International Journal of Environmental & Science Education
Vol. 7, No. 4, October 2012, 581-598

Two approaches for analyzing students' competence of 'evaluation' in group discussions about climate change

Timo Feierabend • Marc Stuckey • Sarah Nienaber • Ingo Eilks

Received 3 May 2011; Accepted 23 February 2012

Up until now, very few models conceptualizing students' competence in evaluation, argumentation and discourse in the context of science education have been proposed. Most suggestions for analyzing this particular competence in students are normative and the empirical support for them remains weak. The problem becomes even more severe when such evaluations include ethical and societal perspectives as part of the analytical parameters. In support of this topic, this paper presents two approaches for handling students' evaluation capabilities in the context of multidimensional discussion situations. One approach focuses on the quality of learners' arguments concerning levels of justification; the second reflects upon the quality of pupils' complexity of argumentation. Both approaches were created using group discussion data collected for evaluation purposes. The data stems from a curriculum innovation project focusing on the teaching of climate change in four teaching domains: Biology, Chemistry, Physics and Politics. Participants from 20 different learning groups conducted semi-structured, pre- and post-group discussions on the issue of climate change. Analysis of a total of 76 group discussions showed positive potential in both evaluation grids on the topic.

Keywords: competence of evaluation, group discussion, assessment, climate change

Introduction

For the first time ever, Germany produced nation-wide standards for science education in the year 2004 (KMK, 2004). The standards were normatively outlined for all three science subjects in German secondary schools, namely Biology, Chemistry and Physics. In parallel, four different domains of competencies were defined. In addition to describing subject matter knowledge and content matter, three process-oriented domains were outlined: knowledge generation in the sciences, communication ability, and evaluation competency. All four domains were expanded upon on three levels. These levels can roughly be described using the labels: reproduction, simple application, and application as a transfer to more complex tasks (KMK, 2004).

The content domain was well-known to teachers and curriculum developers from previous science syllabi structures. In opposite, new process-oriented domains challenged both these groups and the related network of education assessors greatly. That is, the higher levels in the communication and evaluation domain proved to be a very uncommon element for many of them. The reason for this was the multi-dimensional view applied to both competency domains.

ISSN 1306-3065

http://www.ijese.com/

Understanding communication and evaluation encompasses aspects borrowed from the fields of valuation, argumentation and decision-making. These aspects should not only be applied to science itself, but also to its technological applications in society and include aspects of argumentation and decision-making dealing with the ethical and societal implications of issues taken from both science and technology (KMK, 2004).

This paper is grounded in one of the very few approaches in Germany which explicitly addresses socio-scientific reflections in the science classroom. In the year 2000, Eilks and co-workers (Eilks, 2000, 2002) began developing the socio-critical and problem-oriented model of science teaching. For a more thorough overview see Marks and Eilks (2009). This model structures lesson plans around actively learning about the societal implications of science. It also focuses on the interaction of science, technology and society and allows learners to directly experience the societal mechanisms for handling scientific issues in public debate.

In our case, the lesson plans dealt with the issue of climate change (Feierabend & Eilks, 2010). Duschl and Osborne (2002) suggested this topic as a promising field for interdisciplinary learning about the interplay of science with other domains, including its societal implications. They considered climate change to be a prototype field when it comes to learning about multi-facetted argumentation and decision-making.

In a Participatory Action Research project (Eilks & Ralle, 2002), four groups of science teachers accompanied by educators from the university began to structure domain-specific lesson plans on climate change. These four groups worked in the domains of Biology, Chemistry, Physics, and Politics teaching. Insights into the lesson plans are given in Feierabend and Eilks (2010). Reflection on the process of the participants' cooperation is discussed in Feierabend and Eilks (2011).

A large amount of data was collected as part of the process of curriculum innovation. The teachers' group discussions, student feedback questionnaires, and videotaped role-playing activities, which were embedded in all the lesson plans, provided insights into the feasibility of the teaching scenarios and gave initial indications on their effectiveness (Feierabend & Eilks, 2010; 2011). Also pre- and post-group discussions were conducted in the final phase of testing the lesson plans. These discussions focused the students' attention on the problem of climate change and asked the learners to discuss the transfer tasks of evaluation and decision-making within the framework of climate change.

This paper discusses one evaluation aspect taken from part of the group discussion data. The focus is the development of and reflection upon potential evaluation grids for measuring students' evaluation and communication competence in the means of students' abilities to discuss and argue about the socio-scientific issue of climate change. Two evaluation grids were developed and applied to different parts of the data. The first grid focuses on the quality of students' arguments in terms of levels of justification with regard to the content. The second grid differentiates the quality of argumentation with respect to its internal complexity. Both grids should be compared in order to pinpoint their potential for evaluating student discussions on socio-scientific issues with respect to the learners' evaluation competence. Thus the research questions of this study are:

- How can the evaluation and communication competence of students' be characterized in the means of students' abilities to discuss and argue about the socio-scientific issue of climate change?
- What level of evaluation and communication competence in argumentation on the socio-scientific issue of climate change do German student have at the end of lower secondary education?

- In which way can students' competency in argumentation be affected by a lesson plan about climate change including a role play exercise?

Theoretical Framework

Justifying Societal Relevant Science Education

Evaluating science and technology within societal applications has long been an accepted goal in any developed version of scientific literacy (Bybee, 1997). Although societal-oriented science teaching is still insufficiently developed and implemented in many countries, its importance has been widely acknowledged (Hofstein, Eilks, & Bybee, 2011). Many long traditions deal with the development of Science-Technology-Society (STS) type curricula (Holbrook, 1998; Holman, 1986; Sadler, 2004; Solomon & Aikenhead, 1994; Marks & Eilks, 2009). But this issue has also been dealt with under theoretical considerations in the fields of argumentation (Erduran & Jiminez-Aleixandre, 2007), discourse (Duschl & Osborne, 2002), and decision-making (Bell & Lederman, 2003).

