International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Student performance on argumentation task in the Swedish National Assessment in science

Anders Jönsson

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2016
http://dx.doi.org/10.1080/09500693.2016.1218567

Routledge
Taylor & Francis Group

Student performance on argumentation task in the Swedish National Assessment in science

Anders Jönsson

Department of Science, Kristianstad University, Kristianstad, Sweden

ABSTRACT
The aim of this study is to investigate the influence of content knowledge on students' socio-scientific argumentation in the Swedish National Assessment in biology, chemistry and physics for 12-year-olds. In Sweden, the assessment of socio-scientific argumentation has been a major part of the National Assessment during three consecutive years and this study utilizes data on student performance to investigate (a) the relationship between tasks primarily addressing argumentation and tasks addressing primarily content knowledge as well as (b) students' performance on argumentation tasks, which differ in relation to content, subject, aspect of argumentation and assessment criteria. Findings suggest a strong and positive relationship between content knowledge and students' performance on argumentation tasks. The analysis also provides some hypotheses about the task difficulty of argumentation tasks that may be pursued in future investigations.

ARTICLE HISTORY
Received 3 September 2015
Accepted 26 July 2016

KEYWORDS
Argumentation; assessment; scientific literacy

Introduction

A major aim of contemporary science curricula is to develop students' scientific literacy, which includes socio-scientific reasoning and decision-making. Such reasoning and decision-making involves negotiating science knowledge and evidence alongside other forms of knowledge and beliefs, such as societal, economic and political knowledge, and values. Thus, an important focus of school science has become to aid students in developing interest and engagement in science discourse, being able to identify questions and draw evidence-based conclusions, being skeptical and questioning, and being able to make informed decisions about the environment and socio-scientific issues (SSIs) (Morin, Simonneaux, Simonneaux, Tytler, & Barraza, 2014).

Current research, however, indicates that opportunities for students to participate in such authentic scientific argumentation inside the classroom are rare, not least because many teachers seem to lack the pedagogical knowledge necessary to design lessons that may foster student engagement in scientific argumentation and teachers also seem to have limited resources to assist the students in this area (Sampson & Blanchard, 2012; Simon, Erduran, & Osborne, 2006). There is also evidence to support the claim that

CONTACT Anders Jönsson anders.jonsson@hkr.se Department of Science, Kristianstad University, SE-291 88 Kristianstad, Sweden

reasoning and argumentation is difficult for many students and that they often base their decisions on values or personal experiences rather than on scientific knowledge (e.g. Chang & Chiu, 2008). Obviously, this is an area in great need of reform, so that all students are given the opportunity to develop the necessary knowledge and skills for active participation as citizens in democratic societies.

In Sweden, teaching towards scientific literacy has been emphasized by including the assessment of students' skills in socio-scientific reasoning and decision-making as a major part of the National Assessment in biology, chemistry and physics. One of the challenges when assessing students' socio-scientific argumentation is the intimate relationship between content knowledge in science and the structural knowledge of argumentation (i.e. argumentation skills, such as how to formulate arguments, justify claims, etc.). In order to participate in discussions and make informed decisions, students need relevant content knowledge, but also knowledge about how to formulate a valid argument. Consequently, when interpreting scores, it is difficult to know whether poor performance is a result of a lack of content or of structural knowledge, or even deficits in other aspects of students' knowledge, such as reading comprehension.

A question often raised is whether the opposite may also be true, i.e., that high performance on argumentation tasks can be attained without relevant content knowledge. The findings reported by Lewis and Leach (2006) seem to contradict such a possibility. According to these authors, the amount of relevant knowledge students need to engage in discussions and to justify their views may be relatively small, but when lacking, students' responses in their study were generally 'ill-considered and emotional' (p. 1275). But what about knowledge above that basic level: does more content knowledge improve students' argumentation skills?

The performance of science and non-science majors on argumentation tasks has been compared in a number of studies, assuming greater content knowledge in science on behalf of the former group. But while the science majors may outperform the non-science majors in one study (Chang & Chiu, 2008), the opposite may be true in another (Christenson, Chang Rundgren, & Zeidler, 2014). It has also been noted that the kind of scientific knowledge that students use for argumentation may differ for different topics, such as global warming, genetically modified organisms (GMOs), nuclear power or consumption (Christenson, Chang Rundgren, & Höglund, 2012). Taken together, the inconclusive results from current research make it difficult to draw any firm conclusions about the influence of content knowledge on students' argumentation skills.

This study therefore aims to contribute to the research in this area by investigating the relationship between students' structural knowledge of argumentation and their content knowledge in science when solving argumentation tasks in the Swedish National Assessment in biology, chemistry and physics for 12-year-olds.

Background

Argumentation skills

Argumentation may be seen as a form of discourse, which is based on the process of reasoning. This reasoning may be either formal, where all premises are fixed and the conclusion follows logically from the propositions (as in syllogisms[1]), or informal. Informal reasoning is typical for SSIs, such as how to deal with the conservation of endangered species, and involves using information from multiple sources in order to arrive at one out of several possible conclusions. Which conclusions are drawn, may be influenced by both cognitive and affective features. Needless to say, such conclusions are more tentative in nature as compared to conclusions from formal reasoning (Chang & Chiu, 2008).

