Examining the Relation between Domain-Related Communication and Collaborative Inquiry Learning

Anjo Anjewierden, Hannie Gijlers, Bas Kolloffel, Nadira Saab*, and Robert de Hoog
Department of Instructional Technology, Faculty of Behavioural Sciences,
University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands
Email: {a.a.anjewierden, a.h.gijlers, b.j.kolloffel, r.dehoog}@utwente.nl
* Center for the Study of Education and Instruction, Faculty of Social and Behavioural Sciences,
Leiden University, PO Box 9555, 2300 RB Leiden, The Netherlands
Email: nsaab@fsw.leidenuniv.nl

Abstract: Research has suggested that providing elaborated explanations is often more beneficial for learning than receiving explanations (e.g., Webb, 1989). Applied to chat communication in a collaborate inquiry learning environment, we would expect that in a dyad the learner with more domain-related contributions than his partner would learn more. In the paper we develop a method to examine the relation between domain-related chats and learning outcome for intuitive knowledge. We describe how we automatically extract domain-related messages, and score them based on the expected cognitive effort to produce the messages. The analysis confirms that there is a positive relation between a high score on domain-related chats and the learning improvement as measured by the difference between a post-test and a pre-test on intuitive knowledge.

Introduction

In a collaborative learning setting two or more students share and construct knowledge while they work towards the solution of a problem or assignment. Research has shown that collaboration between students may enhance learning (Lou, Abrami, & d'Apollonia, 2001; Slavin, 1994; van der Linden, Erkens, Schmidt, & Renshaw, 2000). Inquiry learning environments are very suitable for collaborative learning. In a simulation based inquiry environment students learn through experimentation and scientific reasoning. The interface of the learning environment allows students to change input variables and observe the effects of their actions. Students learn about the principles and rules of scientific phenomena through processes like hypothesis generation, experimentation, and conclusion (e.g., de Jong & van Joolingen, 1998). During inquiry learning, students must make many decisions (e.g., which hypothesis to test, what variables to change) and in a collaborative setting, the presence of a partner stimulates students to make their plans and reasoning about these decisions explicit.

To maintain a successful collaborative working relationship, ideas and theories must be externalized and explained in a mutually understandable way for the partners in the collaborative learning group (Teasley, 1995). Through externalization students express and explain ideas, ask for clarifications or arguments and might generate new ideas. The process of making ideas public through externalization and explanation, stimulates students to rethink their own ideas and might even make them aware of possible deficits in their reasoning (van Boxtel, van der Linden, & Kanselaar, 2000). Research indicates that the degree of participation in collaborative activities is related to group performance as well as students' individual learning.

Protocols recording collaborative learning can be analyzed from several perspectives. They have been analyzed in terms of students' degree of participation, communicative activities such as arguments and elaborations (van Boxtel et al., 2000), different learning processes (Hmelo-Silver, 2003; Saab, Van Joolingen, & Van Hout-Wolters, 2005), or the exchange of domain-related information (Van Drie, Van Boxtel, Jaspers, & Kanselaar, 2005). In this study we focus on the collaborative construction of domain-related knowledge, by examining the communication protocols of dyads who together interacted with a collaborative inquiry learning environment. Within a dyad, students may not only differ in their overall degree of participation (Cohen, 1994), the characteristics of students' contributions may also differ. For the collaborative construction of domain related knowledge it seems important that students not only actively participate in the chat discussion but also share and exchange domain-related information in their dialogue.

Being interested in domain-related knowledge construction, our focus is on students' externalizations of domain-related conceptions and their interpretation of information obtained from the learning environment or provided by their partner. More specifically, we are interested in the degree of domain-related information each partner contributes to the dialogue. From earlier studies on collaborative learning we learn that, for example, providing elaborated explanations is often more beneficial for learning than receiving explanations (e.g., Webb, 1989). This can be explained by the fact that students who are providing elaborate explanations are actively engaged in externalization processes, which probably stimulates acquiring new knowledge. However, not only explanations can be beneficial for learning together.

Research question and hypothesis

Based on the considerations presented above, this study examines how students' domain-related contributions are related to their individual learning outcomes. The simulation-based inquiry learning environment used in this study requires students to focus on relations between variables in the domain. It is expected that within a dyad, a student who externalizes a higher proportion of domain-related knowledge in that dyad than the partner, reaches a higher post-test score than students who externalize less domain-related knowledge during the learning session.

