This is the pre-proof version of


Running head: ARGUMENTATION AND EXPLANATION

Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialogue

Christa S. C. Asterhan and Baruch B. Schwarz

School of Education

The Hebrew University of Jerusalem

Mt Scopus, Jerusalem 91905, Israel

Email: asterhan@huji.ac.il, msschwar@mscc.huji.ac.il

Keywords: argumentation, explanation, conceptual change, peer interaction, peer dialogue, dialogue analysis
Abstract

In this paper we attempt to identify which peer collaboration characteristics may be accountable for conceptual change through interaction. We focus on different socio-cognitive aspects of the peer dialogue and relate these with learning gains on the dyadic, as well as the individual level. The scientific topic that was used for this study concerns natural selection, a topic for which students' intuitive conceptions have been shown to be particularly robust. Learning tasks were designed according to the socio-cognitive conflict instructional paradigm. After receiving a short instructional intervention on natural selection, paired students were asked to collaboratively construct explanations for certain evolutionary phenomena while engaging in dialectical argumentation. Two quantitative coding schemes were developed, each with a different granularity: The first assessed discrete dialogue moves that pertained to dialectical argumentation and to consensual explanation development. The second scheme characterized the dialogue as a whole on a number of socio-cognitive dimensions. Results from analyses on the dyadic as well as the individual level revealed that the engagement in dialectical argumentation predicted conceptual learning gains, whereas consensual explanation development did not. These findings open up new venues for research on the mechanisms of learning in and from peer collaboration.
1. Introduction

Research into the benefits of collaborative work on learning have shifted from questions regarding whether it promotes learning, to investigations into the conditions that promote learning and the identification of processes that make collaborative group settings particularly effective in these conditions. For example, Schwartz (1995) showed that problem-solving performance benefits from peer collaboration, since the coordination of different perspectives promotes the construction of abstract representations. In a later study, Shirouzu, Miyake & Masukawa (2002) showed that it is the frequent role exchange between task-doing and task-monitoring in particular which allows for such abstractions to emerge. Such investigations are important since they enable adequate task design that optimizes learning results by purposefully promoting and scaffolding those socio-cognitive processes considered to be responsible for learning from collaboration in a certain topic domain.

Peer collaboration has also been identified as a potentially powerful means for increasing individual conceptual understanding of subject matter (e.g., Kruger, 1993; Phelps & Damon, 1989). In particular, several studies have shown that the engagement in explanatory activities, such as elaboration and the collaborative development of explanation, are particular beneficial for improving conceptual understanding in several domains (e.g., van Boxtel, van der Linden & Kanselaar, 2000; Brown & Palinesar, 1989; Chi, deLeeuw, Chiu, & Lavancher, 1994; Coleman, 1998; King & Rosenshine, 1993; Webb, Troper & Fall, 1995).

In this paper, we will focus on a specific form of learning, namely conceptual change, which involves the changing of prior misconceived conceptual knowledge to
correct knowledge (e.g., Chi, 2008). Even though there is still much debate and research on different types of conceptual change, most researchers distinguish between at least two different types of change (e.g., Chi, 2008; de Leeuw & Chi, 2003; Gentner, Brem, Ferguson, Markman, Levidow, Wolff, & Forbus, 1997; Thagard, 1992; Vosniadou & Brewer, 1994): Some misconceptions may be locally repaired by replacement or correction of knowledge, or what has been referred to as *incremental* conceptual change. Others, however, have been found to be extremely resistant to instructional interventions and require *radical* conceptual change (Carey, 1985; de Leeuw & Chi, 2003), a substantive re-organization of the knowledge structure. For example, a topic in which misconceptions are known to be particularly robust and resistant to instructional interventions (e.g., Brumby, 1984; Ohlsson & Bee, 1992) and which require this type of conceptual change, is the notion of natural selection (Chi, 2005).

Research on instructional interventions to induce conceptual change through peer collaboration has traditionally been conducted in a Piaget-inspired task design paradigm: the socio-cognitive conflict paradigm (e.g., Amigues, 1988; Mugny & Doise, 1978). It is based on the idea that by pairing students with different initial conceptions or by presenting collaborators with information that contradicts their intuitive conceptions they will experience *cognitive conflict*. This, in turn, is thought to lead them to seek equilibrium and, hopefully, to accommodate their naive theories into more sophisticated ones (Piaget, 1985). However, as the literature on peer collaboration has extensively and repeatedly shown, simply putting two people together is not sufficient (e.g., Coleman, 1998; King & Rosenshine, 1993; Teasley, 1995).
Research from post-hoc analyses of peer dialogues seems to suggest that in order for these types of task designs to be effective, students will have to engage in peer argumentation (Schwarz, Neuman & Biezuner, 2000; Schwarz & Linchevski, 2007). Argumentation is an activity in which interlocutors attempt to decrease or increase the acceptability of one or more ideas by reasoning (Baker, 2002; 2003; Walton, 2006). The conjecture that argumentation is essential in learning that requires intentional attempts to induce conceptual change was experimentally tested in two recent studies on conceptual understanding in evolutionary theory in which argumentation was treated as a condition (Asterhan & Schwarz, 2007a): Undergraduates were assigned to dyads and collaboratively tried to explain an evolutionary phenomenon (i.e., the evolution of webbed feet of ducks). Half of the dyads were instructed to engage in argumentative dialogue on their respective explanations and received some written examples of argumentative moves. The other half was merely instructed to collaborate. Individual evolutionary understanding was assessed as the quality of the explanatory schemas (Ohlsson, 2002) they used to explain different evolutionary phenomena on three separate test occasions: Prior to, immediately after and a week following the intervention.

When controlled for pretest performance and nested effects, delayed posttest explanations of students in the argumentative condition were found to reflect better conceptual understanding than those of control students. Furthermore, whereas peer collaboration by itself led to immediate gains in conceptual understanding, only those students that were explicitly instructed to engage in collective argumentation preserved these gains to later test occasions. The advantage of argumentation observed in collaborative dyadic situations was then replicated in a confederate study in which experimental subjects were elicited to engage in scripted dialectical
argumentation directed at and prompted by a peer confederate (Asterhan & Schwarz, 2007).

However, in spite of these mean differences between conditions, not all experimental subjects attained conceptual change and not all experimental dyads engaged in a *dialectical* argumentative discussion, in spite of the instructions and task design. In the present paper, we will adopt a *phenomenological*, instead of an experimental, approach and attempt to trace socio-cognitive processes and interaction features that predict learning gains from collaboration. Based on our experimental findings we will focus on the role of argumentation in peer-to-peer dialogue.

