education sciences

MDPI

Article

How Digital and Oral Peer Feedback Improves High School Students' Written Argumentation—A Case Study Exploring the Effectiveness of Peer Feedback in Geography

Michael Morawski * and Alexandra Budke

Institute of Geography Education, University of Cologne, Albertus-Magnus-Platz, 50923 Köln, Germany
* Correspondence: michael.morawski@gmx.de

Received: 18 April 2019; Accepted: 28 June 2019; Published: 10 July 2019

Abstract: This article approaches written argumentation as a concept of promoting geographical literacy. It is argued that student-centered peer feedback is an effective method with which to improve individual students' argumentative texts. This research uses a case-study design, which analyzed how high school students in different pairs improve their argumentation text under subject-specific criteria. For the feedback process, a subject-specific feedback sheet for students has been designed for them to review their partner's argumentative text. The findings mainly suggest two outcomes: Different kinds of feedback in terms of interaction, content and argumentative integration of text material lead to text improvement, and that there are varying complexities of feedback acceptance in terms of subject-specific criteria. The results provide a deeper insight into how students can be prepared and rewarded for producing qualitatively high and effective feedback on argumentative texts in socio-scientific contexts with a strong focus on the (linguistic) skills they need for these procedures.

Keywords: argumentation; geography; feedback; peer review

1. Introduction

Should I buy clothes from *Primark*? Should I get the book "*Turtles all the way down*" from *Amazon* or from the shop around the corner? Should I develop my knowledge of climate change and is it really all about China? Why is my city bulldozing the park nearby? Such socio-scientific, cultural, political and personal issues are faced by students every day. The role of geography education is to make students aware of these issues and capable of acting on them, by enabling them to make, justify, understand and reflect on their decisions in social processes of negotiation. In a digitalized and globalized world, students need argumentation competences more than ever to obtain, evaluate and communicate information, in order to become democratically active, autonomous and empathic social individuals in cooperative societies [1].

1.1. Research Background

There has been a paradigm shift in education towards "teaching Science as an argument/practice" in the United States or "competence-based subject learning" in parts of Europe over the last couple of years, which may help to pave the way to addressing new and old challenges [2–4]. These shifts not only develop a more challenging and cognitive demanding practice in schools, they also emphasize the necessity to make teachers more aware of "disciplinary literacy" when teaching science, in which the abilities "to decode and interpret more complex forms of text, to recognize the nature and function of genre specific to the discipline, and to use author intents as a frame for a critical response" are put into focus [5,6]. In this discussion about disciplinary literacy, which is held under a comparable goal of

subject-specific language awareness in Europe and the fruits of which are already implemented in its schools there, argumentation and writing competences are seen as core activities for understanding and judging [6–10].

A need for a more prominent role of subject-specific language or disciplinary literacy in teaching subjects is rooted in the broader goal of enabling students to use academic (written) language confidently, a form of language that is used to attain "conciseness, achieved by avoiding redundancy" [11–14]. Another aspect that makes awareness of disciplinary literacy worth consideration is that different teaching methods resulting from the debate support students who do not speak the classroom language as L1, such as migrants. These children participate in classrooms that are becoming increasingly heterogeneous. The development of these children's linguistic abilities in monolingual settings is necessary to prevent institutional discrimination in the education system.

However, reality displays two things: Firstly, there is a lack of knowledge among teachers about the vital role argumentation as a form of literacy plays in enhancing rather than replacing (social) science learning and teaching [5,15]. Secondly, schools and teachers often struggle to meet this new complexity and, to say it with the utmost caution, might occasionally tend to just scrape the surface of new and upcoming demands [5,16–18]. Argumentation is discussed broadly in educational research but is rarely implemented and promoted in everyday teaching [18–22]. The mentioned gap should therefore encourage educational researchers to use intervention studies to test new methods, which might support and encourage teachers and schools in addressing incorporation of argumentation into disciplinary literacy awareness in their lessons. Peer feedback could be one of these methods—investigated here in the area of individual argumentative writing. Peer feedback is mediated by the provision of a structured worksheet that helps peers to find weaknesses in the arguments written by their peers. Argumentation here is seen as an argumentation assessable by a Toulmin-based structural rather than discursive approach. Peer feedback has been studied extensively with regard to its effects on language learning and teaching (writing) in (foreign or second) languages [23–26]. It has also, under the flag of cooperative and collaborative learning, brought to light potential alterations in fostering students' argumentation skills [17,27–29]. This article presents a case study in which the method of peer feedback, using a designed feedback sheet for argumentation within a socio-scientific issue, was tested with two classes of ninth graders in geography classes in Germany.

1.2. Research Questions

(Digitally) written and oral, interactive peer feedback based on a empirically- and theoretically-guided designed feedback sheet for written argumentation can be considered to be a meaningful method to promote (written) argumentation and as a way to engage students of high/secondary schools in constructive criticism during the writing processes in (social) science education. This article aims to deliver merit primarily for geography education but also an outlook for teaching subject-specific language in the (social) sciences as well. Consequently, this article has two objectives: the first is to approach a concept of disciplinary literacy in terms of written argumentation in geography education and lessons; the second is to determine the extent to which peer feedback can improve written argumentation in the geography classroom, which is addressed using the following sub-questions and objectives:

When is peer feedback for improving argumentative texts addressing a socio-scientific issue successfully and effectively?

1. What do successful and effective peer feedback groups do differently or better than less effective groups?
2. What deficits and potential in written argumentation and feedback can be identified?
3. How are different media (digital or written) used for feedback?
4. To what extent do students accept feedback and how do they optimize their texts due to the feedback of their peers?

2. Theoretical Perspectives

Subject-specific content is taught through language. Media and material are assessed by using language, discussions about the subject-specific coherences held, and the results formulated. Language is the central basis, a matter of learning and reflection and a medium of learning and a means of communication and evaluation [30]. Due to ongoing educational reforms, the empirical examination of students' argumentation has become more relevant [31].

2.1. Argumentation (Competence) in Educational Contexts

In terms of education policy, argumentation is legitimized on a number of levels. In US settings, in the American National Science Education standards, argumentation is among the main requirements of scientific inquiry for Grades 5 to 12 [32]. In the European Parliament's recommendation of key competences for lifelong learning [33], argumentation skills are included in three of the eight key competences presented in the reference framework. In German national standards for geography education, argumentation is highlighted as a central element of the competence areas for communication and evaluation [2]. All legislating curricula and reforms share a fundamental belief in the beneficial effects of argumentation. These effects result from several perspectives, such as the assumed relationship between argumentation practice and conceptual change, because though argumentation meanings are negotiated, solutions are co-constructed and the epistemic status of concepts is change [34]). Argumentation is, therefore, considered to be a way of constructing specific knowledge and refers to individuals' epistemological beliefs [35–37]. Argumentation is related to informal reasoning mechanisms that become activated only through the practice of argument [38,39]. Argumentation is connected to critical thinking [40] and people learn better when they argue [41,42]. A number of German Ph.D. theses have shown students' argumentation skills are crucial for understanding spatial conflicts [43], for problem solving [44] and understanding complex systems [45] and that they have an important impact on civic education an learning from sustainable development [46] All curricula constructing processes and reforms have also been undergoing an extended competence discourse, which lack (interdisciplinary) unique definitions and understandings of what argumentative competence and its constituent skills actually are [27]. This discourse has generally contained a lively discussion about what students should actually be able to perform after a certain grade instead of what should be taught to them in terms of content.

There is a general consent to understand argumentation as a means to solve problems by confirming or disproving a critical thesis in order to, by logical reasoning, get the partner of interaction to approve the represented position [18,47,48]. Central to our work of argumentation as a geographical literacy approach and, in it, peer feedback as a method of constructive criticism to improve its written argumentation skills, is how argumentation competence can actually be constructed and assessed. Rapanta, Garcia-Mila and Gilabert [27] summarize three main definitions of what argumentation is and how it can be assessed: argument as a form, a strategy, and a goal. Argument as structure or form implies that an argument is a unit of reasoning in which one or more propositions, which are the premises, are combined to support another proposition, which would be the conclusion. Toulmin's function of warrant has been respected here, which becomes explicit only when the argument is challenged or when the producer needs to make warrants explicit [49,50]. Defining an argument or argumentation as a procedure or strategy calls for special attention to the dialogical aspects of argument, such as the use of reasoning in a context [51]. Finally, argument as a process both involves and addresses the whole person and his or her context—that is, taking into consideration the particular circumstances in which the argument is used [52]. Kuhn's' description of the main argument skills—that is, argument construction, justification, counterargument construction and refutation—has been confirmed in both interpersonal [53–56] and intrapersonal contexts [57–59]. These findings add strength to the assumption that all argumentation is dialogical [60], including the one expressed in written form. Specifically, argumentative writing is therefore understood here as an epistemic and strategic process of problem solving, which depends on and is influenced by social, affective and cognitive factors, such as context

of tasks, recipients, the structural and content prior-knowledge, the topic, the intention of writing and motivation [61–63]. Our research interest particularly lies in recursive and social construction and revision of argumentative texts.

The quality of argumentation is interdisciplinary and is most commonly analyzed using the grade of structural completeness, e.g., by structural competence models [64]. Toulmin's model [65] for argumentation differentiates between data, warrants and conclusion and is used as a framework, including here. However, the quality of these elements can only be assessed by the different, subject-specific background of the subjects. Consequently, subject-specific criteria also have to be considered. For a qualitative approach and analysis of argumentation, we therefore need to take a closer look at the subject specifics of argumentation, such as content, questions and issues, evidences and criteria of quality, and how they appear in the different subjects.

2.2. A Disciplinary Literacy Approach: Written Argumentation Competence in Geography Education

Using a superordinate descriptive theoretical approach, Budke [18] concluded that, based on the CEFR (Common European Framework of Reference for Languages) for communication [66], written argumentation competence in geography means that students can understand and evaluate geographic argumentation from different media (reception), that students can interact on a written basis to geographical issues in a subject-specific and recipient-referring suitable way (interaction), and that students can develop opinions to geographical issues and back them up in written tasks (production). Argumentation competence in geography is understood as an umbrella term, which received practical relevance through differentiation in content.

Argumentations in geography are commonly open ended with multiple solutions and norms in addition to facts. If students argue, for instance, about a central urban planning issue or sustainability they mainly refer to norms. In the context of the subject and according to different criteria, argumentation can therefore be identified as qualitatively high or not. These criteria are the consideration of spatial perspectives, multi-perspective analysis of different actors and time frames and a complex justification based on different geographically approved media, which basically refers to sources that are credible and trustworthy. Socio-scientific argumentation can therefore be seen as differing from scientific argumentation, since scientific argumentation is "the connection between claims and data through justifications, or the evaluation of knowledge claims in light of evidence, either empirical or theoretical" [67], while socio-scientific argumentation depends not only on knowledge of science, but also on the application of moral and ethical values, and personal identity [68–71]. In order to develop qualitatively high argumentation products, students need to acquire skills that enable them to evaluate argumentations in different contexts. Furthermore, they need transferable knowledge concerning the general structure and form of arguments, subject-specific knowledge and awareness of the subject-specific criteria of argumentation. Eventually, they need knowledge of the required means of subject-specific language in order to formulate qualitatively appropriate argumentation in the subject.

2.3. Developing High School Students' Argumentation Skills in Geography

Kuhn [72] states that argumentative discourse occasions may present the danger of the cognitive overloading of young adolescents, a fear that has been supported by disciplinary research in geography and has to be considered. The design of the feedback sheet, the arrangement of the instruction of the classes in the case study, and the focus of the empirical evaluation try to consider problems students commonly have that have been identified in prior studies. Further research, in that matter, has shown that the reception of argumentations, e.g., concerning spatial conflicts is challenging for pupils [43,73]. High school students have problems with plausibly connecting information found in literature with a certain cause [74]. Other studies have shown that students successfully work with opinion and justification for these opinions in simple argumentation structures, but struggle with more stages and structures [75–77]. Another observation is that, since many issues in geography refer to the

everyday lives of students, students tend to have on opinion on issues but commonly have problems justifying this opinion using complex evidence and links with validity [75–77]. Furthermore, they tend to evaluate given arguments as more positive when these fit their own opinion. The lack of competence of many students in areas of argumentation has been shown to be due to argumentation rarely being implemented in geography. This has been proven by monitoring studies of lessons and the analysis of schoolbooks at international and national scales [16,18]. An analysis of (national) schoolbooks in Germany, which was concerned with enabling students to participate in social discourses through the use of maps, has shown that pupils rarely work with tasks that involve argumentation, critical thinking or high-order thinking tasks. Hence, students rarely get training in the evaluation of social discourses and to take a critical stand [78].