About ten years ago, Duschl and Osborne (2002) described the entire framework as a field of study which still requires extensive research and curriculum development although there have been many approaches towards structured teaching of argumentation and decision-making, and the need for respective assessment is widely acknowledged (Erduran, Simon & Osborne, 2004; Jiménez-Aleixandre, Bugallo Rodríguez & Duschl, 2000). Dawson and Venville (2010) claimed that we are continually faced with problems and dilemmas requiring us to make decisions and choices, many of which specifically center around questions concerning science and technology. Therefore, school science education should contribute to producing students who are able to both participate in societal debates on socio-scientific issues and to consciously make balanced decisions on such issues. They need to understand not only argumentation beyond single context domains (as in science itself), but also learn about using argumentation across multidisciplinary, socio-scientific issues which transcend the boundaries of school science subjects, e.g. the causes and effects of global warming (Duschl & Osborne, 2002). In the end, science education should aim at helping pupils develop their decision-making skills by practicing different forms of argumentation (Dawson & Venville, 2010; Marks & Eilks, 2009; Hofstein et al., 2011).

Understanding Argumentation and Decision-Making

Concerning argumentation, Duschl and Osborne (2002) suggested clearly distinguishing between the process of argumentation and the use of an argument as such. They prefer using the word 'argumentation' to denote the process of constructing an argument. The word 'argument' is used to refer to the specific content of an argument. This distinction is in line with Dawson and Venville (2010), who referred to Kuhn (1991) when defining an argument as "an assertion with accompanying justification" (p. 12) and Means and Voss (1996) when describing an argument as "a conclusion supported by at least one reason" (p. 141).

On the other hand, argumentation (Dawson & Venville, 2010) is referred to in many papers in the sense found in the works of Toulmin (1958). For example, Erduran et al. (2004) state that scientists always use arguments to support the claims they favor through the use of warrants and backings and their relation to evidence. This is why students of science should learn about this process. This approach is closely connected to Duschl and Osborne (2002), who value the use of argumentation and discourse in science education, since they stimulate the process of reflection through which students can acquire conceptual understanding. In the end, the rationality of science is explained as the ability to construct persuasive and convincing arguments which relate explanatory theories to observational data and use them for sound and convincing argumentation.

Research on students' argumentation and decision-making skills is still a work in progress. Different studies have revealed interesting insights. Kolstø (2006) stated that there are always two sources for the emergence of any argument: a personal, ethical, societal side on the one hand, and science itself on the other. Flemming (1986) found that most students tend to prefer arguments stemming from their social world (Kolstø's first domain) when arguing about socio-scientific issues. Students only rarely use specific knowledge from the science domain (see Solomon (1992) for further information). Tytler, Duggan, and Gott (2001) or Yang and Anderson (2003) identified three types of evidence used by students: informal evidence, evidence from the wider framework of the socio-scientific issue, and scientific evidence. But even in this case the use of the scientific evidence was quite rare among most students.

Argumentational distinction is not generally based on the source of the information. But rather on its quality. Mitchell (1996) has suggested the following separation: regular and critical arguments. Regular arguments are rule-applying arguments. This style puts forward the application of theories without challenging the theories as such. In contrast, critical arguments try to challenge already existing theories. They are necessary for the refinement of theories in the sense that they constructively aid in the development and polishing of an existing theory.

In the area of decision-making, Sadler and Zeidler (2005) characterized students' decision-making skills in discussions about genetic engineering dilemmas with the aid of three modes: rationalistic, emotive, and intuitive informal reasoning. Bell and Lederman (2003) derived their view of decision-making skills from the Nature of Science perspective. They concluded from an experts' survey that it is necessary to re-examine the goals of any Nature of Science instruction and to add more value-based instruction, including paying attention to intellectual/moral development. These components were seen as necessary for learning about decision-making in science education which is connected to the real needs of future citizens. This is because social/political issues, ethical considerations, and personal values were also dominant in experts' decisions on socio-scientific issues, although the decision-makers stemmed from the science field.

Analyzing and Modeling Argumentation and Decision-Making

The field of analyzing argumentation seems to be much more difficult than simply categorizing single arguments. Characterizing argumentation demands analysis of entire chains of arguments, including their interrelatedness to one another. Models are also available in this area. Inch and Warnick (2002) described two types of conceptual models for analyzing argumentation. One type they named "standard models", which analyze how various claims are structured in order to create arguments, counterarguments, and rebuttals. They view Toulmin's (1958) models as being contrary to this type. Toulmin-based models seek to categorize supporting claims - including implicit ones - into grounds and warrants. But these models will not be discussed at length in this paper, since the main focus here is analyzing individual arguments and smaller pieces of argumentation rather than evaluating entire patterns of discourse and decision-making processes.

The theoretical field of structuring and analyzing students' competency in dealing with socio-scientific issues is a very broad one. There are many definitions and research studies available, which describe students' patterns for coping with socio-scientific debate and decision-making. Aikenhead (1985), Kortland (1996) and Ratcliffe (1996) all suggest the use of structured decision-making models based on evaluating the quality of students' decision-making skills.