In addition, and based on the aforementioned findings on the beneficiary effects of explanatory activities in peer collaboration, we also assess acts of explanation development.

1.1 Argumentation and explanation

It is important to distinguish between acts of argumentation and explanation development, since these two quite different epistemic actions (de Vries, Lund & Baker, 2002; Keil, 2006; Walton, 2006) reflect different (socio-)cognitive processes each of which may prove to be beneficial in different task designs, different domains and different types of concepts. For example, de Leeuw and Chi (2003) have suggested that whereas explanation-driven discourse may indeed promote *incremental* conceptual change, it may not suffice when a radical reorganization of conceptual knowledge is required.
However, they are not always easily distinguishable by the untrained eye (Schwarz & Asterhan, in press): In the same conversation, acts of explanation development and argumentation frequently occur interchangeably and even have similar syntactic and formal structures. They are made up of at least two propositions, in which one is presented as the starting point which leads to the other (the end point). They may also make use of similar indicator words, such as 'because', 'therefore', 'as a result of', and 'since'. Moreover, both explanation and argumentation are verbal and social acts of reasoning. Explanations, like arguments, are of a transactional nature: They have recipients, whether this occurs on the inter-personal plane between two individuals, or on the intra-personal level where an individual explains something to the self. In both cases, the goal of the explanation is to expand the recipient's understanding (Keil, 2006).

In spite of these resemblances, argumentation and explanation differ in at least one important aspect: their purpose. An explanation has a clarifying function within a dialogue, in the sense that the recipient should come to understand something better as a result of the explanation. It is often, but not always, preceded by requests for clarifications. In argumentation, on the other hand, the proponent proposes reasons for the recipient to come to accept or refute a certain thesis (Walton, 2006). Baker (2002; 2003) defines argumentation as an activity that involves establishing specific types of relations between the propositions being discussed and other sources of knowledge, the establishment of which is meant to influence the epistemic statuses of these propositions. Accordingly, acts such as elaboration and clarification are not argumentative dialogue moves. Then the distinction between argumentation and explanation should be made based on the context and goals of the dialogue, both in a local and general sense (Schwarz & Asterhan, in press).
In conclusion, theoretical considerations as well as experimental findings led us to explore the role of explanatory and of argumentative acts in peer dialogues and their relation with individual conceptual learning. For this purpose, we will first explore the factors that distinguish between the dialogues of dyads that achieved substantive conceptual gains (i.e., conceptual change) and those that did not. We will then relate features of individual behavior within the dialogue with subsequent individual learning gains.

1.2 Assessment procedures for argumentation in conversation

One of the reasons for the relatively small number of quantitative research works conducted on the relation between argumentative dialogue features and conceptual gains may be the methodological hurdles for analyzing natural argumentative dialogue in face-to-face settings. In a written argumentative text, such as an essay, the underlying argumentative structure is often explicitly laid out. Argumentative dialogues, on the other hand, are characterized by sudden shifts in content, references to previous statements in an un-sequential manner, simultaneous development of different discussion threads and complementing non-verbal forms of communications. The argumentative function of a contribution in natural dialogue is, therefore, not easily determined.

A number of coding schemes have been developed to describe, what Chinn and Anderson (1998) termed, the *macrostructure* of natural argumentative discourse (e.g., Chinn & Anderson, 1998, Resnick, Salmon, Zeitz, Wathen, & Holowchak. 1993). In such schemes, the dialogue as a whole is regarded as a text which is analyzed
according to the deep structure of the argument that emerges from the dialogue. Such schemes often present the argument in a graphical way, do not include a chronological dimension and often ignore social distribution of contributions (Baker, 2002).

Even though these techniques reveal an important facet of argumentative dialogue, namely the argument that was constructed in the course of the dialogue, they are time-consuming and satisfying inter-rater reliability may prove to be an almost impossible task to achieve. More importantly, they may not be able to capture the dynamic nature of face-to-face dialogue and may therefore be more appropriate for the assessment of more static products of argumentative activities, such as essays or collaboratively constructed argumentative maps. In order to capture how argumentative interaction can lead to learning, the inclusion of methods of assessment that are based on dialogical and pragmatic models of argumentation seems imperative (Baker, 2002). Moreover, these methods should preferably include categories that are quantifiable. We will present a coding scheme that attempts to capture collaborative argumentative reasoning as it happens, by identifying interlocutory argumentative moves in situ.

An additional methodological hurdle is the determination of the right level of reference for argumentative moves in dialogue on scientific concepts. Whereas argumentative moves may be more easily identified in discussions on social and moral dilemmas, they are less obvious in collaborative learning activities that focus on conceptual understanding of scientific explanations. In the former, the activity's goal is often persuasion and attitude change, since the object disputed concerns personal opinions and attitudes. In the latter, the intended goal is to learn and understand and the object that is discussed and disputed is what constitutes the most appropriate explanation to the phenomenon. In the latter settings, the level of
reference, or the claim, is explanation $s_i$ instead of a personal attitude towards an issue.

So as to distinguish between processes of argumentation and explanation development, we also assess non-argumentative interlocutory moves that introduce new information to the discussion, such as those that develop or expand the content of preceding contributions. The first distinction we propose is then between dialogical moves that refer to the epistemic status of an idea from those that develop ideas. The second distinction is between two types of argumentative moves: Non-dialectical moves that intend to strengthen and validate the epistemic status of a certain thesis within a consensual dialogue constellation (such as supports and agreements); and Dialectical moves that intend to dispute the validity and strength of a certain thesis or reason and are proposed within a dialectical-critical constellation that (such as, for example, challenges and rebuttals). This distinction enables us to separately examine the role of consensual and dialectical-critical processes of collaborative knowledge construction and their role in learning that involves conceptual change.

2. Method

The dialogues that were used for analyses were drawn from a corpus of data gathered as part of an experimental study on the effects of dialogical argumentation on concept learning in evolutionary theory (see Asterhan & Schwarz, 2007a, for details of the complete experimental study). The analyzes presented here concern the data gathered from dyads that participated in the experimental condition in which they were encouraged and instructed to engage in dialectical argumentation while collaboratively solving two evolutionary phenomena.
2.1. Participants

A total of thirty-eight undergraduates (11 males and 27 females) were randomly assigned to the argumentative condition (mean age = 23.49). Recruitment occurred through announcements on campus note-boards. Students were enrolled in Social Sciences (32%), Humanities (16%) or Education (52%) and received either financial rewards (36%) or course credit (64%) for their participation.