2.4. Peer Feedback in the Context of Written Argumentation

Peer feedback could be a method with which teachers can react to such deficits. Peer feedback is a method that allows students to mutually review their language product on the basis of a feedback sheet or guideline, give feedback to improve their product in a best-case scenario. In empirical teaching research peer feedback in this context can be referred to as cooperative writing to promote textual competences [79]. Research into the effects of peer feedback to develop argumentation competencies within language-aware subject teaching or disciplinary literacy are poorly understood [26,29,80]. In this article's context, peer feedback is to be distinguished from collaborative or cooperative argumentative writing as the first writing process, the first creation of text, which students execute on their own (see chapter methods). The peer feedback method presented here has collaborative elements, since collaborative learning can be viewed as the process of engaging in a mutual discussion or as the shared engagement of people in an effort to solve a problem—in this case, creating a better, more criteria-approved argumentative text [17,80]. Collaboration is distinguished here from cooperation, since cooperative work is the division of the task among the participants, whereas in collaboration the task as a whole is completed by all members of the group [29]. International research into (social science) education considers the intensity and complexity of negotiations of students in the interaction as crucial factors, which influence the quality of mutual students' feedback and collaboration regarding the improvement of argumentation or argumentative texts [81–83]. Berlan and Hammer [84] highlighted that students refer to their argumentation in different ways to achieve certain functions and goals in order, for instance, to win a discussion, exchange ideas or being able to act in a flood of "fake news". Evagorou and Osborne [84], who analyzed collaborative methods for argumentation within socio-scientific issues, mention similar results. The more intense a team was focused on the issue of the debate, the higher quality the mutual feedback tended to be, in terms of oral and written argumentation. The students' ways of contextualization and the intensity of interaction are crucial and was an essential element of the construction of this case study (s. methods). Moreover, the discussion on peer feedback and its fruitful role in improving learning in terms of self-explanation is to mention here as well as the addressing of an recipient due to the feedback process [85–88], which basically concluded that the general principles of peer education would enhance all varieties of schooling.

The effects of peer feedback in geography are a gap in research. However, (positive) effects of peer feedback in educational contexts of language and writing teaching are well analyzed. An important result is that peer feedback as a student-centered method is already efficiently implemented in content and language integrated learning (CLIL) settings [89]. In a symbiosis of content and language learning peer feedback is used to improve more complex textual skills, such as argumentation [90,91]. Research into language teaching found a number of positive effects of peer feedback, for instance, when considering influencing factors such as adequateness [92]. Diab [93], Zhao [94] and S'Eror [95] identified a couple of advantages peer feedback has over other types feedback forms. For instance, such feedback is especially helpful for students who struggle with teacher/expert feedback or self-learning. Diab [24] further detected that peer feedback is relevant in order to overcome "common errors" in language, which are therefore difficult for pupils to eliminate. Berggren [25] and Lundstom and Baker [96]

identified that even feedback givers gain from such feedback, because they become more aware of textual components and structure as well as the addressing a recipient due to the feedback process.

In addition to the empirical appraisal of peer feedback, another thought that emerges when it comes to its legitimization is that we now teach students of the "Millennials and Z" generation. If one takes, disputable, socio-economic approaches into account, it could be that these students tend to want less hierarchy, more autonomy and a construction of knowledge based on their networks [97,98]. One can argue that feedback from students' peers will therefore become increasingly influential, compared to the expert-teacher feedback, in terms of the relevance and acceptance of information that helps them, even in these instructional and often still teacher-centered educational contexts, to optimize their own performance.

3. Methods

3.1. Classroom Contexts: Participants, Procedure of the Teaching Unit

Specifically, two ninth-grader classes (n = 47 students, 23 boys, 24 girls between 14 and 16 years old) were selected. Teachers of these classes had already cooperated in workshops previously, so a fruitful university–school environment of cooperation had already been established. The ethical approval of the collected data was established by parents, teachers and the school's headmaster. First, the general language skills of students were measured by using a validated c-test. This is a valid and broadly used method of analysis by which to diagnose general language skills in German educational contexts for different grades, in which students are supposed to fill in gaps in several texts using the correct spelling, word form and contextualization [99,100]. Then, the correct words are counted, and the researcher receives a valid overview of the scoring. The test proved heterogeneity in the classes but only identified two students, which were under the average validated skills of 70% of the score for L1 speakers.

Students needed three things in preparation for writing the argumentative text: rich and authentic content matter, and methodical and linguistic support, integrated into one teaching unit. Consequently, the case study was arranged in three stages: the preparation stage, the intervening case study stage including the process of peer feedback concerning argumentative texts, the empirical stage of reflection and the evaluation and discussion of the study.

The first stage mainly consisted of a teaching unit of 10 lessons of 45 min each using a cooperative and activating set of methods, which was supposed to prepare students for the task in terms of content. The teaching unit was taught in cooperation with the local teachers. Several research assistants were hired for the process of data collection and interpretation to minimize research bias. Analysis and interpretation were based on criteria of interrater reliability as mentioned in the following chapters on methods and results. A political geographical socio-scientific issue was chosen, namely "To what extent does Turkey belong to Europe these days?". There were plenty of reasons for choosing this issue: first, we assume that argumentative skills can only be promoted and analyzed in context, in which students are engaged in rich and authentic and relevant content [2,101,102], which this issue offered. Germany has a rich history with Turkey. Many students are second- or third-generation immigrants from Turkey, so many classes consist of students of Turkish background. Turkey is understood to be a central partner of Germany and Europe during the refugee "crisis", since Turkey is considered to be politically understood as the "buffer zone" between Europe and Syria. Therefore, the issue had been frequently and controversially discussed in the media. Moreover, Turkey has strong economic relationships with Germany and the status of Turkey becoming an EU member has been discussed frequently. Moreover, Deniz Yüzel, a prominent Turkish-German journalist of German newspaper "Die Welt" was kept hostage by the Turkish government, which led to an intensive discussion of Erdogan's and AKP's art of reigning, even amongst students in the schoolyard. Finally, it is a topic that combines physical and human–geographical topics, and the construction of Europe and its borders are regularly discussed in (political) debates. The teaching unit, therefore, dealt with the following exemplary topics in the lessons,

such as the students' view of Turkey, Turkey's economic, cultural and political development, Turkey's relations to Europe and the several ways Europe can be constructed and understood, e.g., as economic union, as fellowship of values etc. The teaching unit was cooperatively designed with the Institute of Geography Education and held in teams teaching with all the cooperating teachers. After an engaging teaching sequence of eight lessons, during which students debated on the issue in talks—before which, they worked in detail on the social, cultural, political and economic development of the EU and Turkey—students received argumentation training and language support for writing argumentative texts, in order to establish a more comparable basis of knowledge before the interventional case study. The students were heavily encouraged to refer to all the materials they used during the sequence to support their opinion. The training basically was a collaborating collection of criteria for good argumentation, which was also aimed to introduce the criteria of the feedback sheet. First of all, most of the criteria of the feedback sheet were therefore collected collaboratively with the students. Nevertheless, a whole lesson was dedicated to the explanation and establishment of the transparency of the rubrics using examples given by the students. Language support in that matter was a collection of macro-scaffolding strategies and techniques from which pupils can benefit because they can apply them to difficult complex tasks and thus acquire target language and subject content [30,103,104]. In this case, it was a support sheet for the students that contained a repetition of text structuring and examples of linguistic wording, sentence and text level. Moreover, the students received training in *Microsoft Word*, with particular focus on the use of the highlighting and comment functions.

The second stage, the actual case study of the process of peer feedback, began with students writing their argumentative texts. They received a provocative article from German newspaper "Die Zeit—Times", in which the author reflects on the relationship between Europe and Turkey. The article was chosen as it allowed the students to reflect on the established and discussed criteria of complex argumentation in geography. As mentioned, a c-test had been executed. This test is understood to be a diagnosis of general language competence. Based on that diagnosis, the text was chosen. The text was also chosen based on an analysis of curricular demands that are reasonable for the pupil's age. Furthermore, the text was chosen because it offered a setting, in which argumentative writing is not heavily constrained and to not put students whose opinions are against the integration or the belonging of Turkey to Europe in a difficult situation, since both sides are reflected. The students were asked to comment on this article within the next 45 min, which we analyzed as a pre-test. This amount of time was selected by the teachers, and the students were asked to write the text in *Word*, the use of which is explained subsequently. Following this, the feedback sheet was introduced and explained to the students. In this stage, as well as in all other stages, students were encouraged to ask questions with regards to understanding and clarity. The feedback pairs were then selected. A number of studies suggest that groups of learners with similar abilities seem to learn better than groups of widely varying abilities [105,106], while others suggest that mixed-ability groups or groups selected in terms of friendships are better [107,108], and in these groups the low-ability students benefit the most from interacting with the high ability students [56]. More recent studies are moving towards understanding the conditions under which collaborative learning is effective. Based on this literature, we decided on two factors that we expected to present merit to our interest in the conditions of effectiveness of reciprocal peer feedback concerning argumentative writing: (1) we wanted to know whether students with different performances in the pre-test work effectively together and, (2) whether students with fundamentally different opinions to the issue work more effectively together. These two factors were used to select the pairs and led to 23 varying pairs with mixed abilities. The feedback began with the students exchanging their texts, reading them and making remarks in the text file and on their feedback sheet. For us, it was important to arrange the classroom situation in a way that the feedback pairs were seated next to each other, so that interaction and questions were possible at every stage. After 30 min, again selected by the experience of the teachers, the pairs were asked to present their feedback to each other and to agree on central aspects for review and correction of the texts. Then,

students had another 30 min to rewrite the text. The final text served as a post-test. The classroom stage ended with students' feedback and description of their experience.

3.2. Theoretical and Empirically-Guided Design of a Feedback Sheet for Written Argumentation in Geography

The designed feedback sheet, which was the core of the case study, was developed using three influencing factors: Firstly, on the above mentioned research, secondly, on a workshop with teachers concerning disciplinary literacy/language awareness and CLIL strategies and, thirdly, on the interaction with teachers who monitored the university/school cooperation during the feedback project. Furthermore, Yu and Lee [23] suggest in their extensive meta-analysis of the effects of peer feedback on the writing competences of students to focus on the structuring of content in the texts, which influenced the constructional process of this feedback sheet. The feedback sheet addresses high school/senior high school students (15–16 years old) in linguistically heterogeneous (geography) classrooms. In the chosen grades, the execution of more complex and written tasks becomes more important in geography (in Germany), since the students need to be prepared for their final exams [2]. The structure of the sheet needed to be carefully and sensitively designed, since Gielen and DeWever [109] proved that the structure of such a sheet immensely influences the quality and effectiveness of peer feedback, by showing that an overwhelming, cognitively overloading structure can lead to superficial feedback.

The sheet was progressively designed, which means the students were provided with three different stages of feedback tasks and impulses. The first stage contained questions about the general means of language, e.g., spelling and grammar. The second stage offers questions about more subject-specific linguistic means, e.g., a problem-generating introduction or phrases to integrate spatial reference (s. feedback sheet). The third stage referred the highest level of complexity—the quality of argumentation, e.g., the quality of evidence for justifying the opinion. Furthermore, the students in the study were allowed to give feedback via three types of media: firstly, using comments and annotations in the sheet; secondly, using comments and annotations in the text file of their partner and; thirdly, using a short oral presentation in an interaction stage, in which the feedback was supposed to be presented to the partner. This combination proved effective, as exclusive annotations emerged in the different types [110], so that the spectrum of feedback was widened through this variation. The introduction of the sheet or the task description, respectively, provides the students with a first impression of the text and instructs them to read the sheet carefully. In the second part of the description, students were instructed concerning the relevancy of the task, in order to establish the basis for an intensive negotiation [17]. If students felt insecure during the revising process, they were encouraged to refer to the expert/teacher knowledge. A chosen combination of peer and expert feedback has been proven to be effective in similar studies [111]. Another important aspect lay in the last sentence of the task description, where students are motivated to justify their feedback. This justification has been proven as an essential factor in increasing the effectiveness of peer feedback [112].

During the first stage of the sheet, students had been asked to check the general intelligibility of the text, where illogical or contradictory text arrangements had to be identified. The students were able to use the Likert scale to evaluate their feedback and the free space under the question to write their own comments. Stage two referred to aspects of structure, with the aim of guiding the reader and establishing text coherence by using, e.g., connectors on sentences levels. This stage demanded an analytical evaluation of the text using a combination of quantitative and qualitative processes, since students estimate the appropriateness of language, content and structure, respectively. These criteria stemmed from the basis catalog for text evaluation and level of expectation in these terms in school contexts of secondary schools. These have been validated and established extensively for the national context, especially for language subjects, and have been adopted here for argumentative texts in geography [113,114]. We argue, therefore, that this structure is transferable for and useful in international educational and interdisciplinary contexts and is a suitable instrument to evaluate criteria of the linguistic discourse function of argumentative writing in geography.

