

Article

Argumentation within Upper Secondary School Student Groups during Virtual Science Learning: Quality and Quantity of Spoken Argumentation

Marko Telenius ^{1,2,*}, Eija Yli-Panula ¹ , Veli-Matti Vesterinen ³  and Marja Vauras ¹ 

¹ Department of Teacher Education, The University of Turku, 20014 Turku, Finland; eija.yli-panula@utu.fi (E.Y.-P.); vauras@utu.fi (M.V.)

² Upper Secondary School, 21420 Lieto, Finland

³ Department of Chemistry, The University of Turku, 20014 Turku, Finland; veli-matti.vesterinen@utu.fi

* Correspondence: mjtele@utu.fi or marko.telenius@lieto.fi

Received: 27 November 2020; Accepted: 16 December 2020; Published: 19 December 2020



Abstract: In many studies, the focus has been on students' written scientific argumentation rather than on their spoken argumentation. The main aim of this study was to relate the quality of spoken argumentation to groups' learning achievement during a collaborative inquiry task. The data included video recordings of six groups of three upper secondary students performing a collaborative inquiry task in a virtual learning environment. The target groups were selected from a larger sample of 39 groups based on their group outcome: two low, two average, and two high-outcome groups. The analysis focused on argumentation chains during the students' discussions in the planning, experimentation, and conclusion phases of the inquiry task. The core of the coding scheme was based on Toulmin's levels of argumentation. The results revealed differences between the different groups of students, with the high-performing groups having more argumentation than the average and low-performing groups. In high-performing groups, the students asked topic-related questions more frequently, which started the argumentative discussion. Meanwhile, there were few questions in the low-performing groups, and most did not lead to discussion. An evaluation scheme for the quality of the arguments was created and the spoken argumentation was analyzed using a computer-based program. The results may be used to benefit subject teacher education and to raise teachers' awareness of their students' scientific, topic-related discussions.

Keywords: spoken argumentation; virtual learning environments; collaborative learning

1. Introduction

Argumentation is part of the critical discourse through the natural sciences process [1]. Competency in argumentation is a crucial component of scientific literacy and is a central goal of science education [2,3]. In science education, argumentation involves understanding texts, analyzing and evaluating critical material, and generating hypotheses [4]. It also includes the ability to process and discuss educational topics [5]. Argumentation in science education has been shown to support the development of higher order and critical thinking, enculturation in scientific culture, engagement in science learning, as well as learning the core contents of science [6,7].

Argumentation is a basic skill in scientific literacy [2]. In school, goal-based guided assignments, social activities, and argumentative discussion have been shown to enhance higher-order thinking skills [8]. Further, when given an active role to produce knowledge based on their own ideas [9], students have demonstrated scientifically oriented productive disciplinary engagement (PDE) instead of classroom-like PDE [10].

Several studies have focused on the analysis of argumentative discourse in various educational contexts [6,11–14]. Typically, in these earlier studies, the students were instructed to argue and were given instructions for argumentation. The present study focuses on students' spontaneous discourse in learning and thus presents a fresh view on the nature of students' argumentation in real-life collaborative learning, where it arouses and is enacted by the features of the learning environment, the adopted role of students, and the need to achieve a common goal. As far as we know, there are no studies concerning these issues, and a novel analysis method is introduced here specifically for this kind of argumentation. In preliminary studies, it has been found that students' argumentative interdisciplinary discussions occurred spontaneously in the context of collaboratively generating hypotheses and study plans and that the students were able to integrate biological and chemical concepts when studying a virtual environment, which was a thoroughly new working environment for the students [15]. The students were also able to independently transfer their knowledge learned in the virtual environment [16].

1.1. Argumentation in Science

Argumentation can be spoken or written [17]. There is also a difference between spontaneous and scientific argumentation. Scientific argumentation refers to a person's ability to support or criticize an idea, explanation, or conclusion based on values and criteria supported by the scientific community [18]. According to Duschl and Osborne [7], scientific argumentation is a method and process of reasoning in which the argument aims to guarantee the correctness of the presented claim [19]. Argumentation is also a way to convince the listener of the issue at hand and its implications. According to Belland et al. [20], argumentation involves drawing conclusions through logical scientific reasoning and is based on the use of key concepts [6].

Argumentation plays an important role in natural sciences because it is commonly used to present, support, refine, and evaluate scientific knowledge. It is about making conclusions with the help of logical reasoning [20]. To support the argument, evidence is presented, and this evidence is thoroughly tested [21]. Argumentation in science education [22] can have a positive impact on at least five educational areas, including motivation, content learning, general argument skills, specific argumentation skills, and knowledge building practices. In general, there are two theoretical frameworks for the benefits of argumentation, the first being a cognitive account and the second a sociocultural account [13,23]. A cognitive account focuses on individual learning and accounts for learning in terms of improved individual representations. A sociocultural account treats learning through argumentation as the adoption of social practices.

The benefits of argumentation demonstrated in earlier studies include the following:

- It can enhance motivation. A possible reason for this is that students have more autonomy over what they say in argumentation-based discussions than in traditional teacher-led situations, where teachers limit students to answering their several questions [24].
- It may simply offer more interaction with peers, which in itself can be very motivating [25].
- Students' engagement in argumentation allows them to become aware of their peers' ideas and in this way to become interested in which of the ideas are the most defensible ones [26].
- It can support content learning in several ways, such as engaging in explicit elaborative processing (e.g., explanation, which is known to promote learning) and modification of ideas. When engaging in argumentation, students engage in a variety of explicit elaborative processes [23] or receive learning support from others. In the process of argumentation, students learn from their peers or give them reasons to believe. This is because the process of providing evidence for claims can give learners not only a better understanding of the ideas they are learning but also more reason to believe the claims they are developing [27].

- It can improve general argumentation skills, specific argumentation skills, and knowledge building practices. This benefit is explained by students engaging in argumentation and therefore learning to argue and evaluate arguments better [28].

Argumentation is also one of the fundamental aspects of productive disciplinary engagement (PDE), which has been one of the key starting points in this study. In this study, “PDE is assumed to occur when learners use the language, concepts, and practices of the discipline in authentic tasks to ‘get somewhere’ (e.g., develop a product or improve a process) over time” [29]. The part of this definition concerning discipline-related language or reasoning to accomplish a task connects argumentation tightly to the structure of PDE. Engle and Conant [30] explained the term PDE by saying that students’ engagement is productive to the extent that they make intellectual progress. They attempted to foster PDE by introducing a special FCL (Fostering Culture of Learners) case in which discourse and argumentative discussions are basic principles. In their FCL case, discourse involves a dialogic base, shared discourse, common knowledge, mutual appropriation, as well as seeding, migration, and appropriation of ideas. The guiding principles for this particular stage of the FCL case are problematizing, resources, and accountability to others, authorities, and contributors. Engle and Conant [30] discovered that the students’ arguments for their claims became increasingly sophisticated over time, and their discussion gave rise to new questions. Furthermore, to learn the basics of any of the natural sciences students need to understand how scientific arguments and knowledge can help to solve problems in society [31]. One way of seeing scientific argumentation is the notion that argumentation sets the standards for accountability, which relates to PDE [32]. Thus, argumentation can be seen as a distribution of authority and accountability across the discussion. Ong, Duschl, and Plummer [33] viewed scientific argumentation as an epistemic practice benefitting from the PDE-critique and construction (PDE-CC) model. Their results showed that students’ critique skills in the science classroom were improved using this approach.

1.2. Theoretical Frameworks to Analyze Argumentation and Argument Quality

Toulmin’s [34] perspective on argumentation has had a substantial effect on science education research. In his framework, Toulmin distinguished between the idealized notions of logical-formal arguments as used in mathematics and the use of arguments in linguistic contexts. Thus, this model has been used to evaluate the quality of reasoning behind the argumentation. Toulmin’s argument framework suggests that the statements that form an argument have different functions that can be classified into one of six categories: claims, data, warrants (main categories independent of the field of science), backings, qualifiers, and rebuttals (optional categories). Thus, claims are assertions supported by data, and they can either be the conclusion of the argument chain or the basis for accepting a claim. Claims answer to four questions—is the phenomenon existing or not, what is it, what is its value and what should be done with it. Data are the foundations for those claims, the material, fact, or opinion that can be called evidence. Without data, the claim would not have any informative part where the discussion could start. Warrants are comments that are used to justify why data are relevant to the claim, and the strength of a warrant is indicated by the inclusion of a modal qualifier. The warrant is the part of an argument that allows the transition from claim to data or controversially from data to conclusion. Thus, it is an intermediary between the claim and accepted data. The backings of an argument are the comments that are used to establish the general conditions that strengthen the accessibility of the warrants so that the connection between the data and the claims will not be questioned. A rebuttal indicates the circumstances in which the warrant is set aside [34]. The rebuttal is a way to contradict other evidence.