Work on respective models for Germany started after the science education standards were put into practice in 2004. Based on the definitions built into the German standards, two models recently attempted to provide guidance for characterizing students' evaluation competence. The first model stems from the ESNaS-project (Evaluation of the Standards of Science Education for lower secondary schools) (Kauertz, Fischer, Mayer, Sumfleth, & Walpuski, 2010). It tries to differentiate all four areas of competence outlined by the Educational Standards using the two joint axes of cognition and complexity. Complexity in this model means the number of facts mentioned in one thought and the number of relationships existing between the different facts. A second approach was developed by Eggert and Bögeholz (2006), which characterizes students' evaluation competence within the framework of Education for Sustainable Development (ESD). In this model, evaluation competence is divided into four sub-domains, which each contaning four levels. The domains are: A) generation and reflection of subject matter information, B) valuing, decision-making and reflection, C) knowing and understanding about values and norms, and D) knowing and understanding sustainable development. All dimensions are superimposed onto four levels, which can be described in general as those containing: I) intuitive reason, II) poorly justified and unconnected arguments, III) three or more criteria which are related to and partially compensated by each other, and IV) at least three criteria, which compensate for and reflect upon the limitations of the stated decision. Nevertheless, both models must still be viewed as works in progress when it comes to their state of growth and testing. Broad application and final testing in the competency domain of evaluation are still underway.

Data background and sample

Within the project "The Climate Change before the Court" (Eilks et al., 2011), which was funded by the German Environmental Fund (Deutsche Bundesstiftung Umwelt, DBU), four groups of roughly 10 teachers each were accompanied by university educators. These groups covered the fields of Biology, Chemistry, Physics, and Politics education. Group work was structured by the model of Participatory Action Research (PAR) in science education as described by Eilks and Ralle. (2002). The aim of each group was to cyclically develop one domain-specific lesson plan of roughly 10-12 periods (45 min.) duration, which was applicable for lower secondary education and based on the topic climate change. Guidance for the lesson plans was provided by the theoretical framework of the socio-critical and problem-oriented approach to chemistry and science teaching (Marks & Eilks, 2009). All units were planned to provide a clear focus on evaluation competency, and also employed role-playing exercises or business games. In this particular case study, a special focus was added allowing later networking between school subjects and also on adapting the teaching materials for other, informal educational settings (Feierabend & Eilks, 2010).

An essential component of the PAR developmental process is the cyclical testing process, including refinement of the lesson plans (Eilks & Ralle, 2002). During testing, the lesson plans were applied to a large number of different learning groups in grades 9-11 (age range 14-17) from different middle, comprehensive and grammar schools in northern Germany. The developmental process was accompanied by different research interests. The basic focus of the accompanying research was to collect evidence reflecting on the lesson plans' feasibility and teaching effects, thus providing input for further series of cyclical refinement. Different sources of data were collected. Feedback and group discussions were taken to get the teachers' viewpoints. Questionnaires, videos, and pre- and post-group discussions were applied to record both student feedback and information about their *a priori* conceptions and learning progress.

Group discussions are considered to be a good way to get students to discuss many different questions (Solomon, 1993). Nevertheless, several problems with group discussions are also well-known from research experience, in particular the influence exerted by the interviewer (Gilbert & Pope, 1986). Another hang-up is the fact that some students tend to participate in the discussions, while others remain quiet. This tends to skew the conversation away from the actual opinions held by the silent members of the class (Loos & Schäffer, 2011). This did not tend to be a major problem within this study, since the main goal was to achieve an initial exploration of students' argumentation competency when discussing within a group because public debate as well as classroom debate is always taking place in group situation.

A semi-structured mode was chosen for the group discussions. A manual was developed to focus on different aspects like students' prior knowledge, their attitudes, and the evaluation/ consideration of learners personal responsibility when it came to climate change and potential courses of action. The decision was also made to include pre- and post-group discussions with learning groups in all the subjects. In order to reduce the total number of discussion participants, each learning group was split into half-groups of roughly 12-15 pupils each.

Pre-group discussions began by asking the students about their spontaneous reactions towards 'climate change'. A second focus was the pupils' interest in the topic and their ideas about the meaning of climate change in their own lives. These questions led into a discussion about the potential causes and responsible parties, including which avenues of action remained open for students to personally react to climate change. In the final phase of the pre-group discussion, learners were presented with one possible scenario: their school had forbidden all students to come to school by car in an attempt to reduce greenhouse gas emissions. The participants were asked to weigh the pros and cons, reflect upon potential decision-makers, and elucidate the desired outcome. In the post-group discussions, learners were again asked about their spontaneous opinions. The section dealing with personal responsibility and the responsible parties for climate change was also repeated. Then the pupils were confronted with a new scenario. For this, the recent EU-wide ban of conventional light bulbs was chosen on order to provoke discussion. As in the pre-discussion phase, participants were asked to weigh the pros and cons of this decision and to reflect upon alternate routes which might have led to a different decision.

For the purpose of this study, two parts of the data from the group discussions were evaluated. One passage was used in both the pre- and post-discussion. This passage was asking for reasons and responsibilities concerning climate change. A second part of the discussions started from an impulse asking for a decision on a fictive scenario. In the pre-discussion the scenario on a "car-free-school" was used, in the post-discussion a report on the new European law for the compulsory use of energy saving light bulbs was operated. Both passages were selected because these were the most prominent passages of the discussions where the students were asked to evaluate and argue about a socio-scientific issue.

Data was collected in a total of 20 classes in various state schools in northern Germany, with five classes for each of the four above-mentioned school subjects. Half of these classes were from grammar schools, the rest stemmed from comprehensive and middle schools. Of the total number of roughly 400 students, most came from 9th grade classes, with the rest from grade 10 and 11. The classes were randomly organized in half-groups of 12-15 students for the group discussions. Overall, data from 76 audio- and video-typed group discussions was collected (39 pre-group and 37 post-group recordings). Each discussion lasted an average of 25-30 minutes.

For the purpose of this paper, two phases of the group discussions were selected based on their potential for exploring students' evaluation and argumentation skills. The respective passages dealing with 1) student opinions about the responsible parties for and the causes of the phenomenon of climate change, and 2) the above-outlined scenarios of the car-free school and the conventional light bulb ban were accordingly analyzed. Both were analyzed independently from one other in order to allow for different methods of characterizing students' skills in the evaluation competency.

