

Argumentative Peer Discussions Following Individual Reading
Increase Comprehension

Abstract

Although recent research has documented that classroom discussions may enhance students' reading comprehension, limited knowledge exists regarding the 'working ingredients' that extend students' comprehension of texts through peer discussion. This observational study examines the pre- and post-discussion comprehension of fifth graders, investigating whether particular features of peer discussions immediately following individual reading can support comprehension outcomes. The sample consisted of 102 fifth graders (many of whom were second-language learners) in 21 different multilinguistic classrooms in Norway. Altogether, 25 peer discussions were videotaped, transcribed, and coded according to eight categories of talk moves. The findings showed that the number of talk moves offering relevant arguments and counterarguments predicted significant variance in the students' post-discussion comprehension scores (controlling for pre-discussion comprehension, prior topic knowledge, vocabulary, and word-decoding skills). These findings suggest that the argumentative qualities of the peer discussions were linked to improvements in the students' comprehension outcomes.

Key words

Peer discussion

Reading comprehension

Vocabulary

Multilinguistic classrooms

Elementary education

Introduction

Reading comprehension has been described as a process of simultaneously extracting and constructing meaning through interaction and involvement with written language (Snow and RAND Reading Study Group 2002). Although it is well established that reading comprehension builds on students' oral language skills (e.g., vocabulary) to a large extent, less is known about how knowledge acquired through oral communication and dialogue supports reading comprehension. There are distinct differences between the communication and comprehension processes involved in reading and speaking. In written texts, words are typically chosen with greater preciseness and arguments constructed more systematically. Readers can move back and forth in the text to make sense of the perspectives and ideas that the author is attempting to convey. At the same time, critical evaluation of the text content during individual reading may be difficult for many young students. Oral communication in dialogue typically allows for more fluid exchanges of meaning and understanding. The speaker can adjust to the verbal and nonverbal cues conveyed by the listener, as well as draw on a wide repertoire of signals to explain a phenomenon or express his or her own perspective. Additionally, when peers communicate about a topic, the power relationships are relatively symmetrical and may allow for questions and disagreement to surface. As such, opportunities for participation in text-related discussions may potentially support the process of constructing meaning and identifying different perspectives during reading.

This article is based on an observational study, conducted in the fifth grade that investigated whether added effects on comprehension resulted when peers negotiated text after reading. Schools in Norway have become increasingly multilinguistic, serving children of immigrants from all over the world whose first language is not Norwegian. In Norway, as

in other countries, many children of immigrants struggle with developing proficient reading comprehension skills in the language of schooling, even after several years of exposure (e.g., Melby-Lervåg and Lervåg 2014). Researchers have also expressed a concern that students in classrooms with many second-language learners may have fewer opportunities to engage in rich language discourse related to the academic content that is being taught (see discussion in Cazden 2001; Vaughn et al. 2017). Thus, there is a pressing need to know more about the instructional mechanisms that promote reading comprehension for all students in multiethnic and multilingual classrooms. Previous studies have acknowledged the important role played by peers in students' learning and development (Asterhan and Schwartz 2007, 2009; Chinn, O'Donnell, and Jinks 2000; Kuhn 2015); however, we still know very little about whether and how the qualities of peer dialogue may affect reading comprehension.

The present study was prompted by the growing attention being paid to the effects of discussion and argumentation on students' academic performance (e.g., Applebee et al. 2003; Cazden 2001; Wolf, Crosson, and Resnick 2006; Driver, Newton, and Osborne 2000). Microanalytic studies of students' argumentative talk moves have addressed the impact of participating in discussions about texts (see, for instance, Chinn, Anderson, and Waggoner 2001). However, few studies have assessed the immediate comprehension outcomes among groups of elementary school students participating in task-based discussions of texts. We designed an observational study to address this need for more knowledge by examining whether comprehension outcomes after reading can be improved through peer discussion of texts in multilingual classrooms.

Moving from Individual Skills to Classroom Talk as Predictors of Comprehension

Outcomes after Reading

To date, empirical research on reading comprehension has been characterized by an emphasis on individual processes and characteristics. This extensive literature has thoroughly

demonstrated that students' decoding skills, oral language skills (e.g., vocabulary), and prior topic knowledge are crucial to reading comprehension outcomes (Melby-Lervåg and Lervåg 2014; Rydland, Grøver Aukrust, and Fulland 2012). However, as demonstrated by recent studies, reading comprehension in school also draws on a wide repertoire of academic language skills, such as reasoning skills and the ability to identify different perspectives in texts (LaRusso et al. 2016; Uccelli et al. 2015). These academic language skills are typically not dealt with nor assessed in studies that address elementary students' reading comprehension.