The main focus thus far of research involving Toulmin’s [34] argument pattern in science education contexts has been on structural issues. The model has provided a great deal of insight into the ways students structure an argument and the nature of the justification they use to support their ideas. One complication encountered by researchers in applying Toulmin’s framework involves reliably distinguishing between claims, data, warrants, and backings because the comments made by students

can often be classified into multiple categories. This is also the main reason why the model has been simplified in this study. The other reason is the fact that, although the student argument would be considered relatively strong structurally according to most Toulmin-based frameworks, the content may be inaccurate from a scientific perspective. To be more precise, all that matters in Toulmin's model is the presence or absence of data, warrants, and backings, not their accuracy or relevance. To sum up, this framework must be supplemented with other measures since it is not accurate enough from a scientific perspective and does not take the sense of an argument as a whole into account.

Kolstø [35] interviewed students to better understand their informal reasoning on controversial socio-scientific issues. The students presented the following five types of arguments: the relative risk argument, the precautionary argument, the uncertainty argument, the small risk argument, and the pros and cons argument concerning their decision-making and the interplay between knowledge and personal values. The interviews with the students were also video recorded, which gave the researchers the ability to analyze the pattern of decision-making. In earlier studies on argumentation, it has been found that teachers' argumentative support during science lessons leads to improvement in students' argumentative skills [36]. Science education is seen as a combination of cognitive, social, and emotional phenomena, of which argumentation systems often exclude the latter two phenomena [37]. In the present study, the argumentation coding category attempts to account for socio-emotional aspects as well. When analyzing studies using video-taped material, it has been found that claims, rebuttals, and justifications are the critical features for developing and evaluating argumentation in classroom situations [36]. With younger students, it has been found that in video-taped sessions only one pair typically engages with the socio-scientific topic during the science lesson [38]. Furthermore, it was shown that a socio-scientific topic on its own does not lead to engagement in argumentation. In another study setting, groups of three students prepared a report for other students as part of a video-taped peer review practice, and it was demonstrated that students' argumentation skills were improved as a result of this argumentative cycle [39].

1.3. Argumentation and Virtual Learning Environment in Connection to PDE

Researchers in science education have highlighted numerous deficiencies in students' scientific argumentation [40] as well as difficulties teachers encounter in attempting to organize productive classroom-type argumentation [41]. Sandoval et al. [32] studied spoken argumentation in a classroom endeavor to achieve productive argumentation among the whole class, where small groups of students tried to interpret and explain some activities of their own. According to the authors [23], argumentation requires changes in students' mindsets and conceptions related to classroom practices, such as sharing their thinking and engaging with each other's ideas productively and with accountability. Sandoval et al. [32] also used video records when analyzing teacher-led whole-class discussions at the elementary school level. They determined segments and episodes of arguments and focused on arguments arising from any aspect of science activity. In this way, they followed Manz's [42] idea that argumentation is part of the activity of knowledge production, in which disagreements may naturally arise and must be resolved. Teacher utterances in each argument episode were coded in terms of their function in the discourse. In the study, Sandoval et al.'s [32] focus was on how teacher talk promotes student argumentation.

The present study follows the idea that argumentation requires shifting discourse practices in classroom culture but differs in that the key person who is responsible for the specific argumentative culture is not the teacher; rather, the responsibility lies with the students in the small groups. Due to changes in classroom activities, the argumentation occurs in the small groups' discussions, and in virtual learning environments teachers only share the students' learning processes as a guide if needed. To the best of our knowledge, there is no research similar to the present study using videotaped research material in connection to spontaneous group argumentation in a virtual learning environment, and a new coding system (on-task and off-task coding) has been developed to study spontaneous argumentation in upper secondary school.

Originally, the virtual learning environment was developed to facilitate student group learning, and it enables a large amount of study data to be shared and used [43], for example, real research data. Two virtual learning environments for secondary school purposes have been introduced to educate experimental marine scientists about ocean acidification [44], as well as integrative biology and chemistry in connection to the Baltic Sea and its environmental threats [45] (Virtual Baltic Sea Explorer, ViBSE). These web-based learning environments were designed to offer realistic content for learning key science concepts and content, and they both include an introductory part and a laboratory to do experiments. The ViBSE-laboratory is based on real data collected by marine scientists and on data that have been published in scientific peer-reviewed journals. Both of these virtual learning environments support students' PDE in science learning [29].

1.4. Research Questions

The aim of the study was to analyze the quality and quantity of students' spoken argumentation in interdisciplinary science. This aim is related to groups' learning achievement during a collaborative inquiry task and relates to understanding the influence of argumentation on the group's joint learning achievement. Argumentation has been shown to support the development of higher-order and critical thinking, enculturation in scientific culture, and engagement in science learning [6,7,46]. In addition, argumentation fosters learning of the core contents of science tasks during interdisciplinary science lessons. Here, a novel coding scheme was developed to analyze the amount, quality, and quantity of spoken argumentation. The coding scheme was introduced for the three phases (planning, experimenting, and conclusion) concerning students' group argumentation. Four research questions with expectations were generated:

1. How is the duration of verbal communication, on-topic talk, and argumentative discussion related to the quality of group outcomes of a collaborative inquiry task?
 - Expectation: Time spent on a certain step of collaborative inquiry creates a situation where students have more interactions with each other and become more interested in their peers' ideas, which ultimately leads to better group performance [47]. Thus, the groups that use a bigger proportion of their time on argumentation are expected to have higher quality outcomes.
2. How is the amount of argumentation related to the quality of group outcomes of a collaborative inquiry task?
 - Expectation: Previous research (e.g., [48]) has shown that students in collaborative conditions enforced with argumentation should perform better in application problems. Thus, more frequent use of argumentation should lead to higher quality outcomes.
3. How is the quality of the argumentation related to the quality of group outcomes of a collaborative inquiry task?
 - Expectation: Student groups using argumentation in which the claims fit well with the theory and strong evidence is provided should have higher quality outcomes.
4. How do the amount and quality of argumentation change during the three argumentative phases (planning, experimenting, and conclusion) of a collaborative inquiry task?
 - Expectation: The longer the students discuss and share ideas on a certain topic, the more equal they become in terms of outcome quality. Therefore, the quality of the outcomes should be higher for student groups in which all members participated in the discussion [49]. Thus, the phases that include the most argumentative discussion should have the highest quality of argumentation.

2. Materials and Methods

2.1. Group Argumentation as Part of the Study Design

As described above, while the argumentation frameworks diverge regarding the numbers and types of structural components, many of them agree on the structure of an argument as a pair of claims and justification. Thus, in this study argumentation is also seen as a chain from a claim through observation and evidence to the interpretation of the claim. Accordingly, we attempted to achieve a balance between the different approaches in our model. Moreover, the material consists of videotapes, and thus spoken argumentation instead of written arguments, which also had to be taken into account. The argumentation model is emphasized in more depth in this section. In this study, argumentation is defined as a way that students reach a common understanding about a certain scientific topic.

Much of the earlier research on argumentation has focused on written arguments or arguments constructed by a single speaker [34]. Therefore, the focus has usually been on the evidence and reasons with which writers or speakers support their claims [50] rather than on the collaborative nature of argumentation. This so-called collaborative argumentation refers to an interchange of statements, questions, or replies in a dialogue involving two or more participants [19]. Participants in this kind of dialogue usually make claims and support them with reasons. The participants may disagree and try to resolve this disagreement, but not necessarily. Therefore, in this study students' argumentation was analyzed using video data. It is also more appropriate to discuss collaborative argumentation dialogue in which the participants do not disagree with each other [50]. For example, students may collaboratively develop reasons to support a scientific explanation that all can agree with, or they may collaborate to use evidence to develop and refine their scientific explanation. Evidence is also naturally used in this situation to support their developing ideas. Thus, there is collaborative argumentation because the students are jointly participating in the construction of arguments.