Method

Learning environment and task

Students worked in dyads with an inquiry learning environment that was based on a computer simulation of colliding objects. The main task for the students was to discover the laws of physics underlying the simulation. The learning environment consisted of four simulations accompanied with assignments that presented the learners with small research questions to guide their inquiry learning process. A total of 35 assignments were available in the learning environment. Dyads worked collaboratively on two separate computers with a shared interface and communicated through a chat channel (based on Microsoft Netmeeting technology).

Tests

Students' individual learning outcomes are assessed with two domain knowledge tests, a "definitional knowledge" test focusing on facts and formulae, and a "what-if" test for intuitive knowledge on relations. Each question of the "what-if" test consists of three parts: condition, action and prediction. First a condition/situation before a collision is presented to the students. Subsequently, the action (for example, a collision against a fixed wall) is presented. Finally, three predicted states are presented to the students either in text or pictures. Students are asked to select the state that follows from the action in the given condition. The definitional knowledge test as well as the "what-if" test was computer administered and pre- and post-tests were parallel versions of the same test.

Participants

Dyads were heterogeneous with respect to students' school achievement in the domain of physics (this information was provided by the participating schools). This grouping was based on the finding that heterogeneous grouping is beneficial for both high and low achieving students (Webb, Farivar, & Mastergeorge, 2002). Students were paired with a student from their own class. Participants attended two sessions. In the first session dyads were composed and students practiced with a SimQuest practice simulation that allowed them to explore the features of the interface and work with the chat tool. The second session started with the two pre-tests, followed by 90 minutes of interaction with the simulation environment. At the end of the session students completed the post-test versions of the knowledge tests.

Determining learner contribution in dyads

The question we are addressing is whether there is a relation between the nature of the domain-oriented contribution of a learner and learning improvement. In this section we first give some examples of the interaction between learners, and then motivate our choice of how to develop a method that allows answering the research question. An example of an excerpt of the interaction between two learners called X and Y is:

1 12:58:48 X: if the mass becomes higher, the momentum decreases, I think
2 12:59:43 Y: no, that is not true
3 13:00:02 Y: p also is higher, have a look
4 13:00:03 X: experiment
5 13:00:16 *** *Running a simulation*****

Here X thinks there is a qualitative relation between mass and momentum: if the mass increases, the momentum decreases (line 1). Y, after about a minute, realizes that the relation is incorrect (line 2) and suggests that the momentum, the symbol p stands for momentum, increases when mass increases (line 3). X suggests to do an experiment to find out more (line 4) and runs the simulation. This brief example illustrates learners X and Y share their individual understanding of the domain and try to reach a joint understanding.

Excerpts like the one above are not very common. Below is an example, between learners P and Q, which follows a pattern that is much more frequent:

6 13:31:33 P: speed is the same after the collision
7 13:31:37 Q: yes

 © ISLS

8 13:32:04 Q: 4
9 13:32:12 *** *Answer 4 is selected (it is incorrect)* ***
10 13:33:11 P: 1
11 13:33:25 Q: ok
12 13:33:29 *** *Answer 1 is selected (it is correct)* ***

P communicates a domain-related observation (line 6). Q immediately agrees (line 7) and after about half a minute proposes 4 as the correct answer (line 8), the answer is wrong (line 9). Then P proposes 1 as the answer (line 10), Q agrees (line 11) and the answer turns out to be correct (line 12). The pattern is that one of the learners exchanges a domain-related finding (line 6) and the partner only acknowledges this without referring to domain-related terms. Often the entire discussion then switches to what answer to give. If this happens repeatedly, P may at some point decide to give up formulating domain-related messages all together as Q does not appear to do anything with them.

In order to determine what the level of domain related contribution of each learner is, we introduce some abstractions. The level of domain-related contribution of a learner is denoted as the value D(learner). Based on the example excerpts above, we have $D(X) \approx D(Y)$ because X contributes one domain related statement (statement 1) and Y does the same (statement 3) and $D(P) > D(Q)$ because P contributes one domain related statement (statement 6) and Q does not contribute a domain related statement. It would perhaps be tempting to conclude that $D(X) > D(Q)$ and $D(Y) > D(Q)$, however, not allowed to conclude this. The reason is that the value of D(X) is dependent on the collaborate setting with Y, which we denote as D(X|Y), while the value of D(Q) depends on the collaborative setting with P. If we would like to estimate the "true" value of D(X) in other collaborative settings then a better approximation is:

$$D(X) = \text{average}(D(X|Y) + D(X|P) + D(X|Q))$$

Given that we have no values for D(X|P) and D(X|Q), a better estimate of D(X) cannot be computed. The same goes for the other three learners in the example, and we can therefore not compute an estimate for any D(*learner*) that can be meaningfully compared with D estimates for other learners.