Another line of research embedded within contemporary sociocognitive and sociocultural theory has documented the ways in which knowledge-building and reasoning develop powerfully through discussions in which students move between presenting an understanding, evaluating others', and refining their own (see, for example, Driver, Newton, and Osborne 2000; Kuhn 2015; Reznitskaya 2011). There is considerable evidence that instruction approaches that explicitly emphasize dialogic classroom discussions may promote students' ability to state and support arguments in various subject areas, such as science education (Berland and Reiser 2011; Briker and Bell 2008; Lemke 1990; Mercer et al. 2004) and humanities education (Chinn et al. 2001).

There is also evidence to suggest that students' reading comprehension may be supported through enriched classroom dialogue, especially in multilingual classrooms in which there are many second-language learners. Recent intervention studies have, for instance, demonstrated that students in multilingual classrooms may benefit from explicit instruction approaches to teaching academic content and vocabulary in the context of dialogic discussions (Lawrence et al. 2012; Vaughn et al. 2017). Moreover, in an observational study of teacher talk in predominantly urban multilinguistic schools, Gamez and Lesaux (2015) showed that teachers' use of sophisticated vocabulary in the classroom was significantly

related to sixth-grade students' reading comprehension outcomes at the end of the school year when controlling for their reading comprehension and vocabulary knowledge at the beginning of the year. This finding suggests that more implicit aspects of teaching (the classroom language environment) also play a significant role in students' reading comprehension development.

Investigating Argumentation in Peer Talk

Although teacher talk and support in the classroom is inevitably imperative for students' learning, there has been a call for more research into the interactive features of reasoning among children (see, for instance, Cekaite et al. 2014; Mercier 2011). In a meta-analysis, Murphy et al. (2009) analyzed several instructional approaches to classroom discussions about texts and the impact of these approaches on students' comprehension outcomes. The authors found that although various discussion approaches produced increases in student talk, few effectively promoted students' critical thinking and argumentation about and around texts. Thus, increases in student talk did not necessarily result in improved student comprehension.

Researchers on dialogic classroom discussions (see, for instance, Chinn et al. 2000; Driver, Newton, and Osborne 2000; Kuhn 2015; Leitão 2000; Mercier 2011; Reznitskaya et al. 2009, 2011) have specifically pointed to the types of talk moves that may impact student comprehension. This research emphasizes the relevance and accuracy of student claims and views arguments and counterarguments, in which students defend or attempt to weaken the epistemic status of a previous statement, as important 'working ingredients' in scientific reasoning. Researchers have also pointed out that such interactions are rare in most classrooms (Berland and Reiser 2011; Lemke 1990; McNeill 2011).

Whereas argumentative moves may be relatively easy to identify in conversations related to an overarching social and moral dilemma (an argument either supports or contradicts a position), they are less obvious in dialogues in which students are required to put forward their conceptual understanding of a phenomenon or discuss many different dilemmas (for further discussion, see Asterhan and Schwartz 2009; Rapanta, Garcia-Mila, and Gilabert 2013).

Kuhn (2015) labeled the shared-objective argumentation in which participants work together to construct a common understanding *coalescent argumentation*. She distinguished this from argumentation involving opposing viewpoints in which students respond to peers with counterarguments and rebuttals. Kuhn argued that this more adversarial argumentation may, to a greater extent, move the discussion forward, but she suggested that the two types may build on each other instead of being alternative modes of argumentation. Nonetheless, investigating how peer discussions might influence the ways in which students develop their understanding in the classroom is imperative (Cekaite et al. 2014). To our knowledge, a very limited number of quantitative studies have investigated peer talk that occurs in the classroom after students have read multiple texts about a topic.

In summary, we know a great deal about individual student skills (such as decoding, vocabulary, and prior topic knowledge) that significantly predict students' reading comprehension. However, little research has been conducted to examine how interactional qualities among peers may influence comprehension after reading. The present study, conducted in Norwegian multilingual classrooms, was designed to identify the talk moves of fifth-grade students during peer discussions of texts they had read individually. The following research question guided the study: Which talk moves in peer discussions about text content predict students' comprehension outcomes above what can be explained by their oral

language skills, decoding skills, prior topic knowledge, and initial reading comprehension outcomes?

Method

Background of the Present Study and Recruitment of Participants

The present study was conducted in the same 21 fifth-grade classrooms as another ongoing research project in which we conducted a longitudinal follow-up study of 25 second-language learners. Because we wanted to conduct the present study on peer collaboration in a multilingual setting that included second-language learners, we decided to build peer groups around the 25 students from the longitudinal project.¹ The 21 fifth-grade classrooms were located in multiethnic neighborhoods in the larger Oslo area, and approximately 60% of the students in these classrooms reported using a language in addition to Norwegian (such as Turkish, Urdu, Somali, and various Eastern European languages) at home. Norwegian was the language of instruction and communication at the schools.