2.2. Participants, Group Activities, and Learning Context

Students ($n = 120$) comprising 39 groups from three upper secondary schools participated, and six groups with a total of 18 students (4 boys and 14 girls) were ultimately chosen for this study. The goal of this study was to carry out systematic, in-depth analysis of spoken argumentation and its relation to collaborative learning outcomes. The main criterion for selection of the groups was that they had differing science outcomes (two high, two average, and two low achieving groups). Additional inclusion criteria were: (a) How intact the groups remained over the whole activity, that is, over three different lessons; and (b) that the quality of the video data was sufficiently high to allow reliable analysis. In authentic educational circumstances, only the "best possible" option could be used, which meant that in one group one member was absent from the final session. Data collection was carried out during biology and chemistry lessons. This included a biology specialization, biology compulsory, and chemistry compulsory course, and they were all arranged according to the Finnish national core curriculum. All their activities for this study involved performing experiments in the way scientists typically conduct them.

The students studied in pairs or groups of three in the virtual laboratory (Virtual Baltic Sea Explorer, ViBSE, built in the dominant science language, i.e., English) [45,51], on a shared laptop. In these student groups, different learners were mixed in collaboration with the teachers to diminish the individual differences. They had to carry out a virtual experiment on the effects of pH on copepods in the Baltic Sea's food chain. ViBSE was a platform installed on students' laptops, consisting of a library explaining the structure of the research and a scientific background of the phenomena, photos, interviews with the crew and researchers of the real research vessel Aranda, laboratory tasks, and links to external web pages providing real-time data of the Baltic Sea. This virtual exploration acquainted the students with the way real scientists conduct research. The software emphasized hypotheses formation, the experimental phase, and interpreting the results. All of the experiments were based on studies by real marine biologists [52,53]. After completing the experiment, the students were asked to

prepare a PowerPoint presentation including the research plan, results, and conclusions, which was considered the outcome of collaborative learning. Thus, the collaborative inquiry project had three distinct argumentative phases: first, planning; second, experimenting; and third, drawing conclusions and reporting the results. In the beginning of the ViBSE studies, the teachers were encouraged to promote student groups' scientific argumentation, but no explicit guidelines were given to the teachers.

The virtual laboratory task was designed to foster argumentation. In planning and generating their written study and experimental plans, students had to discuss topics that they were not familiar with before the lessons as well as to decide on several key points of the experiment. The teachers intervened and provided help only if the project was not progressing properly and the group was having difficulty in a certain phase of the virtual study.

Two qualified science professionals in biology and chemistry evaluated the overall quality of the groups' PowerPoint presentations, taking into account the students' written research plan, hypotheses, understanding of the task and presentation structure, actual presentation, conclusions, and quality of the scientific language used in the presentation. To ensure reliability and quality of outcome assessment, established assessment criteria [54] were used to categorize presentations into six levels (graded from 1 to 6). According to the assessments, the student groups were divided into high-, average-, and low-performing groups (high = grades 5–6, average = grades = 4–3, and low = grades 2–1). From each level of performance, videos of two groups were selected for closer analysis in this study. All the selected groups consisted of three students.

As part of the study courses, the actual virtual learning environment portion was conducted in three to five lessons. The one upper secondary school with the biology specialization course was able to spend five lessons on this study due to the advanced-level nature of the course, but other schools could use only three lessons. However, the student experiments and data collection were designed so that the students' experiments could be carried out in either three or five lessons. This spontaneous argumentation was captured on video, and the three phases of the Virtual Baltic Sea Explorer lessons were then analyzed.

2.3. Video Analysis and Argumentation Coding

The video analysis proceeded so that three distinct, meaningful, and self-contained interaction segments (approximately 10–16 min) were selected from each of the six groups. Criteria for the selection were that the chosen segments were crucial for task performance and completion, demanded student collaboration, and were representative for all three phases. Therefore, a science expert selected the video segments from the same steps of each phase of the study to ensure their comparability with each other. Videos were first edited using "MovieMaker" software. Subsequently, videos were added into the Observer XT 12 software program as new observations. As the analysis was focused on verbal communication between the students, the first stage of the analysis involved differentiating the verbal communication from the non-verbal communication. Instances of verbal communication were marked as "verbal", and the silent parts were marked as "non-verbal". The whole video was coded this way from beginning to end. After this categorization, the parts of the video with verbal communication were analyzed further. Thus, one verbal comment is regarded as one turn, which is then coded according to the coding scheme.

The second stage of analysis involved distinguishing argumentation from other verbal communication. Each instance of communication was categorized as "off-task", "on-task non-argumentative", or "on-task argumentative". The "off-task" category contained student discussions of issues that were totally irrelevant to the topic at hand. For example, students discussed why the roof of a house across the street was red or what they had done on the weekend. The "on-task non-argumentative" category consisted of students' dialogues that concentrated on scientific content and the task but without scientific argumentation. For example, students talked about how to prepare a presentation, how to cut and paste a picture, or where to save their materials. The instances of verbal communication categorized as "on-task argumentative" were chosen for closer analysis. All of these

codes were mutually exclusive, meaning that the codes could not occur at the same time. In practice, inserting a new code automatically ended the previous one.

The third stage involved analyzing on-task argumentative verbal communication, and the initial evaluation scheme was based on Toulmin's argumentation theory [34]. The main components of Toulmin's model were data (the facts students appeal to in support of their claim), claims (conclusions whose merits are to be established), warrants (the reasons justifying the connections between the data and the knowledge), and backing (basic assumptions that are commonly agreed upon). These categories were modified in the direction of the framework illustrating the components of a scientific argument presented by Sampson and Schleigh [55]. The framework places emphasis on the nature of the evidence and the claim and also on the interpretation of the data and the analysis method. The depth of this model was sufficient in this context of spontaneous, spoken argumentation. Additionally, these categories are in line with the three phases of the task the students carried out.

During the initial analysis of the students' spontaneous on-task argumentation, it was noticed that the chains of argumentation usually began with a question, which often included a claim. Then, the argumentation continued with one or more observations, which provided evidence or offered an interpretation linking the claim or evidence to the theory. Thus, three data- and theory-driven components of spontaneous verbal argumentation were recognized: questions (Q), observations (O), and interpretations (I).

The quality of argumentation in each of these components was evaluated based on the nature and function of the argumentation. The quality of argumentation in "questions" was evaluated based on how well the claim included in the question fit the theory as well as on whether or not it led to the continuation of the argumentative chain. Instances of communication categorized as "observations" were evaluated based on how strong and relevant the evidence was as well as whether or not it led to the continuation of the argumentative chain. "Interpretations" were analyzed based on how strong and relevant the link to theory was as well as on whether or not it led to the continuation of the argumentative chain. The indicators of quality level in each category of argumentation and examples from the students' verbal communications are presented in Table 1.

Table 1. Evaluation scheme for the quality of an argument used in the video analysis based on Toulmin's [34] model. All examples are from discussions of one of the high-performing groups.

| Argumentative Component | Quality Level | Indicators | Examples |
|---------------------------------|---------------|---|--|
| Question (Claim) | 1 (Qlv1) | No claims are given | "Isn't hatching the thing where they are getting out of their eggs?" |
| | 2 (Qlv2) | Claim fits poorly with theory | "So, what's happening to the eggs (in case of pH changing drastically)?" |
| | 3 (Qlv3) | Claim fits well with theory | "So, the eutrophication . . . Wasn't it the result of some changes in pH values?" |
| | | Leads to the continuation of the argumentative chain | |
| Observation (Evidence) | 1 (Olv1) | No evidence is provided | "Yes, it [eutrophication] was connected to it [pH] somehow". |
| | 2 (Olv2) | Weak evidence is provided | "Oh, it's the flow coming out from the sediment". |
| | 3 (Olv3) | Evidence is relevant and strong | "It's the ecosystem and all that belongs to it that is threatened by this occurrence". |
| | | Leads to the continuation of the argumentative chain | |
| Interpretation (Link to theory) | 1 (Ilv1) | No link to theory is given | "It decreases [number of the eggs] when it's bubbling". |
| | 2 (Ilv2) | Link to theory is weak and factious | "Some species can't live there". |
| | 3 (Ilv3) | Link to theory is relevant and strong | "The pH of seawater . . . is probably in which the planktons are normally living" |
| | | Can lead to the continuation of the argumentative chain | |

Using this final framework of analysis, the videos were coded by the first author, and parallel coding was done by another science teacher with pedagogical competence. The percentage agreement of the coding was 81%, and the Cohen's kappa was 0.713. Thus, the inter-rater agreement can be considered substantial [56]. Differences in analysis were discussed until a common agreement was achieved.