2.2. Procedure

All subjects participated in the following sequence of activities: (1) pre-test to assess prior evolutionary understanding; (2) instructional intervention, which was identical (screening of instructional movie excerpt); (3) experimental intervention in which dyads solved two items on evolutionary phenomena according to two different conditions (control and argumentative); (4) immediate post-test which assessed evolutionary understanding after the intervention; (5) delayed post-test administered a week later. The total length of each session was approximately 1.5 hr. The pre- and delayed post-test were administered to individuals in group format (size ranging from 2 to 8) as paper-and-pencil tests. They contained three questions each: One warming-up item that was identical in both test occasions and two target items, each relating to a different evolutionary phenomenon but similar in set-up.

2.2.1. Instructional intervention.

Following the pre-test, the educational movie excerpt was screened. It showed that a number of modern Galapagos species were different from their continental
ancestors (iguanas, turtles, sea-lions), without an account of how that change had occurred. In addition, changes in a population of Galapagos finches (also known as Darwin's finches) were discussed in detail and explained in terms of the Darwinian account of evolutionary change. Darwinism was presented as the scientifically proven and accepted theory of evolution today. No alternative explanations were mentioned.

2.2.2. Collaborative interaction and immediate post-test.

Following the movie, subjects were randomly assigned to a partner. Each dyad was placed in either separate rooms or separate, shielded working corners within the same room, depending on the group size of a session. Each participant received a four-page booklet with written instructions, two test items and two individual, separate answer sheets. The experimenter read the instructions aloud with the participants and left the room immediately after that. Students received detailed instructions on how to conduct a collaborative critical discussion, including descriptions and examples to guide them in such an endeavor (e.g., "Try to think why a certain idea or solution is incorrect", "Can you provide proof for your claim?" "What is the weakness in your own or your partner's arguments?").

Dyads then collaboratively solved the two open-ended items on evolutionary change processes (one warming-up and one target item) without time limits, while being audio-taped. The warming-up item referred to a phenomenon that was explained in detail in the movie (the beak size of Darwin's finches). The target item requested of students to transfer this explanatory schema to a new, more complex, phenomena, namely the evolution of ducks' webbed feet. Students were told that the feet of the ancient ancestor of modern ducks were like those of pigeons and that as a result of climatic changes their living grounds became flooded. After discussing the
solution to an item, subjects wrote the solution separately, each one on their own personal answer sheet.

After approximately thirty seconds into the discussion on the target item, each dyad received a short excerpt of a critical discussion of four turns between two (hypothetical) subjects who, they were told, had participated in the experiment a year earlier. The excerpt was as follows:

A: Then the ducks had to change their feet so that they could swim. The area was flooded with water, and because of the new environment webbed feet developed.
B: What do you mean “developed”? How did that happen?
A: Hmmm. In the beginning they did not know how to swim. But slowly they learned to do it and that caused some sort of development in their feet. I mean, webs developed between their fingers. And that’s how it was passed on to the next generation.
B: Well, if that were true, then Olympic swimmers should also develop webbed feet, since they also swim all day long!

The thirty seconds delay was intended to allow them to articulate their own solutions to the Duck item. They were told that the experimenter forgot to give them the item in time and they were asked to read it and consider it in their discussion. The discussion in the excerpt modeled a critical dialogue on the Duck item without actually revealing or hinting at the correct solution.

2.3. Coding procedures

2.3.1. Conceptual understanding in evolutionary theory

Based on and inspired by previous works, ten qualitatively different explanatory schemas were identified in students' explanations of evolutionary change (Asterhan & Schwarz, 2007a). These different schemas were then quantitatively assessed on four
different dimensions: Whether evolutionary change was considered at all, whether this change was explained, whether some sort of selection mechanism was used and whether existing intra-species variation was considered. Based on the appearance of each of these four dimensions, the ten qualitatively different explanatory schemas were assigned to one of five different categories. The score for each schema category was based on the number of dimensions that featured in the schema in that category (see Table 1). An additional null-category was added to distinguish between explanations that did not consider evolutionary change (the lowest category) and those responses that simply did not answer the question at all (by stating that they did not know the answer or by repeating the data given in the item without providing additional information). This procedure thus yielded six explanatory schema categories with scores ranging between 0-5. Examples of student responses and their classification to the schema categories are presented in Appendix A.

This classification scheme was then used to arrive at an operational definition of conceptual change: Whereas the generative schemata that underlie two explanatory schemas from the same category are qualitatively different, one cannot necessarily be considered superior to the other. In alignment with existing distinctions, only shifts from one explanatory schema category to another were regarded as substantive reorganizations that are described in conceptual change theory and research.

However, previous research has shown that students' responses to different test items are often not consistent in the explanatory schemas they applied on a given test occasion. Asterhan and Schwarz (2007b) found that more than a quarter of undergraduates were inconsistent in their use of explanatory schemas on the same test. An operational definition of conceptual change has to take this instability into account. Whereas a student who applied an explanatory schema of a one-point higher
category on only one of the two test items (i.e., a mean pre- to delayed post-test increase of .5 points) has indeed shown improved conceptual understanding, we argue that this does not provide *sufficient proof* for a substantive change. Sufficient proof for a more profound reorganization in conceptual understanding, i.e. conceptual change, was therefore defined as an increase of at least one point from the mean pre-test to mean post-test score.

2.3.2. Coding dialogue protocols

The emphasis of transcription was on accuracy of verbal content and sequence of turns, rather than other discourse properties. Turns were parsed based on speaker-continuous speech. When a speaker was interrupted but continued talking, then this was considered as one turn, and the interruption as another. If an interruption caused the speaker to stop from speaking and pick it up later on, then the resumed content was considered an additional turn.

Two complementary coding schemes with different granularity were developed: The first focused on the identification of dialogical moves (micro-level), whereas the goal of the second was to characterize the nature of the interaction as a whole on a number of dimensions (macro-level). Both schemes were developed and validated according to a procedure similar to the verbal analysis method described by Chi (1997):

(1) We first read through the entire corpus of protocols and examined them for impressions and trends, and developed the two methods of coding to capture these impressions;

(2) This was followed by a period of coding scheme refinement during which two human raters independently coded two or three dialogues at a time, compared
outcomes, discussed disagreements and fine-tuned the scheme accordingly. This process was repeated until a sufficient amount of agreement was reached and a satisfactory number of stable and distinctive coding categories emerged.