We asked the lead teacher in each classroom to select a group of students (boys and girls) with whom those from the original study would normally interact. These peers, comprising both monolingual and second-language speakers of Norwegian, were sent home with consent forms requesting permission to video-record them in a group discussion and conduct a set of language and reading assessments. We recruited a total of 102 10-year-old students to participate in the video-recorded peer discussions, and they represent the focal students in the present study (46% girls). As described in more detail below, we videotaped one (or, in a few instances, two or three) peer group(s) in each classroom during a time slot devoted to peer discussion. As many as 83% of the focal students had parents who were immigrants to Norway; this was likely because of the procedures that we used to recruit the

¹ Two classrooms were attended by two students, and one classroom was attended by three students from the original longitudinal study.

participants. All the focal students had grown up in Norway, and the vast majority had attended preschool (97%).

Procedure

The focal students participated in a classroom test session followed by an individual test session. Both sessions were administered by us, the researchers. Our research team developed the text material and assessments that were used in the classroom test session. This session included all the students in the class and was designed to generate peer discussions after they read different texts about a topic. We chose to focus on global warming for three reasons: First, global warming is addressed in texts within numerous subject areas (e.g., Norwegian language arts, social science, and natural science). Second, we believed that students in the fifth grade would still have considerably more to learn about global warming, because the sixth-grade science curriculum provides a thorough introduction to the topic. Finally, although the issue of global warming is often discussed in an atmosphere of moral acceptability (e.g., in terms of what we should and should not do to prevent it), the relevant media debates are characterized by the fact that different authoritative figures (e.g., politicians and researchers) present divergent perspectives on the reasons for it, as well as its consequences. Thus, we believed that the students would learn from the texts and that the topic would elicit further discussion.

The classroom test session lasted for approximately two hours and was conducted in five consecutive steps, which are illustrated in Figure 1 and explicated below.

--- Insert Figure 1 approximately here ----

Instruments

Step 1. Prior topic knowledge. To assess the students' knowledge of the topics of the texts before they were read, the participants were asked to respond to a questionnaire assessing prior topic knowledge on global warming. The questionnaire consisted of 14

multiple-choice tasks (four response alternatives, of which three were distracters) and five concept definition tasks. For the concept definition tasks, the students were asked to write down the meanings of relevant concepts, such as *atmosphere* and *climate* (scored 0–2). Level of abstraction and preciseness were considered when categorizing the answers. To illustrate, a two-point answer for *atmosphere* would state that it is a layer of gases or a shield of air surrounding the earth, while a one-point answer would typically mention ‘space’ or ‘air.’ Cohen’s kappa for the scoring of the concept definition tasks was .84 (interrater agreement) based on the scoring of 20% of the questionnaires (number of questionnaires = 20). Students could obtain a maximum score of 24.

Because previous studies conducted with university students have reported that males display higher levels of prior knowledge on the topic of global warming (e.g., Bråten and Strømsø 2010), we investigated whether such gender differences were also present in our sample of younger students. We found no significant gender differences in the students’ performances on the measure of prior topic knowledge (boys, $M = 6.40$, $SD = 3.48$; girls, $M = 5.92$, $SD = 3.69$; $t(100) = 0.67$; $p = 0.50$). Because the prior topic knowledge assessment was conducted with the focal students’ classmates across the 21 fifth-grade classrooms, we were able to compare the focal students’ scores on the prior topic knowledge test with the mean scores of their classmates. This analysis revealed that 58% of the focal students scored below the mean scores in their class. The scores on the prior topic knowledge measure of the 102 focal students were normally distributed, typically ranging from five points below to five points above the mean score in their class.

Step 2. Individual reading. The students read three texts on global warming: one narrative text (665 words), one peer letter to an editor (435 words), and one expository text (376 words). The narrative text was about Mulu, a young girl flying home to Norway with her family after a visit with their relatives in Sudan. While flying over the Sahara Desert,

Mulu's parents tell her about the consequences of pollution for desert expansion and the hardship that people in Sudan experience as a result of it. In the second text, the peer letter to an editor, 10-year-old Ahmed reflects on the reasons for and consequences of global warming and how it affects his daily life in Norway. The first two texts deal with the moral aspects of pollution and introduce divergent perspectives on how people respond to global warming issues. Finally, the expository text, assembled from two different sixth-grade science textbooks, explains the relationships between pollution and global warming and how the greenhouse effect works.

The students were informed that they were going to work with the topics of the texts in small groups immediately following their reading. To increase the students' investment in individual reading prior to the group task and to ensure collaboration through dialogue in the group task, they were informed that the texts would not be available during their discussions.