Cross-tabulations and chi-square tests were performed to investigate the significance of the differences between groups with different outcomes and between the phases of task. Statistical significance and effect size are reported using p -value as well as Phi coefficient for 2×2 cross tables and Cramer's V coefficient for larger tables. It should be noted that the effect size is context-dependent and external criteria for effect size values should be interpreted more as guidelines than hard-set rules [57]. For example, it has been suggested that when degrees of freedom is 2, a value of Cramer's V within the range of 0.07–0.21 indicates only a small effect, but in the context of this data, much smaller differences in the instances of communication might be statistically significant.

3. Results

3.1. The Duration of Verbal Communication, On-Topic Talk, and Argumentative Discussion in Relation to the Quality of the Group Outcome

In this study, the focus is on argumentative communication and comparisons between the groups (Figure 1). As expected, the amount and types of verbal communications in the low-, average-, and high-outcome groups, two of each, differed from each other. In the low-outcome groups, a large portion of the discussion was non-argumentative in nature. In contrast, the high-outcome groups had more argumentative talk during the collaborative group work than the average- and low-outcome groups. Although the average quantity of verbal communication was lowest for one of the average-outcome groups, these groups had considerably more argumentative discussion than the low-outcome groups.

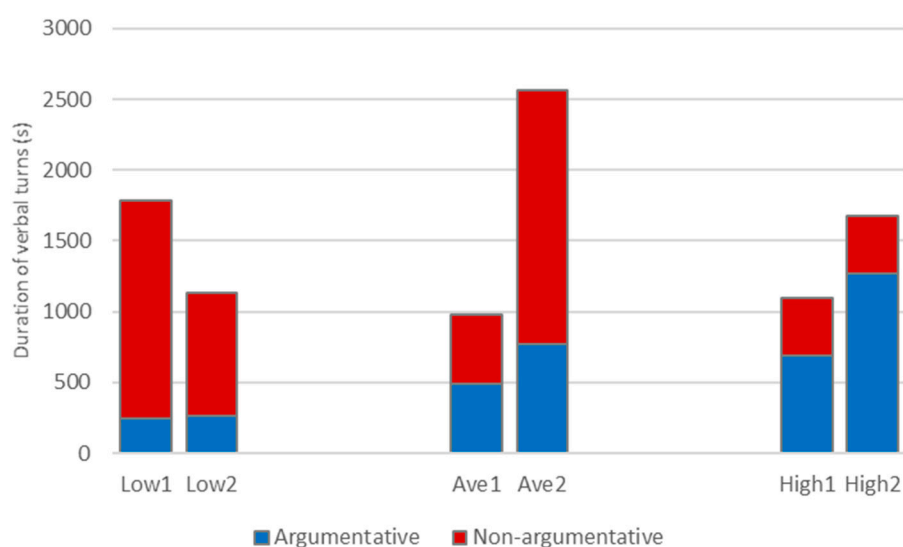


Figure 1. Total summative durations of verbal communication for low-outcome, average-outcome, and high-outcome groups divided into argumentative and non-argumentative components across all argumentative phases (planning, experimenting, and conclusion).

3.2. The Amount of Argumentation in Relation to the Quality of the Outcome of the Collaborative Inquiry Task

Altogether, 6768 instances of verbal communication were analyzed. Of these instances, 1348 (19.9%) were categorized as “on-task argumentative”. The proportion of argumentative instances of verbal communication was lowest for the low-outcome groups (Table 2). There was a significantly lower proportion of on-task argumentation in the low-outcome groups than in the average- and high-outcome groups ($\chi^2(1, N = 6768) = 93.8, p < 0.00001, Phi = 0.12$). There was no statistical difference between the average- and high-outcome groups ($\chi^2(1, N = 4432) = 0.094, p = 0.759$).

Table 2. Instances of verbal communication categorized as “on-task argumentative” or “non-argumentative”.

| Group | On-Task Argumentative | | Non-Argumentative | |
|-------|-----------------------|------|-------------------|------|
| | Frequency | % | Frequency | % |
| Low1 | 194 | 12.5 | 1362 | 87.5 |
| Low2 | 120 | 15.4 | 660 | 84.6 |
| Ave1 | 294 | 22.0 | 1040 | 78.0 |
| Ave2 | 152 | 25.5 | 444 | 74.5 |
| High1 | 358 | 34.8 | 672 | 65.2 |
| High2 | 230 | 15.6 | 1242 | 84.1 |

The frequencies of on-task argumentative instances categorized as “questions”, “observations”, and “interpretations” are presented in Table 3. Compared to the average- and high-outcome groups, the low-outcome groups had a significantly lower proportion of instances categorized as questions ($\chi^2(1, N = 6768) = 30.24, p < 0.00001, Phi = 0.07$) and observations ($\chi^2(1, N = 6768) = 44.98, p < 0.00001, Phi = 0.08$). Again, there were no statistically significant differences between the average- and high-outcome groups, but it should be noted that the two high-outcome groups differed from each other (Table 2).

Table 3. Instances of verbal communication categorized as questions, observations, and interpretations.

| Group | Question | | Observation | | Interpretation | |
|-------|-----------|-----|-------------|------|----------------|-----|
| | Frequency | % | Frequency | % | Frequency | % |
| Low1 | 18 | 1.2 | 128 | 8.2 | 48 | 3.0 |
| Low2 | 24 | 3.1 | 42 | 5.4 | 54 | 6.9 |
| Ave1 | 60 | 4.5 | 164 | 12.3 | 70 | 5.2 |
| Ave2 | 32 | 5.4 | 96 | 16.1 | 24 | 4.0 |
| High1 | 60 | 5.8 | 204 | 19.8 | 94 | 9.1 |
| High2 | 42 | 2.9 | 124 | 8.4 | 64 | 4.3 |

Compared to the high-outcome groups, all of these argumentative components were present in the average-outcome groups but in lesser numbers. Particularly, the most qualified components (e.g., observation level 3 and interpretation level 3) were missing. Although they were not common, even in the high-outcome groups, they still could be found there. The low-outcome groups had few if any demonstrations of these average-outcome categories and were most commonly in the non-argumentative zone. Only a small portion of the topic-relevant discussion was argumentative, and the quantity of argumentation also seemed lesser compared to the average- and high-outcome groups Table 3.

3.3. The Quality of the Argumentation in Relation to the Quality of the Group Outcome in a Collaborative Inquiry Task

The quality of argumentation in each category was evaluated using three levels of quality (Figure 2). There was a significant relationship between the quality of the argumentation and the quality of the group outcome ($\chi^2(4, N = 1348) = 30.46, p < 0.00001, Cramer's V = 0.11$). As expected, student groups using argumentation in which the claims fit the theory well and where strong evidence was provided had higher-quality outcomes. The level of argumentation was highest in the high-outcome groups (see Table 1), although the difference with the average-outcome groups was not statistically significant ($\chi^2(2, N = 1034) = 1.333, p = 0.514, Cramer's V = 0.04$) (Table 4).

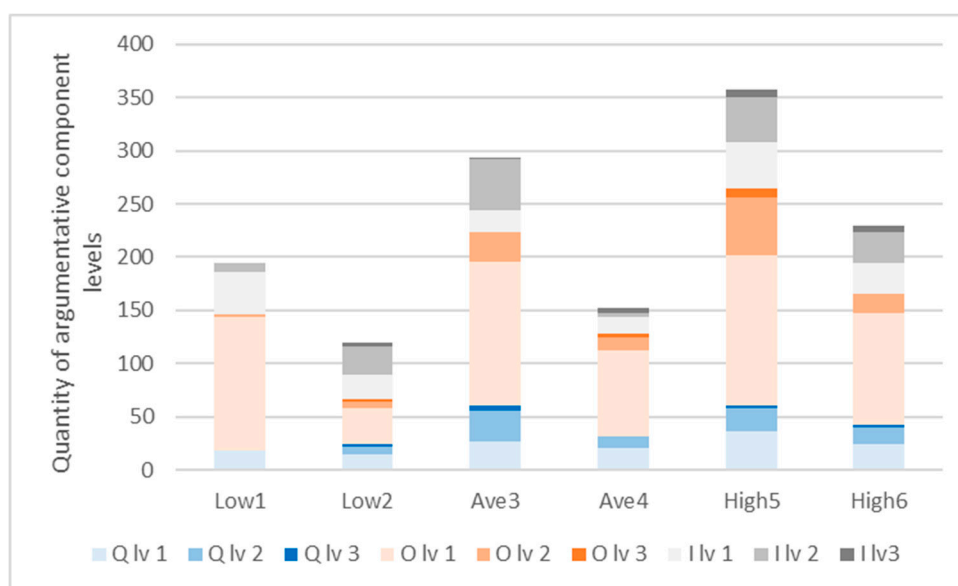


Figure 2. Quantity of argumentative components in each of the six groups (Q lv1–3 is the component question, O lv1–3 is the component observation, and I lv1–3 is the component interpretation with its level).