During the first part of the third feedback stage, students were required to deal more intensively with text structure. Students had to identify whether their partner used a problem-generating introduction, a summarizing conclusion and a reasonable usage of key terms. Moreover, stage three was about the analysis of the qualitative criteria of argumentative texts in geographical contexts (s. chapter before about classroom context). Again, Toulmin's basic structure of arguments [65] was considered, which was adapted as an interdisciplinary accepted approach to design feedback questions that allowed students to identify complex arguments written by their partner based on the form. This structure is specified as follows: students should reflect during feedback on used spatial references, different media of evidence, such as maps, that are used to underpin the solution to the problem and therefore the answer to the central task of their writing. Evaluations tended to be more complex and difficult for students, such as the reflection of spatial conditions and conditions of time in which the respective argument should be valid. Because of that, referring to the central issue, students received text examples on a sentence level in order to make them aware of the potential usage of these means. Other criteria of quality, which students need to reflect on are the text's addressing of recipients, the reflection and consideration of counter arguments contrasting the opinion stated in the text and the multi-perspective observation of problems, e.g., the considering of different actors involved in the issue [75] (Figure 1).

Level of feedback: 3 (Expert) - How good is the argumentative text?				
	Yes	**Partly**	**No**	**Don't know**
Subject-specific structure of the text				
Can you find an **introduction** that presents to the reader the **main problem** and opinion?				
Is there are a last **paragraph**, which **summarizes** the main ideas and has a **conclusion**?				
Do you think that relevant **key terms** are used (correctly) in the text?				
Criteria of good argumentative texts in geography				
Is there an **opinion held** in text?				
Is this opinion **justified** and supported by **evidence**?				
Does your partner explain **where** the arguments are valid, e.g. in what countries or areas? *Example: Concerning Turkey, the argument is to be relativized, since there are massive spatial and economic disparities in the country.* *This might be correct for a country like Turkey, but not for countries, e.g. in the EU.*				

Figure 1. Exemplary extract of feedback stage 3 (own design).

With particular concern to normative types of argumentation [48,67–71], norms and values are included—the relevance and validity of which were to be reflected in the text. An objective should be that the students learn to conceive and justify an opinion. Therefore, both aspects of presenting an opinion and its justification are more intensively considered in the fourth stage of the feedback sheet. In this fourth and final feedback stage, students are asked to evaluate the quality of (single) arguments. Previous studies have shown that students work successfully with evaluations of opinions and justifications at a pre-stage of complex argumentation [10,48,75]. Hence, students evaluated their partner's opinions and justification in this part of the sheet. To enhance this process, students were guided to work with visualizations, such as the highlighting of both elements in the text and the evaluation of the justification via different rows in the chart. Concerning the latter, students examined relevant criteria in terms of the validity and suitability of the justification, to the extent to which the

justification refers to the superior issue and the used evidence is from correct and trustworthy sources. At the end of the sheet students are asked to respectfully present their feedback to their partners using the text file and the completed sheet.

3.3. Methods of the Case Study: Quantitative and Qualitative Mixed-Method Approach and Data Analysis

The influence of the feedback was measured and observed in three ways: (1) the change in text quality, (2) changes in the text file and (3) oral feedback in the interaction/conversation between the students in the presenting stage. The data was triangularly related [115].

The first central empirical task was to determine the quality of the students' before and after the peer review texts, since the influence of peer feedback on this quality was to be assessed. The approach used was the text analytical method [113,114]. Based on valid level of expectation and evaluation sheets for secondary schools in language subjects [113,114], we specified an evaluation sheet for argumentative texts in geography. This specification was a sheet, with which teachers assign points to different criteria and sub-points to these criteria based on the students' texts, specifically based on the assignment of criteria and points of the subject language. The three main criteria in this sheet used were: (1) Structural and linguistic features of the text, (2) Quality of the text genre argumentation in geography in terms of subject-specific linguistic means and (3) Quality of arguments in the text. The first referred to spelling, grammar and structuring of the text. The second included the usage of linguistic means to construct the argumentative texts, such as the integration of counterarguments or spatial reference, wording of introduction and conclusion, which are based on the subject-specific criteria concerning argumentation mentioned previously. The third referred to assessments of argumentation in terms of the form of an argument such as Toulmin's and its application to geographical argumentation [65]. A Toulmin-based framework was therefore used to evaluate the quality of arguments and reasonable reliability is attained. This framework is known for its difficulties in precisely identifying the elements of an argument. How did the author(s) deal with this issue?

For this part of the qualitative text analysis, statements in the students' text that contained an opinion and justifications of the opinions were analyzed with regards to appearance of relevant elements, such as opinion, evidence and link of validity. If one of these elements was missing, the act was not counted as a completed argument. The criteria of argumentation quality, which were then analyzed once the act could be counted as a complex and completed argument, were relevance "Is the argument suitable for the central issue?", validity "are the given (factual or normative) evidences correct and are sourced correctly quoted and summarized" and complexity "does the student integrate spatial and temporary conditions, exceptions and counterarguments" [1]. The four people rating the text with points, two teachers and two researchers, tested the sheet on reliability between different markers. The acceptable accordance of assignments to points in the criteria was supported, despite some debatable assignments and minor changes to the assignment of points in criteria 2 and 3, between 0.60 and 0.75 cohen's cappa (k), due to the indication value in the literature [116–118]. The four markers then evaluated the pre- and post-texts, providing data for the whole text, certain categories and changes between the scoring between pre- and post-texts.

1. The second task was the handling of feedback in the text itself. The students were supposed to write in *Word*, since *Word* offers the function "versions". Due to the function, different versions of text file could be stored, and the changes could be visualized. Moreover, the students' annotations could be digitally saved and highlighted. This allowed a procedural document analysis of the three stages of the texts; the pre-text, the re-viewed text and the final post-text [119]. Here, we could document what kind of feedback appeared in the feedback sheet and in the texts, and which of the feedback suggestions appeared to be accepted and changed and/or corrected in the final file.
2. The final task was the analysis of the interactional stage of the feedback in which mutual feedback was presented. First of all, we recorded the students' interactions and transcribed it on the basis of conversation and sociolinguistic analysis [120–122]. We rejected the idea of videography, because

of our lack of interest in visual aspects of feedback and in favor of less organizational effort and less pressure for the students. Using the transcripts, we documented the quantity of speech acts [123], the quantity of the switching of speakers as a grade for interaction and the content of each of the speech acts. The latter we analyzed by assigning the acts, which contained references to content in the feedback's interactional process, to inductive partially overlapping categories using a number of fundamental frameworks (Table 1) [28,124,125].

Table 1. Coding in the feedback's interactional process (own design).

Discourse Code	Definition	Example
Contribution to discourse by reviewer		
Criticism/negative feedback with text reference	A text passage is evaluated as optimizable and is referenced.	*"Unfortunately, here in the second paragraph you did not write whether the argument would be valid, like 100 years ago".*
Criticism/negative feedback without text reference	A text passage is evaluated as optimizable but is not referenced.	*"What makes the text less structured is that you do not use paragraphs really."*
Praise/positive feedback with text reference	A text passage is evaluated as good and is referenced.	*"The reader gets a good overview, because you use a good structure, for example here in paragraph one and two."*
Praise/positive feedback without text reference	A text passage is evaluated as good but is not referenced.	*"You give good reasons for your opinion. This makes the text professional."*
Precise suggestion without example	Student clearly suggests what is to be changed but does not offer an example.	*"You could use more connectors here."*
Precise suggestion with example	Student clearly suggests what is to be changed and offers an example.	*"It would be a good idea, to use the connector "however" here. This would make the structure clearer."*
Checking criteria of sheet without evaluation	Student says, that one criterion is met or not met in the text but does not evaluate it.	*"No sources are used. The opinion is justified."*
Question, demand by reviewer	Reviewer needs more information about the process of writing.	*"No, where is that from? Are these your emotions?"*
Favorable comment on mistake	Reasons were discussed as to why the mistake is not sensible.	*"But no worries, it is alright. We did not have enough time anyway."*
Request from reviewer to react to feedback	A reaction of the author is demanded, after feedback from reviewer has been mentioned.	*"What you're saying?"*
Contribution to discourse by author of the text being discussed		
Explanation or justification of text by author	Author explains his or her decisions regarding writing the text.	*"Of course, I use counterarguments. Take a look..."*
Approval of feedback	Feedback is accepted.	*"Yes, you're right. I will do that."*
Contradiction leads to discussion	Point is disputable and discussed.	*A1: Of course, I mention Erdogan. A2: No, where?*
Contraction does not lead to discussion	Point is disputable, but not discussed.	*"No, but go on."*

4. Findings

4.1. Descriptive Findings

Figure 2 displays the different criteria of the analysis sheet and the number of pupils who improved in these criteria after the feedback. There is a wide range of scores and improvements. Most students improved in general linguistic skills (spelling and grammar), then in the quality of arguments (points given for complete complex arguments), and then in the quality of the argumentative text in geography in terms of subject-specific language. Addressing the audience by argumentative means and the reflection and integration of counterarguments were the least improved criteria. Most students tended to improve given arguments instead of finding new ones through peer feedback (Figure 2).

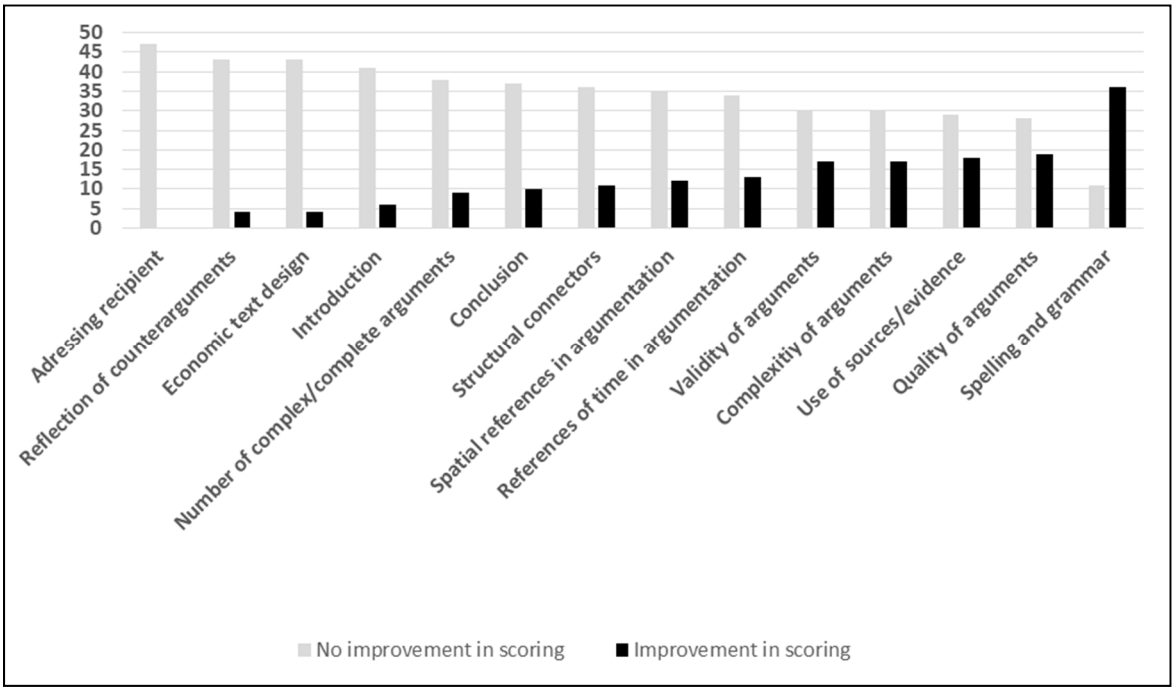

Figure 2. Number of pupils (n = 47) who improved their texts in different categories of argumentative text quality (own design).

Concerning the criteria of quality of argumentation, 65 complete arguments were counted in the pre-test and 81 in the post-test. Figure 2 illustrates the average percentage improvements in areas of quality of argumentation for all student in the class. Students used a maximum of five arguments in the texts. Only complete arguments, (those that containing an opinion, evidence and a link of validity) were counted. A relatively large number of students improved in numbers (Figure 2), although one point increased was deemed as an improvement. However, the average improvement was relatively low when the improvement was not high in quality (Figure 3), e.g., complexity. The percentage here means that students' scoring increased in quality of evidence by approximately 9% between the pre- and post-test.

Figures 3–5 relate to Figure 2 in that they show how the average individual improvement in the class was in comparison to the absolute numbers presented in Figure 2.

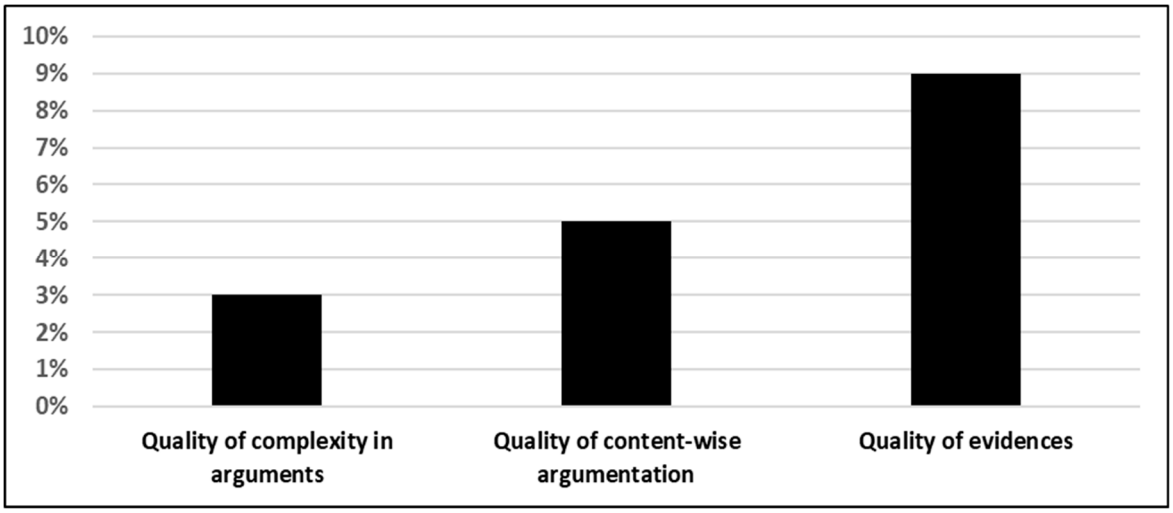

Figure 3. Average improvement of individual students in quality of argumentation (own design).