Step 3. Pre-discussion comprehension. To assess the students' reading comprehension prior to the peer discussion, they were asked to fill out a researcher-created questionnaire. The questionnaire, Form A of the Global Warming Test (hereafter Form A), consisted of 14 multiple-choice tasks (the questions prompted students to synthesize information and make inferences based on the information in the texts) and seven open-ended questions (such as 'What is global warming?' and 'Why is the desert expanding?'). The open-ended questions were scored on a scale from 0–2 based on the level of preciseness and abstraction in the answers. For example, in response to the question about global warming, a student would attain a score of 2 if he or she was able to explain that the Earth's average surface temperature is increasing due to the effect of greenhouse gases and a score of 1 if he or she mentioned that the ice in the North Pole is melting rapidly. Students could obtain a maximum score of 28.

Step 4. Peer discussions. For the time allocated to peer discussion, the teachers in each classroom separated the students into small groups of three to five. The students received no initial training in how to perform argumentative discussions and were prompted only to ‘discuss with peers.’ Although we aimed to intervene as little as possible, we provided each group with eight guiding questions to help the students focus on the task of discussing the topics of the texts. All questions emphasized relevant aspects of the texts that the students had read and were designed to elicit explanations (e.g., What is the greenhouse effect?) or different perspectives on an issue (e.g., What is your response to Ahmed, who wrote the peer letter?). Furthermore, the questions encouraged the students to not only summarize information from the text but also relate new ideas from it to their prior knowledge (e.g., Why is it important that we lower the temperature in our houses?). The groups were given 20 minutes to discuss the eight guiding questions. Only the groups comprising the focal students were videotaped during the peer discussion. All these groups included at least one second-language learner (see recruitment of participants above). A camera was located on a tripod close to the selected peer group, and a tape recorder was positioned on the desktop in the middle of the group (as a supplement to the microphone on the camera). The researchers who conducted the video recording ensured that the groups remained focused on the assignment, and all groups completed the guiding questions within the allotted 20 minutes. Twenty-five different peer discussions were videotaped and transcribed. The majority of the groups had four participating students, but group size ranged from three to six students ($M = 4.08$, $SD = .49$). The observed discussions were peer-driven. Teachers were present in the classroom and helped explain the task to the groups but did not direct the discussions nor provide answers.

Step 5. Post-discussion comprehension. To assess comprehension of the texts after the peer discussion, the students were asked to respond to Form B of the Global Warming Test

(hereafter Form B). Forms A and B of the Global Warming Test were developed to be equivalent measures of comprehension: Paired items were randomly selected for both forms, which were piloted in two other fifth-grade classrooms (46 students) to confirm that the two measures were equivalent in terms of their level of difficulty. In the pilot study, the forms were counterbalanced so that half of the students responded to Form A before Form B, while the other half responded to Form B before Form A. There were no significant differences in the pilot students' pre- and post-test scores between conditions. Note that we did not counterbalance Form A and Form B in the present study. An interrater agreement was calculated for scoring the open-ended questions in Forms A and B of the Global Warming Test (see description above). Based on 20% of the questionnaires (number of transcripts = 20 from Form A and 20 from Form B), this coding rendered a Cohen's kappa of .86.

We wanted to assess the focal students' oral language and decoding skills, because these competencies would presumably influence their reading comprehension, as well as their ability to benefit from discussions with their peers. In an individual test session, the participants' receptive vocabulary skills were assessed using a translated version of the Peabody Picture Vocabulary Test, third edition (PPVT-III, Dunn and Dunn 1997). Each item on this test consists of four pictures. The students were presented with a word orally and were then asked to identify the one picture that corresponded to the word. The maximum possible score was 144.

Productive vocabulary was measured using the vocabulary subtest from the Wechsler Intelligence Scale for Children, third edition. In this task, the students had to explain the meanings of words (e.g., knife and brave) that are read out loud by the examiner. Thirty items were scored on a scale from 0–2. The maximum possible score was 60.

As a third measure of students' vocabulary, we employed the Text Cohesion Task, which was developed by Crosson, Leseaux, and Martiniello (2008) and translated into

Norwegian by the present authors. Text cohesion words are conjunctions that construct the semantic relationship between sentences or clauses in a text. Students were presented with 22 written items, each containing two sentences. The conjunction (e.g., ‘or,’ ‘because,’ or ‘but’) that linked the ideas between the two sentences was missing, and the students selected the missing conjunction from four alternatives. The examiner read the sentences and four alternative conjunctions aloud. The students completed all the items. The maximum possible score was 22.

To measure the students’ word-decoding skills, we employed the Test of Oral Word Reading Efficiency, Form A (TOWRE-A; Torgersen, Wagner, and Rashotte 1999), which is a combined measure of the fluency and accuracy of decontextualized word reading. For the TOWRE-A, students are presented with a list of real words and asked to read them aloud as quickly and accurately as possible. According to the manual, the test–retest coefficient exceeds .90. The students’ scores on the TOWRE-A were calculated based on the number of words read correctly within the allotted time of 45 seconds. The maximum attainable score was 104.

Analysis of the Peer Discussions

The transcription conventions developed within the Codes for the Human Analysis of Transcripts were used to identify students’ unfinished and disrupted utterances, as well as all types of claims, confirmations, and questions based on pauses and intonation (MacWhinney 2000).