(3) The remainder of protocols, twenty-one in total, was then independently coded by the two raters and inter-rater reliability was established. Disagreements were resolved through discussion.

**Micro-level assessment scheme: Dialogical moves.** The unit of analysis in this coding scheme is what we call a *dialogue unit.* Maximally, these consist of a complete speaker turn. It is not uncommon, however, for a speaker turn to contain different statements that refer to different topics or have different interlocutory intentions. For this reason a turn may be subdivided in different units, based on pragmatic features, that is: when one turn contains more than one of the dialogical moves described below. Thus, segmentation within a turn is in many ways also the product of the coding scheme application. When a dialogue unit is interrupted by the other speaker but immediately picked up again, it is considered a continuation of the same dialogue unit.

The scheme comprises a number of non-exhaustive, mutually exclusive categories which are presented in Table 2. The scheme includes both argumentative moves (claims, request for claims, simple agreements, supports, challenges, rebuttals, simple oppositions and concessions), as well as non-argumentative moves (elaborations, requests for information and information providing). Of the eight different argumentative moves, two (supports and agreements) intend to strengthen and validate the epistemic status of a certain explanation in a non-dialectical constellation and are therefore referred to as *non-dialectical* argumentative moves. Four others
(challenges, rebuttals, concessions, and oppositions) intend to explicitly dispute the validity and strength of a thesis or attack on that thesis or are embedded in a dialectical dialogue constellation. They are therefore referred to as dialectical argumentative moves. Acts of simple repetition or correction of the content that appeared in preceding contributions, without significantly adapting or adding new information to it (e.g., paraphrases, corrections of mistakes), are coded separately.

Attempts to rely solely on the content of each dialogue unit as the basis for coding were unsatisfactory, since a contribution has no argumentative function by itself, unless it is in relation to the content and function of the preceding dialogue. Therefore, the nature of the relation between a dialogue unit and the immediately preceding contributions or the dialogue excerpt it was relating to, was coded instead. When the content or intent of a certain turn was not completely clear, the following dialogue units are also used for purposes of interpretation. Thus, coding of dialogue units was based on both the content and the interlocutory function of the contribution, given the dialogical context. It is important to emphasize that this coding scheme does not consider the objective, canonical validity or strength of a contribution, but only its function within the conversational context and the speaker’s intent. Examples of the coding categories can be found in two short excerpts presented in Table 4 and 10.

Coding procedures involved both segmentation into dialogue units, as well as assigning each unit to a category. Default for segmentation was a complete speaker’s turn which was based on the appearances of long pauses and switches between speakers. Deviations from this procedure were the product of category assignment and are therefore incorporated in the fourteen dialogue move categories: 1) when one speaker’s turn included more than one separate dialogue move; or 2) when a
dialogue unit was continued despite attempts to interrupt a speaker’s turn, this
continuation was separately coded as a continuation. Establishment of inter-rater
reliability was therefore based on assignment of units to the different categories,
Cohen’s Kappa = .81.

Macro-level assessment scheme: Interpersonal and socio-cognitive properties.
The aim of this coding scheme was to examine certain macro-properties of students’
dialogues that could be accountable for differences in subsequent individual learning
gains. Operationally, we relied on the identification of these features in the dialogue
as a whole and segmentation was therefore unnecessary. Each dyadic interaction was
characterized on a number of socio-cognitive and interpersonal features:

(1) Argumentative structure of the dialogue: The overall structure of the collective
argument that was constructed in the course of the dialogue. We distinguished
between two types of argumentative structure (Asterhan & Schwarz, 2007): (a)
Dialectical: The dialogue contains more than one solution which Ss feel obliged to
choose from, or the dialogue contains a single proposed solution that is both
contested as well as defended; (b) One-sided: Ss only engage in reasoned
argumentation that intends to strengthen a certain proposed solution.

It should be noted that the appearance of rebuttals to the challenge that appeared
in the written excerpt that dyads were given (see Procedure) do not necessarily change
the nature of the argumentative discussion to a dialectical one, even though a rebuttal
as a dialogical move is dialectical in essence. If Ss proposed only one solution,
provided supports in favor of that solution and responded to the excerpt with technical
rebuttals that were intended to strengthen the epistemic status of that same solution,
then the dialogue still remains one-sided in nature.
(2) Key issue of change: Whether students discussed the key issue of how the ducks’ feet could have changed from “feet like those of pigeons” to webbed feet. It should be noted that presenting a Darwinian-type solution is not required on this dimension; it only relates to whether they gave the issue some explicit consideration, or not.

(3) Interpersonal repartition of solutions: Whether the dialogue moves that referred to different explanatory schemas were contributed by different interlocutors, or not. Baker (2003) has suggested that one of the advantages of collective argumentation is that it enables the objectification of perspectives and their representation by actual persons defending them. Note that a dyad might conduct a dialectical argumentative discussion without their views being distributed among the dyadic partners. For example, they might collaboratively explore the pros and cons of a number of collaboratively proposed solutions, or the same interlocutor that proposed a solution may also be the one to attack that solution. An additional distinction was made between two different types of distributed dialogs: Those in which interlocutors overtly confront the different solutions and relate to the differences between them (Distributed: overt juxtaposition), and those in which they do not overtly discuss these differences (Distributed: no overt juxtaposition). For example, the dyadic partners may each allude to a different solution at some point in the dialog, without actually comparing the two. In such a situation, the different views are personified by different persons, but no attempt is made to resolve this incongruence. Whereas the former may be important in its own right, the collaborative attempt of resolving may prove to be crucial for conceptual change.

(4) Closure: Whether the dyad agreed upon a certain solution at the end of the dialogue or not.
(5) **Contribution symmetry**: The extent to which the dyadic partners contributed evenly to the discussion, or not (Baker, 2002b). Operationally, symmetry was defined as the number of significant contributions that were proposed by the most contributive partner in the dyad, divided by the total number of such statement units in the dialogue. Significant contributions are those dialog units that contain newly asserted information, i.e., any of the contributions categorized as claims, supports, challenges, rebuttals, information and elaborations.

Measures of inter-rater reliability were as follows: *Cohen’s Kappa* = .84 for argumentative structure; *Cohen’s Kappa* = .79 for key issue of change; *Cohen’s Kappa* = .78 for interpersonal repartition of solutions; and *Cohen’s Kappa* = .84 for closure.