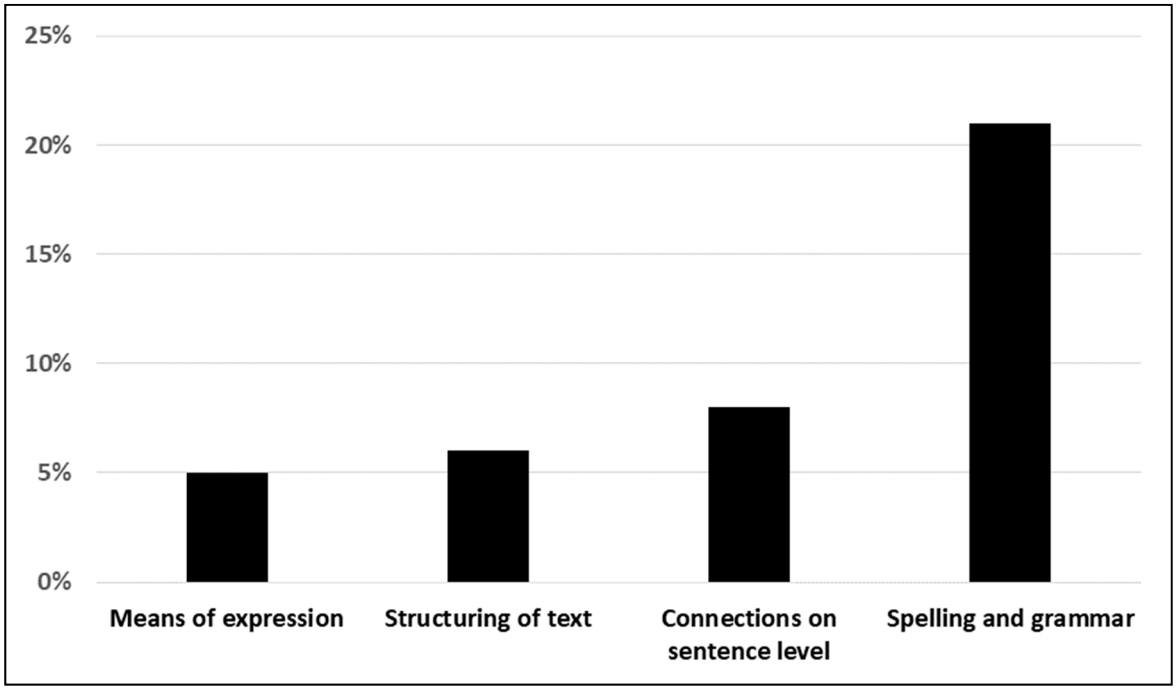

Figure 4. Average individual improvements in text quality: General linguistic means (own design).

In the general linguistic criteria (Figure 4), 47% of the students received the maximum nine points in the pre-test. In the post-test, this increased to 63%. However, both were counted here. The average improvement was 15%. Special attention was placed on spelling and grammar, where the greatest improvements could be identified (Figure 6)—43% in the pre-test and 74% in the post-test. The average grade of improvement was 21%.

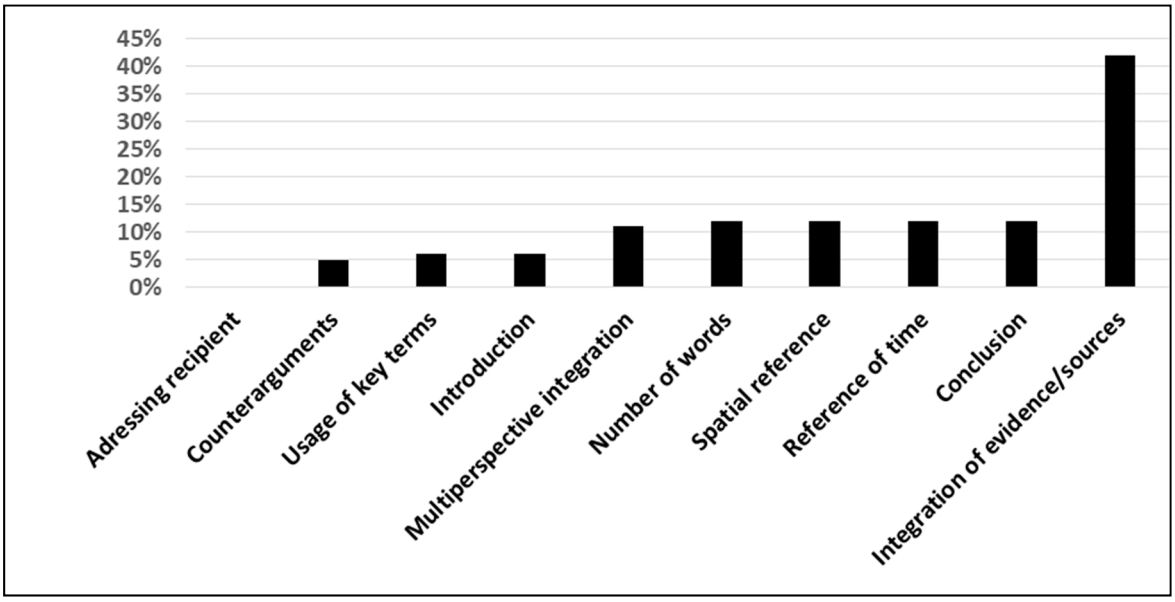

Figure 5. Average individual improvements in text quality: Subject-specific means and text quality (own design).

In the criteria subject-specific means (Figure 5) and text quality, students achieved 46% of all points in the pre-test and 57% in the post-test, with a potential of 29 points on offer. The average improvement was 11%. The integration of evidence from sources increased by 42% (Figure 5). The ability to express spatial references and references of time increased by 12% on average. Relatively high scores of improvement were displayed in the formulating of a conclusion (12%). The expression of the integration and reflection of counterarguments improved by 5%. No improvement was observed in the reference of a recipient (Figure 5). We found a significant level of correlation between improvement between subject-specific criteria and the quality of argumentation. On average, the strongest improvements in this class with regards to the use of the evaluation sheet were made in the area of general linguistic ability, then in subject-specific criteria and finally in quality of argumentation. Students received the most points for argumentation, then subject-specific means and then general linguistic means.

Figure 6 allows a stronger focus on the individual performance of students. It shows the percentage degree of improvement in quality of the text from pre-test to post-test, which are highlighted through different grey scales and the anonymized codes of the pupils. The more the bars extend to the right, the stronger the percentage improvement in one certain area for the student. There was a broad spectrum of improvement results, from students who improved significantly in all areas to students who only improved in one area of the text, or even none. Spelling and grammar improved most, but pupils received more points for argumentation and less for grammar, which influenced the percentage degree of improvement.

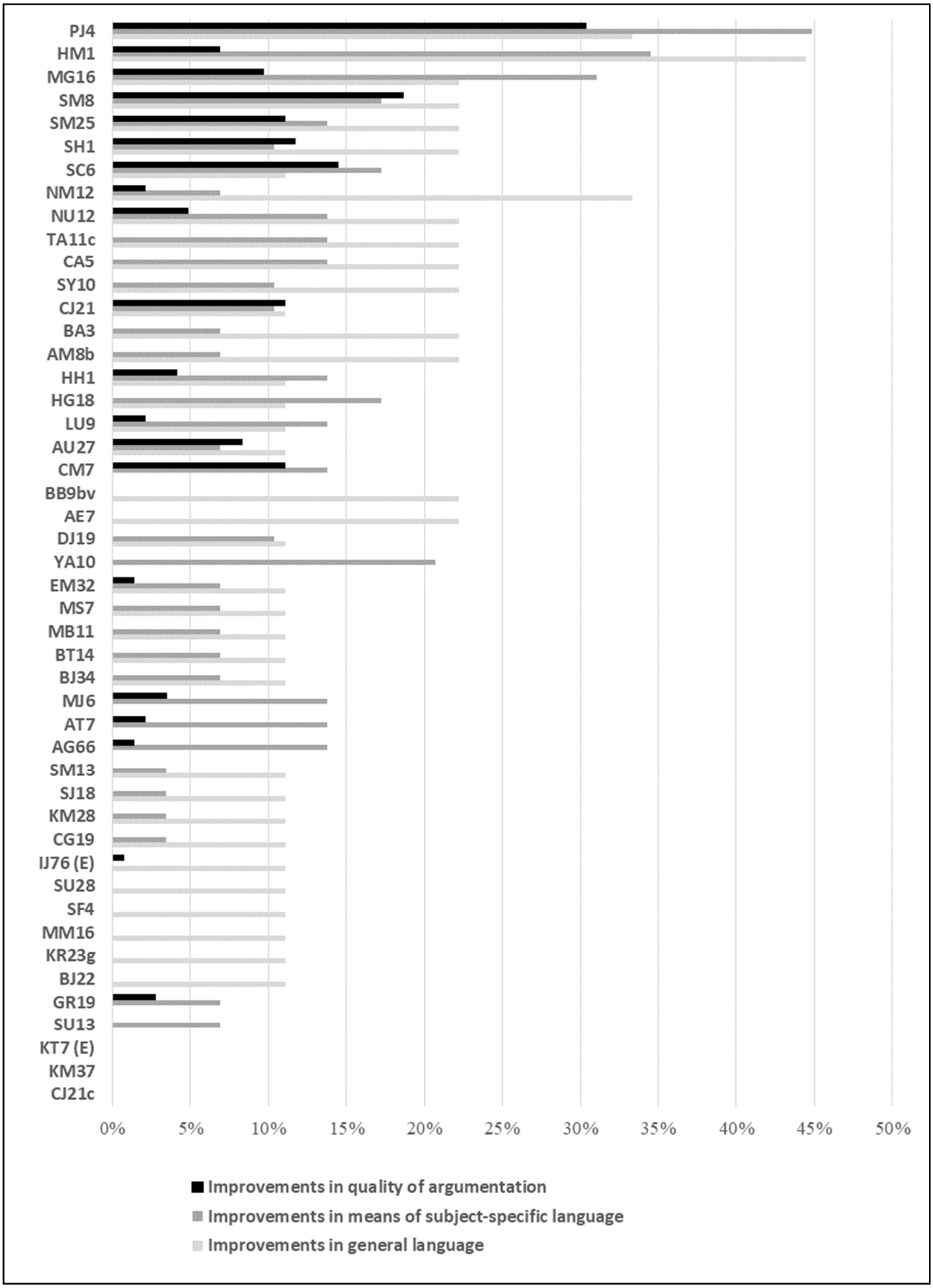

Figure 6. Improvements in the scoring of individual pupils (n = 47) (pre-/post-test comparison) according to text quality. (Own design: The combination of letters and numbers (e.g., KM37) are the anonymized codes for the students.)

4.2. Explanatory Findings

The method used had a predominantly positive impact on the students' argumentative texts in geography in these classes, based on the results presented, but with varying degrees of effectiveness for every student (see Figure 6). The question is what factors could have made the feedback more or less effective, and which should to be considered and reflected on in terms of further implementations of peer feedback in such a context. In order to identify such factors, we first related the interval scaled scores of individual conditions the students brought with them to their different performances by screening potential correlations between them.

A moderate significant correlation (0.323*) between the performance in the pre-test and the c-test was identified. The c-test measured abilities in general language performance in reception and production, and consequently it would be expected that students with higher linguistic abilities in the c-test would be in the higher stages of argumentative text quality in pre-evaluation. However, this correlation did not continue to the grade of improvement and students who improved significantly during peer feedback had different scores in the c-tests. Different general skills were, therefore, no obstacle for improving the text.

The correlation between scores in pre-test and post-test (0.95**) showed that no impairment of text quality occurred, since a high scoring in the pre-test was still a high or higher scoring in the post-test. However, the grade of improvement, measured by the difference between post- and pre-texts scores, was not found to correlate with the score in the pre-test. Hence, it was not only students with qualitatively high text scores in the pre-test that further improved their texts, but students with poorer ratings also benefitted from peer feedback. Therefore, the pre-evaluation did not have an effect on the grade of improvement in this study. Moreover, the quantity of feedback (number of speech acts that refer to geographical content) and degree of interaction (speaker changes) had no correlating influence on the grade of improvement either. However, pairs who gave a lot of feedback to each other interacted a lot due to speaker changes. In summary, pre-evaluation and quantity of feedback and interaction had no correlating effect on the degree of improvement of text quality based on peer feedback. Thus, other factors needed to be considered as an explanation for improvement in text. For that, we will focus on the qualitative part of the analysis and specific aspects of interaction.