In the present study, we chose to analyze talk moves at the utterance level as they unfolded during the peer discussions, thereby emphasizing the ways in which the students grounded their claims in reasons and justifications for their beliefs. The production of peer-group talk is a collective and distributed endeavor in the sense that many participants contribute to constructing meaning across turns. In the peer discussions, all the students

comprising the present samples contributed on-topic talk to some degree, offering a mix of relevant and misleading points in response to the guiding questions.

An utterance was coded whenever a new element was introduced—for instance, when a student offered a new idea regarding a previously formulated claim or when a previously formulated claim was repeated in a new interactional context (for example, to underscore one's own position). We did not code utterances that were repeated because another student did not pay attention. Although we coded talk moves at the utterance level, it was not always possible to ascribe the idea to one speaker (students would, for instance, finish each other's points or help a peer with the correct label for a difficult concept). Because of this, we coded the utterances that best captured the ideas being discussed, and we counted the talk moves produced by the group. We also decided to base the analysis on a definite number of different talk moves instead of computing density measures. This decision seemed most reasonable, as all groups were video-recorded for the same length of time (20 minutes), focusing on the same guiding questions.

The analytical work related to the identification of talk moves was inspired by the reviewed research. We distinguished between two main categories of talk moves: 1) Arguments—that is, facts or claims that were backed up with a reason or justification for a belief, such as explaining a causal relationship ('The CO₂ gas that is let out by the cars makes it much warmer, so the issue is not that it [the gas] mixed with other gasses')—and 2) Information—that is, facts or claims that were not backed up with a reason or justification for a belief ('The gas is carbon dioxide'). Within the two main categories, we coded utterances according to whether they were *requests* for arguments/information; the *relevance* and correctness of the argument/information, coded as either relevant arguments/information or misleading arguments/information; and whether claims were intended to weaken the epistemic status of a previous claim by opposing a peer's position on a matter, which was

categorized as either counterargument or simple opposition. This coding resulted in eight categories of talk moves, which are described and exemplified in Table 1. The categories were mutually exclusive. Opposition was typically observed after a peer had offered a wrong or misleading claim. Although simple-opposition statements had argumentative qualities, in that they opposed a peer's position, we distinguished these from opposition statements that provided additional reasons and justifications for beliefs (counterarguments). Note that wrong and misleading claims were categorized together and separated from other categories of talk moves. Rebuttals that responded to a peer's opposing position were very rare; thus, we chose not to extricate rebuttals in the coding. The few instances of rebuttals are included in the categories of simple opposition or counterarguments. Cohen's kappa was calculated for the scoring of 20% (i.e., five) of the transcripts, resulting in interrater agreement of .86.

--- Insert Table 1 approximately here ----

Analytical Plan

We present the results by looking at the correlational relationships between the language and reading comprehension variables in the study. Aggregated group scores for the different categories of talk moves were used as predictors of individual performance on Form B of the Global Warming Test (controlling for the students' prior performance on Form A). Because the students were nested within groups and we expected a dependency of scores within these groups (e.g., due to superior instructional quality and/or the presence of stronger students in certain classrooms), we employed two-level, mixed-model SPSS analyses to investigate the research questions. The mixed-effects models allowed us to account for group differences when predicting students' comprehension outcomes after their peer discussions.

Results

Table 2 presents the means, standard deviations, Cronbach's alpha, and two-tailed bivariate correlations between the students' assessment scores. There were marked

differences in the sampled students' vocabulary levels, as seen in the relatively high standard deviations. As expected, there was a high positive correlation between the students' performance on the pre- and post-tests (Forms A and B), and these scores were significantly related to their prior topic knowledge, vocabulary, and word-decoding skills. Thus, the students who were more knowledgeable about the topic and those with more advanced vocabulary and word-decoding skills appeared to surpass their less skilled peers on Forms A and B of the Global Warming Test.

--- Insert Table 2 approximately here ----

Table 3 presents the means and standard deviations of the talk-move variables of the present study. In almost all groups, the students requested further arguments from their peers, and all groups provided relevant arguments. However, counterarguments, in which the students opposed a peer's position using relevant reasoning, were low-frequency talk moves in the observed peer groups. In 9 of the 25 groups, the students did not oppose each other's positions using counterarguments.

The students actively requested their peers to offer information on the topic. In the group with the highest frequency, the students asked their peers to provide more information 17 times during the discussion task. All groups provided relevant information during the discussion, and the students in the vast majority of the groups also offered simple oppositions to their peers' claims. Misleading information was present in all peer groups during the discussion, and all except five groups also offered misleading arguments. One group contributed as many as eight misleading arguments, indicating that a high degree of misconception was introduced in some groups.

