is removed.

• 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.

• 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.

• Argumentation might be content specific - for the same structure of the two tests, Test A had a lower Cronbach’s alpha, and that was higher when Question 1 (knowledge of electric circuits) was removed (also see Means & Voss, 1996).
Implications

- **Implications for research:**
  
  - 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.

- **Implications for practice:** finding ways to support students when deciding upon convincing arguments, since this is a skill useful in their everyday life, and appears to be more difficult to achieve.
Paper and presentation available at:

http://www.mariaevagorou.weebly.com
or
evagorou.m@unic.ac.cy