An intuitive example, for comparison purposes, is the idea of a marathon run. Two runners X and Y decide to beat the world record. They agree that during the first 35 kilometers Y acts as a pacemaker and runs in front. X wins the race in a record 2.03:00, Y finishes in 2.04:00. A year later, under precisely the same circumstances, P' and Q' also want to beat the world record, but P' and Q' do not make any prior agreements, each runs his own race, P' finishes in 2.03:30 and Q' in 2.04:30. All other things being equal, it is not allowed to conclude that P is faster than P' as their performance depends on having a runner they as a pacemaker or not. Though the conclusion is drawn in practice, otherwise there could not be a world record for the marathon, it is hardly justified and proposals to ban pacemakers from marathon is based on these considerations.

To summarize: it is not possible to obtain a good estimate for the value of the contribution of a learner in a dyad (or larger group of learners) that can be compared to other learners in other dyads in the same experimental setting. This result may be important for CSCL in general, as it points to a major methodological issue, and might also explain why previous research has not related individual performance to individual learning outcomes in a collaborative learning setting and why researchers carefully consider the composition of the dyads or groups (Webb, Nemer, Chizhik & Sugrue, 1998).

Does this prevent us from finding out whether the contribution measure, D(learner), is related to learning outcome? We think this is still possible by using an indirect method. Within a dyad, the learner with the highest value for the level of contribution measure is added to a group of learners called A, the other learner in the same dyad is added to a group called B. This is like a knock-out competition, the A's would be the winners the B's the losers. If the average learning outcome of the A's is significantly different from that of the B's, the level of contribution measure is the probable cause, because this was the reason for partitioning the learners into A's and B's.

Domain-related contributions

Partitioning learners into A's and B's, as proposed above, requires a measure of a learners' contribution to the dialogue, that allows us to do the assignment as being A or B. Inquiry learning environments, like the one used in this study, stimulate the acquisition of knowledge about relations in the domain. For example, in physics momentum is defined as $p = m * v$, where p is momentum, m is mass, and v is velocity. Learners can find out about these kind of relations by changing one of the variables and (graphically) inspecting the effect on the others.

Learners can share their thinking on the relations with their partner through the chat tool. To determine the level of domain-oriented content of a message, we distinguish three types of contributions:

- *Domain terms*. The use of domain terms, such as velocity, increases and the abbreviation v, transfers at least a certain domain focus by the learner. The message *shall we look at mass* contains one domain term.
- *Qualitative statements*. These are phrases containing both a quantity and a qualitative relation, for example *speed increases*, or *momentum is lower*. We do not make a distinction between qualitative statements that result from observation (*speed increases*) and qualitative statements that suggest future action (*shall we increase the mass*). All such statements demonstrate a clear domain focus.
- *Conditional sentences involving qualitative statements*. These are evidence of interpretations or hypotheses related to the domain. A conditional sentence is the grammatical construct which relates a condition to a consequence. In the collision domain, an example is *if the mass becomes higher, the momentum decreases*.

Analysis of the messages

Identifying the above three types of contributions in chat protocols, requires an analysis, or semantic interpretation, of the message. A human can, if the message is *velocity increases,* reason that *velocity* is a quantity, and *increases* a qualitative relation. It therefore may be concluded that the message is a qualitative statement. Interpretations of this kind are very different from categorical coding, the usual method of analysis in the behavioral sciences in general and CSCL in particular (automatic support for categorical coding is discussed by Rosé et al., 2008; Anjewierden & Gijlers, 2008). Erkens & Janssen (2008) focus on determining the communicative function of messages in online discussions. Their MEPA tool automatically segments messages into one of 29 predetermined categories based on the theory of dialogue acts. Erkens & Janssen state that their automatic procedure can only be used for content that can be indicated by specific marker words, phrases, or actions, which is a limitation of the approach. In *velocity increases* there are no marker words, both words convey meaning. For the automatic analysis of the messages, fitting our research question, a tool which provides more flexibility, in some sense generalizing the idea behind MEPA, is required.