Table 4. Quality of instances of on-task argumentation (N = 1348) for the low-, average-, and high-outcome groups.

| Group | Level 1 | | Level 2 | | Level 3 | |
|-------|-----------|------|-----------|------|-----------|-----|
| | Frequency | % | Frequency | % | Frequency | % |
| Low | 256 | 81.5 | 50 | 15.9 | 8 | 2.5 |
| Ave | 298 | 66.8 | 134 | 30.0 | 14 | 3.1 |
| High | 380 | 64.6 | 182 | 30.9 | 26 | 4.4 |

Deeper analysis of student discussion (Figure 2) revealed that the high-outcome groups asked more questions than their counterparts and had an urge to interpret their observations and link them to the theory (see examples in Table 1). Meanwhile, the low-outcome groups were poor at asking questions and thus at promoting argumentation within their group. This might also be the reason that the low-outcome groups had few if any questions of the highest quality. The average- and high-outcome groups made high-quality observations. There was a trend of increasing quantity of argumentative components from the low-outcome to high-outcome groups, but there were differences between the groups.

3.4. The Changes in the Amount and Quality of Argumentation during the Three Argumentative Phases (Planning, Experimenting, and Conclusion) of a Collaborative Inquiry Task

There was a significant relationship between the phase of the collaborative inquiry task and the amount of argumentation ($\chi^2 (2, N = 6768) = 41.20, p < 0.00001, \text{Cramer's } V = 0.08$). The proportion of argumentation decreased in each consecutive phase (Table 5). There was also a significant relationship between the phase and the quality of argumentation ($\chi^2 (4, N = 1348) = 43.31, p < 0.00001, \text{Cramer's } V = 0.13$) as well as between the phase and the components of argumentation ($\chi^2 (4, N = 1348) = 20.72, p = 0.00036, \text{Cramer's } V = 0.09$).

Table 5. Instances of verbal communication categorized as “on-task argumentative” or “non-argumentative” during each phase of the collaborative inquiry task.

| Phase | On-Task Argumentative | | Non-Argumentative | |
|---------------|-----------------------|------|-------------------|------|
| | Frequency | % | Frequency | % |
| Planning | 518 | 22.8 | 1750 | 77.2 |
| Experimenting | 388 | 22.0 | 1376 | 78.0 |
| Conclusions | 442 | 16.2 | 2294 | 83.8 |

The planning phase seemed to offer the best environment for rich argumentative and topic-centered discussion. It contained the highest proportion of verbal communication categorized as on-task argumentation (Table 5). The proportion of instances with level 2 and 3 argumentative quality was also significantly higher than during the other phases ($\chi^2(2, N = 1348) = 18.58, p = 0.000093$, Cramer’s $V = 0.12$) (Table 6).

Table 6. Quality of instances of on-task argumentation ($N = 1348$) during each phase of the collaborative inquiry task.

| Phase | Level 1 | | Level 2 | | Level 3 | |
|---------------|-----------|------|-----------|------|-----------|-----|
| | Frequency | % | Frequency | % | Frequency | % |
| Planning | 324 | 62.5 | 174 | 33.6 | 20 | 3.9 |
| Experimenting | 316 | 81.4 | 58 | 14.9 | 14 | 3.6 |
| Conclusions | 294 | 66.5 | 134 | 30.3 | 14 | 3.2 |

The proportion of argumentative instances of verbal communication during the experimenting phase was almost as high as during the planning phase. However, most of these instances were observations and questions (Table 7). The number of interpretations was lower during the experimenting phase than during the other phases. Compared with the other phases, the quality of argumentative instances was also significantly lower ($\chi^2(2, N = 1348) = 41.50, p < 0.00001$, Cramer’s $V = 0.18$).

Table 7. Instances of verbal communication categorized as questions, observations, and interpretations during each phase of the collaborative inquiry task.

| Phase | Question | | Observation | | Interpretation | |
|---------------|-----------|------|-------------|------|----------------|------|
| | Frequency | % | Frequency | % | Frequency | % |
| Planning | 98 | 18.9 | 276 | 53.3 | 144 | 27.8 |
| Experimenting | 82 | 21.1 | 230 | 59.3 | 76 | 19.6 |
| Conclusions | 56 | 12.7 | 252 | 57.0 | 134 | 30.3 |

In theory, the conclusion phase could be a good phase for argumentative discussion, but the results showed that most groups just repeated the results they received from the software without much discussion. During the conclusion phase, the argumentation focused on interpretations, and the proportion of questions was lowest of all the phases (Table 7). The proportion of instances of on-task argumentation during this phase was significantly lower than during the planning and experimenting phases ($\chi^2(1, N = 6768) = 40.76, p < 0.00001, Phi = 0.08$), but the quality of argumentation was nearly as high as during the planning phase (Table 5).

Some connections could also be seen between the differences in argumentation within each phase and the outcomes of the groups (Figure 2). This could be due to the different natures of the phases, as the second phase primarily involved carrying out the experiment and focused little on planning and interpreting the materials provided by the virtual laboratory, as was the case in the first and last phases. For example, the instances of argumentation for the second low-outcome group went down as the project proceeded, even though the instances and quality of argumentation were in line

with the higher-outcome groups during the first step of the collaborative inquiry task. Although the downward trend in the quality of argumentation could be seen in the average- and high-outcome groups as well, the change was not nearly as sharp.

4. Discussion

The aim of this study was to analyze the quality and quantity of upper secondary school student groups' spontaneous argumentative talk and to understand their relation to the groups' joint learning outcomes during a collaborative interdisciplinary science learning inquiry task. This aim was achieved by creating a coding scheme to code spoken argumentation from video-taped material. Thus, the present study contributes to the current literature both conceptually and methodologically. The coding scheme implemented for this purpose contained the essential aspects of Toulmin's [34] argumentative components: claim, observation, and interpretation. This decision was supported by the literature [36,38] and the goal to code argumentation fluently yet rigorously. The coding scheme worked in the sense that all the argumentative components were present in the findings and in students' discussions. It also showed the differences between each working phase (e.g., generating a hypothesis or writing a study plan) in a clear way.

The main findings of this study were that there was a difference in the argumentative discussions of the low-, average-, and high-outcome groups, while the group rating was solely based on the outcome. This is supported by the literature, which has shown that longer and more intense student discussion on a topic leads to a deeper understanding and learning [47,48]. Additionally, when classroom activities change toward argumentation practices students can generate better learning outcomes [58], as found in the present study. The other important finding was the longer duration of argumentation in the high- and low-outcome groups and its relation to better outcomes. This finding is in line with previous research. Thus, a longer working phase helps students generate more elaborate arguments and interpretation based on scientific data [59]. This is also true for the duration of the intervention, which could have been longer in this study. Even so, the results are promising. In another similar study in which students needed to construct a representation individually, discuss the topic, and then write a representation together, the students who engaged in deep discussion and argumentation learned more than their counterparts who did not [60]. Notably, in that study the setting consisted of three phases (construction of a presentation, discussion on the topic, and consolidation), similar to the present study. In conclusion, at the end of this study a novel coding scheme was introduced for spoken spontaneous group argumentation in argumentative phases (question, observation, and interpretation).

Distinct differences in argumentation chains between variously performing groups of students were observed. The high-performing groups had high-end episodes more frequently and for longer times than the average- and low-performing groups. In the high-performing groups, the students also asked topic-related questions more frequently, which started the argumentative discussion, whereas in the weak-performing groups there were few questions asked, and most of them did not lead to a longer discussion, which could have been valuable in terms of the final product and argumentation [24,59,60]. An evaluation scheme for the quality of the arguments used concerning spoken argumentation was created, and verbal and non-verbal categories were detected and modified using a computer-based analysis program. Sandoval et al. [32] demonstrated teachers' importance and accountability in spoken classroom argumentation processes and in guiding elementary students through this process.

The upper secondary students had to take the responsibility in small groups to resolve issues in the virtual learning environment. They were asked to be active (e.g., generate a question/claim) and to keep the focus on the key issues to resolve and create a collaborative culture, which obviously begins by framing science activity. The students could improve the criteria and framework for their learning as well. They needed to take accountability for the issues in the small groups, which the teacher normally does, as in the study of Sandoval et al. [32]. These issues are tasks that are demanding for individual learners, but in groups the students manage to work toward the learning goals in science using argumentative discussion.