Figure 7 shows the number of speech acts of critical feedback (speech acts that claim some sort of change in the text) that were communicated by the students in different feedback media.

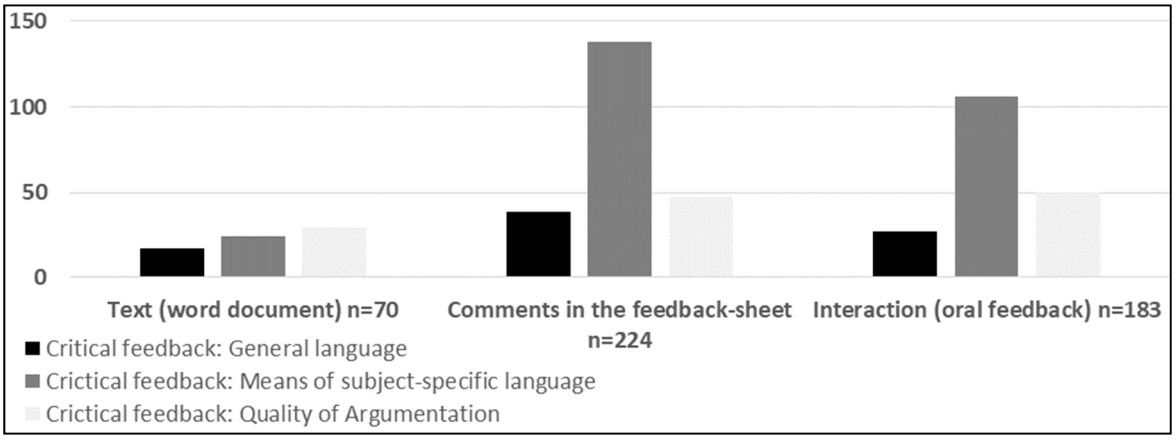

Figure 7. Number of speech acts (critical feedback) according to medium of feedback (own design).

The feedback sheet formed the basis of feedback and provided structuring for the oral feedback, which is why the high numbers of comments in this area are comprehensive. A comparison of the acts of improvement in the text and mentioned acts in the interactions showed that the students' oral comments were altered in the text, especially in debate situations. Moreover, the combination of media was relevant, since feedback acts that were then improved in the text were not always found

in every media, so there were exclusive aspects the students related to. The text file was rather used for illustrations and specific explanations, as Example 1 shows. The student NM 12 directly refers in their feedback to the marking in the text to prove their point. In the text, different passages were marked with colors and highlighted with comments, in which the reviewer demanded more evidence for sentences, for instance in the suppression of minorities and spatial disparities.

Example 1.

> *NM12: "And … (hesitating) your lines of argumentation were very good. What I missed were evidences, facts, you know, (unintelligible), so what you relate to. Here, he (Erdogan) never mentioned, this is why you have … too, and you have to give evidence for that. This is why I marked that passage, you understand?"*

Figure 7 further reveals that students talked less about aspects of general linguistic skills and more about subject-specific ones and quality of argumentation like in that case of example 1—the evidence. This is interesting, since the grade for improvement in the text does not completely mirror that. The aspects the students talked more about were less improved, supported by Figure 8, which visualizes the quantity of speech acts referring to the medial communicated criteria of text quality. Criteria of argumentation and subject specifics, such as sources, evidence, reference of time and conclusion were quantitatively well considered. Students communicated less about structure or spelling. Therefore, it seems that it was more complex and challenging to improve scores in argumentation than in other criteria like spelling.

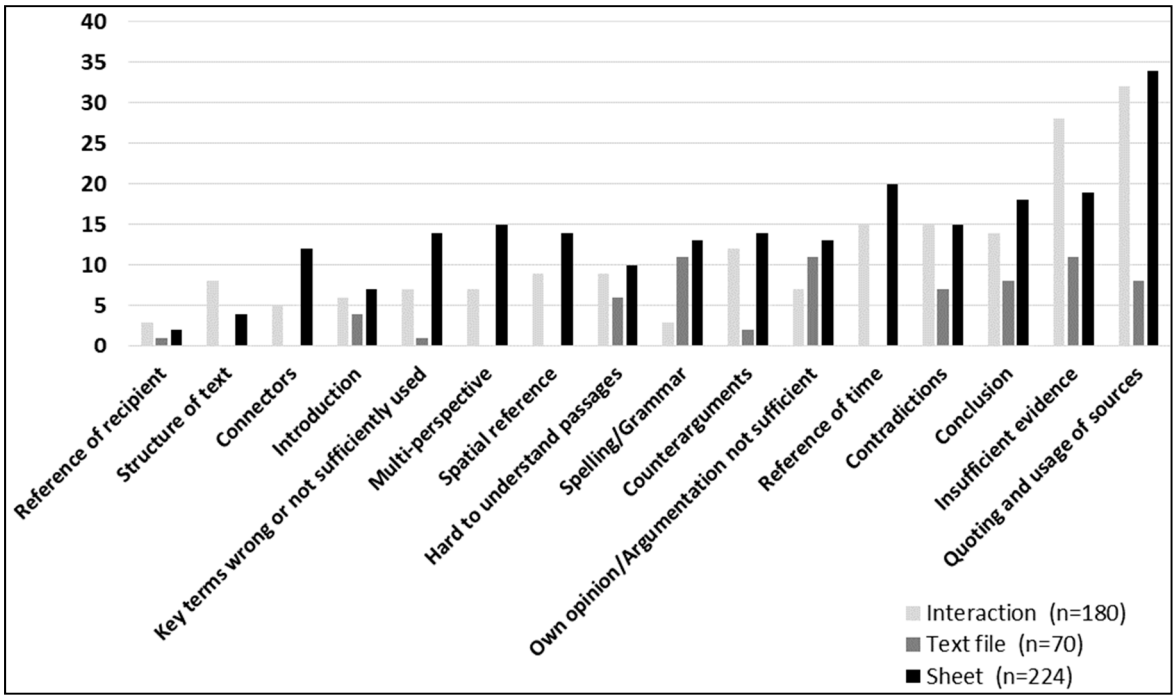

Figure 8. Criteria of text quality in different forms of feedback (speech acts) (own design).

The students talked intensively about argumentation quality, but in absolute numbers (Figure 2), only half of the students improved the quality of argumentation in their texts. We identified two possible explanations for this. The first is that the explanation and transmission of mutual, critical feedback of students according to criteria of argumentation led to different grades of improvement. The second one is that the way the feedback was performed in the pairs led to different improvements. Figure 9 shows how critical feedback was given and what was missing in these speech acts.

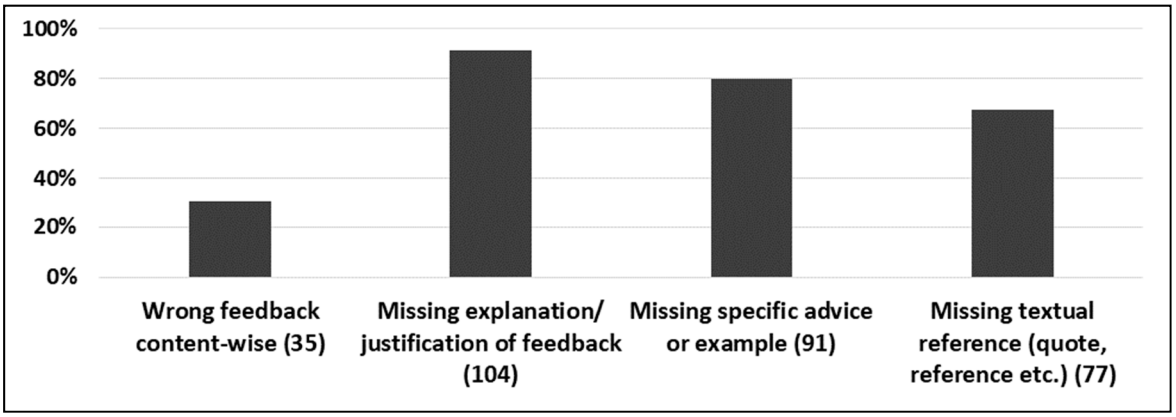

Figure 9. Quality of critical feedback transmission (n = 114, number of speech acts concerning critical feed back towards argumentation) (own design).

Incorrect feedback was feedback that the teachers would not have given since it was either linguistically or factual wrong. Out of 114 feedback points, only 35 were incorrect. Students rarely explained and justified their feedback, in terms of why they thought a particular change in the text is necessary and how it would improve the text. Furthermore, students rarely used specific suggestions or examples of how the text could be improved in that particular matter, for instance, as to how the conclusion could be improved. Finally, in the majority of feedback, students did not refer directly to passages in the texts when giving feedback, so it tended to be more general and unspecific. Not all students' partners who considered every feedback criterion necessarily improved but the average scores indicate the majority of students who addressed the feedback did improve, as those that met the criteria were more frequently in higher scoring groups.

We could separate the ways in which students gave each other feedback into two extremes, a best practice and worst practice: The first example was executed by students who achieved high scores in improvement and whose way of feedback could best be described as "appreciating negotiations". A lot of speaker switches occurred in these pairs and, during these switches, students tended to be more involved, which was revealed by several counter questions and negotiations of critical points, as is observed in example two:

Example 2.

NM12: "That, here, he never said it that way, so I think … you have to give evidence for that. This is why I marked it, you know?"

HH1: "Yes, but, I thought … I used a different phrase. I can do that, can't I?"

NM12: "Yeah, but still. You have to … you know … prove it somehow, give evidence."

HH1: "Yes, I could have used a quote from the text. Ok, I agree. I will do that."

Another characteristic of students in these pairs was that they were using more text references and appreciated the material provided:

Example 3.

KM28: "Well, I really liked how you justified and explained, when you write here, that the EU would be able to take action in the crisis. That's a really good reason."

Finally, these pairs found a way to comment on errors in an appreciative way by, for instance, justifying mistakes or even calming their partner down. Additionally, these pairs were all pairs that had differing opinions to the central issue.

Example 4.

MG16: "What you could have done is, you could've used more paragraphs, like, you know for structure."

BJ22: "True, I think this would have helped to make the text easier to read. Yes, I haven't thought of that, actually."

MG16: "Yes, but we haven't had enough time, I know. So it's alright."

The other extreme could be described as "listed and non-reflective critique". This pair had lower scores in improvement, critical feedback was dominant, listed and often general and superficial. The style of feedback seemed to be more of a speech than an interaction and left less time for questions or clarification. Speaker switches, which were counted as a grade for interaction, did occur, but rather more for one-syllable agreements or recognitions. Moreover, the feedback and the comments tended to be less appreciative and they used less individual-specific feedback, e.g., aspects should have been improved, which were, e.g., already in the text.

Example 5.

AD185: " ... ok, the text has many grammar and spelling errors, like generally, many errors, it is not written appropriately, many parts cannot be understood, because of spelling ... ehm ... the bad structure of the text and general a bad structure makes it hard to follow ... ehm ... it is difficult to follow ideas and points of the author."

Example 6.

BJ 22: " ... and then there is the main part with arguments, which is badly structured, everything is mixed up, really bad."

To sum up, therefore, the communicative and content-related way the students constructed their critical feedback seemed to affect the beneficial implications of the method.

5. Discussion and Implications

5.1. Factors for Success in Feedback and the Role of Socio-Scientific References in Feedback

The following section intends to discuss the relevance of the findings for answering the central research questions "When is peer feedback for improving argumentative texts addressing a socio-scientific issue successful and effective?" and "What do successful and effective peer feedback groups do differently or better?"

This study can be considered as a successful exploration into understanding the conditions under which collaborative feedback, with the goal to improve written argumentation of a socio-scientific issue, is effective in disciplinary literacy contexts. Several observations of other studies can be used to discussion this disciplinary literacy approach.