If using informal reasoning as a base for argumentation, where several (and sometimes conflicting) conclusions may be drawn from the same information, it is evident that argumentation skills need to include more than pure logic. Toulmin (1958) is often claimed to be the first to present a model of such argumentation, and this model has been extensively used in the field of science education research. As pointed out by Nielsen (2013), however, several scholars agree that the Toulmin model may not provide proper guidance when, for instance, empirically trying to distinguish between the different elements in the model. He also suggests that 'interesting discursive aspects may become lost in translation when the Toulmin model is used to reduce the dialogic nature of students' argumentation into passive patterns of arguments' (pp. 371–372). As a consequence, several attempts have been made to find other models for analyzing students' argumentation, such as the 'Lakatos' Scientific Research Programmes' (Chang & Chiu, 2008), the SEE-SEP model ('SEE-SEP' refers to the subject areas Sociology/culture, Environment, Economy, Science, Ethics/morality, and Policy) (Chang Rundgren & Rundgren, 2010), the S^3R Model (Morin et al., 2014), and the 'normative pragmatics analysis' (Nielsen, 2012). Most of these models have a number of features in common, but there are also some notable differences. For instance, in the model used by Nielsen (2012) emphasis is placed on the dialogical features of argumentation and analyzing the argumentation 'in situ', rather than analyzing individual utterances in isolation from the specific context or without connection to the preceding dialogue. The SEE-SEP model includes an elaborated framework for analyzing the grounds for different arguments (i.e. whether arguments are based on personal experiences, values, or 'knowledge', and in which field of knowledge).

In the research presented here, which draws on data from standardized tests, the operationalization of argumentation skills most closely resembles a framework presented by Christenson and Chang Rundgren (2014). This framework is a fusion of the SEE-SEP model and the 'Lakatos model' by Chang and Chiu (2008), but it has been simplified so that it may be used in assessment situations, where students argue about SSI task. The starting point of the analysis or assessment is a claim or decision made by the student, for which there may be a justification, including both arguments in favor of the claim and against the claim. The arguments in favor and/or against the claim/decision may then be based on either 'knowledge' or values, and this knowledge may differ in terms of how relevant and scientifically correct it is. Important to notice is that by evaluating knowledge in terms of relevancy and correctness, as well as arguments in favor and/or against the claim/decision, this model of analysis/assessment takes both the structure of the argumentation and the content into account. This feature distinguishes the model from many other frameworks that are either 'structure oriented' or 'content oriented' (Christenson, 2015). There is also a differentiation between two sub-categories of value-based argument, where the main distinction seems to be whether the students are able to explicate the grounds for their arguments or not.

Research about students' socio-scientific argumentation

Student performance on argumentation tasks has been investigated in a number of studies. For instance, in a study by Christenson et al. (2012) the informal argumentation of 80 upper-secondary students was analyzed. The main aims of this research were to investigate students' use of supporting reasons and to what extent the students used scientific knowledge in their arguments. Data consisted of students' written argumentation in response to four SSI scenarios. Results from this study show that the students used values for supporting their argument to a greater extent, as compared to scientific knowledge. This was true for all four SSI topics. However, what kind of scientific knowledge they used differed among the topics. For instance, knowledge from the natural sciences were particularly prevalent in scenarios concerning GMOs and nuclear power.

That scientific knowledge may not be the primary choice for students when formulating arguments is also shown by Chang and Chiu (2008). In their study, 70 Taiwanese undergraduates participated; 40 science majors from a medical school and 30 non-science majors from a national art university. In this study too, data consisted of students' written argumentation in response to four SSI scenarios and results show that personal experiences were the main resource for both groups when formulating arguments. Science majors performed better, as compared to the non-science majors, which, according to the authors, suggest that background knowledge may be related to the ability of informal reasoning and argumentation. Still, students in both groups had difficulties providing supporting reasons and presenting counterarguments, as well as evaluating others' arguments. In fact, for some SSI topics, only a handful of students were able to achieve the complete set of indicators for informal argumentation used in this study.

Some of the results from the above-mentioned studies, for instance, that students' arguments are primarily based on value judgments or personal experiences instead of scientific evidence, are typical for research in this area (e.g. Grace & Ratcliffe, 2002; Jiménez-Aleixandre & Pereiro-Muñoz, 2002; Sadler & Zeidler, 2004). It is also typical that students are generally not particularly skilled at providing arguments and especially poor at providing counterarguments (e.g. Perkins, 1985).