Method

Two different approaches were chosen for handling the data from the group discussions. The first approach measured the quality of student arguments by comparing their levels of justification with regard to the argumentational content. The first approach was chosen as an explorative approach for finding out the level of general argumentation skills in the means of the connection of arguments. It intends evaluating about the connection between facts and justifications as essential part of students' competencies in communication and evaluation.

This approach was applied to the group discussion parts dealing with students' opinions about who is responsible for climate change and which courses of potential action remain open to them. The second approach reflected upon the quality of students' evaluation in the sense of overall argumentational complexity. This second focus was applied to the group discussions asking students to make a decision on the car-free school and European conventional light bulb ban scenarios.

The analysis began with the initial steps of qualitative content analysis (Mayring, 2002) in order to compare students' justifications for the responsible parties of climate change with regard to the learners' actual argumentational content. Analytical coding revealed many different categories expressing pupils' opinions about the main causes and responsible parties of climate change. The same was true for the potential countermeasures open to the participants. It became quite clear during the coding process that the quality of justifications covered a wide range. Some students responded by simply repeating keywords borrowed from the lessons. Others tried to provide evidence for single arguments. And some of the learners attempted to use reason in constructing their their statements.

Consequently, a pattern was developed for rating students' answers by comparing the quality of justification and the content level of argumentation. The rating system was inspired by the work of Jungermann, Pfister and Fischer (2005), who previously suggested gradations in both the manner and the extent to which cognitive effort is undertaken for decision-making. The authors suggested using four categories, namely I) experienced decisions, II) stereotyped decisions, III) reflective decisions, and IV) constructive decisions. A total of five categories was constructed by an approach near to Grounded Theory (Glaser & Strauss, 1967) taking the ideas of Jungermann et al. (2005), and Eggert and Bögeholz (2006) into account for cyclically processing the data. The emerging evaluation grid offers the possibility of rating arguments on a nominal scale presenting increasingly sophisticated levels of justification (Table 1). Levels 1 and 2 of this model represent rather low levels of evaluation competency. As discussed in Dawson and Venville (2010) and Means and Voss (1996), these lower levels may not even represent a full-fledged argument, since they do not necessarily contain to formal justification or support taken from the content side (see above). However, we decided to retain Levels 1 and 2, because they best represent the "Level I" defined by the German science education standards (KMK, 2004; see above). Level 3 of our grid can be considered to express a medium level; it has parallels to "Level II" in the German standards. Levels 4 and 5 can be considered rather high levels of evaluation competence when compared to "Level III" of the German standards. Jungermann et al. (2005) would describe these higher levels as expressing more elaborated arguments and more complex evaluations. They can thus be interpreted as representing higher achievement in the respective competence domains.

During cyclical checking of the data, this description showed a good fit with the collected data in sense of data saturation. As the final step, the entire set of selected passages from the 76 evaluated group discussions were coded according to the above-mentioned scheme. Rating was performed by two independent raters. Inter-rater reliability was calculated using Cohen's κ and percentage agreement in order to ensure quality control. The calculated values were κ = 0,80

(85,3% agreement) for the pre-discussion and κ = 0,82 (89,5%) for the post-discussion, thus evidencing high levels of overall agreement.

Table 1. Rating grid inspired by Jungermann et al. (2005)

Level		Description
1	Keyword arguments	Arguments are given by repetition of a short keyword or a fixed expression. A prediction of whether a valuation process has occurred or not is impossible. In most cases one can guess that the argument is either memorized or totally spontaneous without any reflective thought behind it.
2	Intuition arguments	Arguments are a little more elaborated but are still based on intuition, perception or stereotyped decisions, not on facts.
3	Justified arguments	Arguments are based on facts and scientific knowledge.
4	Reflective arguments	Arguments connect several pieces of information. Arguments contain the weighing of information from at least two different perspectives.
5	Constructive arguments	Arguments include information from different perspectives, are interconnected and recommendation avenues of action which were not imposed on the students by the task at hand.

Another approach was created for dealing with the two scenarios in the pre- and post-discussion (car-free school and light bulb ban). The focus in this area was the complexity of students' arguments. The analysis also began with detailed evaluation of various categories according to qualitative content analysis (Mayring, 2002). Great variety was found in the complexity of the arguments employed, ranging from mentioning only one fact to using two and more facts and statements in a connected and justified fashion.

An evaluation grid was developed to show the level of complexity contained in the arguments. This development was inspired by the modification and level-combination ideas of valuation competence as presented in Haidt (2001) and Wilson and Sloane (2000). A combined model incorporating Kauertz et al. (2010) model of complexity and its various suggestions for categorizing student tasks was derived in order to cyclically evaluate the group discussion data. The final evaluation grid offers researchers the possibility of ranking student answers on a nominal scale of increasing levels of complexity (Table 2). In this model, Levels 0-2 represent rather low levels of evaluation competence. Level 3 can be considered to be a medium level, and Levels 4-5 embody quite high levels of personal evaluation competence. Here Level 5 is the highest level because it includes a reflective component or conclusion in the end. This component is seen of higher level because it adds an additional crirical quality beyond justification, as it is discussed by Mitchell (1996). As in the first grid discussed above, the lower, medium and higher levels roughly coincide with Levels I-III of the German national science education standards (KMK, 2004). As in the first grid, Kauertz et al.'s (2010) higher levels are also valid here as interpretations of higher achievement in the domains of argumentation and evaluation.