--- Insert Table 3 approximately here ----

As illustrated by the mean scores of Forms A and B (see Table 2), the students as a group did not improve significantly on the measure of reading comprehension conducted

after the peer discussion. Further inspection of the data revealed that some students' scores decreased, while some increased from the pre- to post-test. This finding led us to hypothesize that certain aspects related to the quality of the peer discussion impacted participants' performance on Form B.

The mixed-effect analyses with Form B (post-discussion comprehension) as the outcome variable are summarized in Table 4 (successive analytic steps are presented from Model 1 to Model 4). The first analytical step is presented in Model 1. This is the intercepts-only model with no predictors (the value of Form B, when the x-variable is set to 0), and it serves as the baseline by which to evaluate other models. This analytical step revealed that the students' scores on Form B varied between peer groups (as seen in the significant intercept) and among the students within these groups (as seen in the significant residual). The overall mean score of the 25 groups on Form B was 10.91. In the next analytical step (Model 2), reading comprehension measured before the peer discussion (Form A) was included as a predictor of students' scores on Form B. As expected, there was a strong association between the students' pre- and post-discussion comprehension scores. However, the statistically significant residual indicated that there was still room to improve the model. Thus, to explore whether certain aspects of the peer discussions could predict students' post-discussion comprehension scores, we included the different talk moves in our analysis. As demonstrated in Model 3, only two types of talk moves—that is, relevant arguments and counterarguments—predicted significant variance in the students' scores on Form B when students' scores on Form A were accounted for. As the final step, we included the different measures of prior topic knowledge, vocabulary, and word-decoding skills to determine whether the relevant arguments and counterarguments in the peer discussions would still predict post-discussion performance. Prior topic knowledge and word-decoding skills did not predict students' scores on Form B when accounting for Form A. However, the students'

vocabulary skills appeared to predict the degree to which they benefitted from participating in the peer discussions. Although receptive and productive vocabulary was only marginally statistically significant ($p = 0.06$ and $p = 0.07$, respectively) when the relevant arguments and counterarguments were included in the model, the text cohesion vocabulary remained as a predictor. To illustrate this, the final model (Model 4) included the students' text cohesion vocabulary scores. The association between the argumentative talk moves in the peer discussions (relevant arguments and counterarguments) and the students' performance on Form B remained. The final model was significantly better than the intercepts-only model (as seen in the $-2LL$). The findings indicate that although the students' scores on Form B differed among the peer-discussion groups and the students within these groups, the students' scores increased on average when they discussed the topic of global warming with peers who offered a high degree of relevant arguments and counterarguments. Merely offering relevant information or simple opposition was not associated with the same increment in students' scores on Form B of the Global Warming Test. It is noteworthy that being in peer groups that offered more misleading arguments or information did not seem to lower the students' scores on Form B of the Global Warming Test in a statistically significant manner.

--- Insert Table 4 approximately here ----

We conducted a set of initial analyses to investigate whether the potential differential effects on Form B could be ascribed to other group-characteristics, such as the presence of high-performing peers (assessed based on the measure of prior topic knowledge). We also looked at whether the sheer amount of talk (counted as the total number of words produced by each participant during the discussion) predicted students' scores on Form B of the Global Warming test. None of these variables appeared to contribute uniquely to the relationships presented here.

Discussion and Conclusions

Although the extensive research literature has revealed that too many students in multilinguistic classrooms struggle with reading comprehension in the content areas, little research has specifically investigated the collaborative processes that may support students' text-based learning in these classrooms (but see Vaughn et al. 2017). The present study investigated whether student comprehension after reading multiple texts about a topic could be supported through oral dialogues in multilinguistic classroom environments. The present study builds on the theoretical and empirical framework developed within research aiming to enhance the quality of argumentation in classroom dialogue; however, we took a different focus by observing the dialogue features of the task-based classroom interactions of peers with no initial training in how to apply argumentative schemas in discussions. Moreover, instead of prompting the students to discuss one overarching question (see Reznitskaya 2011) or evaluate different conclusions to a problem (see Chinn et al. 2000), they were asked to discuss guiding questions related to the text material that they had read individually. The main finding of this study is that specific argumentative qualities of the peer discussions immediately following individual reading were positively associated with the students' scores on the post-discussion comprehension measure (controlling for their pre-discussion reading comprehension scores). More specifically, peer discussions that were characterized by a high degree of relevant arguments and counterarguments (that opposed a peer's position) appeared to support the development of participants' comprehension of the text topic.