The influence of content/domain knowledge on students' argumentation skills is, however, a contested issue. As mentioned above, Chang and Chiu (2008) claim that the content/domain knowledge of science majors influenced student performance in their study. Still, Christenson et al. (2014) found that social-science majors were able to provide more justifications for their arguments, as compared to natural science majors. Furthermore, in the study by Perkins (1985), where arguments from 320 participants at various positions in the school system (as well as non-students) were analyzed, the differences in performance were small, even when comparing students in high school and graduate school. Means and Voss (1996), who studied the performance of students in grades 5, 7, 9, and 11, found that performance did increase with grade level, but that students' 'ability level' (i.e. IQ) generally had a much stronger effect.

Another contested issue is how context-dependent reasoning is regarding SSIs. As was mentioned above, in relation to the study by Christenson et al. (2012), the use of knowledge from the natural sciences differed among topics, but were particularly prevalent in scenarios concerning GMOs and nuclear power. Other studies present similar results (e.g. Christenson et al., 2014), but offer no explanations as to why some topics may be easier or more difficult than others.

There is also a question about whether or not developmental differences affect decision-making (Sadler & Zeidler, 2004). This question is important, since not only the SSI scenarios may vary in different studies, but the age of the students as well. For instance, a number of studies analyze the reasoning of 15- to 17-year-old students and college students, while research about students in compulsory school seems to be less prevalent.

Taken together, there is evidence to support the claim that reasoning and argumentation is difficult for many students and that they often base their decisions on values or personal experiences rather than on scientific knowledge. Less is known, however, about the influence of content/domain knowledge on students' argumentation, about student performance in relation to different topics and about the argumentation skills of younger students. This study therefore aims to contribute to the research in this area by investigating the relationship between students' structural knowledge of argumentation and their content knowledge in science when solving argumentation tasks in the Swedish National Assessment in biology, chemistry and physics for 12-year-olds. Specifically, this article aims to answer the following questions:

(1) What is the relationship between student performance on items addressing argumentation skills and items addressing content knowledge?
(2) How do students' performance on argumentation tasks differ in relation to different content categories, subjects, aspects of argumentation and assessment criteria?

The Swedish National Assessment in science

The Swedish National Assessment in science for 12-year-olds will be described in some detail, since these tests may differ in several respects from other science tests around the world. It should also be noted that although national testing has a long history in Sweden, these particular assessments, targeting this age group and these subjects, have been provided only during three consecutive years. In the first year (2013), the tests were trialed nationwide with the whole cohort of 12-year-olds (approximately 100,000 students); during the second year (2014), all students in the country performed the tests and during the third year (2015), the tests were made voluntary for the schools. As of 2016, no more national tests in science are planned for this age group.

First of all, there are three tests, one in each subject (biology, chemistry and physics), but each individual student only do one of them. Some weeks before the tests are scheduled, each school is randomly assigned one of the tests. Each test, in turn, consists of three parts, for which one hour of testing time is allocated and they focus on: Argumentation skills (Part A), Investigations (Part B) and Content knowledge (Part C). This structure is similar for all three subjects.

Part A, which targets students' argumentation skill, and therefore is of primary interest here, consists of three tasks, each focusing on a particular subskill or 'aspect of argumentation'. First, there is one task addressing students' skills in using scientific knowledge in discussions about SSIs. For instance, in the year 2013 biology test, the context was children discussing the advantages and disadvantages of taking part in a soccer competition if you have a cold and are not feeling well. In the task, three fictional characters give their opinions in speech bubbles and the students are expected to suggest how to continue the conversation.

The second task in the Part A tests is about choosing sources. For instance, in the year 2013 physics test, the context was a group of students trying to find information about how clouds are formed. In the task, the students are presented with an email from a fictional student, suggesting a number of different sources to use in a presentation about cloud formation. The students are expected to choose which of these sources are relevant for this purpose and to justify their choices.

The third, and last, task in the Part A tests, is about using scientific knowledge in order to produce texts, figures, tables, etc. for different audiences and purposes. For instance, in the year 2013 chemistry test, the context was providing reasons for sorting waste at a sports club. In the task, the students watched a movie about sorting and recycling household waste and they were asked to write a letter to the sports club, arguing for the club to start sorting the waste (Figure 1).

As can be seen from the descriptions and examples of tasks in the Part A test, these tasks address different aspects of students' argumentation skills, which are similar across subjects and across years. The tasks also cover different topics and scientific knowledge, where the topics may be characterized as being more to the 'socio end' (for example, discussions about participation in a soccer tournament or the sorting of household waste) or more to the 'science end' (for example, cloud formation) on a 'SSI gradient' (cf. Vision I and II in Roberts, 2007).

While Part A addresses a fairly well-defined set of subskills, Part C aims to assess a broad range of factual and conceptual knowledge, such as identifying predatory animals in the biology test or describing the properties of magnets in the physics test.[2] Furthermore, while the tasks in the Part A tests are all open-ended, the tasks in Part C are a mix of both selected- and constructed-response items. Since all parts of tests are allocated

In school, you have been talking about why it is important to sort garbage. But at the local sports club all waste is thrown in the same can. You do not think that this is a good idea and decide to write a letter to the club. The letter must explain why the club should start sorting their garbage. Use information from the text below and from the movie you are about to watch. You will watch the movie twice.