This second grid was also used to evaluate this study's selected aspects of the 76 individual group discussions. The two coders were also used to code the data and inter-rater reliability was calculated using Cohen's κ and percentage agreement. The final values were κ = 0,93 (95,3%

agreement) for the pre-discussion and κ = 0,95 (96,7%) for the post-discussion, which indicates high levels of grid reliability in this case, too.

Table 2. Rating grid inspired by Haidt (2001), Wilson and Sloane (2000), and Kauertz et al. (2010)

Level		Description
0	Not related	Students provide arguments that do not correspond to the question.
1	One argument	Students provide one relevant argument but do not provide any justification for it.
2	Two arguments	Students provide two or more relevant arguments without logical relation or sound justification.
3	One or two arguments, one justification	Students provide one relevant argument with well-founded justification either by facts or personal experience. Or they give two or more arguments with at least one justification.
4	Two and more connected arguments with justification	Students provide two or more relevant arguments connected in a logical chain, justified by facts and/or personal experience.
5	One and more connected arguments with justification and reflection	Students provide one or more relevant arguments, provide justifications for them and draw sound conclusions from their argument's interconnectedness.

Examplary quotes for both grids are given in Table 3.

Findings

Applying the quality grid for student arguments, which compares the arguments' justification to their content matter, showed that the group discussions could more or less be completely rated. When looking in the data, some 60% of the arguments in all four subjects are located in the first two levels, which correspond to low levels of evaluation competence. Almost 25% of the arguments reached medium levels of quality, but only roughly 10% of the arguments could be considered to be at the highest level of achievement (Table 4).

When comparing the pre- and post-discussion results, we notice a large increase in the number of arguments from roughly 400 to over 650. But, the largest increases took place at the two lower argumentation levels. Students learned many facts and not well-supported arguments within the lesson plan and mentioned them in the group discussions. But, they didn't use them in a form which gave reasonable justifications for their choices. At the medium level there was a small increase in the total number of arguments presented; at the two highest levels a small decrease was even seen (Table 5).

The application of the second rating grid for argument complexity led to a similar picture. Overall the lowest levels of evaluation competence (Levels 0 to 2) incorporated roughly 50% of the total arguments. At the medium level (Level 3) we see a proportion of nearly 40%, whereas the most complex arguments make up less than 10% of the total. One more piece of information is offered here in comparison to the first grid: almost one-fifth of the arguments landed at Level 0, which shows that many of the given statements did not even refer~~red~~ back to the question at all (Table 6).

Table 3. Examplary quotes for both grids and the different levels

Grid 1	Topic: Discussion about responsibility for climate change			
1: Keyword arguments	2: Intuition arguments	3: Justified arguments	4: Reflective arguments	5: Constructive arguments
The politicians.	Hm, I think that politics has also a big responsibility.	Yes, the politicians are responsible, because they have to decide about the laws which protect the nature.	The politicians are responsible for acting on Climate Change, because the citizens won't change their attitudes by their own, but they [the politicians] also have to take care that they will be re-elected.	If the politic makes no decision, there wouldn't be any change. So that is why they are responsible. But it is also important that the citizens and the industry will take place on it. Because if not, nothing would help.
Grid 2	Topic: Discussion about a ban of conventional light bulbs			
1: One argument	2: Two arguments	3: One or two arguments, one justification	4: Two and more connected arguments with justification	5: One and more connected arguments with justification and reflection
Energy saving lamps are expensive.	They [energy saving lamps] are more expensive. But they are running longer.	I think they [energy saving lamps] need a special disposal system because they contain toxic substances.	In principle, a referendum would be very good because more people are included then. A problem is that many people would vote for the normal bulbs, because it is more comfortable and they also might not know much about climate change.	Energy saving lamps are dangerous because they contain mercury. Mercury is environmentally dangerous. Therefore, we would need special waste treatment and recycling systems for them.

Table 4. Categorization by grid 1 (justification of arguments) according for the pre- and post-discussions

Level	Biology	Chemistry	Physics	Politics	Total
1: Keyword arguments	106 (43,6%)	110 (43,1%)	116 (39,3%)	127 (41,5%)	459 (41,8%)
2: Intuition arguments	51 (21,0%)	66 (25,9%)	77 (26,1%)	58 (19,0%)	252 (22,9%)
3: Justified arguments	71 (29,2%)	58 (22,7%)	75 (25,4%)	76 (24,8%)	280 (25,5%)
4: Reflective arguments	11 (4,5%)	13 (5,1%)	18 (6,1%)	34 (11,1%)	76 (6,9%)
5: Constructive arguments	4 (1,6%)	8 (3,1%)	9 (3,1%)	11 (3,6%)	32 (2,9%)
Total	243	255	295	306	

Table 5. Categorization according to pre- and post-post-discussion in Grid 1 (justification of arguments)

Level	Pre-discussion	Post-discussion
1: Keyword arguments	137 (31,4%)	322 (48,6%)
2: Intuition arguments	98 (22,4%)	154 (23,3%)
3: Justified arguments	136 (31,1%)	144 (21,8%)
4: Reflective arguments	52 (11,9%)	24 (3,6%)
5: Constructive arguments	14 (3,2%)	18 (2,7%)
Total	437	662

An increase in the total number of arguments was also observed in this case. Nevertheless, the quality of argumentation in the sense of increasing complexity did not evidence much change. The largest increase occurred in the fields representing arguments of lower complexity. There was, however, a slight increase in quality at the medium level, and even a small increase at the two higher levels of argumentation (Table 7).

Discussion

This paper presents two different grids for evaluating students' arguments in a discourse situation regarding the case of climate change. Both grids proved themselves to be feasible, reliable and easily applied to group discussion data. These grids analyze students' argumentation skills either as an expression of personal argumentation competence or as evaluation competence.