Thus, this study being positioned within the context of the Baltic Sea, there is an interesting connection between scientific argumentation and solving the problems presented in the virtual learning environment. In relation to socio-scientific issues, students can generate different types of arguments: relative risk arguments as a prototype for decision-making, precautionary arguments prioritizing the avoidance of risks and decisions, and impossible arguments based on static factual knowledge about the possible risk [35]. In the present study, this kind of decision-making could also explain the results, which included a discussion on ways to enhance the situation in the Baltic Sea based on real statistical data gathered using the software. The collaborative virtual learning environment made it possible for students to take an active role in argumentation and carry out the study tasks, thus supporting scientific argumentation instead of classroom-type argumentation. These findings are congruent with those of Meyer [10], and the study environment seemed to support students' accountability in understanding experimental marine science issues and to engage them to work productively.

Argumentation requires changes in discourse regarding classroom practices to support PDE and accountability [32]. The present study supports this idea, as the collaborative virtual learning environment made it possible for students to take an active role in argumentation and carry out the study tasks, thus supporting scientific argumentation rather than classroom-type argumentation. This is in line with the findings of Engle [9] and Meyer [10], who showed that, when students are allowed to work with their own ideas, they exhibit scientifically oriented PDE instead of classroom-like PDE [10].

5. Methodological Considerations

The argumentative types and quality of arguments were analyzed based on Toulmin's scheme [34]. The analyses showed that the high-performing group had more argumentative questions, interpretations, and argumentative evidence in higher levels than the other two groups and spent less time on off-tasks issues. Sandoval et al. [32] showed that teachers have to guide students through the spoken argumentation process. For the upper secondary school students in the study, this means that they needed to be active (e.g., generate a question/claim), stay focused on the key science issues to be resolved, and follow scientific experimental research steps, which obviously begins by framing science activity and creating a collaborative culture in which students need to improve the criteria and framework for their learning. The students had to carry out all this in small-group discussions, reflecting the students' productive disciplinary engagement, especially in the high-performing groups. By combining discourse practices specifically with a collective sense-making goal, Sandoval et al. [32] promoted productive argumentation in classroom-type spoken situations. Their result is in line with the spoken spontaneous argumentation of the small groups in the virtual learning environment in this study.

Here, Toulmin's [34] traditional scheme that somewhat dominates the scientific argumentation research was used as a basis for the coding scheme because of its clear, easy-to-analyze structure. The natural scientific aspect of the scheme was introduced by the authors, which was suitable for analyzing science lessons in a virtual environment. It could not have been too complex or else the data analysis would have been prone to miscoding. As the results showed, this scheme enabled analysis of the argumentation in the argumentative phases (planning, experimenting, and making conclusions) in the ViBSE collaborative virtual learning environment.

The most important goal of using this coding system was to achieve a balance between the depth and complexity of the argumentation-coding model and the fluency of coding. One of the most challenging issues with the frameworks is that they lack a template for spontaneous, spoken argumentation. Driver, Newton, and Osborne [6] introduced argumentation as an important discursive practice, and since then written scientific argumentation has received most of the focus in research papers (e.g., [7,61]). There are hardly any studies on spoken argumentation in science (e.g., [15,32]) and few describing disciplinary engagement as productive [30].

The reliability across different researchers—inter-rater reliability—was good in this study since the coding category was created and discussed together and the coding agreement was relatively high [56,62]. The inter-coder reliability and agreement were therefore good as well. The definitions of codes and components of argumentation were also consistent in this study, and much attention was given to ensure that they were consistent and understandable for both the readers and researchers. The over-time reliability was there in the form of three to five lessons and three phases, but the study was not repeated with the same group after a certain time, which is a drawback. Further, the content validity supported the good reliability because the chosen argumentation coding scheme included critical components from argumentation theory. Criterion validity was intact since the argumentation coding yielded results that were line with the outcomes of the groups. Therefore, the argumentation was also actualized in a real outcome. These results are transferable to other science lessons in which argumentation is not explicitly required and instructed. There should be virtual software to provide students with a common ground for collaborative group work. However, there is a problem with transferability, as there were only six groups for the statistical analyses. So, in the strict sense, the statistical significance is valid only for these six groups and not for all groups. Therefore, there might be individual differences between the students that can affect the outcome of the whole group. This was tackled to some extent beforehand by mixing different learners to groups in collaboration with the teachers. Moreover, the focus was solely on collaborative learning, argumentation, and group outcome. Hence, studies with larger samples are needed in the future.

In sum, the argumentation of the three outcome-level groups differed in correlation to their outcomes, and there was a difference between the low- and high-outcome groups. To demonstrate this, a coding scheme was created, which worked sufficiently within the context of a virtual environment, where the topic of discussion changes quickly according to the software. In a virtual environment and in analyzing spoken, spontaneous argumentation, a more detailed scheme would be too diverse to use and would not provide much more information.

6. Conclusions and Implications

To conclude, the argumentation coding scheme presented in this study yielded results that were in line with the assumption that better argumentative discussion would lead to better qualitative outcomes. Moreover, this study presented a light and useful yet precise way to code student argumentation in a spoken format and in the context of collaborative group work. The virtual tool enabled students to argue even though they were not instructed to do so, which was a finding in itself. This showed that a socio-scientific topic in science could promote argumentation if it is implemented in a lesson (e.g., in a discussion-promoting way that encourages students to collaborate). Then, it could engage all the students in the discussion. However, it must be noted that the spectrum of theories delving into scientific argumentation from the viewpoint of student activity is enormous [42]. Therefore, there is no single theory that could explain all the results shown earlier. On the other hand, this study brings new findings to the discussion on scientific argumentation with its unique study design and focus on spoken argumentation during a collaborative inquiry task.

In the current study, if the turns formed a chain of argumentation around the same issue and on the same quality level they were coded as episodes. Episodes combined with rigorous analyses focusing on shared metacognitive regulation will be the main focus in the next article of the first author of this article. In future studies, the connections between pre- and post-tests, especially the topic-specific epistemic beliefs questionnaire, could be analyzed in relation to argumentation coding. The epistemic beliefs questionnaire is of particular interest since the literature [63] indicates that there should be a strong connection between students' argumentation and epistemic beliefs. These results and argumentation analysis together with the results from this same data but with affect in focus [64,65] can be used to benefit teacher education and to provide teachers with tools to observe and evaluate their students' argumentative discussion, which is at the heart of science education. Therefore, the findings can be utilized in designing the science learning situations, the collaborative tasks during the lessons,

as well as the instructions the teachers give to their students. Teachers should be aware that certain tasks can spontaneously promote discussion and argumentation. Such task can be used to enhance science learning and the adoption of the culture in which the science is carried out. The results of this study support a new learning culture that values a group's joint learning outcome instead of an individual learner's outcome, thus leading to true collaborative learning. This dawn of a new learning culture is inevitably facilitated by a virtual learning environment that allows students to figure and follow the steps from hypotheses to results and conduct research that is carefully crafted on the principles of inquiry learning.

Author Contributions: Conceptualization, M.T., E.Y.-P., V.-M.V. and M.V.; Methodology, M.T., E.Y.-P., V.-M.V. and M.V.; Formal Analysis, M.T. and V.-M.V.; Investigation, planning and design from the beginning M.T., E.Y.-P. and M.V.; Writing—Original Draft Preparation M.T., E.Y.-P., V.-M.V. and M.V.; Review & Editing, M.T., E.Y.-P., V.-M.V. and M.V.; Visualization, M.T., E.Y.-P., V.-M.V. and M.V.; Supervision, E.Y.-P., V.-M.V. and M.V.; Project Administration, M.V.; Funding Acquisition, SciLes-project, principal investigator M.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The Academy of Finland, 274117, SciLes project.