Figure 1. Introduction to a task in chemistry (author's translation).

Notes: The students are asked to write a letter to a sports club, arguing for the club to start sorting their waste. As mentioned in the task, the students are provided with some additional information about sorting and recycling household waste, both from a short text and from a movie. The task also includes a 'Remember this … ' box, where students are reminded to keep to the subject and to write as many arguments as they can. The tests from 2013 can be downloaded (in Swedish) from http://npno6.se/page/2013-ars-prov.php.

one hour of testing time, there are roughly six times as many items in Part C as compared to Part A.

A particular feature of the Swedish National Assessment tests is that they are assessed by the teachers themselves. To that end, each test is accompanied by a comprehensive set of instructions for how to assess each of the items in the tests. There are also instructions about how to summarize the item scores from all parts of the tests (A, B and C), in order to generate a test score (or more truthfully a 'test grade' from F to A, where F is fail, E is pass and A is the highest passing grade) for the test as a whole. To aid the teachers in calculating the test score, the teachers may report students' item scores through a website, where the software automatically calculates the test scores for all students and also provides some summary statistics for the class. In addition, the teacher may compare the performance of her/his own students, with all other classes in the country whose results have been reported through the website.

Finally, it should be recognized that the tests are developed according to a rigorous methodology, involving peer review by researchers in the field of science education, review panels of in-service teachers and trials with hundreds of students from different geographical regions. Furthermore, both quantitative and qualitative analyses of item and test data are performed as part of the test development, before as well as after the tests have been administrated, including differential-item-functioning analyses and estimations of interrater agreement. The development of the Part C tests, in particular, has made use of modern test theory (Item Response Theory) as a tool for item analysis and test equating.

Data and analysis

The data for this study consist of the results (i.e. item scores) reported by teachers through the website described in the previous section. The number of teachers who chose to report their results in this way differs somewhat between years and subjects, but in 2013–2014 the results from roughly one third of the entire cohort of students were reported through the website (Table 1). In 2015, due to the decision to make the tests voluntary for the schools, the number of results reported is considerably lower (approximately 3,400 students per subject).

Students' results are reported as 'indicators' of the letter grades F (fail), E, C and A.[3] These letters are transformed to numbers (0–3) before analyzing the data, but the numbers are still treated as ordinal data.

In order to investigate the relationship between student performance on items addressing argumentation skills and items addressing content knowledge (RQ 1), correlation analyses using Spearman's rho have been performed between Part A (Argumentation skills) and Part C (Content knowledge) for each of the tests, as well as between the items in Part A. A limitation of this analysis is that the Parts A and C of each test are

Table 1. Number of students' results reported through the website.

	2013	2014	2015
Biology	9,415	8,445	3,298
Chemistry	9,667	7,731	3,399
Physics	10,849	6,962	3,576

not linked, for instance, by assessing the same content knowledge in Part C that the students use in Part A to support their reasoning. Instead, Part C will provide a measure of students' general knowledge in science and the analysis will show whether this measure correlates with students' argumentation skills.

In order to investigate students' performance on argumentation tasks in relation to different content categories, score frequencies (i.e. students' scores divided by the maximum score) were calculated for each individual item, providing an indication of difficulty for the items. Each item was also categorized according to both the aspects of argumentation and content addressed, as well as assessment criteria.

For aspects of argumentation, the three subskills outlined above were used: (1) Using scientific knowledge in discussions about SSIs, (2) choosing sources and (3) using scientific knowledge in order to produce texts, figures, tables, etc. for different audiences and purposes. Since these aspects of argumentation are generic for all subjects and content categories, they represent the 'structure-oriented' facet of student performance.[4]

For content, three broad categories were used for each of the subjects: (A) Science in Society (Vision II), (B) Within Science (Vision I) and (C) History and Worldview of Science (Table 2). Regarding the first two content categories, they represent a gradient rather than two distinct categories. Since all items address SSIs in some respect, there are no 'pure' Vision I items. But for analytical purposes, all items have been categorized as representing either the 'society' or 'science end' of the spectrum. Furthermore, the specific focus on SSIs means that the first category occurs more frequently in the material. The third category, History and Worldview of Science, on the other hand, is a more distinct (but infrequent) category as it addresses historical and cultural perspectives on science.

The assessment criteria used to determine the quality of students' responses were categorized according to the amount of content knowledge required. Basically, there are three different categories. In some items, all necessary information (such as facts and concepts) are provided and the students are expected to *use* this information in some way. For such items, the content knowledge requirements are low. In other items, students are expected to provide a breadth of *different* advantages and/or disadvantages, reasons, etc. For such items, some content knowledge is required, but since it is the number of advantages (or

Table 2. The three content categories with examples.