3. Results and Discussion

Two different types of analyses were conducted: The first part of this section contains results from quantitative analyses that were conducted on the dyadic level. In the second, we present results from investigations into the relation between individual learning gains and dialogue moves enacted by each individual learner during the interaction.

3.1. Dyadic level: dialogue features and conceptual change

We first explored which micro-level and macro-level dialogue features distinguished between successful and unsuccessful dyads. In half of the sixteen dyads at least one of the students showed individual pre- to post-test gains that were substantive enough to be an indication of conceptual change (a mean gain of at least
one point, see Coding procedures). The dialogue features of these eight dyadic interactions (hereafter referred to as the *gaining dyads*) were then compared to those of the eight remaining dyads, in which none of the individual dyad partners demonstrated conceptual change (hereafter referred to as the *non-gaining dyads*).

3.1.1. Analysis of macro-level interaction features

All sixteen protocols were analyzed according to the macro-analysis coding scheme. The two types of dyads were then compared on each of these macro-dimensions (see Table 3). The data in Table 3 show that those interactions that resulted in substantive learning gains for at least one of the dyadic partners were more likely to be characterized by interpersonal repartition of explanatory schemas among discussants *during* the interaction. In addition, they were also more likely to have developed a dialectical argument (as opposed to a one-sided argument) during the course of their interaction. However, the dialogues of gaining and non-gaining dyads did not significantly differ in whether the partners reached closure by the end of the discussion and whether they discussed the critical issue of how the duck's feet could have changed.

-Insert Table 3 About Here-

In addition, a Mann-Whitney test for nonparametric samples yielded the following results on contribution symmetry: On average, the highest contributing partners accounted for 55% of the total number of significant contributions in dialogues of gaining dyads (SD = 5%), compared to 68% for non-gaining dyads (SD = 12%), $z = 2.37, p = .018$. In other words, the number of substantive dialogue contributions was more evenly distributed among interlocutors in dialogues of gaining dyads.
3.1.2. An illustration of one-sided argumentation

Surprisingly, five dyads did not engage in dialectical argumentation, in spite of the task design (confrontation with a different account of evolution in the movie, random pairing, and instructions to engage in critical argumentation). This is even more striking in light of the provoking dialogue excerpt they received some 30 seconds into their discussion on the evolution of webbed feet. Following the instructions, all the dyads were found to have responded to the excerpt. However, instead of causing them to critically examine their existing misconception(s), the reaction of one-sided dyads was to defend and strengthen them. This is nicely illustrated in the protocol excerpt presented in Table 4, which shows how two female students respond to the hypothetical excerpt they were shown. The first column represents an ordered list of the dialogue units. They generally coincide with turns; however, one turn may include more than one dialogical move. In the last column, the abbreviations refer to the dialogue moves described in Table 2. The numbers in parentheses refer to the explanatory schema a particular move relates to. The numbers are chronological (1 is the first explanatory schema introduced in the dialogue, 2 is the second, and so on). In the following protocol, '1' refers to an explanation based on a Lamarckian schema that was proposed in the preceding turns (not presented here).

-Insert Table 4 About Here-

After reading the hypothetical excerpt aloud, discussant M reacts to the challenge proposed in the text with three different rebuttals, each intended to prove that this challenge to the Lamarckian explanation is in fact invalid (turns 9, 11 and 15). N
agrees with the rebuttals proposed by M, and in turn 12 (12.2 and 12.3) elaborates the second rebuttal and supports it with a personal example, respectively.

As aforementioned, rebuttals are in essence of a dialectical nature. However, within the argumentative context of this dialogue, the discussants are merely defending and developing their commonly shared explanatory schema by reacting to a challenge that is artificially planted in their interaction by the experimenter. Through these technical (as opposed to actual) rebuttals, they co-constructed a one-sided argument, which eventually led to the consolidation of their misconceptions as it appeared in their post-tests.

It is possible that the failure of these five dyads to develop a dialectical discussion may have been caused by the fact that they were more likely to have similar pre-conceptions on evolution during interaction. An examination of the dyad members' pretest levels of conceptual understanding revealed that one-sided dyads were somewhat more likely to have similar conceptions prior to the intervention (60%) than dialectical dyads (46%); however, this difference failed to reach significance, $\chi^2 (N = 16) = .04$, ns. Moreover, of the six dialectical dyads with similar preconceptions before the intervention, four different dialogues nevertheless showed interpersonal repartition of different solutions during the interaction. In other words, each of the two interlocutors in these four dyads introduced at least one solution that reflected a different explanatory schema. This difference may have been the result of differential learning effects from the educational movie or it may have been created in the course of the interaction.

3.1.2 Analysis of micro-level dialogue features
Analyses on features of the dialogues' microstructure were conducted with Mann-Whitney tests for non-parametric independent samples and are presented in Table 2. Since the average number of dialogue units was not significantly larger for any of the two types of dyad (see Table 5), analyses that were conducted on the relative – instead of the raw, frequencies of the dialogical moves yielded similar outcomes.

-Insert Table 5 About Here-

The results presented in Table 5 show that the dialogues of gaining dyads were characterized by a larger mean number of claims, requests for claims, challenges, technical rebuttals, actual rebuttals, agreements and concessions. They did not differ, however, on the mean number of supports and simple oppositions. As for the dialogical moves in the non-argumentative category, non-gaining dyads were found to have made more elaborations, but a smaller number of requests for information and pure informative moves. However, the within-group variance was relatively large on most measures and only the number of actual rebuttals reached significance, whereas the number of reasoned challenges was marginally significant. What characterizes these two particular types of dialogue moves is that they are both dialectical argumentative moves. Moreover, as opposed to, for example, oppositions, request for claims, and agreements, they are, by definition, *reasoned* argumentative moves (see Table 2).

Overall, gaining dyads proposed a larger number of total argumentative moves that are of a dialectical nature (challenges, actual rebuttals, concessions and oppositions). Agreements and supports, on the other hand, are argumentative moves that by themselves are not of a dialectical nature, whereas propositions of and request
for claims could be interpreted as either. Elaborations have in common with agreements and supports that they are found in processes of consensual explanation construction. Even though an elaboration is an act of explanation *development* and a support or agreement acts of explanation *validation*, they are all of a consensual nature in relation to the solution that is proposed or the explanation that is constructed. Dialogical moves that are of a pure informative nature (requests and information) are by definition not directly related to the construction or validation of an explanation. Based on these observations, two aggregate measures were compiled, each considered to capture certain dialogical moves that are typical of two different socio-cognitive processes: (1) Total number of the dialogical moves that are typical of dialectical-critical argumentation (consisting of challenges, oppositions, actual rebuttals and concessions), referred to as *dialectical argumentation* hereafter; (2) Total number of the dialogical moves that are typical of consensual, non-dialectical construction and consolidation of an explanation (consisting of supports, agreements and elaborations), referred to as *consensual explanation construction* hereafter.