Although previous studies have emphasized counterarguments as an important mechanism for altering understanding (see Asterhan and Schwarz 2009; Leitão 2000), the present study showed that both consensual arguments and counterarguments were important for moving comprehension forward (for a similar finding, see Chinn et al. 2000). Another related point concerns the complexity of the argumentation that are constructed in peer discussions. Chinn et al. (2000) analyzed the quality of argument structures when fifth

graders evaluated different conclusions drawn about electrical circuits. Their findings suggested that positions elaborated with further relevant evidence were associated with post-discussion performance, while the more high-frequency, simple-reason arguments were not. We discovered a somewhat similar finding in the present study: Talk moves that offered relevant information (e.g., *we need to use less electricity*) and simple opposition to a peer's position (e.g., *oil does not make plants grow!*) related to the topic did not appear to alter the students' understanding during the peer discussions. Thus, the present study lends support to the research indicating that enhanced comprehension of texts may evolve as a result of discussions in which students are concerned with supporting their claims and positions with relevant reasoning (Chinn et al. 2000; Reznitskaya et al. 2011; Wolf et al. 2006). The Global Warming Test was developed to measure comprehension in a manner that may favor students' reasoning skills more than their fact-memorizing skills. This is probably one reason why the contribution of relevant information per se did not lead to an increase in students' post-discussion comprehension scores.

The fact that the group mean scores on the Global Warming Test did not increase from the pre (Form A) to the post (Form B) period for the group as a whole is a reminder that peer-driven discussions in and of themselves may not result in enhanced comprehension of texts (see discussion in Kuhn 2015; Murphy et al. 2009). This finding may be attributed to the fact that students may also lead each other astray during a discussion if a more knowledgeable peer or teacher is not present to question misconceptions and ensure rigorous understanding of the topic. In addition, there are significant differences in the ways in which individual students participate in classroom discussions, and this may have implications for students' learning. In the present study, we did not find that the students' prior topic knowledge or decoding skills predicted how they benefitted from peer collaboration. However, we found that the students' vocabulary skills, and particularly their understanding

of conjunctions that construct the semantic relationship between sentences or clauses in a text, were important. This finding is not surprising given that the understanding of text cohesive words may be imperative to construct and understand an argument during peer discussions. Although previous research has documented that knowledge of the functions and meanings of various text cohesion words is important for reading comprehension and particularly for students who are reading in a second language (Degand and Sanders 2002), the present study is unique in demonstrating that insufficient understanding of connectives may affect what students can gain from discussing a text topic with their peers in multilingual classrooms.

There are limitations to the current study related to the relatively small sample size, the lack of a comparative design and the lack of data on teacher beliefs and practices; consequently, caution should be used when drawing implications based on these results. Nonetheless, it should be noted that it was the quality of the peer discussions that appeared to account for the increment in student scores on Form B on the Global Warming Test. Many years of research have demonstrated that instruction that explicitly focuses on how to perform reasoning in dialogues, can help students to engage in discussion in a manner that promotes critical thinking and argumentation (e.g., McNeill 2011; Mercer et al. 2004). The findings of the present observational study point to the importance of supporting students' ability not only to build relevant arguments, but to do so vis-à-vis others' perspectives and positions to enhance young readers' comprehension of texts (see also LaRusso et al. 2016; Uccelli et al. 2015) and that such practices may be a fruitful path in multilinguistic classrooms with many second-language learners (see also Lawrence et al. 2012; Vaughn et al. 2017). Group composition should be planned so that students feel free to share their ideas and disagreement and simultaneously have access to their more knowledgeable peers. In relation to texts, students may need to be exposed to questions and dialogues that help them reflect upon the

intentions of different authors, the perspectives of protagonists and the understandings of their communication partners. Thus, the ways in which teachers organize text conversations in the classroom and the level of reasoning and perspective-taking skills demanded by the questions they pose appear to be of importance. In other words, supporting reading fluency, vocabulary skills and topic knowledge may be highly necessary, but not sufficient, to strengthen elementary students' reading comprehension in the content areas.

References

- Applebee, A., Langer, J., Nystrand, M., and Gamoran, A. 2003. Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle school and high school English. *American Educational Research Journal*, 40, 685-730.
- Asterhan, C. S. C., and Schwarz, B. B. 2009. The role of argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialogue. *Cognitive Science*, 33, 374-400.
- Berland, L. K., and Reiser, B. 2011. Classroom communities' adaptations of the practice of scientific argumentation. *Science Education*, 95, 191-216.
- Briker, L. A., and Bell, P. 2008. Conceptualizations of argumentation from science studies and the learning sciences and their implications for the practice of science education. *Science Education*, 92, 473-498.
- Bråten, I., and Strømsø, H. 2010. Effects of task instruction and personal epistemology on the understanding of multiple texts about climate change. *Discourse Processes*, 47, 1-31.
- Cazden, C. B. 2001. *Classroom Discourse: The Language of Teaching and Learning* 2nd ed. Portsmouth, NH: Heineman.
- Cekaite, A., Blum-Kulka, S., Grøver, V., and Teubal, E. Eds. 2014. *Children's Peer Talk: Learning From Each Other*. Cambridge University Press.
- Chinn, C. A., Anderson, R. C., and Waggoner, M. A. 2001. Pattern of discourse in two kinds of literature discussion. *Reading Research Quarterly*, 36, 378-411.
- Chinn, C.A., O'Donnell, A. M., and Jinks, T. S. 2000. The structure of discourse in collaborative learning. *Journal of Experimental Education*, 69, 77-97.
- Crosson, A. C., Lesaux, N. K., and Martiniello, M. 2008. Factors that influence comprehension of connectives among language minority children from Spanish-speaking backgrounds. *Applied Psycholinguistics*, 29, 603-625.