Category	Examples of content (from the curriculum)
Science in Society (17 items)	Human dependence and impact on nature; sustainable development; ecosystem services (biology)
	Conversion of materials through cultivation of raw materials to products; waste; recycling (chemistry)
	Different types of energy sources and their impact on the environment; the use of energy in society (physics)
Within Science (9 items)	Development of life and adaptation of organisms to different habitats (biology)
	Particle model to describe and explain the structure of matter. Movements of particles as an explanation for transitions between solids, liquids and gases (chemistry)
	Meteorological phenomena and their causes (physics)
History and Worldview of Science (4 items)	Historical discoveries and their importance for people's views on nature (all subjects)
	Descriptions and explanations of nature in earlier science, as well as in fiction, myths and art (all subjects)

Note: Three items have been categorized as addressing content from two different categories, therefore the total number of items in the table is 30 instead of 27 (3×9).

The response is regarded as an indication of level	
E	if **one** of the following applies: • **At least two** arguments are provided, **both** in favor **and** against, but none of them are scientific arguments. • **At least two** arguments are provided, where **one** of the arguments is a scientific argument (either in favor or against).
C	if **all** of the following applies: • **At least three** arguments are provided. • Arguments **both** in favor **and** against are provided. • **At least one** of the arguments is a scientific argument.
A	if **all** of the following applies: • **At least four** arguments are provided. • Arguments **both** in favor **and** against are provided. • **At least two** of the arguments are scientific arguments.

Figure 2. An example of a scoring rubric for a task in physics (author's translation).

corresponding things) that counts, the requirements are lower than items asking students to justify their answers. High content items are therefore items where students are expected to use content knowledge (not provided in the task) to explain or justify their answers.

An illustration of a scoring rubric is shown in Figure 2 and examples of student responses in Figure 3. The rubric is taken from a task in physics asking the students to provide arguments both in favor and against an increase in the use of wind power. In this task, it is the number of arguments (i.e. breadth) that counts, not how well the students are able to use their content knowledge to justify their arguments. The student responses are taken from a task in chemistry, asking the students to add food to a plate with some hard bread, vegetables and a chocolate biscuit, so that the lunch becomes

Add meat because it's good for us. But not too much. (F)
We need proteins, so I'd add for example fish, meat, beans and so on. We also need vitamins, minerals that strengthen the body. There should be food for those who're sensitive to for example milk, bread and so on. Or those who don't eat meat, fish and so on. (E)
It's good that she has a lot of vegetables because they contain a lot of vitamins. It's not so good that she doesn't have any protein at all (add meat, fish, eggs and so on) because it builds up the body. It's not so good that she doesn't have so much carbohydrates (the hard bread) she needs to add more pasta, rice or potatoes in order to make it through the day. You could eat chocolate biscuits now and then but if you eat them too often they can give you tooth decay. (A)

Figure 3. Examples of student responses at different levels (F, E and A) for a task in chemistry (author's translation).

Table 3. Categorized items ranked according to difficulty.

Easier items	More difficult items
57.0 Biology 2bM	45.3* Physics 1aH
58.3 Physics 2cM	45.4* Biology 1aH
58.7 Chemistry 1aM	46.0 Chemistry 1aH
58.7 Physics 1aM	46.0* Chemistry 2bcM
58.8 Physics 2bM	48.9 Chemistry 3aH
59.8 Biology 3aM	49.3 Biology 2aH
60.9 Physics 3bL	50.2 Physics 2bM
61.0 Biology 2abM	51.4 Biology 3cH
61.3 Chemistry 1aM	52.2 Biology 3bH
63.9 Chemistry 2aL	55.4 Physics 3aM
65.5 Physics 3aL	55.7 Chemistry 3aM
65.8 Chemistry 3bL	56.5 Biology 1aM

*Anchor items.

more nutritious. The students are also asked to explain why they think that the lunch would be more nutritious with their additions.

After categorizing the items, they were ranked according to difficulty, representing a range of approximately 22 percentage units from 44.0 (most difficult item) to 65.8% of the maximum score (least difficult item). This range has been split into two equal sections around the median, called easier and more difficult items in Table 3. Besides the estimations of difficulty, subjects and codes for structure- and content-oriented facets, as well as the categorization of content requirements in the assessment criteria (L, M and H for low, medium and high, respectively), are also presented in the table, so that the distribution of these characteristics are placed along a gradient of item difficulty.

A particular remark has to be made regarding the anchor items. For purposes of test equating, Part A in the 2014 and 2015 tests had one item in common. These are called 'anchor items', because they link the tests together. In Table 3, the score frequencies from these items are means (marked by *), so that they are not counted twice. Since the results from the anchor items are very similar between the years (for an example, see Table 4), they also provide support for the possibility to compare results from the tests across cohorts.

Since most statistical procedures are inappropriate for analyzing the limited number of items in Table 3, only non-parametric statistics have been used. The analysis has aimed at revealing any tendencies in the relationship between item difficulty on the one hand and aspects of argumentation, subjects and content on the other. Needless to say, the findings from such an analysis is quite crude and the findings should therefore not be generalized beyond the context of this study. Still, the analysis provides a useful tool for identifying possible tendencies in relation to the items in these particular tests.