We have used a text analysis tool called tOKo (Anjewierden, 2006). This tool is being used by social scientists to study, for example, online communities (e.g., de Moor & Anjewierden, 2007), and by semantic web researchers to create domain vocabularies and extract semantic relations (e.g., de Boer, van Someren & Wielinga, 2007). Automatic (semantic) text analysis is uncommon in CSCL, so we provide only a global overview of how we applied it to the analysis of the chat messages and omit technical detail where possible. The analysis starts with a corpus that contains all the chats and the objective is to define syntactic patterns in these chats that allow the extraction of the types of contributions we are interested in. For this we must identify the features detailed below.

Domain terms. The domain terms have been selected by sorting all words on frequency and then manually selecting domain terms that occur at least five times. The most frequent word, ignoring stop words, is *ok* (1294), the most frequent domain term is *velocity* (snelheid, 440). Selecting the domain terms took about two hours.

Conditional sentences can be found using discourse markers and they correspond to the categories called *condition* and *consequence* in MEPA. We have identified several syntactic patterns in the corpus which mark a conditional sentence. The most frequent is *als ... dan ...* (if, then), also frequent is *hoe ... hoe ... (*the, the; *the higher the speed, the lower the mass)* and sometimes *... wanneer ...* (when) is used. In the latter case, the condition and the consequence are reversed (*speed increases when mass is increased*).

Qualitative statements. The extraction of qualitative statements consists of three steps. The first step is to enumerate all quantities and all qualitative relations. Next, for each quantity and each qualitative relation we need to find the terms learners use for them. For example, the quantity velocity can appear as *velocity*, *speed*, and *v* (the symbol). Similarly, the qualitative relation *increase* can appear as *goes up*, *higher* and so forth. In our Dutch chats we found seven syntactic patterns for *increase* and eight for *constant*, including the negation *does not change.* Finding these variations is not difficult, the easiest method is to use a quantity as a key word and use a concordance index to inspect the surrounding text. The lower right pane of Figure 1 shows the results for *snelheid*. In the central column the concept snelheid is displayed, and to the left and right the surrounding text, for example, the first line “denk je dat je je snelheid moet veranderen” (do you think you must change your velocity).

We now have a set of syntactic patterns for quantities which we call $quan$ and one for qualitative relations which is called $qual$. The third step is to define patterns for qualitative statements as a whole, they are:

$quan$ ... $qual$ and $qual$ ... $quan$

The first pattern finds phrases in which the quantity appears before the relation and the second pattern finds phrases in which the quantity appears after the relation (*increase speed*). .

 © ISLS

In total the selection of the terms and the definition of the patterns took about three man days. Once the patterns are specified, the analysis is automatic. tOKo extracts zero or more clauses from each message. For example, from the utterance *the velocity increases* it extracts:

```
domain_term(velocity).
domain_term(increases).
qualitative_statement(velocity, increase).
```

and from *speed becomes larger* it extracts:

```
domain_term(speed).
qualitative_statement(velocity, increase).
```

Notice that the qualitative statement extracted is identical in both cases, although the two phrases have not a single word in common. It should be noted that the manual part of the analysis appears to be time consuming, but one should realize that this upfront work greatly reduces the amount of work that otherwise would go into a manual analysis of thousands of chat messages (in this research 18.700). In other words: the investment in identifying syntactic patterns is easily recovered by reducing the investment in the analysis of a large number of chat messages.

We have, of course, omitted several details in the above description. Not every conditional that begins with *if* also contains the corresponding *then,* certainly not in chat text, and it still should count as a conditional sentence. Another, relatively common case when mismatches occur is an overrun of one pattern match into another (Dutch readers will recognize an example in the last line in Figure 1, under the heading Pattern search: *snelheid hoe groter*). The pattern language which has been used in the automatic analysis provides a mechanism to suppress such matches.

The reliability of extraction is difficult to assess, the patterns were derived from the corpus and testing on the same corpus provides no information about reliability for other chat corpora. Some of the terms are very domain related, momentum is an example, and it would probably not make any sense to test the patterns on chats from, say, a thermodynamics domain. To provide some idea about reliability we have manually examined 15% of the chats and computed precision (0.94) and recall (0.88) for the conditional sentences and qualitative statements. Domain terms are found using an algorithmic procedure, this can only affect reliability when a domain term is used out of context (*the speed of your typing is amazing*).