The focus of the first grid concerned the quality of justification provided with regard to the content matter. Within this particular teaching situation, it was quickly recognizable that roughly half of the overall arguments presented in the semi-structured group discussions consisted of lower-level justifications, mainly in the form of either keyword and intuitive arguments. About one-third of the arguments could be characterized as medium-level justifications, defended by arguments based on either facts and theories, but without reflection or the use of constructive thought. These two latter kinds of arguments were only rarely mentioned. Connecting these results to understanding the German standards with their differentiation of three levels of reproduction, simple application, and application as a transfer to more complex tasks (KMK, 2004), we must consider that most students are on quite low levels and only rarely reach the highest level.

Table 6. Categorization by grid 2 (complexity of arguments) according to school subjects (pre- and post-discussion)

Level	Biology	Chemistry	Physics	Politics	Total
0: Not related	41 (17,3%)	56 (26,0%)	36 (16,7%)	34 (15,2%)	167 (18,7%)
1: One argument	81 (34,2%)	78 (36,3%)	74 (34,4%)	75 (33,5%)	308 (34,6 %)
2: Two arguments	12 (5,1%)	9 (4,2%)	8 (3,7%)	6 (2,7%)	35 (3,9%)
3: One or two arguments, one justification	88 (37,1%)	66 (30,7%)	84 (39,1%)	83 (37,1%)	321 (36,0%)
4: Two and more connected arguments with justification	4 (1,7%)	2 (0,9%)	5 (2,3%)	9 (4,0%)	20 (2,2%)
5: One and more connected arguments with justification and reflection	11 (4,6%)	4 (1,9%)	8 (3,7%)	17 (7,6%)	40 (4,4%)
Total	237	215	215	224	

The second grid did not judge the content quality of the students' arguments, but rather its structure with respect to overall argumentational complexity. Nevertheless, the picture resulting from this analysis of a different aspect of the same data source led to a quite similar picture. Again, roughly half of the arguments landed in lower-level categories, which generally encompassed either unrelated, unconnected, or unjustified arguments. Even in this case, about one-third of the answers given were rated as only medium-level, characterized by questionable justifications of whichever quality was being discussed. Again, the proportion of highly complex argumentation encompassing several interconnected and skillfully justified arguments was very low. Here we can see a parallel picture to the one from grid 1. In summary, there was no large increase in the number of high-level arguments.

Table 7. Categorization according to pre- and post-post-discussion in Grid 2 (complexity of arguments)

Level	Pre-discussion	Post-discussion
0: Not related	78 (20,6%)	89 (17,5%)
1: One argument	117 (30,9%)	191 (37,5%)
2: Two arguments	12 (3,2%)	23 (4,5%)
3: One or two arguments, one justification	147 (38,8%)	174 (34,1%)
4: Two and more connected arguments with justification	12 (3,2%)	8 (1,6%)
5: One and more connected arguments with justification and reflection	13 (3,4%)	27 (5,3%)
Total	379	512

A larger overall proportion in the highest categories could be seen, however, and the overall numbers of answers appearing in the two highest categories increased slightly. One common feature of both evaluations was an increase in the total number of arguments between the pre- and post-discussion. This was the case for both grids, one repeating a similar task in the pre- and post-discussion and the other introducing a new scenario (the ban of conventional light bulbs) to replace the original task (car-free-school). In any case, increases mainly occurred at the lower levels of arguments and argumentation. It appears that the students learned many new facts and unsupported arguments during the lesson plan. They were able to recognize the value of various arguments and to repeat them. Also at the medium level, the total number of arguments increased slightly. At the highest levels, one grid suggests an increase and one a slight decrease, but both sets of overall numbers were low. Therefore, there are parallels between the two grids which either can be interpreted as being parallel levels of competencies in two different aspects from the field of argumentation or decision-making.

From the data and its evaluation, it is not clear whether this parallel is coincidently or because of a structured interdependence. A hypothetical explanation for a structured interdependence might be that both grids express a similar growth in complexity of argumentation. In the first grid complexity increases by the growing need of referring back a claim to its justification and later to a reflective thought. In the two lowest levels there is only one claim or argument to be mentioned. On the medium level 3 the claim has to be connected to a single justification, on the two higher levels the claim has to be connected to more complex higher order thinking skills in the means of reflection or suggestions for future action. Also in the second grid complexity is increasing and demands higher levels of cognitive skills. In this grid the two lower levels are based on the rote mentioning of one or two facts but without need for making any kind of connection. The medium level 3 asks for at least one connection either between two arguments or one argument and its justification, the two highest levels asks for more complex thinking skills in the means of making chains of claims, facts or thoughts. Anyhow, this interdependence needs to be further researched.

Most of the participants evaluated have competencies corresponding on a low level of repeating isolated facts as arguments. This single lesson plan was possible to support this level through quantitative means. There are also some further indications that progress might be possible at a medium level and, possibly, at a high level. However, such hope of progress at the higher levels seems in this case to be small and less-strongly supported. Maybe one lesson plan of 10-12 periods is a too short run to receive thorough progresses.

Conclusions and Implications

This paper presented two different approaches suggesting possible structures for operationalizing and measuring students' argumentation competence as an expression of their evaluation competence. This was held in line with current German national science education standards (KMK, 2004). Both instruments proved to work well. The results can be interpreted through clarification of the descriptive levels found in the German national standards. Both tools measure a related construct. Although this is a purely qualitative study, the interpretation seems to be justified that the results support each other in the same competency domain. It appears that both the question of the quality of argumentational justification with respect to content matter and the complexity structure of argumentation should be viewed as two sides of the same coin. Thus, applying one of the instruments allows teachers a general consideration of the students' average abilities when discussing and evaluating socio-scientific questions under the inclusion of societal perspectives. Nevertheless, further testing of the grids should be undertaken.