Feedback pairs, who improved and were part of the "negotiating" successful pairs had differing opinions on the issue of Turkey and Europe. Hence, this dialectical aspect in improving argumentative texts is definitely worth mentioning. These results therefore present teachers with a successful and content-based option by which to pair teams compared to options presented in previous studies [105,107]. These previous studies showed, for instance, that friendship groups are better in collaboration than groups assigned by teachers, and that single-sex groups function better than mixed groups [108]. The successful groups in this study also used an intensive combination of a larger number of speech acts, more textual references and more explanations, as well as counter questions of clarity, when compared to the non-benefiting groups. The benefiting groups discussed sources, references, tried to improve the reference to sources, and had less protests in teams, which were ignored, and they argued with regards

to evidence and the explanation of their own opinion [126,127]. These observations are comparable to those of Mercer, Wegerif, and Dawes' [128] and Resnick, Michaels and O'Connor's [129], who suggest that when students explicitly discuss each other's ideas, the reasoning gains are higher. Chin and Osborne [130] can confirm, in that context, that successful groups were characterized by the use of questions which focused on key inquiry ideas and explicit reference to the structure of the argument. What we could not prove is that only high ability groups collaborate more and focus on making sense of their data, but only that the controversial constellation of the groups led to success [131]. The studies of Sampson and Clark [132] add substance to these results as they found that the main differences between high and low performing groups when working on a scientific issue were the number of ideas introduced, how partners responded to these ideas and how often ideas were challenged. Some students, such as PJ4, gained high improvement rates independently of feedback interaction. Research undertaken by Van Amelsvoort, Andriessen, and Kanselaar [133] may deliver an explanation, since they found that some students beneficially focus on finishing the task individually, ignoring the fact that they had to co-construct the product in a group. Finally, an interesting interactional result is that peer feedback seemed to overcome linguistic differences, since students from all levels of general linguistic skills (c-test) profited and improved the texts. In the pre-test, we found a correlation between scoring in the argumentation and the c-test. After the peer feedback, this correlation was did not exist, and the quality of the texts balanced.

5.2. Deficits and Potential in Written Argumentation and Feedback in Terms of Disciplinary Literacy

This section deals with the question "What deficits and potential in written argumentation and feedback can be identified and how are the available feedback media used for feedback?".

Regarding the areas in which students improved, it seems that students had no difficulty in improving more general linguistic aspects of the texts, as feedback on that matter was easily implemented. This might be because knowledge about these kinds of means are taught in other (language teaching) subjects and the transferability is high. Strong improvements in the area of subject-specific criteria could be found in more interdisciplinary aspects such as the integration of sources as evidence (Figures 5 and 8). This result shows that students had basic skills in the reflection of subject-specific criteria. We found significant level of correlation between improvement between subject-specific criteria and the quality of argumentation. This coherence confirms that when students expanded subject-specific criteria, they did it correctly and they improved the quality by meeting criteria of subject-specific argumentation. Moreover, this is one of the results that leads to the understanding that the integrated peer feedback process here can be seen as an explorative success, as students actively communicated and reflected on the construction of argumentation and improved their texts. However, the subject-specific criteria such as the spatial reference for complex argumentation need a stronger focus for promotion.

The results show that many students only partially improved in argumentative quality of their texts. These results of the ratings indicate that students had more difficulty in giving, receiving and using feedback concerned with the quality of argumentation than those concerned with general or subject-specific linguistic criteria. This difference in the handling of feedback could be linked to the fact that the constructive critical teaching approach of argumentation is rarely taught in schools [16,18]. Argumentation in geography, used as an example here for an approach to include disciplinary literacy, seems to be a strong subject-specific competence, for which students can only partially benefit from knowledge acquired in other subjects. Research further suggests that students struggle with the reception of argumentation, which is a basic stage in the feedback process [43,73].

The designed feedback sheet proved useful, as most of the critical feedback speech acts referring to the partner's argumentation could be found in the sheet. It has been shown that arguments were mainly improved rather than new ones found. In future research, it may be worth considering the inclusion of a central task, which would motivate students to look for more arguments in addition to their partner's ones. Moreover, the combination of media was relevant, since feedback acts that were

then improved in the text were not always found in every media, so there were exclusive aspects the students related to.

5.3. Limitations

Concerning the method, one clear limitation is that we could not clearly identify which media of feedback, oral, sheet or file, were responsible for certain improvements. Instead, we attempted to establish a coherence between the final product and quantitative and qualitative aspects of the process. Furthermore, the evaluation of the argumentation texts only considered complete and complex arguments (opinion, evidence and link of validity). This led to a rather strict evaluation of arguments. Another limitation is that this case study did not have a control group, e.g., which worked with only expert-feedback or individual text testing the hypothesis of peer benefits. An addition of a control group design in the future would help to provide greater substance to the implementation of the method in disciplinary contexts. Such an addition would also enable more in-depth research into more variables and lead to understanding of other factors that might have partially contributed to the differences in the quality of the written arguments.

Furthermore, it can be argued that students, because of the newspaper article, might be constrained to arrive at certain conclusions rather than invited to construct arguments integrating personal views and views expressed by others. It might be worth discussing that the worksheet imposes another structure on the writer and that the peer feedback can create the impression that it is mediated by authority. This kind of anticipated limitation can be reduced by pointing out that the structure is merely understood as a tool of support and that students are allowed to overcome this and that they should use the free space to note own ideas. However, the whole teaching sequence and the focus on the individual involvement concerning the issue was helpful to reduce a risk of authoritative mediation, since interdependent cooperation was a key aspect in the classes and the individual participation played a central role.

Nevertheless, the case-study offers relevant insight into how successful peer feedback teams reciprocally cooperate together, how different feedback can be presented among the groups and where actual deficits in argumentation and according feedback exist, which should be responded to with different means of support.

5.4. Implications and Prospects

Similar future studies need to give insights into which stages of receptive, interactive and productive acts of written argumentation could be processed effectively by students, in which environments in order to specify written argumentation, and its improvement through peer feedback as an approach of disciplinary literacy. It might be helpful for the reader to be reminded that the pairs in the study were of mixed abilities (s.p. 9). Final thoughts can be divided into three stages: the first relates to preparation of the feedback process and material design. The structure of the feedback sheet is relevant and intensively reflected concerning the teaching objectives when it comes to its design since it crucially guides students in different stages of the feedback process [109]. A preparing instructional unit or constructive collection of feedback criteria with the class should focus on the relevance of textual references and reasons and justification for feedback [112]. A flipped classroom stage, for instance, with tutorial videos, e.g., "how to give proper feedback for your partner's argumentation" for students may be appropriate. Hovardas, Tsivitanidou and Zacharia [111] showed what non-expert feedback is lacking when compared to that provided by experts. The non-expert feedback by the students here, which is to be stressed again, was predominantly correct. The best practice teams of students here followed what Hovardas, Tsivitanidou and Zacharia [111] found to be typical for expert feedback. This result suggests that one could hand over more responsibility in terms of feedback to the students. Furthermore, the emphatic interaction [134], specific suggestions for text improvements using examples [135] and advice could be discussed in class, to allow time for questions and involvement

during feedback. Students should be prepared for the subject-specific criteria of argumentation so that students achieve a similar intensive access and understanding of the socio-scientific issue [17].

In the second stage, the interactional side, the results showed the difference between appreciating, negotiating teams and listing, and the effectiveness of pairing students with different opinions towards the issue in question. Future studies could reveal more about the importance and influence of particular partners to construct a similar atmosphere in the interactional setting.

The last stage that has to be considered is the stage of the quality and acceptance of feedback. Again, reflecting with the students on the reasons, references and specific advice after the feedback stages will help improve the process. The results showed that there was a different level of acceptance and understanding of feedback, which differed in complexity with the feedback acceptance of spelling to argumentation. Another aspect that is to be considered is how correct content in socio-scientific issues is maintained. A possibility might be to work with hint cards, where further input and information is listed that helps students to figure out a solution when they struggle with diverse aspects of their discourses during their interaction. Furthermore, the opinions on the topic were quite varied among the students.

Finally, we think peer feedback is a beneficial method to improve written argumentation skills as a part of introducing students to disciplinary literacy, and as a constructive process of developing classroom initiatives and critical analysis of empirical evidence. In a world where climate change is denied by political leaders and right-wing populists rise to power, skilled argumentation is an important tool, and researchers and educators have a great responsibility to teach these skills to the following generations.

Author Contributions: Conceptualization, M.M. and A.B.; methodology, M.M. and A.B.; software, M.M.; validation, M.M. and A.B.; formal analysis, M.M. and A.B.; investigation, M.M. and A.B.; resources, M.M.; data curation, M.M.; writing—original draft preparation, M.M.; writing—review and editing, M.M. and A.B.; visualization, M.M.; supervision, A.B.; project administration, A.B.

Funding: This research received no external funding.

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

References

1. Budke, A.; Meyer, M. Fachlich argumentieren lernen—Die Bedeutung der Argumentation in den unterschiedlichen Schulfächern. In *Fachlich Argumentieren Lernen. Didaktische Forschungen zur Argumentation in den Unterrichtsfächern. Learning How to Argue in the Subjects*; Budke, A., Kuckuck, M., Meyer, M., Schäbitz, F., Schlüter, K., Weiss, G., Eds.; Waxmann: Münster, Germany, 2015; pp. 9–31.
2. Deutsche Gesellschaft für Geographie (DGfG)—German Association for Geography. *Bildungsstandards im Fach Geographie für den Mittleren Schulabschluss mit Aufgabenbeispielen. Educational Standards in Geography for the Intermediate School Certificate*; Selbstverlag DGfG: Bonn, Germany, 2014.
3. National Research Council. *New Research Opportunities in the Earth Sciences*; The National Academies Press: Washington, DC, USA, 2012. [CrossRef]
4. Achieve. Next Generation Science Standards. 5 May 2012. Available online: http://www.nextgenerationscience.org (accessed on 16 June 2018).
5. Osborne, J. Teaching Scientific Practices: Meeting the Challenge of Change. *J. Sci. Teach. Educ.* **2014**, *25*, 177–196. [CrossRef]
6. Jetton, T.L.; Shanahan, C. Preface. In *Adolescent Literacy in the Academic Disciplines: General Principles and Practical Strategies*; Jetton, T.L., Shanahan, C., Eds.; Guilford Press: New York, NY, USA, 2012; pp. 1–34.
7. Norris, S.; Phillips, L. How literacy in its fundamental sense is central to scientific literacy. *Sci. Educ.* **2003**, *87*, 223–240. [CrossRef]
8. Leisen, J. *Handbuch Sprachförderung im Fach. Sprachsensibler Fachunterricht in der Praxis. A Guide to Language Support in the Subjects. Language-Aware Subject Teaching in Practice*; Varus: Bonn, Germany, 2010.

9. Becker-Mrotzek, M.; Schramm, K.; Thürmann, E.; Vollmer, H.K. Sprache im Fach—Einleitung [Subject-language—An introduction]. In *Sprache im Fach. Sprachlichkeit und Fachliches Lernen. Subject-Langugae and Learning in the Sciences*; Becker-Mrotzek, M., Schramm, M.K., Thürmann, E., Vollmer, H.K., Eds.; Waxmann: Berlin, Germany, 2013; pp. 7–24.
10. Budke, A.; Kuckuck, M. (Eds.) *Sprache im Geographieunterricht. Bilinguale und sprachsensible Materialien und Methoden. Language in Geography Education. Bilingual and Language-Aware Material*; Waxmann: Münster, Germany, 2017; Available online: https://www.waxmann.com/?eID=texte&pdf=3550Volltext.pdf&typ=zusatztext (accessed on 16 June 2018).
11. Cummins, J. Interdependence of first- and second-language proficiency in bilingual children. In *Language Processing in Bilingual Children*; Bialystok, E., Ed.; Cambridge University Press: Cambridge, UK, 1991; pp. 70–89.
12. Common Core State Standards Initiative. Common Core State Standards for English Language Arts and Literacy in History/Social Studies and Science. 2010. Available online: http://www.corestandards.org/ (accessed on 16 June 2018).
13. Bazerman, C. Emerging perspectives on the many dimensions of scientific discourse. In *Reading Science*; Martin, J.R., Veel, R., Eds.; Routledge: London, UK, 1998; pp. 15–28.
14. Halliday, M.A.K.; Martin, J.R. *Writing Science: Literacy and Discursive Power*; University of Pittsburgh Press: Pittsburgh, PA, USA, 1993.
15. Pearson, J.C.; Carmon, A.; Tobola, C.; Fowler, M. Motives for communication: Why the Millennial generation uses electronic devices. *J. Commun. Speech Theatre Assoc. N. D.* **2010**, *22*, 45–55.
16. Weiss, I.R.; Pasley, J.D.; Smith, P.; Banilower, E.R.; Heck, D.J. *A Study of K–12 Mathematics and Science Education in the United States*; Horizon Research: Chapel Hill, NC, USA, 2003.
17. Evagorou, M.; Osborne, J. Exploring Young Students' Collaborative Argumentation Within a Socioscientific Issue. *J. Res. Sci. Teach.* **2013**, *50*, 209–237. [CrossRef]
18. Budke, A. Förderung von Argumentationskompetenzen in aktuellen Geographieschulbüchern. In *Aufgaben im Schulbuch*; Matthes, E., Heinze, C., Eds.; Julius Klinkhardt: Bad Heilbrunn, Germany, 2011.
19. Erduran, S.; Simon, S.; Osborne, J. TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Sci. Educ.* **2004**, *88*, 915–933. [CrossRef]
20. Bell, P. Promoting Students' Argument Construction and Collaborative Debate in the Science Classroom. In *Internet Environments for Science Education*; Linn, M., Davis, E., Bell, P., Eds.; Lawrence Erlbaum: Hillsdale, NJ, USA, 2004; pp. 115–143.
21. Newton, P.; Driver, R.; Osborne, J. The place of argumentation in the pedagogy of school science. *Int. J. Sci. Educ.* **1999**, *21*, 553–576. [CrossRef]
22. Budke, A.; Kuckuck, M.; Morawski, M. Sprachbewusste Kartenarbeit? Beobachtungen zum Karteneinsatz im Geographieunterricht. Language-aware analysis of teaching maps. *GW-Unterricht* **2017**, *148*, 5–15. [CrossRef]
23. Yu, S.; Lee, I. Peer feedback in second language writing (2005–2014). *Lang. Teach.* **2016**, *49*, 461–493. [CrossRef]
24. Diab, N.W. Assessing the relationship between different types of student feedback and the quality of revised writing. *Assess. Writ.* **2011**, *14*, 274–292. [CrossRef]
25. Berggren, J. Learning from giving feedback: A study of secondary-level students. *ELT J.* **2015**, *69*, 58–70. [CrossRef]
26. Lehnen, K. Gemeinsames Schreiben. Writing together. In *Schriftlicher Sprachgebrauch/Texte Verfassen. Writing Texts. Written Usage of Language*; Feilke, H., Pohl, T., Eds.; Schneider Hohengehren: Baltmannsweiler, Germany, 2014; pp. 414–431.
27. Rapanta, C.; Garcia-Mila, M.; Gilabert, S. What Is Meant by Argumentative Competence? An Integrative Review of Methods of Analysis and Assessment in Education. *Rev. Educ. Res.* **2013**, *83*, 438–520. [CrossRef]
28. Mercer, N.; Dawes, L.; Wegerif, R.; Sams, C. Reasoning as a scientist: Ways of helping children to use language to learn science. *Br. Educ. Res. J.* **2004**, *30*, 359–377. [CrossRef]
29. Roschelle, J.; Teasley, S. *The Construction of Shared Knowledge in Collaborative Problem Solving*; Springer: Heidelberg, Germany, 1995.
30. Michalak, M.; Lemke, V.; Goeke, M. *Sprache im Fachunterricht: Eine Einführung in Deutsch als Zweitsprache und Sprachbewussten Unterricht. Language in the Subject: An Introduction in German as A Second Language and Language-Aware Teaching*; Narr Francke Attempto: Tübingen, Germany, 2015.