Acknowledgments: The authors want to thank the whole SciLes-project team and the teachers and upper secondary school students who participated in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lakatos, I. *Proofs and Refutations: The Logic of Mathematical Discovery*; Cambridge University Press: Cambridge, UK, 1976.
2. Sadler, T.D. Informal reasoning regarding socioscientific issues: A critical review of research. *J. Res. Sci. Teach.* **2004**, *41*, 513–536. [[CrossRef](#)]
3. Jiménez-Aleixandre, M.P.; Puig, B. Argumentation, Evidence, Evaluation and Critical Thinking. In *Second International Handbook of Science Education*; Fraser, B.J., Tobin, K., McRobbie, C.J., Eds.; Springer: Dordrecht, The Netherlands, 2012; pp. 1001–1015. [[CrossRef](#)]
4. Norris, S.P.; Philips, L.M. How literacy in its fundamental sense is central to scientific literacy. *Sci. Educ.* **2003**, *87*, 224–240. [[CrossRef](#)]
5. Sampson, V.; Clark, D.B. Assessment of the Ways Students Generate Arguments in Science Education: Current Perspectives and Recommendations for Future Directions. *Int. Sci. Educ.* **2008**, *92*, 447–472. [[CrossRef](#)]
6. Driver, R.; Newton, P.; Osborne, J. Establishing the Norms of Scientific Argumentation in classrooms. *Sci. Educ.* **2000**, *84*, 287–312. [[CrossRef](#)]
7. Duschl, R.A.; Osborne, J. Supporting and Promoting Argumentation Discourse in Science Education. *Stud. Sci. Educ.* **2002**, *38*, 39–72. [[CrossRef](#)]
8. Erduran, S.; Ozdem, Y.; Park, J.-Y. Research Trends on Argumentation in Science Education: A Journal Content Analysis from 1998–2014. *Int. J. STEM Educ.* **2015**, *2*, 1–12. [[CrossRef](#)]
9. Engle, R.A. The Productive Disciplinary Engagement Framework: Origins, Key Concepts and Developments. In *Design Research on Learning and Thinking in Educational Settings: Enhancing Intellectual Growth and Functioning*; Dai, D.Y., Ed.; Taylor & Francis: London, UK, 2011; pp. 161–200.
10. Meyer, X. Productive Disciplinary Engagement as a Recursive Process: Initial Engagement in a Scientific Investigation as a Resource for Deeper Engagement in the Scientific Discipline. *Int. J. Educ. Res.* **2014**, *64*, 184–198.
11. Duschl, R.; Ellenbogen, K.; Erduran, S. Understanding Dialogic Argumentation. In Proceedings of the Annual Meeting of the American Educational Research Association, Montreal, QC, Canada, 19–23 April 1999.
12. Forman, E.A. Discourse, Intersubjectivity and the Development of Peer Collaboration: A Vygotskian Approach. In *Children's Development within Social Contexts: Metatheoretical, Theoretical and Methodological Issues*; Winegar, L.T., Valsiner, J., Eds.; Erlbaum: Hillsdale, NJ, USA, 1992; Volume 1, pp. 143–159.
13. Jiménez-Aleixandre, M.; Rodríguez, A.; Duschl, R. "Doing the Lesson" or "Doing Science": Argument in High School Genetics. *Sci. Educ.* **2000**, *84*, 757–792. [[CrossRef](#)]

14. Kelly, G.; Takao, A. Epistemic Levels in Argument: An Analysis of University Oceanography Students' Use of Evidence in Writing. *Sci. Educ.* **2002**, *86*, 314–342. [[CrossRef](#)]
15. Telenius, M.; Yli-Panula, E.; Ahtineva, A.; Vauras, M. Collaborative Science Lessons—Learning and Argumentation in an Interdisciplinary Virtual Laboratory. In *Tutkimuksesta Luokkahuoneisiin*; Rautiainen, M., Tarnanen, M., Eds.; Suomen Ainedidaktinen Tutkimusseura: Jyväskylä, Finland, 2019; Series 15; pp. 35–56. Available online: <http://hdl.handle.net/10138/298542> (accessed on 18 December 2020).
16. Yli-Panula, E.; Hiilovaara-Teijo, M.; Vauras, M. High school students' inquiry based collaborative learning in virtual marine science laboratory. In *Suomen Ainedidaktisen Tutkimusseuran Julkaisuja, Ainedidaktisia Tutkimuksia [Finnish Research Association for Subject Didactics, Studies in Subject Didactics]*; Kauppinen, M., Rautiainen, M., Tarnane, M., Eds.; Suomen Ainedidaktinen Tutkimusseura ry: Helsinki, Finland, 2015; Series 8; pp. 135–153. Available online: <http://hdl.handle.net/10138/153212> (accessed on 18 December 2020).
17. Özer Keskin, M.; Keskin Samanci, N.; Yaman, H. Argumentation Based Bioethics Education: Sample Implementation on Genetically Modified Organisms (GMOs) and Genetic Screening Tests. *Educ. Res. Rev.* **2013**, *8*, 1383–1391. [[CrossRef](#)]
18. Norris, S.P.; Philips, L.M.; Osborne, J. Scientific inquiry: The place of interpretation and argumentation. In *Science as Inquiry in the Secondary Setting*; NSTA Press: Arlington, VA, USA, 2007; pp. 87–98.
19. van Eemeren, F.H.; Grootendorst, R. *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*; Cambridge University Press: Cambridge, UK, 2004.
20. Belland, B.R.; Gu, J.; Kim, N.J.; Turner, D.J. An Ethnomethodological Perspective on How Middle School Students Addressed a Water Quality Problem. *Educ. Technol. Res. Dev.* **2016**, *64*, 1135–1161. [[CrossRef](#)]
21. Belland, B.R.; Gu, J.; Armbrust, S.; Cook, B. Scaffolding Argumentation about Water Quality: A Mixed-Method Study in a Rural Middle School. *Educ. Technol. Res. Dev.* **2015**, *63*, 325–353. [[CrossRef](#)]
22. Erduran, S.; Osborne, J.; Erduran, S.; Osborne, J. Developing arguments. In *Analysing Exemplary Science Teaching: Theoretical Lenses and a Spectrum of Possibilities for Practice*; Alsop, S., Bencze, L., Pedretti, E., Eds.; Open University Press: Philadelphia, PA, USA, 2005.
23. Schwarz, C.V.; Reiser, B.J.; Davis, E.A.; Kenyon, L.; Acher, A.; Fortus, D.; Schwartz, Y.; Hug, B.; Krajcik, J. Developing a Learning Progression for Scientific Modeling: Making Scientific Modeling Accessible and Meaningful for Learners. *J. Res. Sci. Teach.* **2009**, *46*, 632–654. [[CrossRef](#)]
24. Andriessen, J.E.; Schwarz, B.B. Argumentative Design. In *Argumentation and Education*; Muller, M.N., Perret-Clermont, A.N., Eds.; Springer: Boston, MA, USA, 2009. [[CrossRef](#)]
25. Weinberger, A.; Stegmann, K.; Fischer, F.; Mandl, H. Scripting Argumentative Knowledge Construction in Computer-Supported Learning Environments. In *Scripting Computer-Supported Collaborative Learning*; Fischer, F., Kollar, I., Mandl, H., Haake, J.M., Eds.; Springer: Boston, MA, USA, 2007; Volume 6. [[CrossRef](#)]
26. Schwarz, B.B.; Baker, M.J. (Eds.) *Argumentation Theory for Education in Dialogue, Argumentation and Education: History, Theory and Practice*; Cambridge University Press: Cambridge, UK, 2017.
27. Chinn, C.A. Learning to Argue. In *The Rutgers Invitation Symposium on Education Series. Collaborative Learning, Reasoning, and Technology*; O'Donnell, A.M., Hmelo-Silver, C.E., Erkens, G., Eds.; Lawrence Erlbaum Associates Publishers: Hillsdale, NJ, USA, 2006; pp. 355–383.
28. Walker, J.P.; Sampson, V. Learning to Argue and Arguing to Learn: Argument-Driven Inquiry as a Way to Help Undergraduate Chemistry Students Learn How to Construct Arguments and Engage in Argumentation During a Laboratory Course. *J. Res. Sci. Teach.* **2013**, *50*, 561–596. [[CrossRef](#)]
29. Koretsky, M.D.; Vauras, M.; Jones, C.; Iiskala, T.; Volet, S. Productive Disciplinary Engagement in High and Low-Outcome Student Groups: Observations from Three Collaborative Science Learning Contexts. *Res. Sci. Educ.* **2019**, *138*, 1–24. [[CrossRef](#)]
30. Engle, R.; Conant, F. Guiding Principles for Fostering Productive Disciplinary Engagement: Explaining an Emergent Argument in a Community of Learners Classroom. *Cogn. Instr.* **2002**, *20*, 399–483. Available online: <http://www.jstor.org/stable/3233901> (accessed on 9 September 2020). [[CrossRef](#)]
31. Lundqvist, E.; Sund, P. Selective Traditions in Group Discussions: Teachers' Views about Good Science and the Possible Obstacles when Encountering a New Topic. *Cult. Stud. Sci. Educ.* **2018**, *13*, 353–370. [[CrossRef](#)]
32. Sandoval, W.A.; Enyedy, N.; Redman, E.H.; Xiao, S. Organising a Culture of argumentation in Elementary Science. *Int. J. Sci. Educ.* **2019**, *41*, 1848–1869. [[CrossRef](#)]