Reflecting on the results, we see that both grids led to a similar picture. Overall, the quality of students' argumentation skills as part of their evaluation competence is generally not well-developed. In any case, the applied lesson plan led to a quantitative increase in the numbers of arguments at lower levels of competency. Students were able to give many more arguments in the post-discussion exercise than in the pre-discussion. This was despite the fact that many of the arguments were neither interconnected with respect to justification, nor were they embedded in complex chains of argumentation. Progress at the higher levels of evaluation competency in the sense of Germany's national standards may require more time and repeated emphasis of this issue.

For science education practices, we can recognize the necessity for increased initiative in educating students with respect to better argumentation and decision-making skills. There appears to be a lack of such efforts thus far. The issue of climate change nevertheless proved itself to be a positive addition to school curricula because of its potential for better focusing students' argumentation and evaluation skills (Feierabend & Eilks, 2010). This topic allows for an increased orientation on interdisciplinary knowledge, adds connections to informal education and promotes the societal aspects of science education (Feierabend & Eilks, 2010). There is hope that if such approaches were to be applied more often, higher levels of student argumentation and evaluation skills would develop as a result.

Acknowledgment

We would to thank the German Environmental Fund (Deutsche Bundesstiftung Umwelt, DBU) for the financial support and the participating teachers and students for making this work possible.

References

Aikenhead, G. (1985). Collective decision making in the social context of science. *Science Education*, *69*(4), 453–475.

Bell, R. L. & Lederman, N. G. (2003). Understandings of the nature of science and decision making on science and technology based issues, *Science Education*, 87(3), 352-377.

Bybee, R. (1997). Toward an understanding of scientific literacy. In W. Gräber & C. Bolte (eds.): *Scientific Literacy – An International Symposium* (pp. 37-68). Kiel: IPN.

Dawson, V. M. & Venville, G. (2010). Teaching strategies for developing students' argumentation skills about socioscientific issues in high school genetics. *Research in Science Education*, *40*(2), 133-148.

Duschl, R. A. & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, *38*(1), 39-72.

Eggert, S. & Bögeholz, S. (2006). Göttinger Modell der Bewertungskompetenz – Teilkompetenz „Bewerten, Entscheiden und Reflektieren" für Gestaltungsaufgaben Nachhaltiger Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, *12*, 199-217.

Eilks, I. (2000). Promoting scientific and technological literacy: Teaching Biodiesel. *Science Educational International*, *11*(1), 16-21.

Eilks, I. (2002). Teaching 'Biodiesel': A sociocritical and problem-oriented approach to chemistry teaching, and students' first views on it. *Chemstry Education: Research and Practice in Europe*, *3*(1), 67-75.

Eilks, I., Feierabend, T., Hößle, C., Höttecke, D., Menthe, J., Mrochen, M., Oelgeklaus, H. (2011). Bewerten Lernen und der Klimawandel in vier Fächern - Einblicke in das Projekt "Der Klimawandel vor Gericht" (Teil 1). *Der Mathematische und Naturwissenschaftliche Unterricht, 64*, 7-11

Eilks, I., & Ralle, B. (2002). Participatory Action Research in chemical education. In B. Ralle & I. Eilks (eds.), *Research in chemical education - what does this mean?* (pp. 87-98), Aachen: Shaker.

Erduran, S. & Jiménez-Aleixandre, M., eds. (2007). *Argumentation in science education.* New York: Springer.

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*, 915-933.

Feierabend, T., & Eilks, I. (2010). Raising students' perception of the relevance of science teaching and promoting communication and evaluation capabilities using authentic and controversial socio-scientific issues in the framework of climate change. *Science Education International*, *21*(3), 176-196.

Feierabend, T. & Eilks, I. (2011). Innovating science teaching by Participatory Action Research ? Reflections from an interdisciplinary project on curriculum development in the field of climate change. *Centre for Educational Policy Studies Journal*, *1*(1), 93-112.

Flemming, R. (1986). Adolescent reasoning in socio-scientific issues, Part I: Social cognition. *Journal of Research in Science Teaching*, 23(8), 677–688.

Gilbert, J. K. & Pope, M. L. (1986). Small Group Discussions About Conceptions in Science: a case study. *Research in Science & Technological Education, 4*(1), 61-76.

Glaser, B. G. & Strauss, A. L. (1967). *The discovery of grounded theory*. Chicago: Aldine.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814-834.

Hofstein, A., Eilks, I., & Bybee, R. (2011). Societal issues and their importance for contemporary science education: a pedagogical justification and the state of the art in Israel, Germany and the USA. *International Journal of Science and Mathematics Education*, *9* (6), 1459-1483.

Holbrook, J. (1998). Operationalising scientific and technological literacy: A new approach to science teaching. *Science Education International*, *9*(2), 15-19.

Holman, J. (1986). *Science and Technology in Society; a general maide for teachers.* Haffield, Herts: Association for Science Education.

Inch, E. S. & Warnick, B. (2002). *Critical thinking and communication.* Boston: Allyn & Bacon.

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

Jungermann, H., Pfister, H.-R. & Fischer, K. (2005). *Die Psychologie der Entscheidung – Eine Einführung.* München: Elsevier.

Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E. & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den naturwissenschaftlichen Fächern der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften, 16*, 135-153.

Kolstø, S. D. (2006). Patterns in Students' Argumentation Confronted with a Risk-focused Socio-scientific Issue. *International Journal of Science Education*, *28*(14), 1689-1716.

Kortland (1996). An STS case study about students' decision-making on the waste issue. *Science Education, 80*(6), 673-689.

Kuhn, D. (1991). *The skills of argument.* Cambridge: Cambridge University Press.