A characterization of the approach we use is that within a limited domain, with a limited vocabulary, it is more or less possible to enumerate the domain terms and use syntactic patterns (Hearst, 1992) to extract meaningful phrases related to the domain. This extraction can be done automatically.

As the result of the analysis outlined, we have for each learner a set of domain statements classified into the three types of contributions. The next question is how to "value" these contributions as they are not equal in their level of cognitive effort to produce them.

Computing the value of domain-related contributions

The level of domain-related contribution, D(*learner*), is computed by assigning a score to each message for any given learner, summing these scores and then dividing by the total number of messages. D(*learner*) is therefore the average domain-oriented content per message. The unit of analysis for this computation (Strijbos et al., 2006) is a single chat message.

The weights used in the scoring function are: +2.0 for a conditional sentence, +1.0 for a qualitative statement, and +0.4 for use of a domain term. These weights have been chosen in such a way that they reflect the cognitive effort of the learner to produce them. The weight -0.5 is given to an agreement. This negative weight penalizes learners who agree more than their peer and can thus be seen as passive, relying on the cognitive efforts of their partner. In all cases, a weight is only assigned once to a single message. If a message contains two domain terms it obtains a score of +0.4 (and not +0.8). The scores for the examples above are +1.4 (*speed increases*) and +3.4 (*if the mass becomes higher, the momentum decreases).* As there is some arbitrariness in the scoring function, we conducted a sensitivity analysis by varying all weights independently in a range of -0.5 to +0.5 from the values defined above and this produced no significant change in the overall outcome: who is A stays A and who is B stays B. This shows that the scoring function is robust, which, of course, does not entail that it is the "correct" function.

Results

The results presented below are based on the data obtained by Saab (2005) on the collision domain as described in the Method section. Saab's original process data, consisting of interactions with the simulation environment and the chats, were first integrated. The chats have been normalized by correcting spelling errors, contractions

and other textual noise. Dyads for which at least one of the learners a pre-test or post-test was not available were removed, and so was a single dyad who communicated in a language other than Dutch. The process data from the 66 dyads that remained contained 18,007 chat messages. Data with respect to the chat communication, particularly the number and nature of chat messages, is displayed in Table 1.