33. Ong, Y.S.; Duschl, R.A.; Plummer, J.D. Scientific Argumentation as an Epistemic Practice: Secondary Students' Critique of Science Research Posters. In *Science Education in the 21st Century*; Teo, T., Tan, A.L., Ong, Y., Eds.; Springer: Singapore, 2020. [\[CrossRef\]](#)
34. Toulmin, S.E. *The Uses of Argument*; Cambridge University Press: Cambridge, UK, 1958.
35. Kolstø, S.D. Patterns in Students' Argumentation Confronted with a Risk-focused Socio-scientific Issue. *Int. J. Sci. Educ.* **2006**, *28*, 1689–1716. [\[CrossRef\]](#)
36. Osborne, J.; Erduran, S.; Simon, S. Enhancing the Quality of Argumentation in School Science. *J. Res. Sci. Teach.* **2004**, *42*, 994–1020. [\[CrossRef\]](#)
37. Polo, C.; Lund, K.; Plantin, C.; Niccolai, G.P. Group Emotions: The Social and Cognitive Functions of Emotions in Argumentation. *Int. J. Comput. Support Collab. Learn.* **2016**, *11*, 123–156. [\[CrossRef\]](#)
38. Evagorou, M.; Osborne, J. Exploring Young Students' Collaborative Argumentation within a Socioscientific Issue. *J. Res. Sci. Teach.* **2013**, *50*, 209–237. [\[CrossRef\]](#)
39. Kim, H.; Song, J. The Features of Peer Argumentation in Middle School Students' Scientific Inquiry. *Res. Sci. Educ.* **2006**, *36*, 211–233. [\[CrossRef\]](#)
40. McDonald, C.V. The Influence of Explicit Nature of Science and Argumentation Instruction on Presentative Primary Teachers' View of Nature of Science. *J. Res. Sci. Teach.* **2010**, *47*, 1137–1164. [\[CrossRef\]](#)
41. McNeill, K.L.; Pimentel, D.S. Scientific discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Sci. Educ.* **2010**, *94*, 203–229. [\[CrossRef\]](#)
42. Manz, E. Representing Student Argumentation as Functionally Emergent from Scientific Activity. *Rev. Educ. Res.* **2015**, *85*, 553–590. [\[CrossRef\]](#)
43. Gibbs, G.R. Learning How to Learn Using a Virtual Learning Environment for Philosophy. *J. Comput. Assist. Learn.* **1999**, *15*, 221–231. [\[CrossRef\]](#)
44. Fauville, G.; Hodin, J.; Dupont, S.; Miller, P.; Haws, J.; Thorndyke, M.; Epel, D. Virtual ocean acidification laboratory. In *The Economic, Social and Political Elements of Climate Change*; Filho, W.L., Ed.; Springer: Berlin, Germany, 2011; pp. 825–836.
45. Kinnunen, R.; Telenius, M.; Scheinin, M.; Lehtinen, E.; Vauras, M. *ViBSE, Virtual Baltic Sea Explorer. Virtual Learning Environment Supporting Interdisciplinary Understanding of Chemistry, Connected to Marine Biology and Environmental Challenges*; University of Turku: Turku, Finland, 2018.
46. Sampson, V.; Grooms, J.; Walker, J.P. Argument-driven Inquiry as a Way to Help Students Learn How to Participate in Scientific Argumentation and Craft Written Arguments: An Exploratory Study. *Sci. Educ.* **2011**, *95*, 217–257. [\[CrossRef\]](#)
47. Zhao, J.; Lin, L.; Sun, J.; Zheng, X.; Yin, J. Students' Engagement in a Science Classroom: Does Knowledge Diversity Matter? *J. Educ. Res.* **2018**, *111*, 756–763. [\[CrossRef\]](#)
48. Sampson, V.; Clark, D. The Impact of Collaboration on the Outcomes of Scientific Argumentation. *Sci. Educ.* **2009**, *93*, 448–484. [\[CrossRef\]](#)
49. Yun, S.M.; Kim, H.B. Changes in Students' Participation and Small Group Norms in Scientific Argumentation. *Res. Sci. Educ.* **2015**, *45*, 465–484. [\[CrossRef\]](#)
50. Chinn, C.; Clark, D. Learning through Collaborative Argumentation. In *International Handbook of Collaborative Learning*; Hmelo Silver, C.E., Chinn, C.A., Chan, C.K.K., O'Donnell, A.M., Eds.; Taylor & Francis: New York, NY, USA, 2013; pp. 314–332.
51. Vauras, M.; Volet, S.; Nolen, S. Supporting Motivation in Collaborative Learning: Challenges in the Face of an Uncertain Future. In *Motivation in Education at a Time of Global Change: Theory, Research, and Implications for Practice. Advances in Motivation and Achievement*; Gonida, E., Lemos, M., Eds.; Emerald: New York, NY, USA, 2019; Volume 20, pp. 187–203. [\[CrossRef\]](#)
52. Bonaglia, S.; Bartoli, M.; Gunnarsson, J.S.; Rahm, L.; Raymond, C.; Svensson, O.; Yekta, S.S.; Brüchert, V. Effect of Reoxygenation and *Marenzelleria* spp. Bioturbation on Baltic Sea Sediment Metabolism. *Mar. Ecol. Prog. Ser.* **2013**, *482*, 43–55. [\[CrossRef\]](#)
53. Engeström-Öst, J.; Holmborn, T.; Brutemark, A.; Hogfors, H.; Vehmaa, A.; Gorokhova, E. The Effects of Short-term pH decrease on the Reproductive Output of the Copepod *Acartia Bifilosa*—A Laboratory Study. *Mar. Freshw. Behav. Physiol.* **2014**, *47*, 173–183. [\[CrossRef\]](#)
54. Peeters, M.J.; Sahloff, E.G.; Stone, G.E. A standardized rubric to evaluate student presentations. *Am. J. Pharm. Educ.* **2010**, *9*, 1–8. [\[CrossRef\]](#)
55. Sampson, V.; Schleigh, S. *Scientific Argumentation in Biology*; NSTA Press: Arlington, VA, USA, 2013.

56. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)]
57. Sun, S.; Pan, W.; Wang, L.L. A comprehensive review of effect size reporting and interpreting practices in academic journals in education and psychology. *J. Educ. Psychol.* **2010**, *102*, 989. [[CrossRef](#)]
58. Chen, Y.-C.; Benus, M.J.; Yarker, M.B. Using Models to Support Argumentation in the Science Classroom. *Am. Biol. Teach.* **2016**, *78*, 549–559. [[CrossRef](#)]
59. Georgiou, M.; Mavrikaki, E.; Halkia, K.; Papassideri, I. Investigating the Impact of the Duration of Engagement in Socioscientific Issues in Developing Greek Students' Argumentation and Informal Reasoning Skills. *Am. J. Educ. Res.* **2020**, *8*, 16–23. [[CrossRef](#)]
60. Van Amelsvoort, M.; Andriessen, J.; Kanselaar, G. Representational Tools in Computer-Supported Collaborative Argumentation-Based Learning: How Dyads Work with Constructed and Inspected Argumentative Diagrams. *J. Learn. Sci.* **2007**, *16*, 485–521. [[CrossRef](#)]
61. Zohar, A.; Nemet, F. Fostering Students' Knowledge and Argumentation Skills through Dilemmas in Human Genetics. *J. Res. Sci. Teach.* **2002**, *39*, 35–62. [[CrossRef](#)]
62. Anastasi, A.; Urbina, S. *Psychological Testing*, 7th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1997.
63. Duschl, R. Science Education in Three-Part Harmony: Balancing Conceptual, Epistemic, and Social Learning Goals. *Rev. Res. Educ.* **2008**, *32*, 268–291. [[CrossRef](#)]
64. Pietarinen, T.; Vauras, M.; Laakkonen, E.; Kinnunen, R.; Volet, S. High School Students' Perceptions of Affect and Collaboration during Virtual Science Inquiry Learning. *J. Comput. Assist. Learn.* **2019**, *35*, 334–348. [[CrossRef](#)]
65. Pietarinen, T.; Volet, S.; Lehtinen, E.; Vauras, M. Affect in Peer Group Learning during Virtual Science Inquiry: Insights from Self-reports and Video Observations. *Front. Psychol.* **2020**, *14*. [[CrossRef](#)] [[PubMed](#)]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).