Table 5 shows that whereas the non-gaining and gaining dyads equally engaged in consensual explanation construction, gaining dyads showed a substantially and significantly larger number of dialogical moves that reflect dialectical argumentation. We then tested for the possibility that even though consensual explanation construction was by itself not related to learning gains, it could prove to be crucial among dialectical dyads (e.g., developing an explanation after having reached an agreement following an episode of dialectical argumentation). However, the mean number of consensual construction moves made by gaining dyads was actually smaller ($M = 17.75$) than those of the three non-gaining, dialectical dyads ($M = 19.00$).
3.1.3. Collaborating with an 'expert' peer

As aforementioned, dyad formation was based on random assignment. As a result, four students were randomly paired with a peer whose pretest responses showed a clear and relatively stable Darwinian explanatory schema. In this section we examine whether the previous findings could have been the result of such pairing effects.

Hereafter we refer to such mixed-level dyads as W-R dyads (one student showed a wrong and the other a scientifically accepted pretest conception of natural selection), as opposed to W-W dyads (both students demonstrated wrong pretest conceptions).

We would like to emphasize that the distinction between W and R is based on the pretest performance of students prior to the educational movie excerpt and the peer dialogue, and not on the solutions that were proposed during the interaction phase. Moreover, it is important to remember that peer members of W-W dyads do not necessarily have similar misconceptions, nor that they consistently applied the same schema on the pretest.

Previous studies have shown the advantages of being instructed by an expert tutor over collaboration with an equal status peer (see Chi, et al, 2008 for an overview). However, in the context of the present study, the Darwinian peer in these four mixed dyads was not presented as an expert or assigned the role of being a tutor in the situation. This may in turn affect the type of dialogue features that are related to learning gains.

-Insert Table 6 About Here-
Table 6 shows the distribution of W-R and W-W pairs among the gaining and non-gaining dyads. The data in Table 6 seem to indicate that being paired with a Darwinian partner somewhat increased the likelihood of substantive learning gains. Even though this difference was not significant (Fisher exact test $\chi^2 (N=16) = 1.33, ns$), it cannot be ruled out that the higher number of mixed-level pairs among gaining dyads in some way affected the pattern of dialogue characteristics that were found to predict conceptual gains. In fact, all four W-R dyads engaged in two-sided argumentation during their interaction. We therefore repeated the micro-level analyses while omitting the data of the W-R dyads (see Table 7).

As is shown in Table 7, the analyses conducted on W-W dyads only ($N=12$) yielded similar results to those conducted on the whole sample ($N=16$): The dialogues of gaining dyads showed a significantly larger number of challenges, rebuttals and overall moves of dialectical argumentation. Thus, even though the learning mechanisms may be different for W-R and W-W dyads (Schwarz, et al, 2000), statistically, the potential confounding effect of dyad compositions on the previous findings could be ruled out.

3.2. Individual behavior and conceptual gains

In the preceding section we referred to a dialogue as the common product of two collaborators. In doing so, we did not consider the individual’s contributions to the interaction. Even though collaborators often commonly construct and sustain a shared
problem-solving space, it does not inevitably follow that the knowledge that has been constructed during the interaction is perceived in the same manner by all participants, nor that will their ideas be more similar following collaboration. Moreover, they may adopt different roles in an interaction (e.g., Shirouzu, et al, 2002), each promoting different learning trajectories. In the present section we will therefore focus on the individual participant and his/her contributions to the dialogue. Other studies have shown that active construction of knowledge predicted learning gains, whereas mere receiving of explanations does not (e.g., Chi, et al, 2008; Webb, Troper & Fall, 1995). Likewise, and in alignment with the self-explanation effect (Chi, et al, 1994) it is possible that self-generated constructive moves in which an individual develops an explanation may be found to relate with learning gains. Observing a collaboration partner doing the same thing, on the other hand, may not lead to gains. Thus, it is possible that a frequent occurrence of a category associated to a pair may be solely contributed to one person. We therefore calculated correlations between a single person’s learning gains and the frequency of each dialogue move enacted by this person, as well as his/her collaboration partner’s. For this analysis we focused on the data of the 12 W-W dyads only³.

All participants were exposed to a detailed explanation of the Darwinian account which was presented to them as the accepted and scientifically proven account of evolutionary change (see Method). Theoretically then, all individual learners could have increased their performance scores of conceptual understanding to a maximum post-test score of 5. For the analyses presented in this section, individual learning gains are defined as the ratio of actual gains from pretest to post-test out of the total potential gains, based on their mean pretest score. For example, a student with a mean pretest score of 3 and a mean post test score of 3.25 would have gained a mere .25 out
of a maximum potential gain of 2 points, yielding an individual learning gain score of \( \frac{.25}{2} = .125 \).

-Insert Table 8 About Here-

Table 8 presents Pearson correlations of an individual student’s learning gain score with (1) the number of times an individual learner enacted each of the different dialogue moves; and (2) the number of times the dialogue partner enacted each of these dialogue moves. As Table 8 shows, individual conceptual gains were related with the number of times that the individual posed challenges, actual rebuttals, and provided simple information that was not directly related to the explanation(s) constructed. It was also related to the partner’s behavior within the discussion, namely the number of claims that the partner proposed and the number of rebuttals.

As for the aggregate measures, neither moves of consensual construction by the self nor by the other were found to relate to learning gains. Thus, similar to the analyses on the dyadic level, processes of explanation development and validation were not found to predict learning gains. Needless to say, this does not imply that explanation construction was detrimental to learning in our settings; it simply did not have an added value with regards to conceptual learning gains. In contrast, individual learning gains did relate to the extent of engagement in dialectical argumentation both by the self as well as the collaboration partner. This seems to suggest that it was the distribution of dialectical-critical reasoning within the dyad that predicted learning gains, and not so much the effect of individual argumentation.