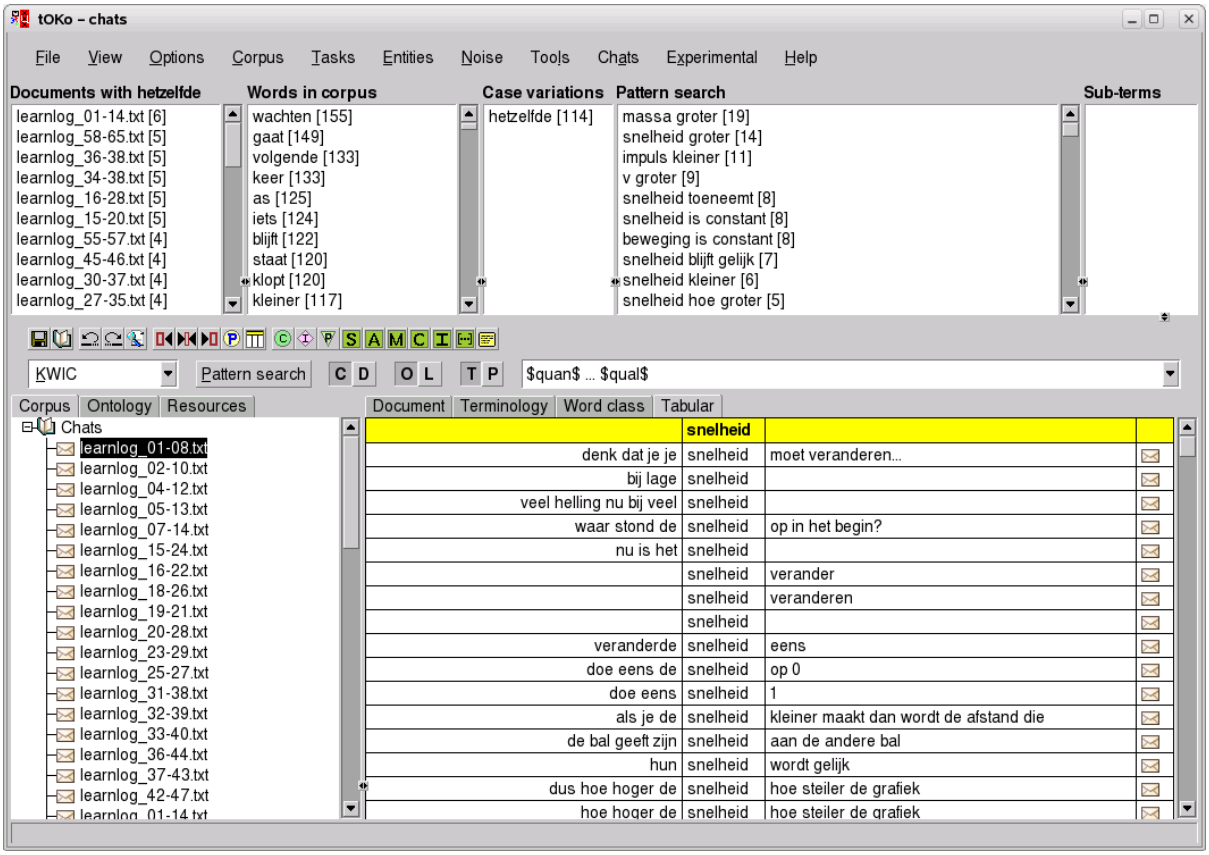


Figure 1. The application of the tOKo text analysis tool on the collision chat corpus.

Table 1: Data chat messages

	Group			
	A (n = 66)		B (n = 66)	
	M	*SD*	*M*	*SD*
Number of messages expressing				
Agreement	25.83	13.70	33.95	16.71
Domain terms	15.08	9.19	12.08	8.53
Conditional sentences	1.82	1.65	0.94	1.38
Qualitative statements	4.38	3.10	2.64	2.64
Total number of chat messages	138.03	58.72	134.20	56.84
Chat score per message	0.10	0.10	0.02	0.08

The data regarding chat messages displayed in Table 1 were analyzed by means of two-sided Wilcoxon signed rank tests. With regard to Agreement it was observed that B's express agreement more frequently compared to A's (z = -4.28, p < .001). On the other hand, compared to B's, A's mentioned more domain terms in general (z = -2.73, p < .01), made more conditional sentences (z = -4.00, p < .001), and more qualitative statements (z = -4.26, p < .001). The number of chat messages did not differ between A's and B's (z = -0.97, p = .33). By definition, the chat scores differed in favor of the A's (z = -7.06, p < .001).

In Table 2 the results of the two knowledge tests (definitional knowledge and intuitive knowledge (what-if)) are displayed. Group A contains the learners with the highest domain-related chat score within a dyad, and group B contains the lowest scoring partners in a dyad.

 © ISLS

Table 2: Results knowledge tests

	Group			
	A (n = 66)		B (n = 66)	
Category	*M*	*SD*	*M*	*SD*
Definitional knowledge				
Pre-test	6.39	2.26	5.98	2.35
Post-test	8.09	2.29	7.36	2.81
Gain from Pre to Post	1.70	3.06	1.38	3.36
Intuitive knowledge				
Pre-test	4.76	2.12	4.97	1.95
Post-test	7.55	2.29	6.91	2.33
Gain from Pre to Post	2.79	2.25	1.94	2.41

In order to gain more insight in knowledge gains within and between Group A and Group B, the data displayed in Table 2 were analyzed by means of paired samples T-tests. Sidak's correction for multiple comparisons was applied to control for chance capitalization.

With regard to *definitional knowledge*, within Group A a significant knowledge gain was observed, that is their post-test scores were significantly higher than their pre-test scores (t = -4.51, p < .001). The same was true for the knowledge gain within Group B (t = -3.33, p < .01). Comparisons between Group A and B showed that their pre-test scores were equal (t = 1.02, p = .31) and so are their post-test scores (t = 1.84, p = .07) and definitional knowledge gain (t =.59, p = .56).

Regarding *intuitive knowledge*, the knowledge gain within Group A was significant (t = -9.01, p < .001) and so was the gain among B's (t = -6.54, p < .001). Furthermore, comparison between A's and B's showed that A's intuitive knowledge gains were greater than those of B's (t = 2.35, p < .05). A's and B's did not differ from each other with regard to pre-test score (t = -0.58, p = .57), post-test score (t = 1.94, p = .06).

What students talk about can influence what they learn, and conversely, what they learn (or already know) can influence what they talk about. In order to investigate how A and B's knowledge, learning, and communication relate to each other, Pearson's product-moment correlations between A's and B's have been calculated with regard to knowledge measures, chat messages, and chat scores, (see Table 3).