31. Duschl, R.A.; Schweingruber, H.A.; Shouse, A.W. *Taking Science to School: Learning and Teaching Science in Grades K–8*; National Academic Press: Washington, DC, USA, 2007.
32. National Research Council. *Inquiry and the National Science Education Standards*; National Academy Press: Washington, DC, USA, 2000; Available online: http://www.nap.edu/openbook.php?isbn=0309064767 (accessed on 16 June 2018).
33. European Union Recommendation of the European Parliament and of the Council of 18 December 2006 on Key Competences for Lifelong Learning. 2006. Available online: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:394:0010:0018:en:PDF (accessed on 16 June 2018).
34. Asterhan, C.S.C.; Schwarz, B.B. Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog. *Cogn. Sci.* **2009**, *33*, 374–400. [CrossRef] [PubMed]
35. Weinstock, M.P. Psychological research and the epistemological approach to argumentation. *Informal Logic* **2006**, *26*, 103–120. [CrossRef]
36. Weinstock, M.P.; Neuman, Y.; Glassner, A. Identification of informal reasoning fallacies as a function of epistemological level, grade level, and cognitive ability. *J. Educ. Psychol.* **2006**, *98*, 327–341. [CrossRef]
37. Hoogen, A. Didaktische Rekonstruktion des Themas Illegale Migration Argumentationsanalytische Untersuchung von Schüler*innenvorstellungen im Fach Geographie. Didactical Reconstruction of the Topic "Illegal Migration". Analyis of Argumentation among Pupils' Concepts. Ph.D. Thesis, 2006. Available online: https://www.uni-muenster.de/imperia/md/content/geographiedidaktische-forschungen/pdfdok/band_59.pdf (accessed on 16 June 2018).
38. Means, M.L.; Voss, J.F. Who reasons well? Two studies of informal reasoning among children of different grade, ability and knowledge levels. *Cogn. Instr.* **1996**, *14*, 139–178. [CrossRef]
39. Reznitskaya, A.; Anderson, R.; McNurlen, B.; Nguyen-Jahiel, K.; Archoudidou, A.; Kim, S. Influence of oral discussion on written argument. *Discourse Process.* **2001**, *32*, 155–175. [CrossRef]
40. Kuhn, D. *Education for Thinking*; Harvard University Press: Cambridge, MA, USA, 2005.
41. Leitao, S. The potential of argument in knowledge building. *Hum. Dev.* **2000**, *43*, 332–360. [CrossRef]
42. Nussbaum, E.M.; Sinatra, G.M. Argument and conceptual engagement. *Contemp. Educ. Psychol.* **2003**, *28*, 573–595. [CrossRef]
43. Kuckuck, M. Konflikte im Raum—Verständnis von Gesellschaftlichen Diskursen Durch Argumentation im Geographieunterricht. Spatial Conflicts—How Pupils Understand Social Discourses through Argumentation. Ph.D. Thesis, Monsenstein und Vannerdat, Münster, Germany, 2014. Geographiedidaktische Forschungen, Bd. 54.
44. Dittrich, S. Argumentieren als Methode zur Problemlösung Eine Unterrichtsstudie zur Mündlichen Argumentation von Schülerinnen und Schülern in kooperativen Settings im Geographieunterricht. Solving Problems with Argumentation in the Geography Classroom. Ph.D. Thesis, 2017. Available online: https://www.uni-muenster.de/imperia/md/content/geographiedidaktische-forschungen/pdfdok/gdf_65_dittrich.pdf (accessed on 16 June 2018).
45. Müller, B. Komplexe Mensch-Umwelt-Systeme im Geographieunterricht mit Hilfe von Argumentationen erschließen. Analyzing Complex Human-Environment Systems in Geography Education with Argumentation. Ph.D. Thesis, 2016. Available online: http://kups.ub.unikoeln.de/7047/4/Komplexe_Systeme_Geographieunterricht_Beatrice_Mueller.pdfNational (accessed on 16 June 2018).
46. Leder, J.S. Pedagogic Practice and the Transformative Potential of Education for Sustainable Development. Argumentation on Water Conflicts in Geography Teaching in Pune, India. Ph.D. Thesis, Universität zu Köln, Cologne, Germany, 2016. Available online: http://kups.ub.uni-koeln.de/7657/ (accessed on 16 June 2018).
47. Kopperschmidt, J. *Argumentationstheorie. Theory of Argumentation*; Junius: Hamburg, Germany, 2000.
48. Kienpointner, M. *Argumentationsanalyse. Analysis of Argumentation*; Innsbrucker Beiträge zur Kulturwissenschaft 56; Institut für Sprachwissenschaften der Universität Innsbruck: Innsbruck, Austria, 1983.
49. Angell, R.B. Argumentationsrezeptionskompetenzen im Vergleich der Fächer Geographie, Biologie und Mathematik. In *Reasoning and logic*; Budke, A., Ed.; Appleton-Century-Crofts: New York, NY, USA, 1964.
50. Toulmin, S. *The Use of Arguments*; Cambridge University Press: Cambridge, MA, USA, 1958.
51. Walton, D.N. *The New Dialectic: Conversational Contexts of Argument*; University of Toronto Press: Toronto, ON, Canada, 1998.
52. Perelman, C. *The Realm of Rhetoric*; University of Notre Dame Press: Notre Dame, IN, USA, 1982.

53. Felton, M.; Kuhn, D. The development of argumentative discourse skills. *Discourse Processes* **2001**, *32*, 135–153. [CrossRef]
54. Kuhn, D. *The Skills of Argument*; Cambridge University Press: Cambridge, UK, 1991.
55. Kuhn, D.; Shaw, V.; Felton, M. Effects of dyadic interaction on argumentative reasoning. *Cogn. Instr.* **1997**, *15*, 287–315. [CrossRef]
56. Zohar, A.; Nemet, F. Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *J. Res. Sci. Teach.* **2002**, *39*, 35–62. [CrossRef]
57. Erduran, S.; Jiménez-Aleixandre, M.P. *Argumentation in Science Education: Perspectives from Classroom-Based Research*; Springer: Dordrecht, The Netherlands, 2008.
58. Garcia-Mila, M.; Andersen, C. Cognitive foundations of learning argumentation. In *Argumentation in Science Education: Perspectives from Classroom-Based Research*; Erduran, S., Jiménez-Aleixandre, M.P., Eds.; Springer: Dordrecht, The Netherlands, 2008; pp. 29–47.
59. McNeill, K.L. Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Sci. Educ.* **2008**, *93*, 233–268. [CrossRef]
60. Billig, M. *Arguing and Thinking: A Rhetorical Approach to Social Psychology*; Cambridge University Press: Cambridge, UK, 1987.
61. Bachmann, T.; Becker-Mrotzek, M. Schreibaufgaben situieren und profilieren. To profile writing tasks. In *Textformen als Lernformen. Text Forms als Learning Forms*; Pohl, T., Steinhoff, T., Eds.; Gilles & Francke: Duisburg, Germany, 2010; pp. 191–210.
62. Feilke. Schreibdidaktische Konzepte. In *Forschungshandbuch Empirische Schreibdidaktik. Handbook on Empirical Teaching Writing Research*; Becker-Mrotzek, M., Grabowski, J., Steinhoff, T., Eds.; Waxmann: Münster, Germany, 2017; pp. 153–173.
63. Hayes, J.; Flower, L. Identifying the organization of writing processes. In *Cognitive Processes in Writing*; Gregg, L., Steinberg, E., Eds.; Erlbaum: Hillsdale, MI, USA, 1980; pp. 3–30.
64. Dawson, V.; Venville, G. High-school students' informal reasoning and argumentation about biotechnology: An indicator of scientific literacy? *Int. J. Sci. Educ.* **2009**, *31*, 1421–1445. [CrossRef]
65. Toulmin, S. *The Use of Arguments*; Beltz: Weinheim, Germany, 1996.
66. CEFR (2001)—Europarat (2001). *Gemeinsamer Europäischer Referenzrahmen für Sprachen: Lernen, Lehren, Beurteilen. Common European Framework for Languages*; Langenscheidt: Berlin, Germany; München, Germany, 2011; Available online: https://www.coe.int/en/web/portfolio/the-common-european-framework-of-reference-for-languages-learning-teaching-assessment-cefr- (accessed on 16 June 2018).
67. Jimenez-Aleixandre, M.; Erduran, S. Argumentation in science education: An overview. In *Argumentation in Science Education: Perspectives from Classroom-Based Research*; Erduran, S., Jimenez-Aleixandre, M., Eds.; Springer: New York, NY, USA, 2008; pp. 3–27.
68. Evagorou, M. Discussing a socioscientific issue in a primary school classroom: The case of using a technology-supported environment in formal and nonformal settings. In *Socioscientific Issues in the Classroom*; Sadler, T., Ed.; Springer: New York, NY, USA, 2011; pp. 133–160.
69. Evagorou, M.; Jime'nez-Aleixandre, M.; Osborne, J. 'Should We Kill the Grey Squirrels? 'A Study Exploring Students' Justifications and Decision-Making. *Int. J. Sci. Educ.* **2012**, *34*, 401–428. [CrossRef]
70. Nielsen, J.A. Science in discussions: An analysis of the use of science content in socioscientific discussions. *Sci. Educ.* **2012**, *96*, 428–456. [CrossRef]
71. Oliveira, A.; Akerson, V.; Oldfield, M. Environmental argumentation as sociocultural activity. *J. Res. Sci. Teach.* **2012**, *49*, 869–897. [CrossRef]
72. Kuhn, D. Teaching and learning science as argument. *Sci. Educ.* **2010**, *94*, 810–824. [CrossRef]
73. Kuckuck, M. Die Rezeptionsfähigkeit von Schüler*innen und Schülern bei der Bewertung von Argumentationen im Geographieunterricht am Beispiel von raum-bezogenen Konflikten. *Z. Geogr.* **2015**, *43*, 263–284.
74. Budke, A.; Weiss, G. Sprachsensibler Geographieunterricht. In *Sprache als Lernmedium im Fachunterricht. Theorien und Modelle für das Sprachbewusste Lehren und Lernen. Language as Learning Medium in the Subjects*; Michalak, M., Ed.; Schneider Hohengehren: Baltmannsweiler, Germany, 2014; pp. 113–133.