However, as in any authentic dialogue, the nature of one particular interlocutor’s utterances is likely to affect the other’s. For example, challenges will often elicit
rebuttals. Therefore high inter-correlations are expected between dialogue moves made by the two individual dialogue partners. For example, the correlation between the occurrence of own and partner’s dialectical moves was $r = .610$, $p = .002$. For this reason, we conducted a hierarchical multiple regression analysis on individual learning gains (see Table 9). In the first model, learning gain scores were predicted by a person’s own and collaboration partner’s intuitive conceptual knowledge (i.e., the pretest scores). As shown in Table 9, neither measure significantly added to the prediction of conceptual gains following the interaction. All the following comparisons of prediction contribution will be compared to this first regression equation.

Following, and based on the pattern of simple correlations presented in Table 8, the two dialectical dialogue move categories (self, other) were entered in the equation. Together these two variables explained an additional 31% of the total amount of variance in learning gain scores, $F$ change $(2, 19) = 5.254$, $p = .015$. The standardized regression coefficients reveal that as a result of collinearity only the relative contribution of the number of own dialectical moves reached significance. Thus, from this analysis it seems that it is the active generation of dialectical argumentative moves by the self that is crucial for learning that involves conceptual change.

Whereas Table 8 showed that moves of consensual construction were not found to correlate with learning gains, this relation may be conditioned by the number of dialectical moves. For example, it is possible that processes of dialectical argumentation leads to learning gains only when in concurrence with processes of consensual construction. To test for this possibility, the two categories of consensual
construction (self, other) were entered in the regression equation. However, as shown in Table 9, they did not improve the prediction of individual learning gains, $F$ change (2, 17) = .907, ns.

Finally, to further investigate which of the non-aggregated micro-level dialogue categories predicts individual learning gains, an additional exploratory stepwise regression analysis was conducted with all the 22 discrete dialogue categories. The cut-off value for predictor inclusion was $p < .05$. Of the different discrete dialogue categories, only the number of own and other’s actual rebuttals were each found to uniquely add to the prediction of consequential learning gains (see Table 8). Given the initial conceptions of both dyadic partners, the number of own and other’s rebuttals predicted an additional 39% to the total amount of variance, $F$ change (2, 19) = 7.52, $p=.004$. Compared to other argumentative moves, rebuttals require reasoning of a more sophisticated kind: By definition they follow the articulation of a solution and the appearance of at least one challenge to that solution. Cognitively, a rebuttal requires of the individual to devaluate this challenge and at the same time reinforce the initially articulated solution. Challenges, by themselves, were not found to significantly add to the prediction power of own and other’s rebuttals, even though they are a necessary condition for rebuttals to occur. It seems to indicate that only when peers are really and deeply involved in high-quality critical reasoning they may gain from peer dialogue on a scientific topic, such as natural selection, that is known to be extremely robust to change.

At this point we can only speculate on the exact mechanism through which learning through dialectical peer argumentation occurs. The short excerpt in Table 10, however, does seem to illustrate how a chain of interchanging rebuttals and challenges may support students in comprehending the essential difference between
their own intuitive misconceptions and the Darwinian account that they attempt to apply to the duck’s question. Both students had applied typological schemas on the pretest. Following the movie excerpt, student A attempted to construct explanations that incorporated several Darwinian features, whereas student B seemed to adhere to a Lamarckian type of explanation throughout the discussion:

-Insert Table 10 About Here-

4. Conclusions

The combined findings of the present study and our previous experimental studies (Asterhan & Schwarz, 2007) have revealed the benefits of dialectical argumentation in fostering learning of the conceptual change kind. The quantitative data presented here allowed us to identify several dialogue and interaction characteristics that predict or do not predict learning gains. This is an important step forward. However, they do not provide us with empirical evidence on the mechanism of learning through peer dialectical argumentation. To uncover the exact dynamics of such processes qualitative in-depth analyses are warranted (see for example, Trognon, 1993, for an application of such an approach). Our findings also outline a number of interesting and important questions concerning learning from and in peer interaction, which should be explored in future research:

Most importantly, the failure to find a relation between dyadic or individual learning gains and processes of consensual construction of explanations is particularly surprising: Acts of constructing explanations and of arguing in favor of that explanation are both epistemic activities that are believed to lead students to
externalize, clarify and organize their knowledge (de Vries, et al, 2002; Ohlsson, 1996). Indeed, empirical research has provided extensive evidence that consensual explanatory and elaboration-based activities are an important instigator of learning from peer collaboration (e.g., van Boxtel et al, 2000; Chi et al, 1994; Coleman, 1998; King & Rosenshine, 1993; Okada & Simon, 1997; Shirouzu, et al, 2002; Webb, et al, 1995). We would like to highlight two important differences between the present and the above-mentioned studies that may account for these different findings:

First of all, even though some of these studies focused on conceptual understanding as a learning outcome, they did not specifically focus on learning that requires the revision of robust misconceptions. In a way, our findings confirm the expectations articulated by deLeeuw and Chi (2003) that whereas (self-) explanation may be sufficient for certain forms of conceptual change that involve local replacement of knowledge pieces, it may by itself not be sufficient to bring about conceptual change of particularly robust misconceptions, such as evolution (Chi, 2005). Even though psycho-educational theories of learning and related research seems to suggest that dialectical argumentation may be beneficial for different types of knowledge revisions (see Schwarz & Asterhan, in press, for an overview), our findings indicate that it might be particularly beneficial for learning that involves the substantive reorganizations of knowledge described in the literature as conceptual change literature of the radical type.

The second contrast that we would like to draw concerns the difference between our task design and those from peer collaboration studies in problem solving and discovery learning (e.g., Okada & Simon, 1997; Shirouzu, et al, 2002; Schwartz, 1995). In the latter type of studies, students are typically asked to perform certain manipulations and draw conclusions concerning an underlying set of rules that govern
the phenomena. Assessment of learning focuses on measures such as number of correct rules, number of trials, or time on task until completion, not on conceptual change. In tasks settings such as these in which students have to make sense of the outcomes of their manipulations in order to refine their predictions and/or next trials, it is not surprising that their attempts to interpret and construct explanations of these outcomes have been proven to be crucial for effective problem solving.

In contrast, our students did not gather any data or perform any manipulations that they needed to reflect on and/or interpret. Instead, the 'data' was given to them as factual information (i.e., a change from pigeon-like feet to webbed feet) and their actions only consisted of providing the best possible explanation for this 'data'. Whether this was particularly evident in our task design which did not involve any data collection, it is always at the heart of instructional interventions for conceptual change within the socio-cognitive conflict paradigm: The aim is to confront misconceived knowledge with contradicting evidence or theories. It is based on the idea that students will come to realize that the explanatory power of their naïve conceptions is not sufficient for these newly presented cases and that a different explanatory schema should be adopted, applied or constructed. In essence, it thus concerns a situation in which (at least) two explanatory schemas or theories compete with each other and the epistemic status of at least one of them is contested\(^5\). Peer argumentation allows for this process to be played out explicitly between two actual persons.