Table 3: Product-moment correlations between group A (rows) and B (columns)

		Group B											
		Knowledge measures						Chat communication					
Group A		1	2	3	4	5	6	7	8	9	10	11	12
Knowledge measures	1 Pre-test definitional						-.28*						
	2 Pre-test intuitive												
	3 Post-test definitional												
	4 Post-test intuitive			.29*	.33**		.26*						
	5 Gain definitional				.27*								
	6 Gain intuitive				.41**		.29*						
Chat communic.	7 Agreement							.62**	.25*			.41**	-.30*
	8 Domain terms	-.27*						.36**	.48**	.33**	.48**		
	9 Conditional sent.	-.26*								.32**	.29*	-.29*	.28*
	10 Qualitative statem.	-.27*							.34**	.45**	.50**		.31*
	11 Number messages							.36**				.41**	
	12 Chat score/message							-.35**		.40**		-.36**	.48**

* Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

The *top-left quadrant* of Table 3 displays the correlations between A's and B's knowledge measures. It can be observed that particularly A and B's intuitive knowledge gains (6) and their intuitive knowledge scores on the post-test (4) are related to each other.

The *top-right quadrant* is empty, indicating that there are no correlations between A's knowledge measures (1-6) and B's communication (7-12).

The *bottom-left quadrant* shows that there are some correlations between B's knowledge measures (1-6) and A's communication (7-12): B's prior definitional knowledge (1) is slightly and inversely related to A's frequency of expressing domain terms (8), conditional sentences (9), and qualitative statements (10).

The quadrant at the *bottom-right* of the table, displays correlations between A's and B's communication. The correlations on the (shaded) diagonal suggest that students within dyads "mirror" each other. There is not only a moderate positive correlation between the number of chat messages uttered by A's and B's (11), but also the nature of the messages (7-10) is found to correlate positively. For example, if one peer expresses agreement (7) often, the other peer is likely to do so as well (r = .62); if one peer frequently makes qualitative statements (10), the other peer is likely to frequently make qualitative statements too (r = .50), and so on.

Conclusion

In this study it was found that within dyads, the students who post more domain-related messages often gain more intuitive knowledge than their partners. Nonetheless, the data also showed that gains in intuitive knowledge of A's and B's are positively correlated. Furthermore, A's use domain terms, conditional sentences, and qualitative statements more frequently, in particular when their partners' prior definitional knowledge was rather weak. These partners in turn often express more statements reflecting agreement ("yes", "ok", and so on). They seem to leave the externalization of knowledge and ideas to their partners, mostly replying by expressing agreement only. This is also called cumulative talk (e.g., Mercer, 1996). In other studies, the acquisition of intuitive knowledge has been found to be fostered by processes of drawing conclusions, interpretation and sense-making (Gijlers & de Jong, submitted; Reid, Zhang, & Chen, 2003; Zhang, Chen, Sun, & Reid, 2004). Students actively attempting to make domain-related contributions to the communication, instead of mainly agreeing with statements of their partner, are possibly more likely to actively engage in these processes and to externalize them, which might explain their higher gains with respect to intuitive knowledge.

The correlation analysis also indicated that the lower the initial definitional knowledge of B's, the more A's posted domain-related messages, which suggest A's explained the domain to their partners. As stated in the introduction section, providing elaborate explanations is often more beneficial than receiving explanations (e.g., Webb, 1989), because students who are providing elaborate explanations are actively engaged in externalization processes. With regard to communication, it was observed that students within dyads seem to "mirror" each other: if one peer posts more domain-related messages, the other peer is also more likely to post domain-related messages (see Table 3). The number of messages posted by A's and B's is positively correlated, the chat scores, which give an indication of domain-relatedness of the chat, of A's and B's were also positively correlated. Moreover, the positive correlations between A's and B's also extend to the level of types of chat messages (e.g., conditional sentences, qualitative statements): if one peer posts more qualitative statements, the other peer is also more likely to post qualitative statements, and so on.