Table 4. Score distributions (in percent) for the anchor item in the physics tests.

	Physics	
Score	2014 ($n = 7{,}731$)	2015 ($n = 3{,}528$)
0	17.3	17.2
1	36.7	36.2
2	32.5	34.0
3	13.5	12.6

Table 5. Correlations between Part A and Part C in the National Assessment.

Biology			Chemistry			Physics		
2013	2014	2015	2013	2014	2015	2013	2014	2015
.590	.622	.600	.515	.575	.506	.477	.455	.505

Note: All correlations are significant at the .01 level.

Results

The first research question in this study addresses the relationship between student performance on test items targeting content knowledge in science and students' argumentation skills. As can be seen in Table 5, the correlation between Part A (argumentation skills) and Part C (content knowledge) shows a moderate-to-strong positive relationship. This means that those students who perform well on the test for argumentation skills also perform well on the test for content knowledge. The correlation is generally stronger for biology and weaker for physics. In particular, the 2014 test shows a weaker correlation for physics. This means that the relationship between student performance on Part A and C is not as strong on the test in physics as for the other two subjects, except for the 2015 test when the correlation is similar for both chemistry and physics.

Table 6 displays correlations between the items in the Part A tests for each of the subjects, respectively. As can be seen in the table, all correlations are weak to medium. As compared to Table 5, all correlations between individual items in Part A are weaker than between Part A and Part C. This means that students performing well on one item in Part A do not necessarily perform well on other items in the same part of the test, but that students who perform well on Part A as a whole also tend to perform well on Part C.

The second research question in this study addresses students' performance on argumentation tasks in relation to different content categories, subjects, aspects of argumentation and content requirements. In this case, analyses showed that the three subjects are very similar in difficulty, but also indicated some tendencies, such as category B (Within Science) being associated with the easier items and category C (History and Worldview of Science) with the more difficult ones. For aspects of argumentation, there also seems to be a tendency for the first subskill to be associated with more difficult items and the third with easier items. However, none of these tendencies showed any statistical significance.

The content requirements in the assessment criteria, on the other hand, were strongly correlated with item difficulty ($r = .778$; $p < .001$), which can also be clearly seen by ocular inspection of Table 3.

These findings suggest that items requiring students to use content knowledge in science are more difficult for students, more or less regardless of which content is

Table 6. Correlations between items in Part A.

	2013			2014			2015		
	1*2	1*3	2*3	1*2	1*3	2*3	1*2	1*3	2*3
Biology	.365	.373	.312	.415	.423	.375	.302	.252	.218
Chemistry	.409	.337	.333	.372	.384	.397	.349	.323	.369
Physics	.377	.364	.347	.333	.342	.390	.354	.344	.379

addressed. However, items addressing the history and worldview of science may be slightly more difficult for students, as compared to the other content categories, and the skill of using scientific knowledge in discussions about SSIs may also be slightly more difficult for students, as compared to the other aspects of argumentation. Furthermore, the findings suggest that items addressing 'within science' knowledge may be easier for students, as compared to the other content categories and the skill of using scientific knowledge in order to produce texts, figures, tables, etc. for different audiences and purposes may be easier for students, as compared to the other aspects of argumentation. Due to the limited number of items, it has not been possible to statistically investigate possible interactions between these factors, but since low-content requirements are overrepresented among items addressing the skill of using scientific knowledge in order to produce texts, figures, tables, etc., it is likely that such interactions do exist.

Discussion

The aim of this study was to investigate the relationship between students' argumentation skills and their content knowledge in science in relation to the Swedish National Assessment in biology, chemistry and physics for 12-year-olds. This was done by analyzing item scores from a total of 63,342 students from the tests in 2013–2015.

The relationship between students' performance on items addressing argumentation skills and items addressing content knowledge was investigated using a correlation analysis. Findings show that student performance on items addressing argumentation skills and items addressing content knowledge was positively correlated. Furthermore, student performance on items addressing different aspects of argumentation was also positively correlated, but only weakly. This means that students who perform well on Part A as a whole also tend to perform well on Part C, but performing well on one item in the Part A does not necessarily predict success on other items in the same part of the test. These findings suggest that students with more solid content knowledge in science succeed on both parts of the tests. This study therefore adds to the assertion made by Chang and Chiu (2008) that content knowledge, and not only structural knowledge, has a positive influence on student performance on argumentation tasks.

That students' performance on items addressing different aspects of argumentation was only weakly correlated suggests that the three items in Part A address related, but slightly different, constructs. Since the items have been designed to address different aspects of argumentation, this finding adds to the validity of the test.

The second research question addressed students' performance on argumentation tasks in relation to different content categories, subjects and aspects of argumentation. Due to the limited number of items, this analysis only involved the categorization and rank order of the items. Findings from this analysis further support the importance of content knowledge for successful argumentation. While the score frequency was high for items requiring less content knowledge, items requiring content knowledge were generally more difficult for the students. Since these items covered a range of different content, it seems like the use of scientific knowledge in argumentation tasks for explanations and justifying choices is difficult, more or less regardless of the specific content.