Loos, P. & Schäffer, B. (2011). *Das Gruppendiskussionsverfahren. Theoretische Grundlagen und empirische anwendung.* Wiesbaden: VS Verlag.

Marks, R., & Eilks, I. (2009). Promoting Scientific literacy using a socio-critical and problem-oriented approach in chemistry education: concept, examples, experiences. *International Journal of Environmental and Science Education*, *4*(3), 231-245.

Mayring, P. (2002). Qualitative content analysis - research instrument or mode of interpretation? In M. Kiegelmann (ed.), *The role of the researcher in qualitative psychology* (pp. 139-148). Tübingen: Ingeborg Huber.

Means, M. L. & Voss, J. F. (1996). Who reasons well? Two studies of informed reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction, 14*(2), 139-178.

Mitchell, S. (1996). *Improving the quality of argument in higher education interim report.* London: Middlesex University, School of Education.

Ratcliffe, M. (1996). Pupil decision-making about socio-scientific issues, within the science curriculum. *International Journal of Science Education, 19*(2), 167-182.

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research Science Teaching, 41*(5), 513–536.

Sadler, T. D., & Zeidler, D. L. (2005). Patterns of informal reasoning in the context of socioscientific decision-making. *Journal of Research in Science Teaching*, *42*(1), 112-138.

Sadler, T. D. & Zeidler, D. L. (2009). Scientific Literacy, PISA, and socioscientific discourse: assessment for progressive aims of science education. *Journal of Research in Science Teaching*, *46*(8), 909-921.

Solomon, J. (1992). The classroom discussion of science-based social issues presented on television: Knowledge, attitudes and values. *International Journal of Science Education, 14*(4), 431-444.

Solomon, J. (1993). *Teaching science, technology, and society.* Buckingham: Open University Press.

Solomon, J. & Aikenhead, G. (1994). *STS Education: International perspectives on reform*. New York: Teachers College Press.

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

Tytler, R., Duggan, S. & Gott, R. (2001). Dimensions of evidence, the public understanding of science and science education. *International Journal of Science Education, 23*(8), 815-832.

Wilson, M. & Sloane, K. (2000). From principles to practice: an embedded assessment system. *Applied Measurement in Education*, *13*(2), 181-208.

Yang, F.-Y. & Anderson, O. R. (2003). Senior high school students' preference and reasoning modes about nuclear energy use. *International Journal of Science Education*, *25*(2), 221-244.

Authors

Dr Timo Feierabend obtained his PhD in chemistry education from the Institute of Science Education (IDN) at the University of Bremen, Germany. His research interests are climate change education and socio-scientific issues in science education. He is now working as a teacher for chemistry and mathematics at a grammar school in Vechta, Germany. E-mail: tfeier@uni-bremen.de

Marc Stuckey did his MEd at the Institute of Science Education (IDN) of the University of Bremen, Germany. He is now working as a researcher at the IDN. His research interests are climate change education and the question of relevance within socio-scientific issues-based science education. E-mail: mstuckey@uni-bremen.de

Sarah Nienaber obtained her MEd from the Institute of Science Education (IDN) of the University of Bremen, Germany. Her research interest is climate change education. E-mail: SarahNienaber@gmx.de

Prof Dr Ingo Eilks FRSC is a professor of chemistry education at the Institute of Science Education (IDN) at the University of Bremen, Germany. His research interests are socio-scientific issues in science education, education for sustainable development, cooperative learning, the particulate nature of matter, and action research in science education. Correspondance: Prof Dr Ingo Eilks, University of Bremen, Department of Biology and Chemistry, Leobener Str NW 2, 28334 Bremen, German. **Correspondence:** University of Bremen, Department of Biology and Chemistry, Institute for Science Education (IDN), Leobener Str. NW 2, 28334 Bremen, Germany. E-mail: ingo.eilks@uni-bremen.de

Öğrencilerin grup çalişmalarinda iklim değişikliğine yönelik yaptıkları tartiışmalarin yeterlik değerlendirmesini analiz etmek için iki yaklaşim

Şu ana kadar, öğrencilerin fen eğitim bağlamındaki kavramsallaştırmalarını açıklayan çok az sayıda model bulunmaktadır. Öğrencilerin bu özel yeterlikleri için yapılan çoğu öneri normatif olup bu yeterliklere yönelik deneysel destek zayıf kalmaktadır. Analitik parametrelerin bir parçası olarak etik ve sosyal perspektifler değerlendirildiğinde bu durum daha da problem olmaktadır. Bu konuyu incelemede çok boyutlu tartışma durumları bağlamında, öğrencilerin değerlendirme kapasitelerine yönelik iki yaklaşım sunulmaktadır. İlk yaklaşım, öğrencilerin gerekçelendirme düzeyleriyle ilgili argümanlarının kalitesi üzerine yoğunlaşmaktadır, ikincisi ise çocukların argümantasyonlarının karmaşıklık düzeyinin kalitesi üzerine yoğunlaşmaktadır. Her iki yaklaşım somutlaştırılarak değerlendirme amacıyla grup tartışmasındaki veriler kullanılmıştır. İklim değişikliğinin öğretimi üzerine yoğunlaşan innovasyon müfredatına dayalı olarak veriler dört öğretim alanına dayandırılmıştır: Biyoloji, kimya, fizik ve siyaset. İklim değişikliği konusuna yönelik 20 farklı grubun her birindeki katılımcıların yarısıyla yarı-yapılandırılmış görüşmeler ve ön-son grup tartışmaları yapılmıştır. Toplam 76 grubun yaptığı tartışmaların analiz sonuçları iklim değişikliği ile ilgili her iki yaklaşımında olumlu potansiyele sahip olduğunu göstermiştir.

Anahtar Kelimeler: değerlendirme yeterliği, grup tartışması, değerlendirme, iklim değişikliği