Dialectical peer argumentation then may be considered as an externalization of these otherwise (assumed) intra-mental processes in a rational, conscious and intentional manner. Moreover, it may be particularly beneficial in task settings that are meant to induce *intentional* conceptual change (Sinatra & Pintrich, 2003) of
robust misconceptions. Approaches that call for consensual elaboration and integration of different explanatory theories seem to be unlikely to succeed in such settings (see also Chi, 2008; Vosniadou, 2007): By asking students to simply engage in acts of explanation, elaboration and justification, they may become aware of flaws in their naïve theories, which in turn may trigger argumentation. However, in the majority of cases, dialectical argumentation on scientific topics will not occur naturally (e.g., Asterhan & Schwarz, 2007a) and it is likely to be contingent on explicit instructions, carefully planned task designs, individual capabilities and cultural norms.

In addition, even though a critical consideration of one's own or a peer's intuitive conceptions is an essential and important step towards the revision of knowledge representations, it cannot be expected to be necessary sufficient to lead to spontaneous conceptual change. It seems that in order to increase the likelihood that learners may benefit from dialectical argumentation, they should have at least some accessibility to new knowledge sources. This knowledge does not necessarily have to be well consolidated or developed yet, but should allow them to construct superior explanations. Prior to the peer interaction phase, all students in our study took part in a rather traditional show-and-tell instructional intervention in which they were shown an application of the scientifically accepted explanatory schema of evolution to a particular phenomenon. As has been repeatedly shown by others (e.g., Brumby, 1984; Ohlsson & Bee, 1992), such instructional techniques have very low success rates by themselves. However, this information is likely to have broadened the student's 'conceptual tool box', and maybe even the range of available explanatory schemas, during the interaction phase immediately following instruction. Through a process of refinement and contrasting with elements of other schemas, including their own
intuitive conceptions, students may come to construct, articulate and consolidate a more sophisticated explanation according to a superior explanatory schema. This suggestion should be further explored, however, by in-depth qualitative analyses of selected student protocols that delve into the dynamics of learning in peer collaboration.

Lastly, whereas acts of consensual development of explanations did not significantly contribute to the prediction of learning gains, gaining and non-gaining dyads equally engaged in such processes. Therefore, it is quite possible that dialectical argumentation in the absence of consensual co-construction of explanations will prove to be detrimental to learning. This prediction could not be tested by the current set of data, since it did not occur in our sample. However, it is likely that it is not only the engagement in dialectical argumentation per se that promotes learning, but also the way in which it is accomplished. For example, competitive debating in peer dialogue may not yield the same learning effects as a collaborative attempt to critically, but jointly examine different solutions. Further systematic research will have to prove whether our insights may be generalized to larger samples and whether the peer collaboration characteristics identified in this study will also be found to support learning in other settings and other content domains.

Acknowledgements

The research reported in this article was supported by the Israeli Foundation Trustees, Grant 27.35. We thank Naomi Miyake and three anonymous reviewers for insightful comments which significantly strengthened the article. We are also grateful to Oded Armoni for his assistance in the dialogue analyses.
Appendix: Examples of two different explanatory schemas each belonging to a
different schema category.

Example of a student explanation according to an explanatory schema termed "individual change followed by selection" (belonging to the Hybrid schema category, grade: 4), which explains change in terms of selection of those individual species members that managed to develop the advantageous trait by themselves:

"Biological theory would explain the process that caused many mosquitoes to be resistant to DDT today, as follows: They developed some kind of defense mechanism against the substance. This mechanism enabled them to defend themselves against it, instead of going extinct, like the other mosquitoes who did not manage to develop such mechanisms. In fact, their bodies and their inner structures created a type of antibody that repels the substance and 'immunizes' the mosquito against it." (student # 53, pre-test)

Example of a student explanation according to an explanatory schema termed "mutations in reaction" (belonging to the Typological schema category, grade: 3), which explains evolution in terms of the appearance of genetic mutations that occur in response to the need for a certain advantageous trait and which occur to the population as a whole:

"In the beginning, a lot of the mosquitoes were killed, but as a result of an evolutionary process mutations were created in the mosquitoes' genetic make-up such that they became resistant to DDT. Since this mutation is inside the mosquitoes' genetic make-up this was passed on to
their descendants, and this is how a new type of resistant mosquitoes was created." (student #76, pre-test)

References


*Journal of Educational Psychology, 98*, 766-776.


Footnotes

1 Chi’s investigations into the self-explanation effect have been conducted by prompting students to provide explanations of learning material to another person (the experimenter), who requires further clarifications if the communication is not sufficiently understandable. Thus, even though these interactions are not necessarily collaborative in nature, the settings are nonetheless cooperative and social in nature since the students’ contemplations or explanations are intended for communication with another person.

2 The few instances of disagreement on segmentation within a turn were all characterized by lack of within-turn segmentation by one of the two human raters. There were no instances of differential segmentation that caused partly overlapping dialogue units between raters.

3 The reason for omitting the data of W-R dyads was two-fold: 1) Students with Darwinian pre-conceptions could not gain and were therefore excluded; 2) Exploratory analyses showed that whereas dialectical argumentation indeed predicted learning gains on the dyadic level, when disentangling the dialogue moves made by each particular discussant we found a distinctively different pattern of learning: In contrast with learning in W-W dyads, individual learning of the partner with initial misconceptions seemed to be negatively related with the number of dialectical moves (s)he made, but positively related with the dialectical moves made by the Darwinian partner. Unfortunately, because of the small sample size ($N = 4$) these relations could not be quantitatively verified.
To test for the possibility that the quantification method for individual learning gains might have affected the outcomes, all analyses presented in this section were also conducted with an alternative measure, namely: the numerical difference between pretest and posttest scores. However, these analyses yielded similar overall outcomes and we therefore do not present them here.

It should be noted that some task designs rely on hypothesis testing devices to create this conflicting evidence (e.g., Howe, Tolmie, Duchak-Tanner & Rattay, 2000; Schwarz & Linchevski, 2007), and thus combine peer dialogue and conflict with testing and data interpretation. It would be interesting to see whether in such task designs both dialectical argumentation and consensual explanation construction will foster conceptual gains.