As for future research, the analysis used in this paper cannot answer the question how these different types of messages are distributed over time and how messages of A's and B's relate to each other. This analysis can shed light on how the interaction between partners in a dyad develops over time. For example, it is interesting to investigate if the number of relatively high-level contributions increases or decreases during the interaction. From this it could be inferred how long a fruitful collaborative learning session should last. If the number of high-level contributions starts to decrease, one could argue that continuing the session will, in general, not contribute much more to better learning results. Another question is if there are differences in learning outcomes for balanced and unbalanced dyads. A balanced dyad is a dyad in which the contribution of each partner is at approximately the same (higher) level, for an unbalanced dyad the number of high-level contributions of one partner substantially exceeds the contributions of the other partner. One of the assumptions behind collaborative learning is that pairing high-level contributors with low-level contributors will benefit both in terms of learning results, but maybe the low-level contributor benefits more. Analyzing the learning results of balanced and unbalanced dyads could confirm or reject this assumption.

References

Anjewierden, A. et al. (2006). tOKo and Sigmund: text analysis support for ontology development and social research. [Software]. http://www.toko-sigmund.org.

Anjewierden, A., & Gijlers, H. (2008). An exploration of tool support for categorical coding. In P. Kirschner, F. Prins, V. Jonker & G. Kanselaar (Eds.). *Proceedings International Conference of the Learning Sciences (ICLS),* Utrecht, The Netherlands, 35-44.

Cohen, E. G. (1994). Restructuring the Classroom: Conditions for Productive Small Groups. *Review of Educational Research, 64*, 1-35.

de Boer, V., van Someren, M & Wielinga, B. J. (2007). A redundancy-based method for the extraction of relation instances from the web. *International Journal of Human-Computer Studies*, *65*, 816-831.

de Jong, T., & van Joolingen, W.R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, *68*, 179-202.

de Moor, A & Anjewierden, A. (2007). A socio-technical approach for topic community member selection. *Communities and Technologies 2007*, 225-244.

Erkens, G., & Janssen, J. (2008). Automatic coding of dialogue acts in collaboration protocols. *International Journal of Computer-Supported Collaborative Learning, 3*, 447-470.

Gijlers, H., & de Jong, T. (submitted). Facilitating collaborative inquiry learning with shared concept maps and proposition tables.

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. *Proceedings of the Fourteenth International Conference on Computational Linguistics,* Nantes, France, 539-545.

Hmelo-Silver, C. E. (2003). Analyzing collaborative knowledge construction: multiple methods for integrated understanding. *Computers & Education, 41*, 397-420.

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71*, 449-521.

Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction*, *6*, 359-377.

Reid, D. J., Zhang, J., & Chen, Q. (2003). Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning*, 19, 9-20.

Rosé, C. P., Wang, Y.-C., Cui. Y., Arguello, J., Weinberger, A., & Stegmann, K. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer Supported Collaborative Learning, 3*, 237-271.

Saab, N. (2005). Chat and explore: the role of support and motivation in collaborative inquiry learning. Ph. D. thesis, University of Amsterdam.

Saab, N., Van Joolingen, W. R., & Van Hout-Wolters, B. H. A. M. (2005). Communication in collaborative discovery learning. *British Journal of Educational Psychology, 75*, 603-621.

Slavin, R. E. (1994). Student teams-achievement divisions. In S. Sharan (Ed.), *Handbook of cooperative learning methods* (pp. 3-19). Westport: Greenwood press.

Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education*, *46*, 29-48.

Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology, 31*, 207-220.

van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning & Instruction, 10*, 311-330.

van der Linden, J. L., Erkens, G., Schmidt, H., & Renshaw, P. (2000). Collaborative learning. In P. R. J. Simons, J. L. v. d. Linden & T. Duffy (Eds.), *New Learning* (pp. 33-48). Dordrecht, The Netherlands: Kluwer.

van Drie, J., van Boxtel, C., Jaspers, J., & Kanselaar, G. (2005). Effects of representational guidance on domain specific reasoning in CSCL. *Computers in Human Behavior, 21*, 575-602.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13*, 21-39.

Webb, N. M., Nemer, K.M., Chizhik, A.W. and Sugrue, B.(1998). Equity issues in collaborative group assessment: group composition and performance. *American Educational Research Journal,* 35, 607–651.

Webb, N. M., Farivar, S. H., & Mastergeorge, A. M. (2002). Productive Helping in Cooperative Groups. *Theory into Practice, 41*, 13-20

Zhang, J. W., Chen, Q., Sun, Y. Q., & Reid, D. J. (2004). Triple scheme of learning support design for scientific discovery learning based on computer simulation: experimental research. *Journal of Computer Assisted Learning*, 20, 269-282.