There may be interactions, however, with certain content areas, which are easier or more difficult in relation to each other. For instance, items addressing the history and worldview of science seem to be slightly more difficult for students, while items addressing 'within science' knowledge seem to be a bit easier. That items addressing the history and worldview of science might be more difficult for students will probably not come as a surprise to anyone, since this topic is often not a major constituent of traditional science teaching (e.g. Hodson, 2009). Items addressing 'within science' knowledge, on the other hand, may be easier for students, just because it is a typical part of traditional science teaching.

There are also possible interactions in relation to different aspects of argumentation, where the skill of using scientific knowledge in discussions about SSIs seems be more difficult for students, as compared to the other subskills. Besides the rank-order analysis above, this finding is further corroborated by student performance in relation to two items in the 2015 tests. These two items, one in chemistry and one in physics, addressed the same topic and provided the same background information. But while one item asked the students to use scientific knowledge in a discussion context (subskill 1), the other asked the students to use it in order to prepare for a presentation (subskill 3). The former turned out to be one of the most difficult items, while the opposite was true for the latter. In line with the analysis of the assessment criteria, what seems to make the item addressing subskill 1 difficult is that the criteria require students to make choices, which they have to justify and, for higher scores, back up with relevant knowledge. This could be compared to the item addressing subskill 3, where students are expected to transform the available information from one format into another, but not to draw upon their own content knowledge.

Conclusions

Can students succeed in formulating valid arguments on argumentation tasks without adequate knowledge in science, using only structural knowledge about argumentation? The findings from this study suggest that this is not the case. Students with more solid content knowledge perform well on both argumentation tasks and items addressing factual and conceptual knowledge. The relative difficulty of the test items also increases with higher demands for content knowledge, almost without regard to the specific content addressed. Taken together, content knowledge seems to play an important part in formulating valid arguments and structural knowledge cannot sufficiently compensate for a lack of adequate content knowledge.

The results also suggest that items involving the history and worldview of science may be more difficult for students, possibly because this topic is not a frequent part of science teaching, while more traditional 'within science' items are relatively easier.

Limitations and implications for future research

One of the important assets of the research reported here is the unique data material. These data are based on items that have been carefully developed, trialed and evaluated with the help of hundreds of students and their teachers across Sweden. Furthermore, a large number of item scores have been reported through the website, allowing for estimations of item difficulty and other test and item properties. However, there are also severe limitations, not least the relatively small number of different items. This has restricted the analyses to quite crude statistical measures and some of the findings are very tentative. This is especially true for analyzing the content of the items and the criteria for assessing student performance, where some of the categories include only a few items. Still, it is argued that some of the findings are more robust, such as the importance of content knowledge for successful performance on argumentation tasks, and that the tendencies observed may be used as a valuable input in future research. In particular, future research could establish whether there are indeed differences in difficulty related to the content categories and aspects of argumentation as identified here, but also investigate the reasons for these differences. While the former has to be addressed with quantitative data and large samples, the latter may be investigated qualitatively with interviews, think-aloud protocols and document analyses. Similar methods may also be used to further investigate how students make use of subject knowledge in argumentation tasks.

Implications for assessing and teaching argumentation skills in school

One of the apparent implications of this research is that formative assessment of students' argumentation skills would need to make use of analytical scoring/assessment, so that structural and content knowledge are assessed as separate (but interconnected) dimensions of student performance. If argumentation tasks are assessed holistically, possible strengths and weaknesses in either structural or content knowledge may not be identified, resulting in less specific feedback. To develop and test scoring rubrics from the framework proposed by Christenson and Chang Rundgren (2014) would be one way to support this analytical scoring/assessment of argumentation tasks and the formative assessment of students' argumentation skills.

Regarding the teaching of argumentation skills, previous research has shown that reasoning and argumentation is difficult for many students and that they often base their decisions on values or personal experiences rather than on scientific knowledge. This study further substantiates these findings, for instance, by showing that most items have a score frequency around 50% (i.e. on average, students are only awarded half of the maximum score) and none of the items exceed a score frequency of two thirds. Findings also suggest that a main factor affecting the difficulty is the content knowledge needed in order to provide justifications and backing them with relevant knowledge.

Since students' argumentation skills do not necessarily improve without instruction (Perkins, 1985), it would seem like an obvious recommendation to allocate more time for teaching these skills. There are some indications that the presence of items in the National Assessment addressing argumentation skills has indeed affected the teaching of science in Swedish schools (Borgström, 2015), but whether this effect will last when the tests are made voluntary is an open question.

Based on the results from the tests, recommendations could also be given to specifically focus on the areas where students perform less well, such as the history and worldview of science, but also to draw students' attention to the need for providing justifications and backing them with scientific knowledge (i.e. supporting their structural knowledge about argumentation). In sum, since neither generic argumentation skills nor content knowledge is sufficient for mastery of successful argumentation, an integration of both structural and content knowledge is needed when teaching argumentation in science to the decision-makers of the future.