- Degand, L., and Sanders, T. 2002. The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing: An Interdisciplinary Journal*, 15, 739–757.
- Driver, R., Newton, P., and Osborne, J. 2000. Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84, 287-312.
- Dunn, L., and Dunn, L. 1997. Peabody Picture Vocabulary Test (PPVT-III). 3rd ed. Circle Pines, MN: American Guidance Service.
- Gamez, P. B., and Lesaux, N. K. 2015. Early-Adolescents’ Reading Comprehension and the Stability of the Middle School Classroom-Language Environment. *Developmental Psychology*, 51, 447-458.
- Kuhn, D. 2015. Thinking together and alone. *Educational Researcher*, 44, 46-53.
- LaRusso, M., Kim, H. Y., Selman, R., Uccelli, P., Dawson, T., Jones, S., and Snow, C. 2016. Contributions of academic language, perspective taking, and complex reasoning to deep reading comprehension. *Journal of Research on Educational Effectiveness*, 9, 201-222.
- Lawrence, J.F., Capotosto, L., Branum-Martin, L., White, C., and Snow, C.E. 2012. Language proficiency, home-language status, and English vocabulary development: A longitudinal follow-up of the Word Generation program. *Bilingualism: Language and Cognition*, 15, 437-451.
- Leitão, S. 2000. The potential of argument in knowledge building. *Human Development*, 43, 332-360.
- Lemke, J. 1990. *Talking Science: Language, Learning, and Values*. Norwood, NJ: Ablex.
- MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk* 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates.

- McNeill, K. L. 2011. Elementary students' views of explanation, argumentation, and evidence, and their abilities to construct arguments over the school year. *Journal of Research in Science Teaching*, 48, 793-823.
- Melby-Lervåg, M., and Lervåg, A. 2014. Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first-and second-language learners. *Psychological bulletin*, 140, 409-433.
- Mercer, N., Dawes, L., Wegerif, R., and Sams, C. 2004. Reasoning as a scientist: Ways of helping children to use language to learn science. *British Educational Research Journal*, 30, 359-377.
- Mercier, H. 2011. Reasoning Serves Argumentation in Children. *Cognitive Development*, 26, 177-191.
- Murphy, P. K., Wilkinson, I. A.G., Soter, A. O., Hennessey, M. N., and Alexander, J. F. 2009. Examining the effects of classroom discussion on students' comprehension of text: a meta-analysis. *Journal of Educational Psychology*, 101, 740-764.
- Rapanta, C., Garcia-Mila, M., and Gilabert, S. 2013. What is meant by argumentative competence? An integrative review of methods of analysis and assessment in education. *Review of Educational Research*, 83, 483-520.
- Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., and Kim, S. 2011. Influences of oral discussion on written argument. *Discourse Processes*, 32, 155-175.
- Reznitskaya, A., Kuo, L., Clark, A., Miller, B., Jadallah, M., Anderson, R. C., and Nguyen-Jahiel, K. 2009. Collaborative reasoning: A dialogic approach to group discussions. *Cambridge Journal of Education*, 39, 29-48.

- Rydland, V., Grøver Aukrust, V., and Fulland, H. 2012. How word decoding, vocabulary and prior topic knowledge predict reading comprehension. A study of language-minority students in Norwegian fifth grade classrooms. *Reading and Writing*, 25, 465-482.
- Snow, C. E., and RAND Reading Study Group 2002. *Reading for Understanding Toward an R&D Program in Reading Comprehension*, Pittsburg: RAND.
- Torgersen, J., Wagner, R., and Rachotte, C. 1999. *Test of Word-Reading Efficiency*. Austin, TX: PRO-ED.
- Uccelli, P., Galloway, E. P., Barr, C. D., Meneses, A., and Dobbs, C. L. 2015. Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. *Reading Research Quarterly*, 50, 337-356.
- Vaughn, S., Martinez, L. R., Wanzek, J., Roberts, G., Swanson, E., and Fall, A. M. 2017. Improving content knowledge and comprehension for English language learners: Findings from a randomized control trial. *Journal of Educational Psychology*, 109, 22-34.
- Wolf, M. K., Crosson, A. C., and Resnick, L. B. 2006. *Accountable Talk in Reading Comprehension Instruction*. CSE Technical Report 670. Center for the Study of Evaluation, University of California, LA.