75. Budke, A.; Uhlenwinkel, A. Argumentieren im Geographieunterricht. Theoretische Grundlagen und unterrichtspraktische Umsetzungen. In *Geographische Bildung. Kompetenzen in der didaktischer Forschung und Schulpraxis. Geography Education. Competences in Educational Research and Practical Implementation*; Meyer, C., Henry, R., Stöber, G., Eds.; Westermann: Braunschweig, Germany, 2011; pp. 114–129.
76. Budke, A.; Kuckuck, M. (Eds.) *Politische Bildung im Geographieunterricht. Civic Education in Teaching Geography*; Franz Steiner Verlag: Stuttgart, Germany, 2016.
77. Uhlenwinkel, A. Geographisches Wissen und geographische Argumentation. In *Fachlich Argumentieren Lernen. Didaktische Forschungen zur Argumentation in den Unterrichtsfächern. Learning to Argue. Empirical Research about Argumenation in the Subjects*; Budke, A., Kuckuck, M., Meyer, M., Schäbitz, F., Schlüter, K., Weiss, G., Eds.; Waxmann: Münster, Germany, 2015; pp. 46–61.
78. Budke, A.; Michalak, M.; Kuckuck, M.; Müller, B. Diskursfähigkeit im Fach Geographie—Förderung von Kartenkompetenzen in Geographieschulbüchern. In *Befähigung zu Gesellschaftlicher Teilhabe. Beiträge der Fachdidaktischen Forschung. Enabling Social Participation*; Menthe, J., Höttecke, J.D., Zabka, T., Hammann, M., Rothgangel, M., Eds.; Waxmann: Münster, Germany, 2016; Bd. 10; pp. 231–246.
79. Lehnen, K. Kooperatives Schreiben. In *Handbuch Empirische Schreibdidaktik. Handbook on Empirical Teaching Writing Research*; Becker-Mrotzek, M., Grabowski, J., Steinhoff, T., Eds.; Waxmann: Münster, Germany, 2017; pp. 299–315.
80. Dillenbourg, P.; Baker, M.; Blaye, A.; O'Malley, C. *The Evolution of Research on Collaborative Learning*; Elsevier: Oxford, UK, 1996.
81. Andriessen, J. Arguing to learn. In *The Cambridge Hand-Book of the Learning Sciences*; Sawyer, R.K., Ed.; Cambridge University Press: New York, NY, USA, 2005; pp. 443–459.
82. Berland, L.; Reiser, B. Classroom communities' adaptations of the practice of scientific argumentation. *Sci. Educ.* **2010**, *95*, 191–216. [CrossRef]
83. Teasley, S.D.; Roschelle, J. Constructing a joint problem space: The computer as a tool for sharing knowledge. In *Computers as Cognitive Tools*; Lajoie, S.P., Derry, S.J., Eds.; Lawrence Erlbaum Associates, Inc.: Mahwah, NJ, USA, 1993; pp. 229–258.
84. Berland, L.K.; Hammer, D. Framing for scientific argumentation. *J. Res. Sci. Teach.* **2012**, *49*, 68–94. [CrossRef]
85. Damon, W. Peer education: The untapped potential. *J. Appl. Dev. Psychol.* **1984**, *5*, 331–343. [CrossRef]
86. Hatano, G.; Inagaki, K. Sharing cognition through collective comprehension activity. In *Perspectives on Socially Shared Cognition*; Resnick, L., Levine, J.M., Teasley, S.D., Eds.; American Psychological Association: Washington, DC, USA, 1991; pp. 331–348.
87. Chi, M.; De Leeuw, N.; Chiu, M.H.; Lavancher, C. Eliciting Self-Explanations Improves Understanding. *Cogn. Sci.* **1994**, *18*, 439–477.
88. King, A. Enhancing peer interaction and learning in the classroom through reciprocal interaction. *Am. Educ. Res. J.* **1990**, *27*, 664–687. [CrossRef]
89. Coyle, D.; Hood, P.; Marsh, D. *CLIL: Content and Language Integrated Learning*; Cambridge University Press: Cambridge, UK, 2010.
90. Morawski, M.; Budke, A. Learning with and by Language: Bilingual Teaching Strategies for the Monolingual Language-Aware Geography Classroom. *Geogr. Teach.* **2017**, *14*, 48–67. [CrossRef]
91. Breidbach, S.; Viebrock, B. CLIL in Germany—Results from Recent Research in a Contested Field of Education. *Int. CLIL Res. J.* **2012**, *1*, 5–16.
92. Gibbs, G.; Simpson, C. Conditions under which Assessment supports Student Learning. *Learn. Teach. High. Educ.* **2004**, *1*, 3–31.
93. Diab, N.W. Effects of peer- versus self-editing on students' revision of language errors in revised drafts. *Syst. Int. J. Educ. Technol. Appl. Linguist.* **2010**, *38*, 85–95. [CrossRef]
94. Zhao, H. Investigating learners' use and understanding of peer and teacher feedback on writing: A comparative study in a Chinese English writing classroom. *Assess. Writ.* **2010**, *1*, 3–17. [CrossRef]
95. S'eror, J. Alternative sources of feedback and second language writing development in university content courses. *Can. J. Appl. Linguist./Revue canadienne de linguistique appliqu'ee* **2011**, *14*, 118–143.
96. Lundstrom, K.; Baker, W. To give is better than to receive: The benefits of peer review to the reviewer's own writing. *J. Second Lang. Writ.* **2009**, *18*, 30–43. [CrossRef]
97. Pyöriä1, P.; Ojala, S.; Saari, T.; Järvinen, K. The Millennial Generation: A New Breed of Labour? *Sage Open* **2017**, *7*, 1–14. [CrossRef]

98. Bresman, H.; Rao, V. A Survey of 19 Countries Shows How Generations X, Y, and Z Are—And Aren't—Different. 2017. Available online: https://hbr.org/2017/08/a-survey-of-19-countries-shows-how-generations-x-y-and-z-are-and-arent-different (accessed on 16 June 2018).
99. Baur, R.; Goggin, M.; Wrede-Jackes, J. *Der c-Test: Einsatzmöglichkeiten im Bereich DaZ. C-test: Usability in SLT*; proDaz; Universität Duisburg-Essen: Duisburg, Germany; Stiftung Mercator: Berlin, Germany, 2013.
100. Baur, R.; Spettmann, M. Lesefertigkeiten testen und fördern. Testing and supporting reading skills. In *Bildungssprachliche Kompetenzen Fördern in der Zweitsprache. Supporting Scientific Language in SLT*; Benholz, E., Kniffka, G., Winters-Ohle, B., Eds.; Waxmann: Münster, Germany, 2010; pp. 95–114.
101. Metz, K. Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. *Cogn. Instr.* **2004**, *22*, 219–290. [CrossRef]
102. Sandoval, W.A.; Millwood, K. The quality of students' use of evidence in written scientific explanations. *Cogn. Instr.* **2005**, *23*, 23–55. [CrossRef]
103. Gibbons, P. *Scaffolding Language, Scaffolding Learning. Teaching Second Language Learners in the Mainstream Classroom*; Heinemann: Portsmouth, NH, USA, 2002.
104. Kniffka, G. Scaffolding. 2010. Available online: http://www.uni-due.de/prodaz/konzept.php (accessed on 11 October 2017).
105. Hogan, K.; Nastasi, B.; Pressley, M. Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cogn. Instr.* **1999**, *17*, 379–432. [CrossRef]
106. Light, P. Collaborative learning with computers. In *Language, Classrooms and Computers*; Scrimshaw, P., Ed.; Routledge: London, UK, 1993; pp. 40–56.
107. Alexopoulou, E.; Driver, R. Small group discussions in physics: Peer inter-action in pairs and fours. *J. Res. Sci. Teach.* **1996**, *33*, 1099–1114. [CrossRef]
108. Bennett, J.; Hogarth, S.; Lubben, F.; Campell, B.; Robinson, A. Talking science: The research evidence on the use of small group discussions in science teaching. *Int. J. Sci. Educ.* **2010**, *32*, 69–95. [CrossRef]
109. Gielen, S.; DeWever, B. Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Comput. Hum. Behav.* **2015**, *52*, 315–325. [CrossRef]
110. Morawski, M.; Budke, A. Förderung von Argumentationskompetenzen durch das "Peer-Review-Verfahren" Ergebnisse einer empirischen Studie im Geographieunterricht unter besonderer Berücksichtigung von Schüler*innen, die Deutsch als Zweitsprache lernen. Teaching argumentation competence with peer-review. Results of an empirical study in geography education with special emphasis on students who learn German as a second language. *Beiträge zur Fremdsprachenvermittlung, Special Issue* **2018**, in press.
111. Hovardas, T.; Tsivitanidou, O.; Zacharia, Z. Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Comput. Educ.* **2014**, *71*, 133–152. [CrossRef]
112. Gielen, S.; Peeters, E.; Dochy, F.; Onghena, P.; Struyven, K. Improving the effectiveness of peer feedback for learning. *Learn. Instr.* **2010**, *20*, 304–315. [CrossRef]
113. Becker-Mrotzek, M.; Böttcher, I. *Schreibkompetenz Entwickeln und Beurteilen*; Cornelsen: Berlin, Germany, 2012.
114. Neumann, A. Zugänge zur Bestimmung von Textqualität. In *Handbuch Empirische Schreibdidaktik. Handbook on Empirical Teaching Writing Research*; Becker-Mrotzek, M., Grabowski, J., Steinhoff, T., Eds.; Waxmann: Münster, Germany, 2017; pp. 173–187.
115. Denzin, N.K. *The Research Act. A Theoretical Introduction to Sociological Methods*; Aldine Pub: Chicago, IL, USA, 1970.
116. Rost, J. Theory and Construction of Tests. In *Testtheorie Testkonstruktion*; Verlag Hans Huber: Bern, Switzerland; Göttingen, Germany; Toronto, ON, Canada; Seattle, WA, USA, 2004.
117. Bortz, J.; Döring, N. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Empirical Methods and Evaluation for the Human and Social Sciences*, 4th ed.; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 2006.
118. Wirtz, M.; Caspar, F. *Beurteilerübereinstimmung und Beurteilerreliabilität. Reliability of Raters*; Hogrefe: Göttingen, Germany, 2002.
119. Bowen, G. Document Analysis as a Qualitative Research Method. *Qual. Res. J.* **2009**, *9*, 27–40. [CrossRef]
120. Gumperz, J.; Berenz, N. Transcribing Conversational Exchanges. In *Talking Data: Transcription and Coding in Discourse Research*; Edwards, J.A., Lampert, M.D., Eds.; Psychology Press: Hillsdale, NJ, USA, 1993; pp. 91–121.

121. Atkinson, J.M. *Structures of Social Action: Studies in Conversation Analysis*; Cambridge University Press: Cambridge, UK; Editions de la Maison des Sciences de l'Homme: Paris, France, 1984.
122. Langford, D. *Analyzing Talk: Investigating Verbal Interaction in English*; Macmillan: Basingstoke, UK, 1994.
123. Searle, J. The classification of illocutionary acts. *Lang. Soc.* **1976**, *5*, 1–24. [CrossRef]
124. Herrenkohl, L.; Guerra, M. Participant structures, scientific discourse, and student engagement in fourth grade. *Cogn. Instr.* **1998**, *16*, 431–473. [CrossRef]
125. Mayring, P. *Qualitative Inhaltsanalyse. Qualitative Analysis of Content*, 12th ed.; Beltz: Weinheim, Geremany, 2015.
126. Chiu, M. Effects of argumentation on group micro-creativity: Statistical discourse analyses of algebra students' collaborative problem solving. *Contemp. Educ. Psychol.* **2008**, *33*, 382–402. [CrossRef]
127. Coleman, E.B. Using explanatory knowledge during collaborative problem solving in science. *J. Learn. Sci.* **1998**, *7*, 387–427. [CrossRef]
128. Mercer, N.; Wegerif, R.; Dawes, L. Children's talk and the development of reasoning in the classroom. *Br. Educ. Res. J.* **1999**, *25*, 95–111. [CrossRef]
129. Resnick, L.; Michaels, S.; O'Connor, C. How (well-structured) talk builds the mind. In *From Genes to Context: New Discoveries about Learning from Educational Research and Their Applications*; Sternberg, R., Preiss, D., Eds.; Springer: New York, NY, USA, 2010; pp. 163–194.
130. Chin, C.; Osborne, J. Supporting argumentation through students' questions: Case studies in science classrooms. *J. Learn. Sci.* **2010**, *19*, 230–284. [CrossRef]
131. Ryu, S.; Sandoval, W. Interpersonal Influences on Collaborative Argument during Scientific Inquiry. In Proceedings of the American Educational Research Association (AERA), New York, NY, USA, 24–29 March 2008.
132. Sampson, V.; Clark, D.B. A comparison of the collaborative scientific argumentation practices of two high and two low performing groups. *Res. Sci. Educ.* **2009**, *41*, 63–97. [CrossRef]
133. Van Amelsvoort, M.; Andriessen, J.; Kanselaar, G. Representational tools in computer-supported argumentation-based learning: How dyads work with constructed and inspected argumentative diagrams. *J. Learn. Sci.* **2007**, *16*, 485–522. [CrossRef]
134. Beach, R.; Friedrich, T. Response to writing. In *Handbook of Writing Research*; MacArthur, C., Graham, S., Fitzgerald, J., Eds.; Guilford: New York, NY, USA, 2006; pp. 222–234.
135. Nelson, M.; Schunn, C. The nature of feedback: How different types of peer feedback affect writing performance. *Instr. Sci.* **2009**, *37*, 375–401. [CrossRef]

© 2019 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/).