Notes

1. The most famous syllogism is probably 'All men are mortal' (premise 1), 'Socrates is a man' (premise 2) ∴ 'Socrates is mortal' (conclusion).
2. Since Part B is not included in this particular study, it is not described in detail. The Part B tests address students' skills in planning, carrying out, documenting and evaluating scientific investigations.
3. In the Swedish grading system, there are only criteria (or so-called 'knowledge requirements') provided for the grades E, C and A. The grades D and B are awarded when all criteria are met for the grade below, but not all criteria for the next grade has been reached (for instance, a B is awarded when all criteria for C is met, but not all for A). Since there are no criteria for the grades D and B, there can be no indicators of these grades. F is awarded when all criteria for E are not met.
4. Even though only a single score is observable for each item, student performance is assumed to consist of different 'facets'. One of these facets represents the structural knowledge of argumentation and another students' content knowledge. There are also other facets, relating to things like reading comprehension and 'test wiseness' (i.e. skills used to improve test scores, which are not related to the construct being assessed).

Acknowledgements

The author would like to thank all students and teachers who have contributed to the development of the National Assessment items and the collection of data in this study. Information about the items is provided with the permission of the Swedish National Agency for Education.

Disclosure statement

No potential conflict of interest was reported by the author.

Note on contributor

Anders Jönsson is Professor of Education at Kristianstad University, Sweden. Together with colleagues at the University of Gothenburg and Malmö University, he has been responsible for the development of the National Assessments in biology, chemistry and physics for year 6 in compulsory school. Anders' main research interest is in classroom assessment.

References

Borgström, E. (2015, May 6). *De nationella provens effekter på skolan och lärares arbete i åk 6 och 9* [The effects of national testing on the work by schools and teachers in year 6 and 9]. Project presentation.

Chang, S.-N., & Chiu, M.-H. (2008). Lakatos' scientific research programmes as a framework for analysing informal argumentation about socioscientific issues. *International Journal of Science Education*, *30*(13), 1753–1773.

Chang Rundgren, S.-N., & Rundgren, C.-J. (2010). SEE-SEP: From a separate to a holistic view of socioscientific issues. *Asia-Pacific Forum on Science Learning and Teaching*, *11*(1).

Christenson, N. (2015). *Socioscientific argumentation: Aspects of content and structure* (Doctoral dissertation). Karlstad University, Sweden.

Christenson, N., & Chang Rundgren, S.-N. (2014). A framework for teachers' assessment of socio-scientific argumentation: An example using the GMO issue. *Journal of Biological Education*. doi:10.1080/00219266.2014.923486

Christenson, N., Chang Rundgren, S.-N., & Höglund, H.-O. (2012). Using the SEE-SEP model to analyze upper secondary students' use of supporting reasons in arguing socioscientific issues. *Journal of Science Education and Technology*, *21*(3), 342–352.

Christenson, N., Chang Rundgren, S.-N., & Zeidler, D. (2014). The relationship of discipline background to upper secondary students' argumentation on socioscientific issues. *Research in Science Education*, *44*(4), 581–601.

Grace, M. M., & Ratcliffe, M. (2002). The science and values that young people draw upon to make decisions about biological conservation issues. *International Journal of Science Education*, *24*(11), 1157–1169.

Hodson, D. (2009). *Teaching and learning about science. Language, theories, methods, history, traditions and values*. Rotterdam: Sense.

Jiménez-Aleixandre, M.-P., & Pereiro-Muñoz, C. (2002). Knowledge producers or knowledge consumers? Argumentation and decision making about environmental management. *International Journal of Science Education*, *24*(11), 1171–1190.

Lewis, J., & Leach, J. (2006). Discussion of socio-scientific issues: The role of science knowledge. *International Journal of Science Education*, *28*(11), 1267–1287.

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.

Morin, O., Simonneaux, L., Simonneaux, J., Tytler, R., & Barraza, L. (2014). Developing and using an S^3R model to analyze reasoning in web-based cross-national exchanges on sustainability. *Science Education*, *98*(3), 517–542.

Nielsen, J. A. (2012). Science in discussions: An analysis of the use of science content in socioscientific discussions. *Science Education*, *96*(3), 428–456.

Nielsen, J. A. (2013). Dialectical features of students' argumentation: A critical review of argumentation studies in science education. *Research in Science Education*, *43*, 371–393.

Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology*, *77*(5), 562–571.

Roberts, D. A. (2007). Scientific literacy/science literacy. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 729–780). Mahwah, NJ: Lawrence Erlbaum.

Sadler, T. D., & Zeidler, D. L. (2004). The morality of socioscientific issues: Construal and resolution of genetic engineering dilemmas. *Science Education*, *88*(1), 4–27.

Sampson, V., & Blanchard, M. R. (2012). Science teachers and scientific argumentation: Trends in views and practice. *Journal of Research in Science Teaching*, *49*(9), 1122–1148.

Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. *International Journal of Science Education*, *28*(2–3), 235–260.

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